

# Principle of Digital Communication

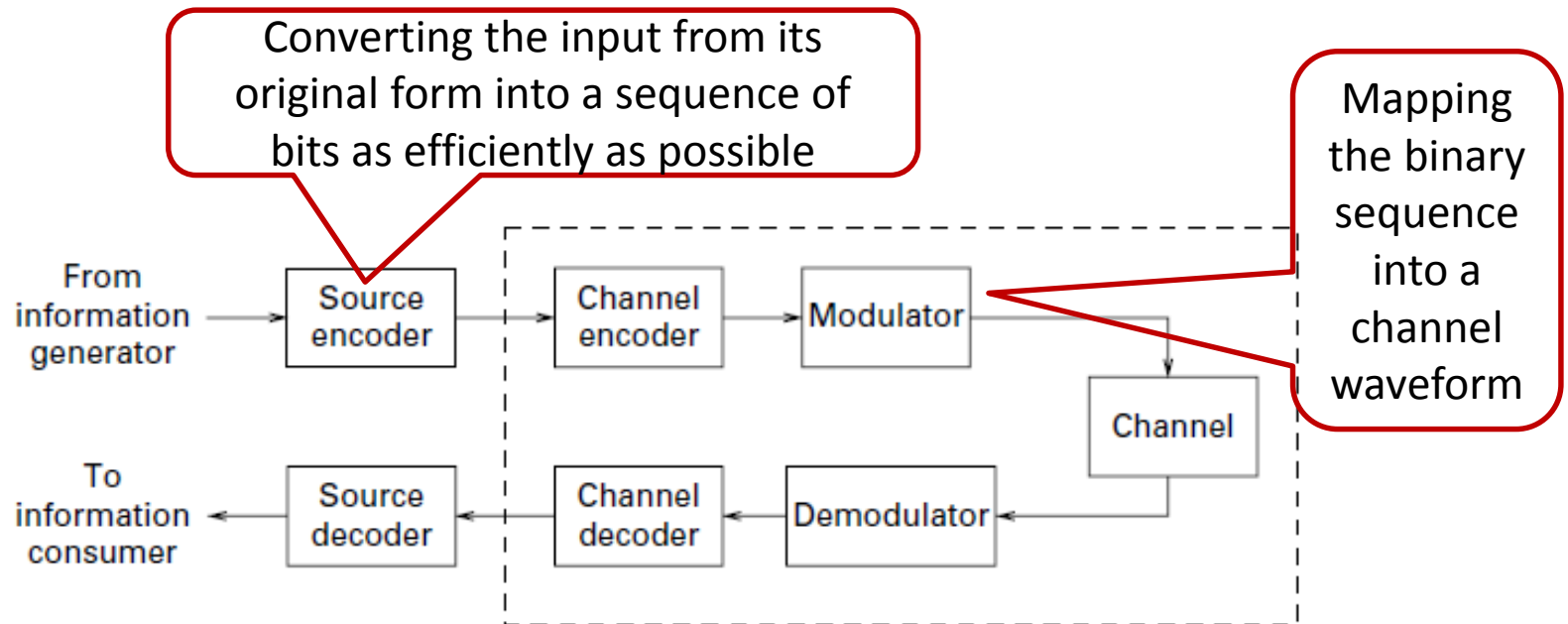
---Lecture 2

Channels, modulation, and demodulation



## Brief review

Chapter 1 discussed the reasons for using a bit sequence as the interface between an arbitrary source and an arbitrary channel.



