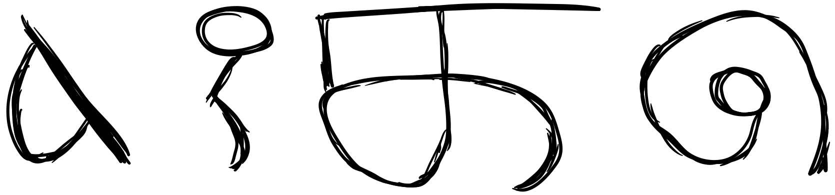
RF Systems Course: Modulation

Introduction

- From perspective of an application programmer, communication link is simple device where bits are fed in at one end and come out the other.
- How this is achieved is not particularly relevant to the application programmer.
- Although implementation details not important to application programmer, there are some attributes of the link that are of interest, including the throughput and the reliability.
- throughput is usually measured bits per second
- Wireless links are notorious for the variability of throughput as the user moves around.
- Reliability refers to either the bit error rate (BER) or the outage probability.
- The BER is probability that transmitted bit is received incorrectly
- outage probability is the probability that the communication link is unusable
- Based on high-level system design, RF system design engineers and DSP engineers decide how to implement individual components to meet system-level functional requirements.
- This design work is typically at the circuit level (for RF components) and software level (for DSP components).

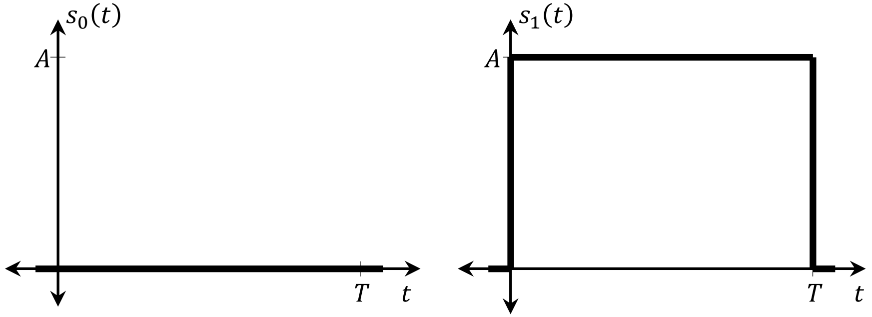
A Simple Digital Communications System

- consider simple e.g. where want to transmit just a single bit.
- To communicate bits, need some physical representation for zeros and ones that is agreed upon by sender and observer.
- E.g. in written English we use 1 to represent a one and 0 to represent a zero.



Cyrillic Hindi Tamil Thai

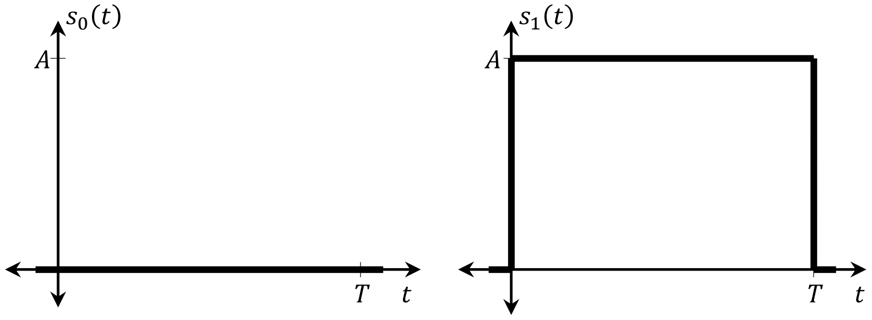
- electronic communication, represent 0s and 1s with different electrical signals.
- E.g. could raise voltage to high level to represent 1, and hold it low to represent a 0.
- known as on-off keying, widely used
- many other schemes are also used
- each has advantages and disadvantages in terms of spectral efficiency, probability of detection error, ease of implementation, etc
- as long as Tx and Rx agree on representation (speak the same language), communication is possible.



(a) (b)

A Simple Digital Communications System

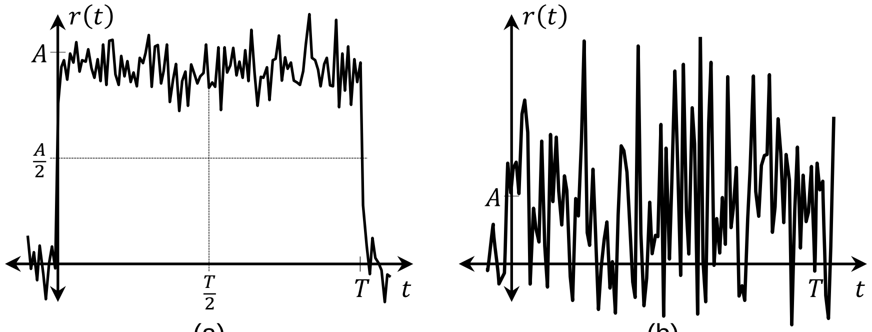
- consider transmission of just a single bit using on-off keying.
- transmitted signal, $v(t)$, is therefore either $s_0(t)$ or $s_1(t)$, depending on whether a 0 or a 1 is to be transmitted.



(a) (b)

- signal propagates over communication medium to receiver.
- During transmission signal is inevitably distorted,

$$r(t) = v(t) + w(t)$$



(a) (b)

- Rx must decide 1 or 0 was transmitted.
- Intuitively $r(t)$ looks more like $s_1(t)$ than $s_0(t)$, so receiver should decide that 1 was transmitted.
- more formal algorithm could be to sample $r(t)$ at some time between 0 and T, such as at $t = T/2$, and then decide whether or not $r(T/2)$ exceeds a certain threshold.
- For example if $r(T/2) \geq A/2$ then the receiver should decide 1 was transmitted, otherwise 0 was transmitted.
- This algorithm, not very reliable.
- If signal is weaker and/or the noise is stronger, quite likely that Rx will make mistake, because any given sample is likely to be below the threshold, even if $s_1(t)$ was transmitted.
- Rx could give better performance by better exploiting all received information.
- Because additive noise, while random, as likely to be positive as negative and changes very quickly over time compared to signal duration, it is better for Rx to calculate average of received signal,

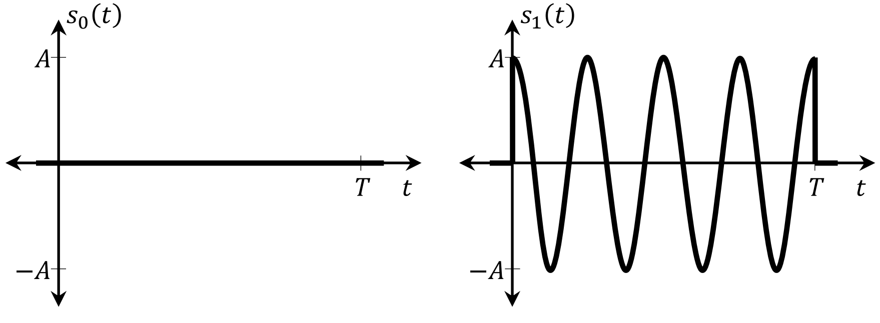
$$R = \frac{1}{T}\int_0^T r(t)\, dt$$

Basic Modulation Schemes

- Simple baseband signaling schemes such as on-off keying, while widely used in wired communication over short distances, are not suitable for wireless transmission.
- must convert signal to higher frequency band prior to transmission using technique known as modulation for wireless
- Demodulation used at Rx to convert signal back to baseband
- By using higher frequency band to transmit signals, possible to use a spectrum sharing technique known as frequency division multiple access (FDMA)
- Modulation achieved by using baseband data-bearing signal to modulate (or vary) the amplitude, phase, or frequency of a carrier wave

Amplitude Shift Keying

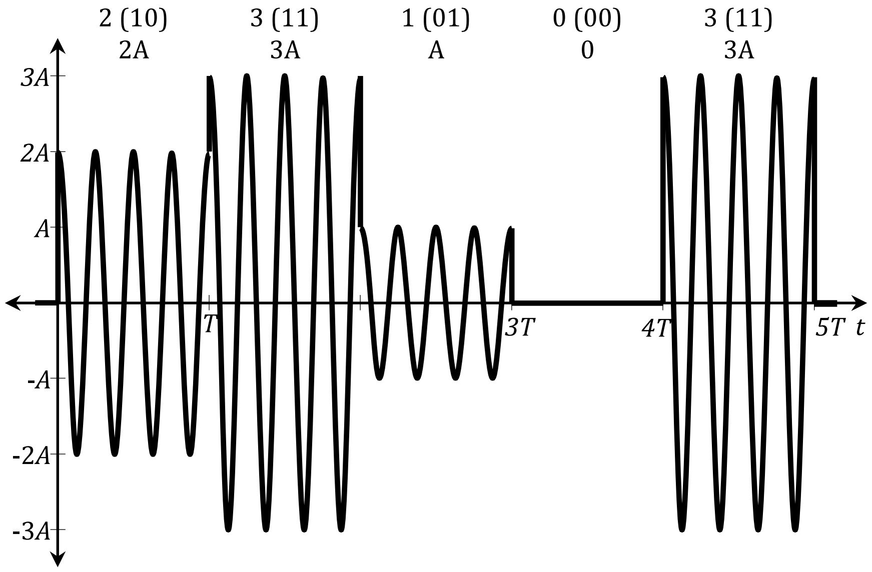
- uses amp of carrier wave to indicate Tx message symbol.
- binary ASK two distinct amplitudes -> 1 bit at a time.



(a) (b)

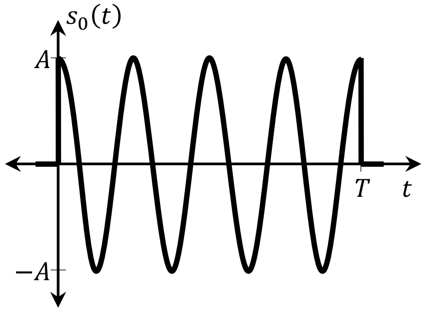
- binary ASK one bit is transmitted every T seconds.
- One way transmit data faster is decrease T.
- However BW inversely related to T, -> decreasing T -> increase BW.
- available BW restricted -> limits to how small T can be
- Another way to increase data throughput is increase #bits that are transmitted in each time slot.
- With M-ary ASK, M different signals (each different amplitudes) are used to represent the M different values.

- Tx of long packet of bits is carried out sequentially over multiple time slots.
- to transmit sequence 10 11 01 00 11 using ASK with M = 4, transmission would be carried out over five time slots, with two bits transmitted per time slot.

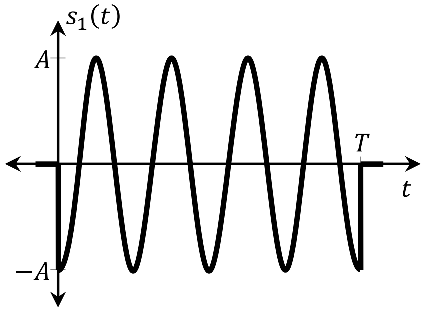


Phase Shift Keying

- Instead of modulating amp of carrier wave, can modulate phase while keeping amp fixed.
- known as *phase shift keying* (PSK).
- When M = 2, known as *binary phase shift keying* (BPSK)
- Although two signals look very similar, possible for Rx to distinguish between two signals, even more so than for binary ASK.

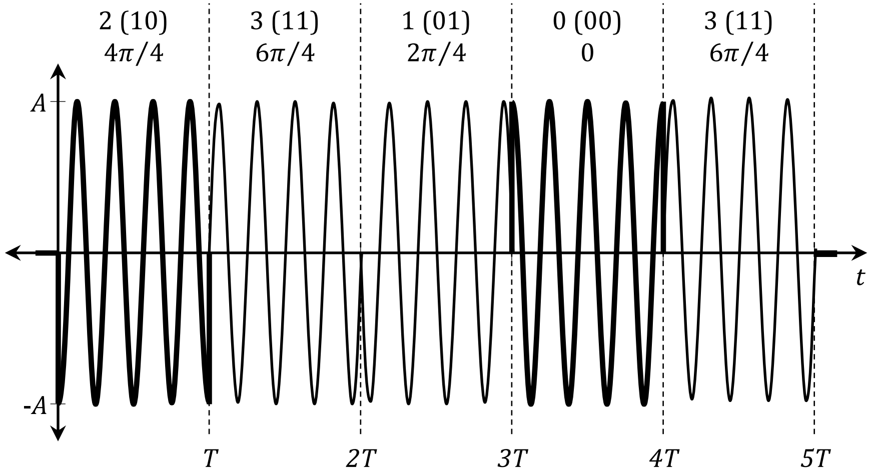


(a)