# Outline

2.1 The Nyquist criterion

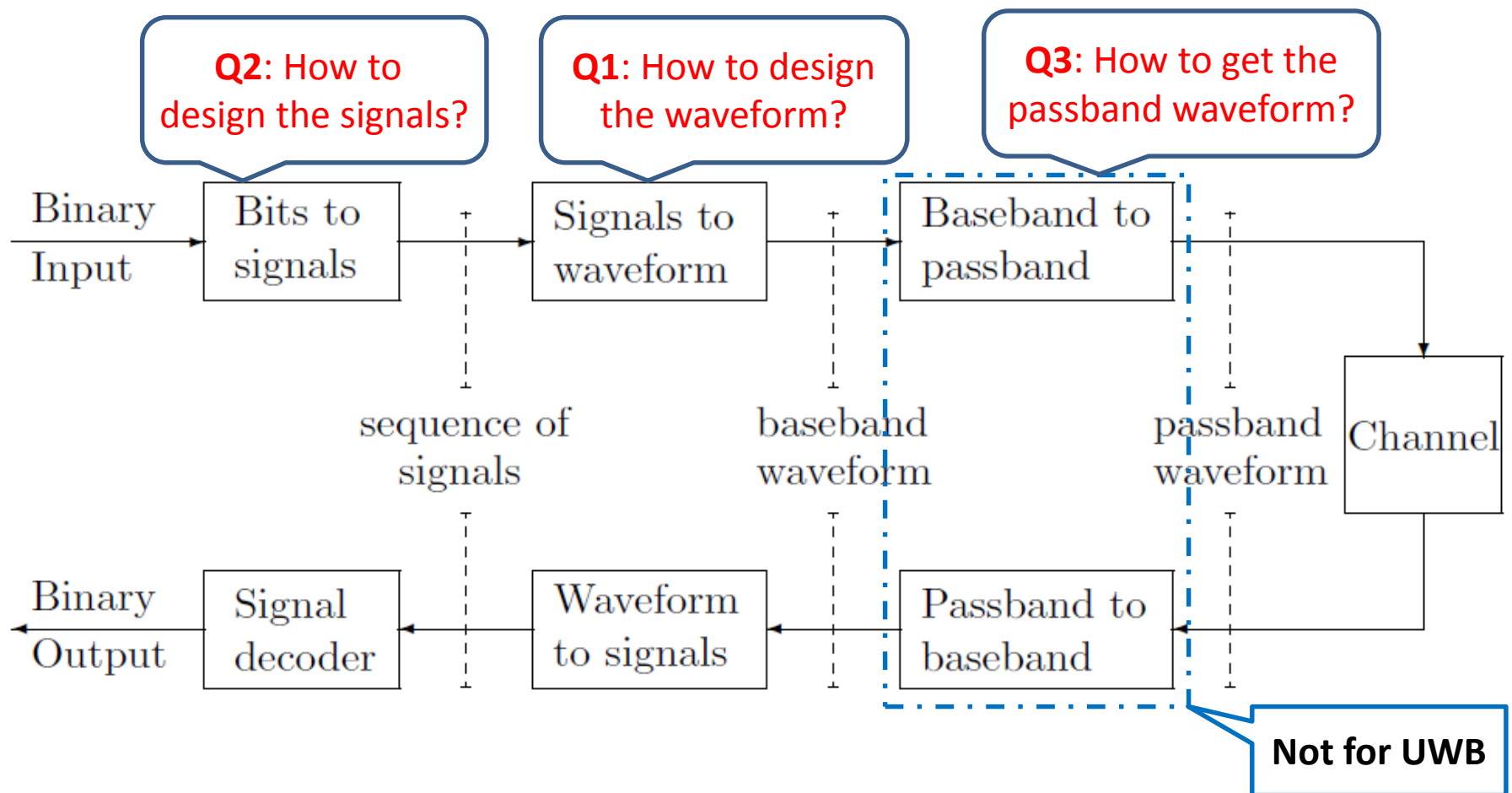
2.2 Signal space (PAM)

2.3 Modulation: baseband to passband and back

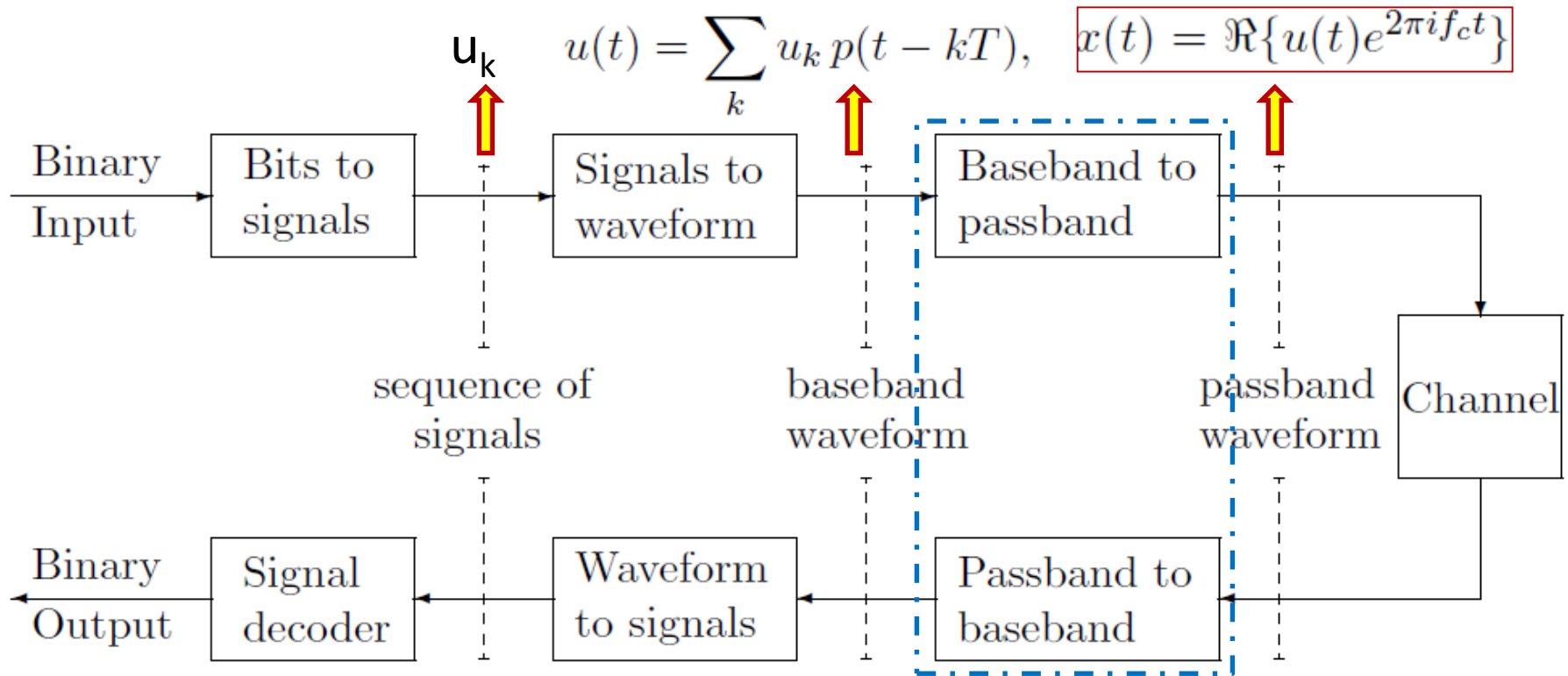
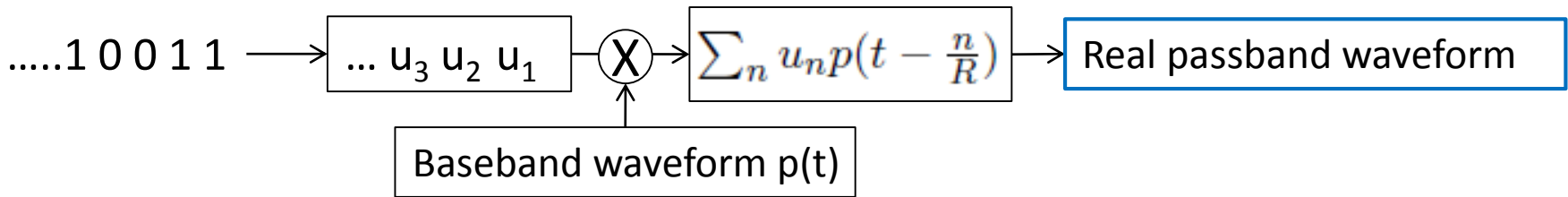
2.4 Quadrature Amplitude Modulation (QAM)