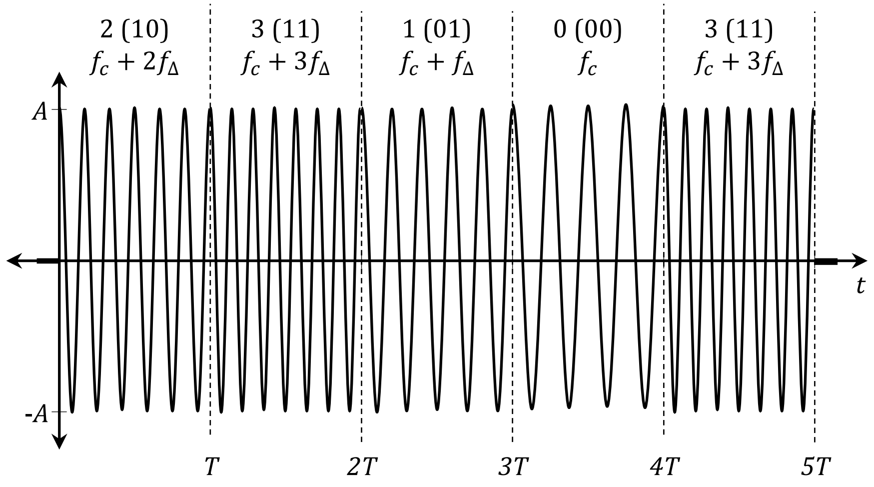
(b)

- For M = 4, called *quaternary phase shift keying* (QPSK),
- phase changes with each symbol interval, but amp. constant.



Frequency Shift Keying

- FSK involves keeping amp and phase of carrier wave constant, while varying freq depending on Tx data.



- Freq offset is typ. chosen to be $f_\Delta = 1/T$.
- Like PSK, FSK exhibits constant amp, but FSK not spectrally efficient so seldom used, except for inexpensive low-data-rate applications with M = 2.

Quadrature Amplitude Modulation (QAM)

- For applications where spectral efficiency is important.

- This mod. scheme involves transmitting two ASK signals simultaneously, one with a cosine wave carrier (the *inphase* carrier) and one with a sine wave carrier (the *quadrature phase* carrier).

- QAM is special case where both the amplitude and phase of a carrier are modulated.

Signal Models

- For mod. schemes that involve modulating the phase and/or the amp, it is convenient to describe signals in terms of their complex lowpass equivalent signal representation.
- The transmitted bandpass signal used to represent symbol $m \in \{0,1,\dots,M-1\}$:

$$s_m(t) = \sqrt{E_m} h_T(t) \sqrt{2} \cos(2\pi f_c t + \phi_m)$$

- More commonly:

$$s_m(t) = s_{I,m} \sqrt{2} \cos(2\pi f_c t) - s_{Q,m} \sqrt{2} \sin(2\pi f_c t)$$

- $s_{I,m}(t) = \sqrt{E_m} \cos \phi_m \, h_T(t)$ is LP signal transmitted on inphase (I) carrier $\sqrt{2} \cos(2\pi f_c t)$
- $s_{Q,m}(t) = \sqrt{E_m} \sin \phi_m \, h_T(t)$ is LP signal transmitted on quadrature phase (Q) carrier $-\sqrt{2} \sin(2\pi f_c t)$.

Signal Space Diagrams

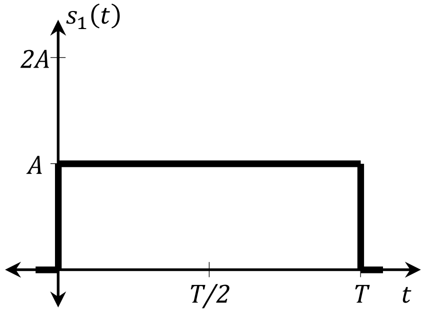
inner product, for 2 signals, $v(t)$ and $w(t)$:

$$\langle v(t), w(t)\rangle \overset{\text{def}}{=} \int_{-\infty}^{\infty} v(t)w(t)\, dt$$

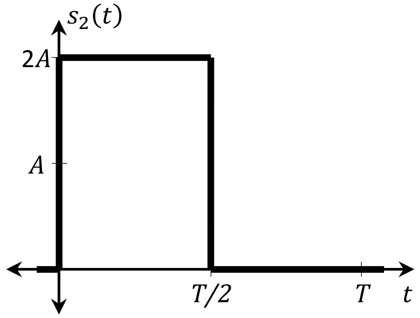
- compare definition of inner product for signals with corresponding definition for vectors.
- The inner (dot) product of two *N*-dimensional vectors, $\mathbf{v}=v1\ v2 \cdots vN$ and $\mathbf{w}=w1\ w\mathbf{2} \cdots wN$ is

$$\langle \mathbf{v}, \mathbf{w}\rangle \overset{\text{def}}{=} \sum_{n=1}^{N} v_n w_n$$

Find the inner product of $s1t$ and $s2t$



(a)



(b)

inner product of two signals is

$$\langle s_1(t), s_2(t)\rangle = \int_{-\infty}^{\infty} s_1(t)s_2(t)\, dt$$

$$= \int_0^{\frac{T}{2}} (A)(2A)\, dt + \int_{\frac{T}{2}}^{\mathrm{T}} (A)(0)\, dt = A^2T$$

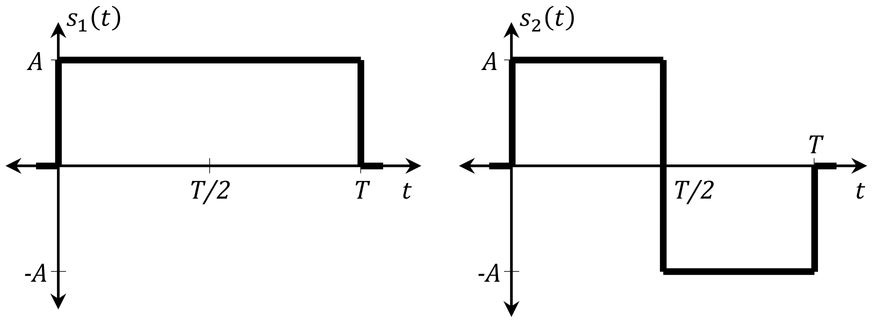
Signal Space Diagrams

norm of signal:

$$\|s(t)\| \stackrel{\text{def}}{=} \sqrt{\langle s(t), s(t)\rangle} \stackrel{\text{def}}{=} \sqrt{\int_{-\infty}^{\infty} |s(t)|^2 \, dt}$$

- norm of a vector is its length, norm of a signal is √ of its energy.
- signal is said to be *normal* (or *normalized*) if $\|s(t)\| = 1$ (has unit energy)
- 2 signals are *orthogonal* if their inner product is 0.
- All pairs of signals that do not overlap in time will have an inner product of zero -> orthogonal.
- there are many pairs of signals that do overlap and are still orthogonal.

Show that *s1t* and *s2t* are orthogonal.



(a) (b)

To show that two signals are orthogonal, we must show that their inner product is zero:

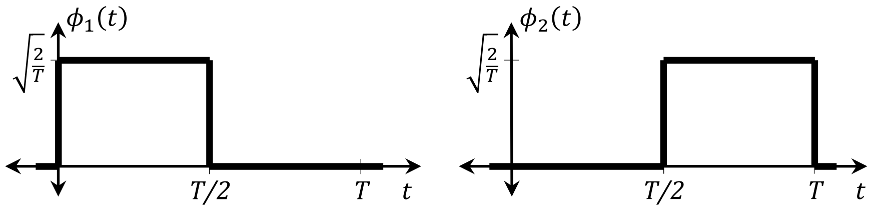
$$\langle s_1(t), s_2(t)\rangle = \int_0^{\frac{T}{2}} (A)(A)\, dt + \int_{\frac{T}{2}}^{\mathrm{T}} (A)(-\mathrm{A})\, dt = A^2\frac{T}{2} - A^2\frac{T}{2} = 0$$

- A set of signals are said to be *orthonormal* if they are all normal, and every signal is orthogonal to the other signals in the set.

- Any set of orthonormal signals can be used as the *basis signals* for some signal space.
- span of signal space is set of all signals that can be expressed as linear combinations of basis signals.
- set of Tx signals used in any mod scheme can always be expressed as linear combinations of a small set of basis signals.

Signal Space Diagrams

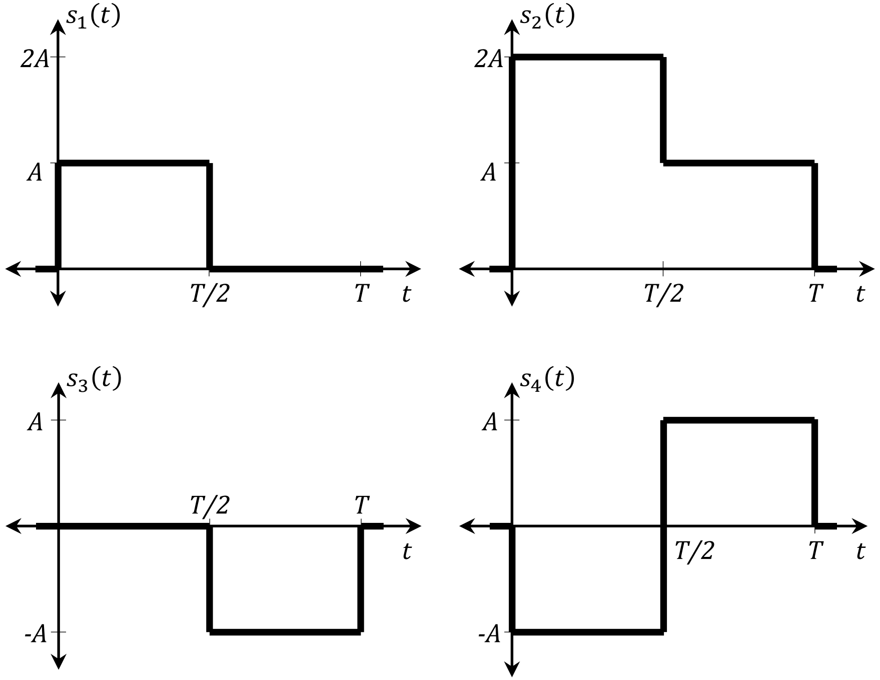
- 4 signals can be expressed as linear combinations of 2 basis signals $\phi 1t$ and $\phi 2t$.
- $\phi 1t$ and $\phi 2t$ are valid basis signals because they are orthonormal.
- They are orthogonal because they do not overlap, so their inner product must be 0, and they are each normalized (specified amp of $2T$ ensures this).
- data signals can be expressed as linear combinations of basis signals:

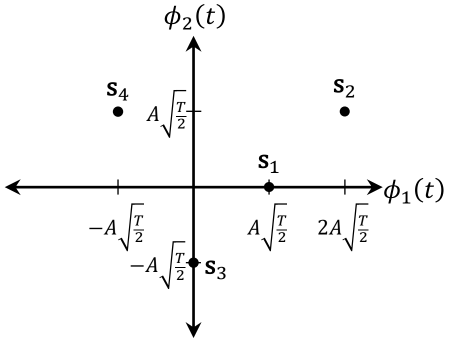


$$s_1(t) = A\sqrt{\frac{T}{2}}\phi_1(t) \qquad s_2(t) = 2A\sqrt{\frac{T}{2}}\phi_1(t) + A\sqrt{\frac{T}{2}}\phi_2(t)$$

$$s_3(t) = -A\sqrt{\frac{T}{2}}\phi_2(t) \qquad s_4(t) = -A\sqrt{\frac{T}{2}}\phi_1(t) + A\sqrt{\frac{T}{2}}\phi_2(t)$$

4 points, $\mathbf{s}_1 = \left(A\sqrt{T/2}, 0\right)$, $\mathbf{s}_2 = \left(2A\sqrt{T/2}, A\sqrt{T/2}\right)$, $\mathbf{s}_3 = \left(0, -A\sqrt{T/2}\right)$, and $\mathbf{s}_4 = \left(-A\sqrt{T/2}, A\sqrt{T/2}\right)$ can be plotted in a signal space diagram



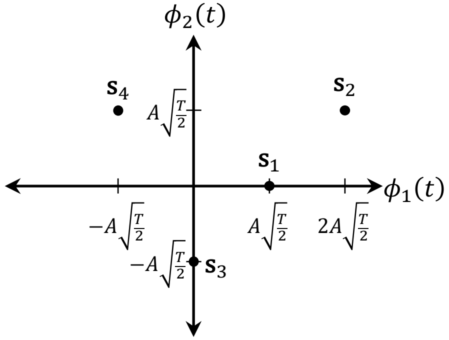


Signal Space Diagrams

- signal space diagram is a very useful tool for describing the signals.
- square of distance between a signal point and the origin gives the energy of the signal, e.g. energy of $s_4(t)$

$$A^2T$$

- distance between any 2 signals is useful for quantifying similarity between 2 signals
- 2 signals that are very close to each other in signal space diagram are very similar.
- receiver's ability to distinguish between two signals depends on how close together they are.
- Ideally want to select our transmitted signals so they are as far apart as possible.
- for any modulation scheme we can increase separation between points by increasing energy of all Tx signals.
- probability that Rx will make an error decreases as average transmitted energy increases.
- due to physical constraints there is limit on max energy that can be transmitted, so for a fixed average energy it is desirable to use a modulation scheme with as large a separation between points as possible.