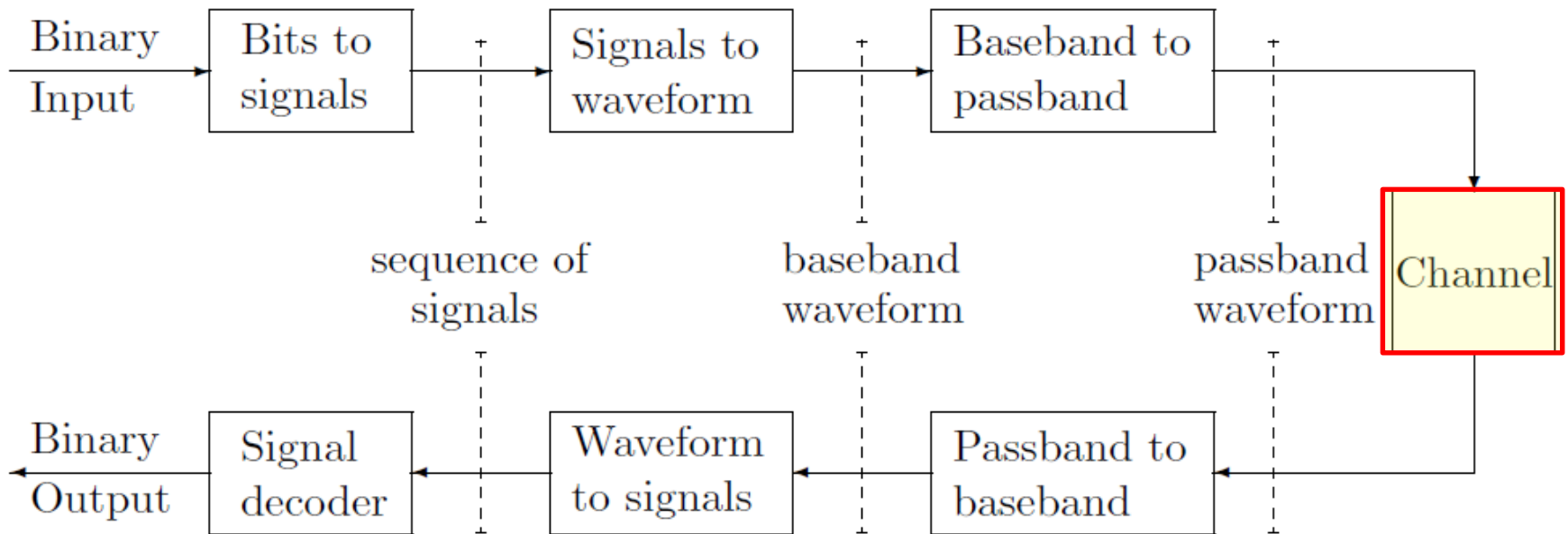
2.5 Differential Modulation (brief introduction)





## 2.1 The Nyquist criterion





### 2.1.1 Channel imperfections: a preliminary view

Physical waveform channels are always subject to propagation delay, attenuation, and noise. Many wireline channels can be reasonably modeled using only these degradations, whereas wireless channels are subject to other degradations such as fading.

This subsection provides a preliminary look at delay, then attenuation, and finally noise.