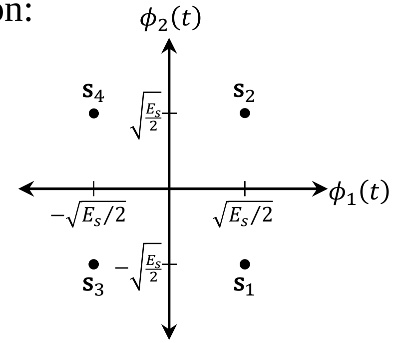


- For the 4 signals, find min separation between the points.
- Express answer in terms of average signal energy.
- Suggest a different set of signals in same signal space that has a larger minimum separation for the same average energy.

closest pair of points are $\mathbf{s}_1$ and $\mathbf{s}_2$ (or $\mathbf{s}_1$ and $\mathbf{s}_3$), they have separation

$$= \sqrt{A^2T}$$

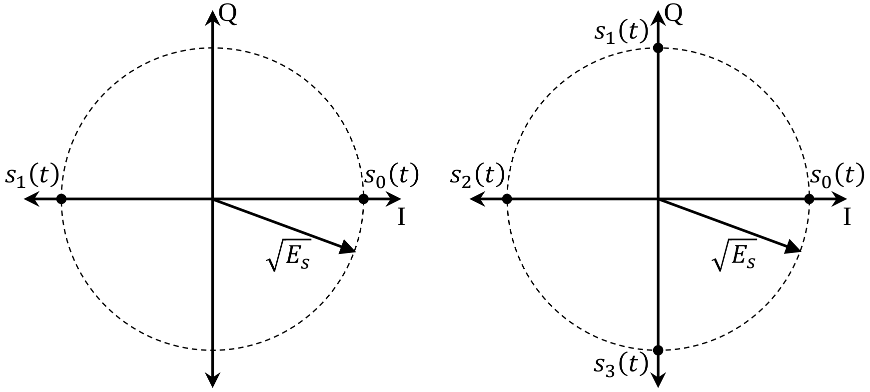
better constellation:



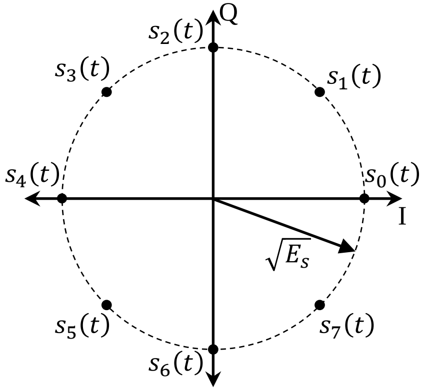
Most common basis set is I and Q

$$s_m(t) = \left(\sqrt{E_m}\cos\phi_m\right)\phi_I(t) + \left(\sqrt{E_m}\sin\phi_m\right)\phi_Q(t) = s_{I,m}(t)\phi_I(t) + s_{Q,m}(t)\phi_Q(t)$$

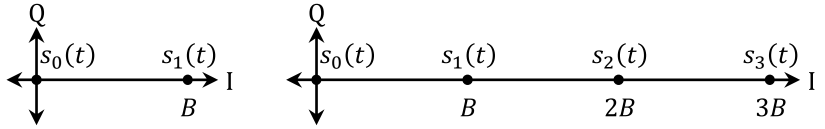
M-ary PSK, ASK, and QAM



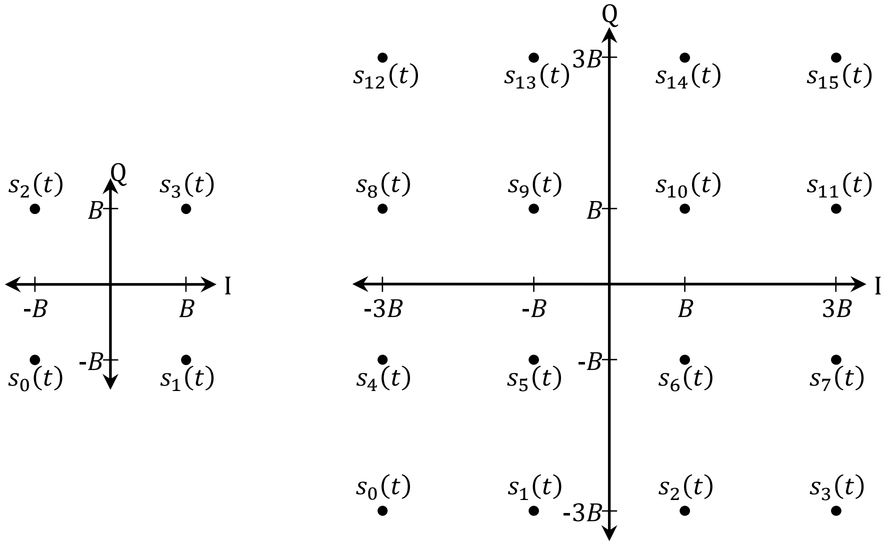
(a) M = 2 (b) M = 4



(c) M = 8



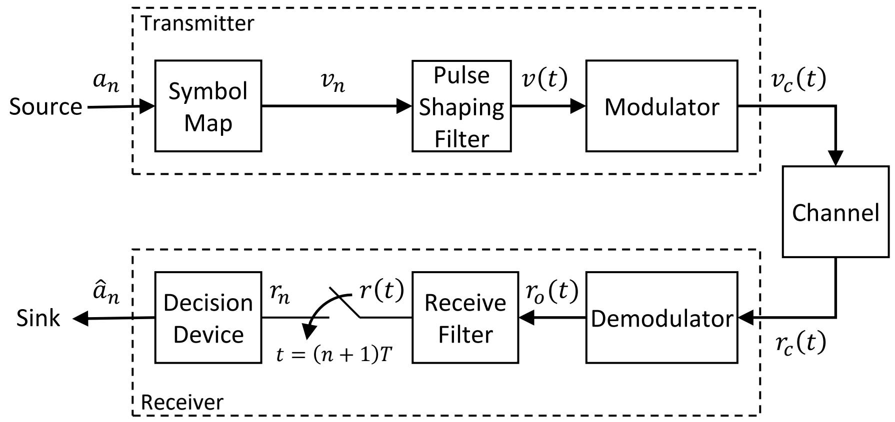
(a) M = 2 (b) M = 4



(a) M = 4 (b) M = 16

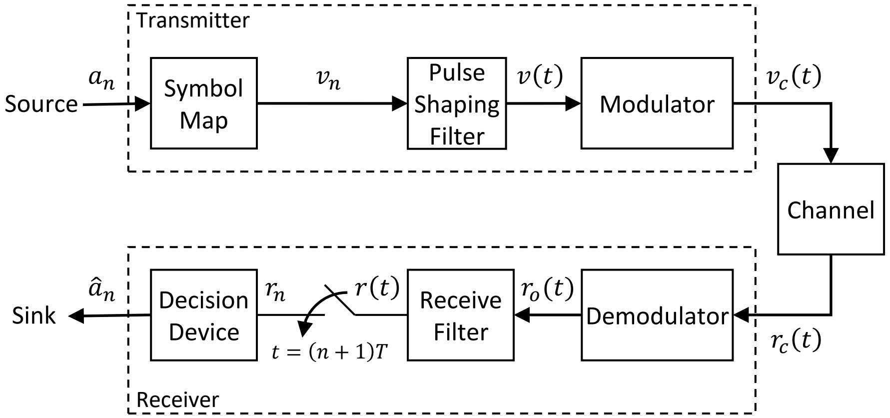
System Model

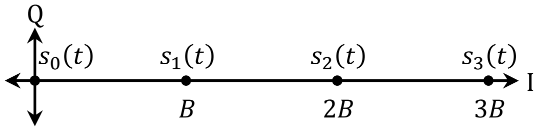
- block diagram suitable for any amplitude and/or phase modulation scheme
- symbol map converts digital data to be transmitted into coordinates of corresponding point in signal constellation.
- pulse shaping filter generates a pulse train based on the desired pulse shape.
- modulator converts the complex LP equivalent signal into BP signal.
- BP signal propagates over the wireless channel to the receiver
- demodulator down-converts signal back to baseband.
- receive filter blocks out unwanted noise and signal is sampled at symbol rate
- Based on samples, decision device tries to determine Tx digital data.



Symbol Map

- can transmit $\log 2M$ bits in every symbol, in practice need to transmit much more information.
- This is achieved by transmitting many symbols sequentially over many subsequent time slots.
- data packet (a block or a frame) containing N_b bits is partitioned into N_S symbols, each containing $\log_2 M$ bits, so $N_b = N_S \log 2M$.
- Depending on value of bits, each symbol mapped to corresponding point in signal constellation.
- consider transmission of data packet $\mathbf{a} = 11\ 01\ 11\ 00\ 10$ using the 4-ASK signal constellation
- M = 4, $N_b = 10$, $N_S = 5$, sequence of transmitted symbols is $\mathbf{v} = 3B, B, 3B, 0, 2B$.

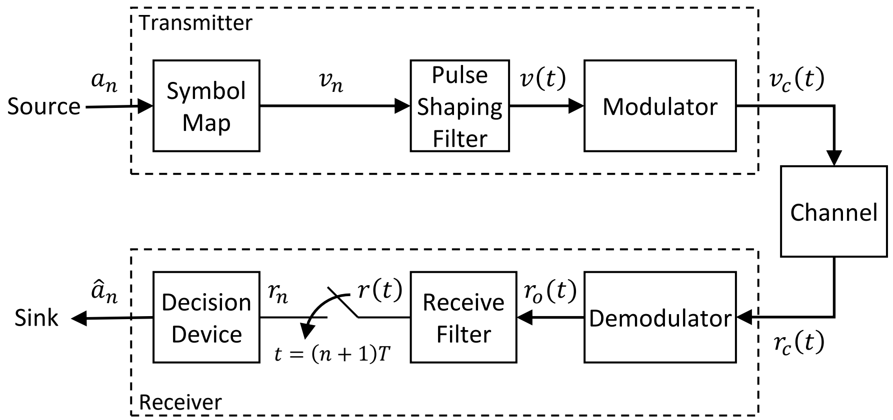


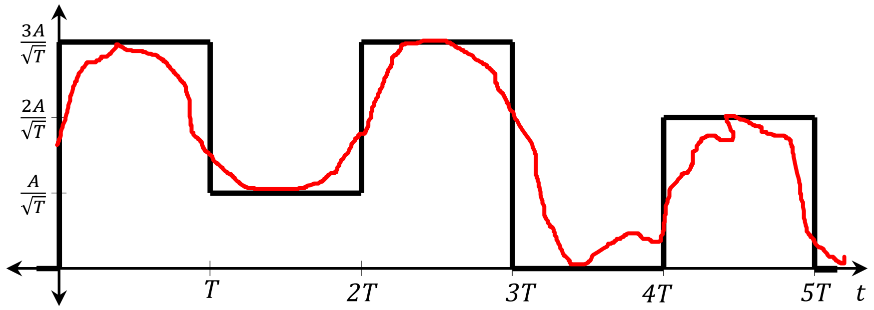


(b) M = 4

Pulse Shaping Filter

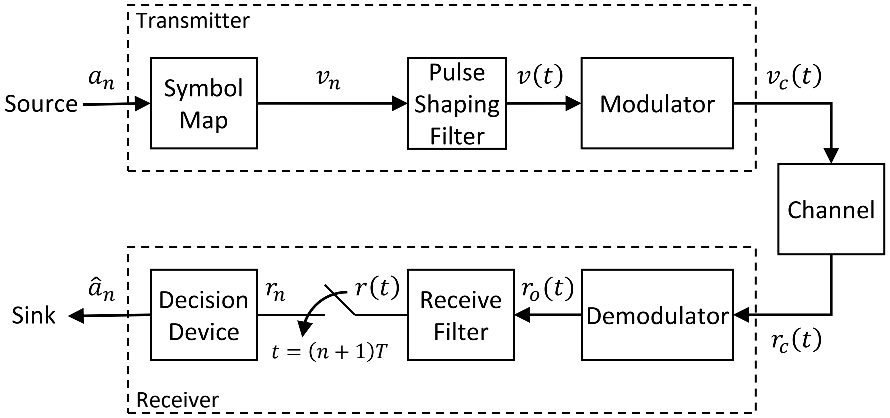
- generates a pulse train using each transmitted symbols as complex-valued amplitude of the desired pulse shape, $h_T(t)$.
- new pulse is generated every T seconds