## 1. Delay

A waveform  $u(t)$  at the transmitter is subject to propagation delay plus various filter delays in the modulator and demodulator. Thus  $u(t)$ , according to the transmitter clock, appears as  $u(t-\tau)$  at the receiver, where  $\tau$  is the overall delay.

By delaying the receiver clock by  $\tau$  from the transmitter clock, the received waveform is  $u(t)$ . With this convention, the channel can be modeled as having no delay, and all equations will be greatly simplified.

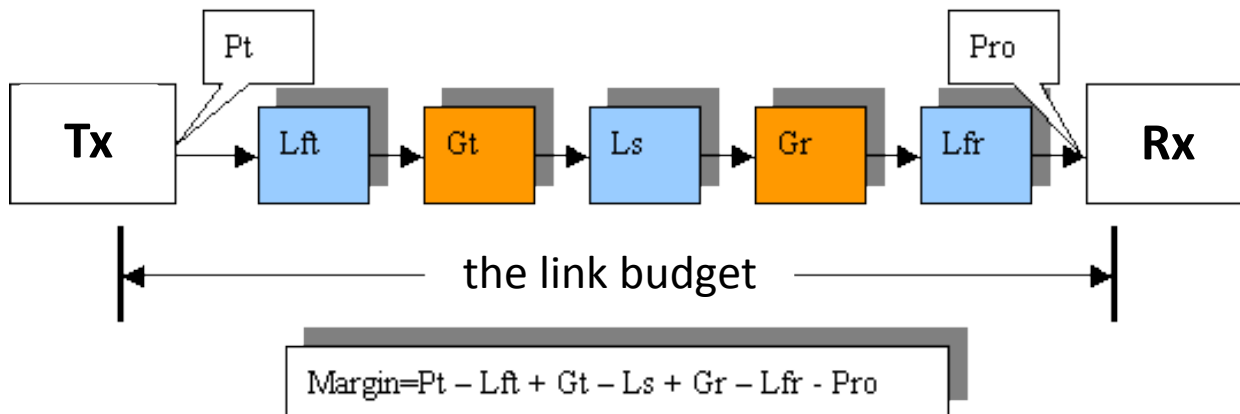
Estimating the above fixed delay at the receiver is a significant problem called **timing recovery**, but is largely separable from the problem of recovering the transmitted data.



## 2 . Attenuation

The actual power attenuation suffered in transmission is a product of amplifier gain, antenna coupling losses, antenna directional gain, propagation losses, etc. The process of finding all these gains and losses is called “the link budget.” Such gains and losses are calculated in decibels (dB).

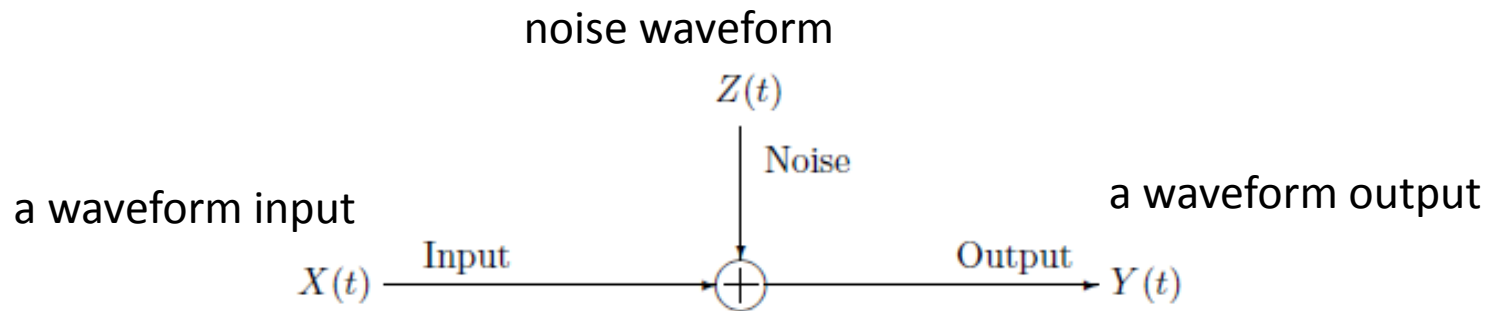
Recall that the number of dB corresponding to a power gain  $\alpha$  is defined to be  $10 \log_{10} \alpha$ .





### 3 . Noise

**This noise is a fundamental limitation to communication** and arises from a variety of causes, including thermal effects and unwanted radiation impact on the receiver.