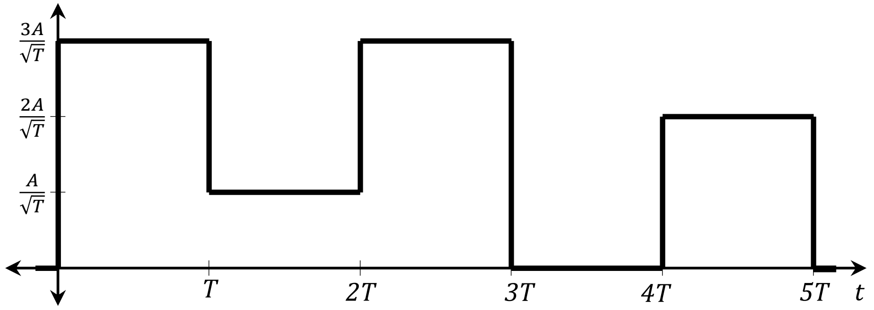


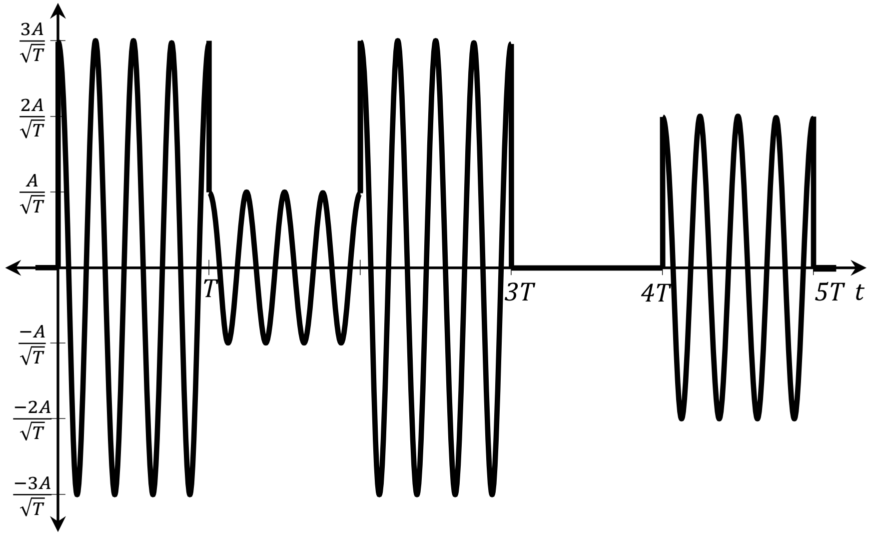


Modulator

Upconverts to the carrier.
Then noise is added during transmission

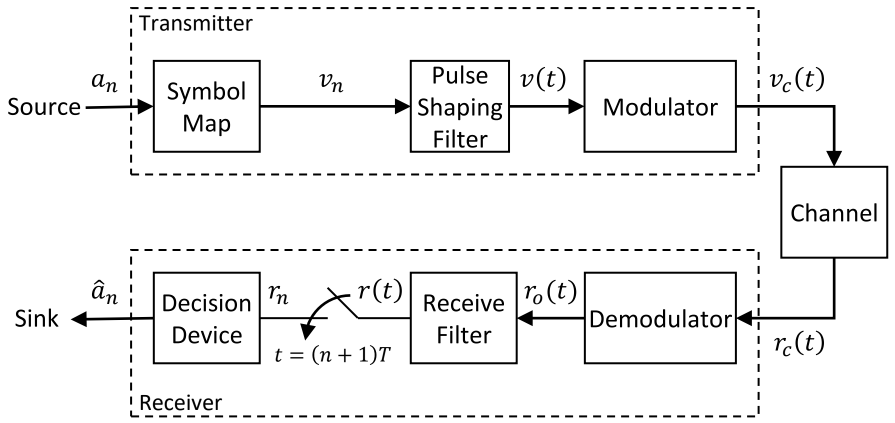






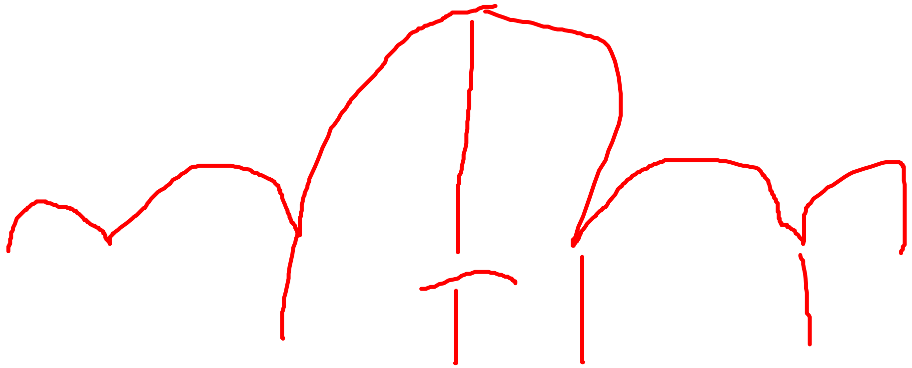
Demodulator/Filter

- Rx signal demodulated to BB by mixing with LO.
- signal should have same freq and nominal phase as carrier signal, but with quadrature phase component reversed in sign.



- After down-converted, must be filtered to remove high-freq component and any other signals in adjacent bands.
- filter should also emphasize data-bearing signal while suppressing noise as much as possible.

Signal Sampling



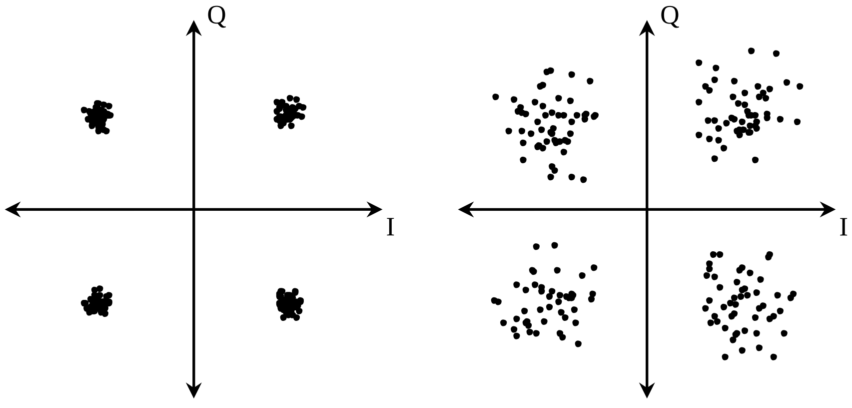
Rx signal sampled once every T seconds starting at $t = T$.

$$r_n = r([n+1]T) = \sum_{m=0}^{N_S-1} v_m h_{TR}([n+1]T - mT) + w([n+1]T) = \sum_{m=0}^{N_S-1} v_m h_{TR}([n-m+1]T) + w_n$$

- r_n depends not only on v_n but potentially on all other Tx symbols -> intersymbol interference (ISI).
- To avoid ISI, choose the pulse shape carefully
- $h_{TR}(t)$ must:

$$h_{TR}([n-m+1]T) = \begin{cases} 1 & \text{if } m = n \\ 0 & \text{if } m \neq n \end{cases}$$

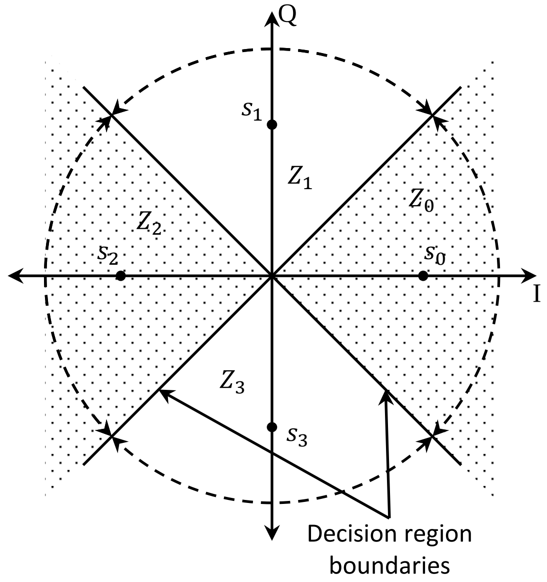
- Any normalized pulse shape, $h_T(t)$, that is non-zero only over interval $t \in [0, T]$ will do this.
- Some other pulse shapes, such as impulse response of a root raised-cosine filter, can do this, and can provide better spectral properties.
- After this Rx signal projected onto signal space.



(a) (b)

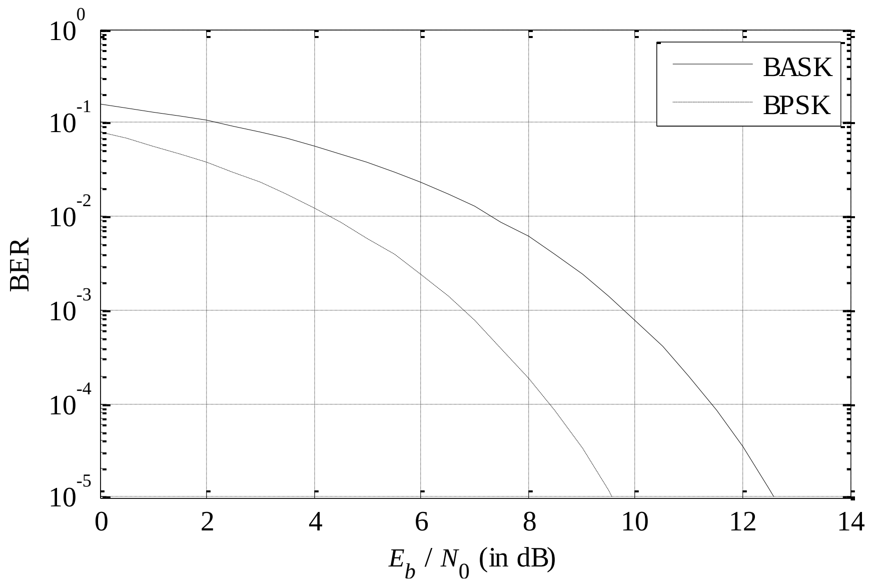
Decision Device

- By dividing signal space into regions corresponding to which signal point is closest, possible to further simplify decision device.



BER BASK/BPSK

- A lot of math in this section
- Not that important
- Resulting curves are though



- BER expressed in terms of ratio of average energy per bit (E_b) to noise power spectral density (N_0).
- In RF SNR more commonly used.
- SNR used in this section: $\gamma_s = P_S/P_N$.
- Since one symbol, with average energy E_s, is transmitted every T seconds, we have $P_S = E_s/T$.
- power of noise, with a PSD of N_0, BW of W Hz, is $P_N = N_0 W$.

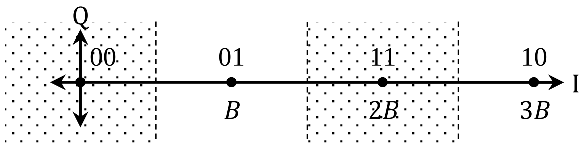
$$\gamma_s = \frac{P_S}{P_N} = \frac{E_s/T}{N_0 W} = \frac{E_s}{N_0} = \frac{E_b \log_2 M}{N_0} = \gamma_b \log_2 M$$

$\log_2 M$ is number of bits per symbol and $\gamma_b = E_b/N_0$ is the SNR per bit.

M-ary Signalling and Grey Mapping

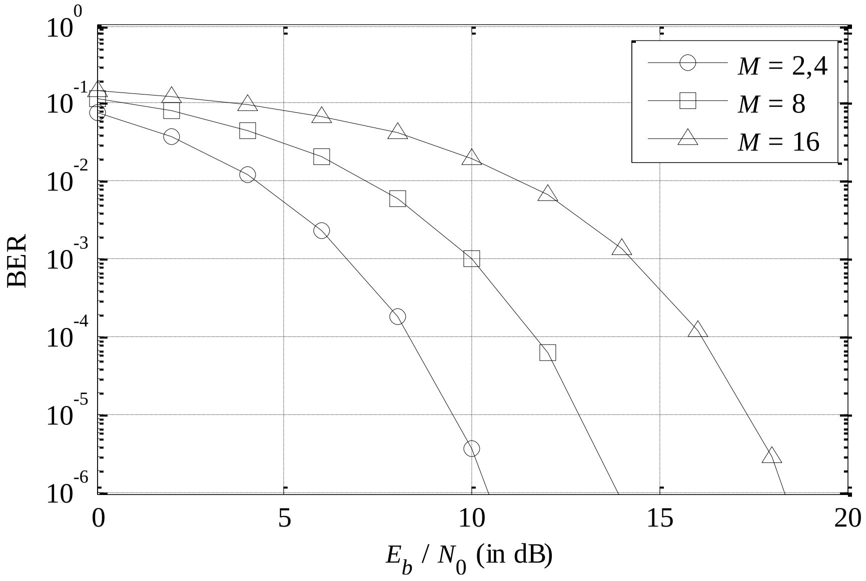
- when BER is small, when errors do occur usually because symbol immediately adjacent to Tx symbol is erroneously selected.
- To ensure BER is small good idea to ensure that adjacent symbols differ in only one bit position
- when a symbol error occurs usually only a single bit is received incorrectly.
- achieved by carefully mapping the Tx bits to signals using a technique known as *Gray mapping*.
- When Grey mapping is used :

$$P_b \cong \frac{P_\varepsilon}{\log_2 M}$$



SNR to Achieve a Target BER

- want to determine the min SNR to achieve BER of 10^{-3} using BPSK
- E_b/N_0 > 6.8 dB
- 1 bit transmitted/symbol -> SNR > 6.8 dB.
- If 8-PSK would require $E_b/N_0 > 10$ dB
- 3 bits/symbol -> SNR = $10+\log_2 3 = 14.8$ dB.
- 8-PSK needs 8 dB more signal power for same BER as BPSK
- must be balanced with the benefits of transmitting bits three times faster.



Curves for PSK without showing all the math.