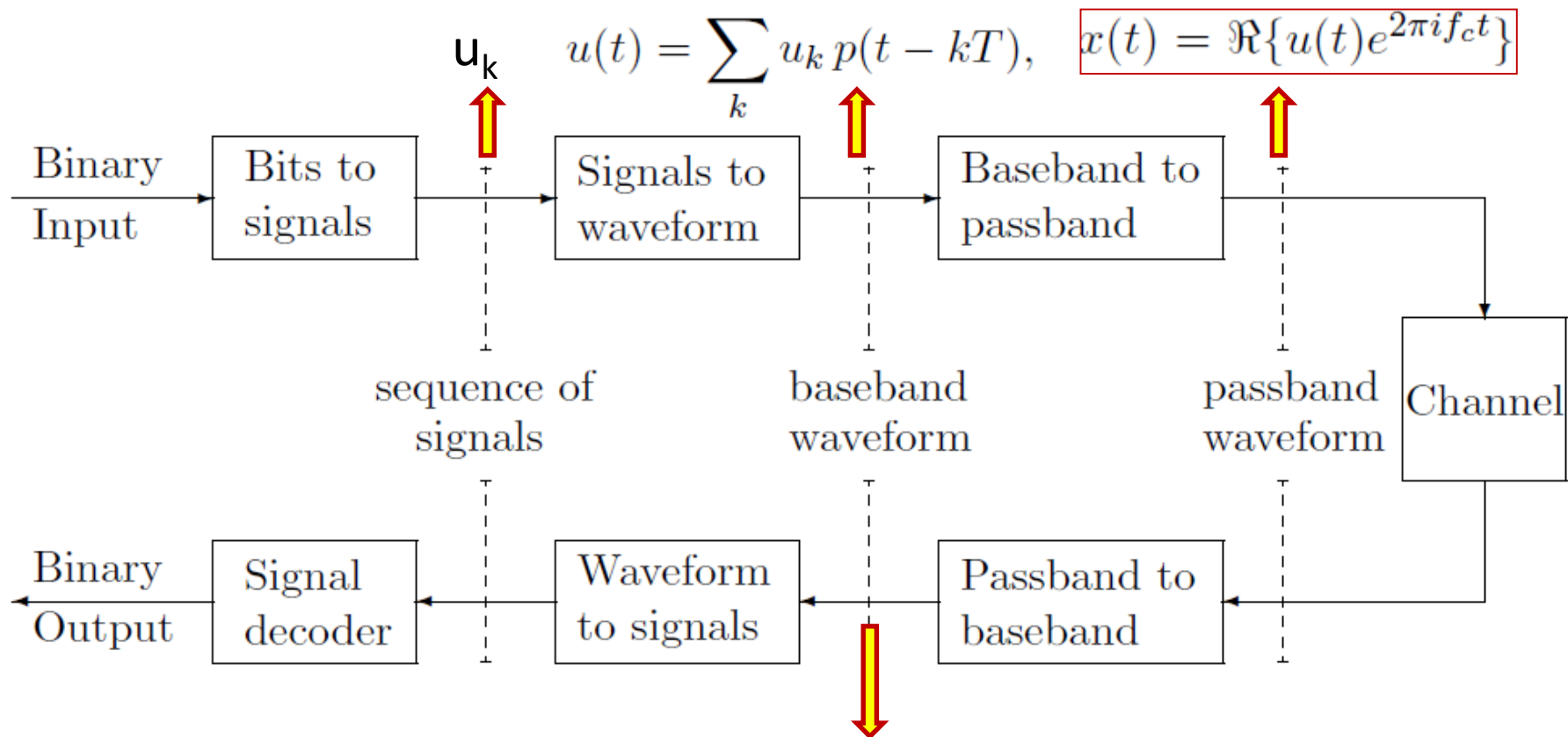
## Q1: How to design the waveform $p(t)$ ?

Assume that the time reference and the amplitude scaling at the receiver have been selected so that the received baseband waveform is the same as the transmitted baseband waveform  $u(t)$ , and no noise has been introduced by the channel.

However, **neglecting the noise is artificial**, since this means neglecting the fundamental limitation on the bit rate.

The reason for posing this artificial problem is, first, that avoiding inter-symbol interference is significant in choosing  $p(t)$ , and, second, that there is a simple and elegant solution to this problem.



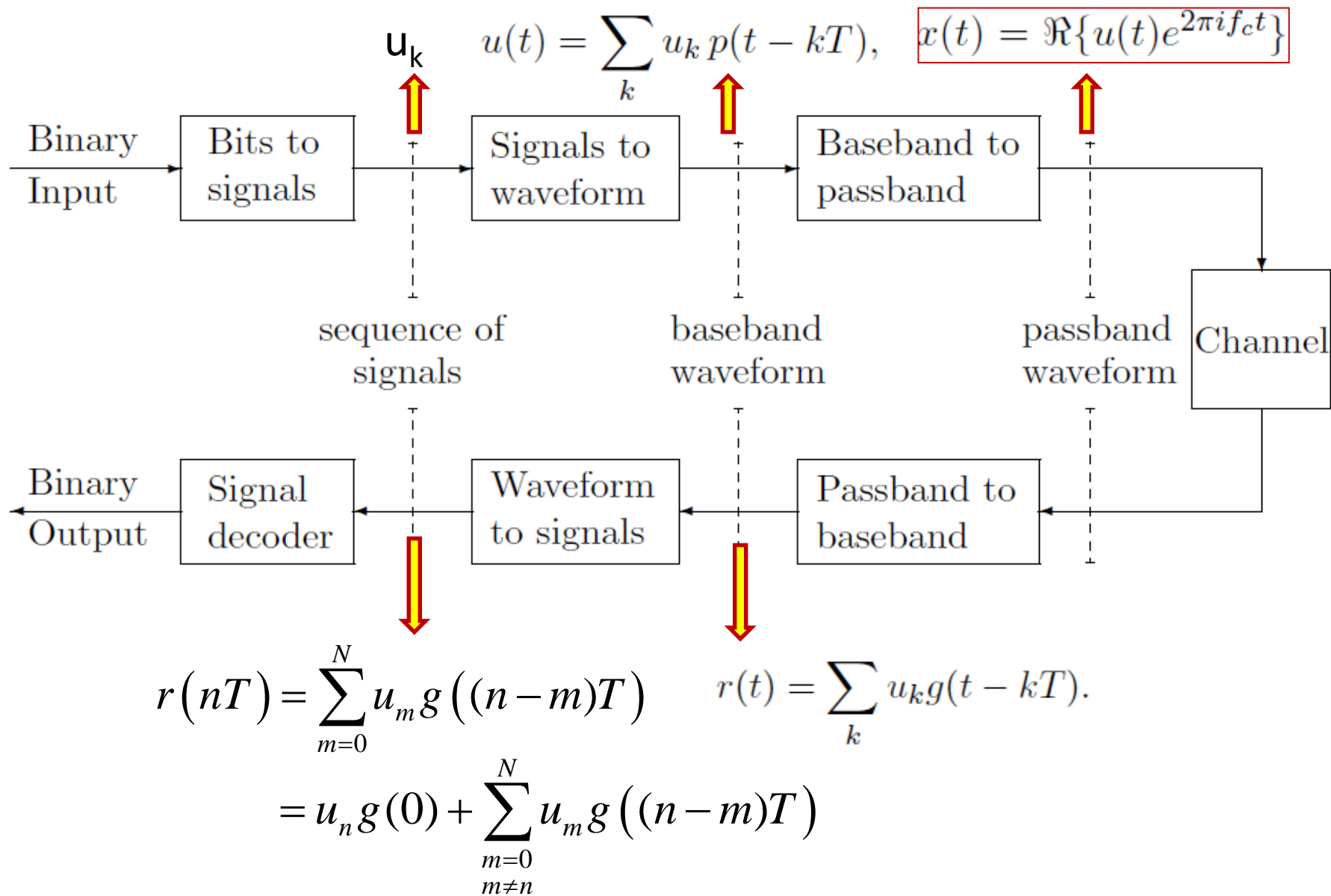


$$r(t) = \int_{-\infty}^{\infty} \sum_k u_k p(\tau - kT) q(t - \tau) d\tau.$$

$$g(t) = p(t) * q(t) = \int p(\tau) q(t - \tau) d\tau$$

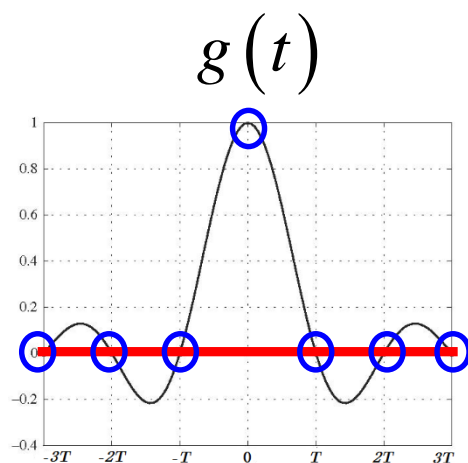
$$r(t) = \sum_k u_k g(t - kT).$$



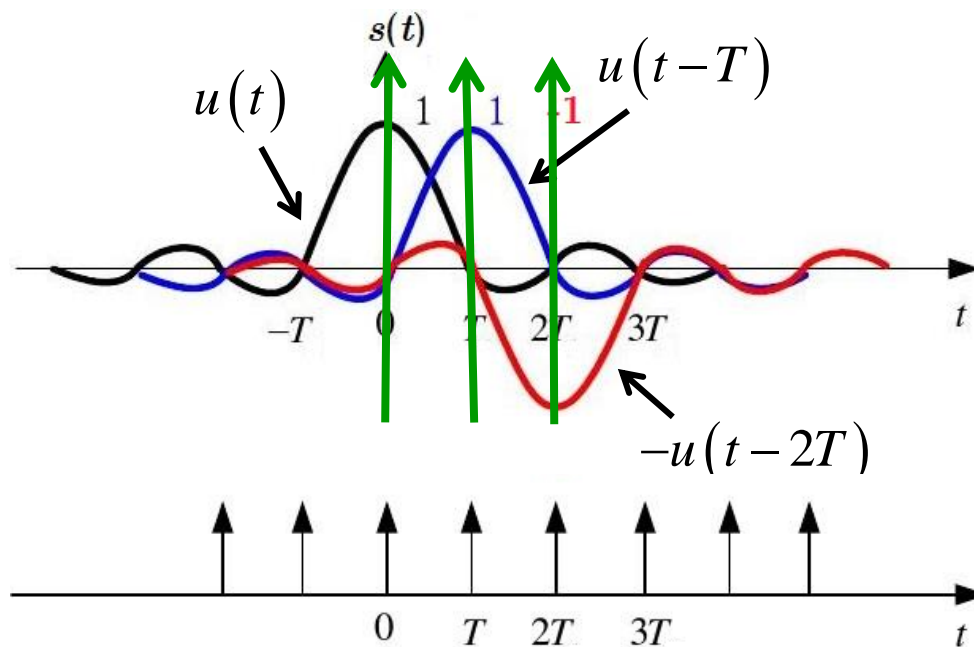


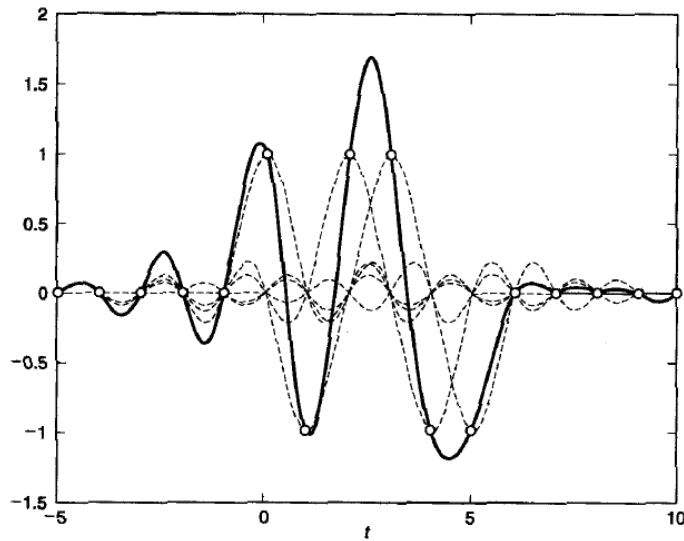
To make sure there is no ISI:

$$g(nT) = \delta(n) = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases} \quad (2.1)$$

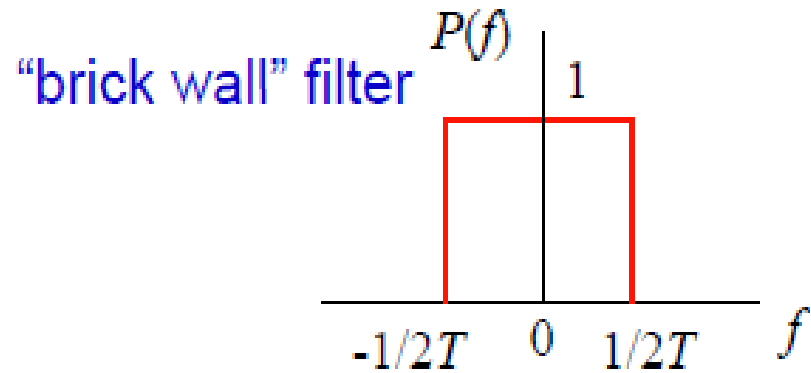


$$\begin{aligned} u_0 &= 1, \\ u_1 &= 1, \\ u_2 &= -1 \end{aligned}$$





Sensitive to timing errors



Too ideal to be implemented

If and only if

$$\text{l.i.m.} \sum_m \hat{g}(f + m/T) \text{rect}(fT) = T \text{rect}(fT) \quad (2.2)$$

is satisfied,  $g(t)$  of  $G(f)$  is ideal Nyquist with interval  $T$  .



# Proof of the Nyquist criterion of NO ISI

$$g(nT) = \delta(n) = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}$$

$$g(nT) = \delta(n)$$

$$g(t) = \int_{-\infty}^{\infty} G(f) e^{j2\pi ft} df$$

$$g(nT) = \sum_{m=-\infty}^{\infty} \int_{(m-1/2)/T}^{(m+1/2)/T} G(f) e^{j2\pi fnT} df B(f)$$

$$= \int_{-1/(2T)}^{1/(2T)} \sum_{m=-\infty}^{\infty} G\left(f - \frac{m}{T}\right) e^{j2\pi fnT} df$$

$$B(f) \xrightarrow{\text{CFS}} Tg(-nT)$$

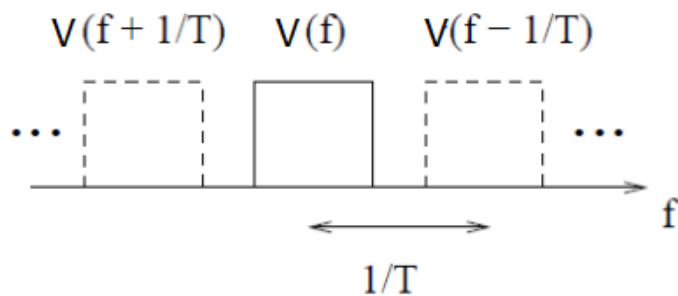
Coefficient of Fourier series

$$\sum_{n=-\infty}^{\infty} G\left(f - \frac{n}{T}\right) = K_0$$

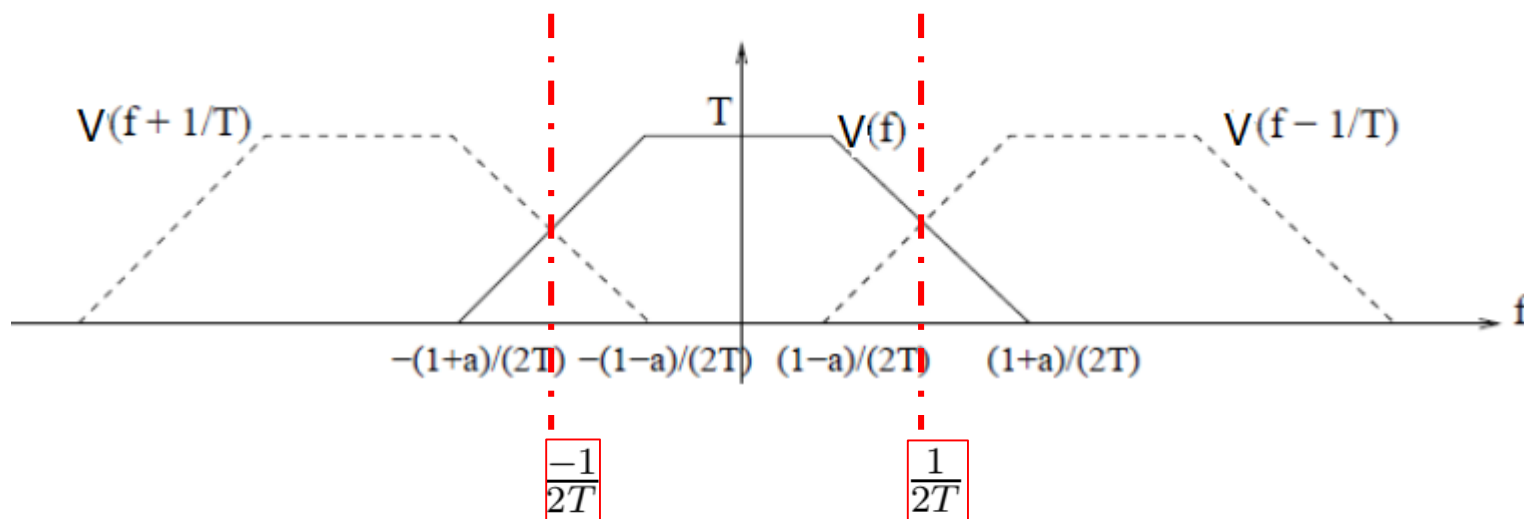
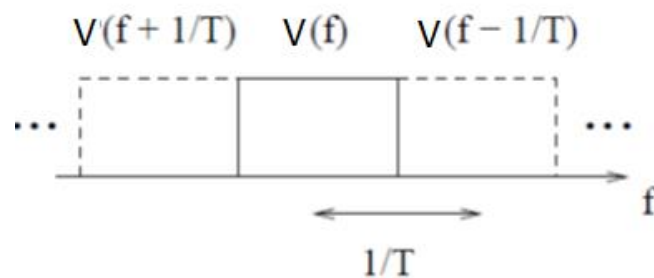
$$\sum_{n=-\infty}^{\infty} g\left(f - \frac{n}{T}\right) = K_0$$



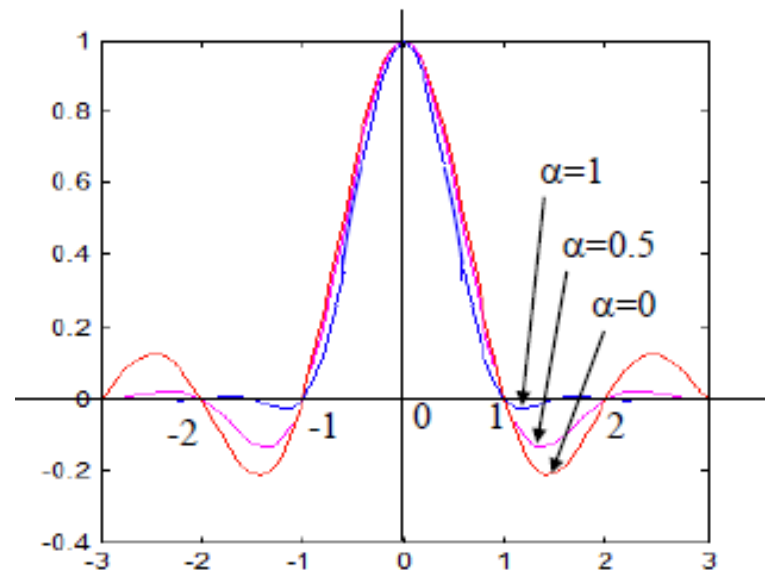
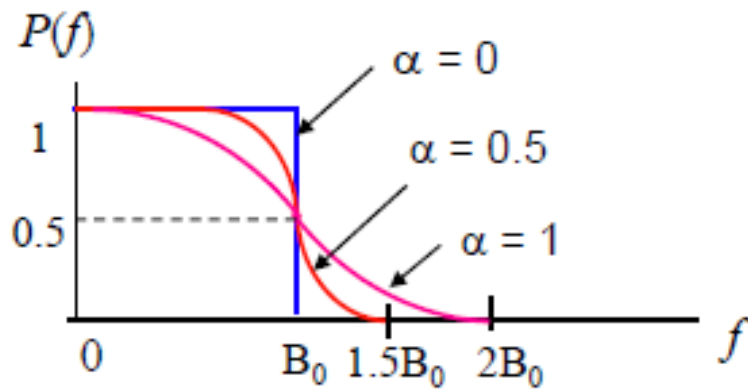
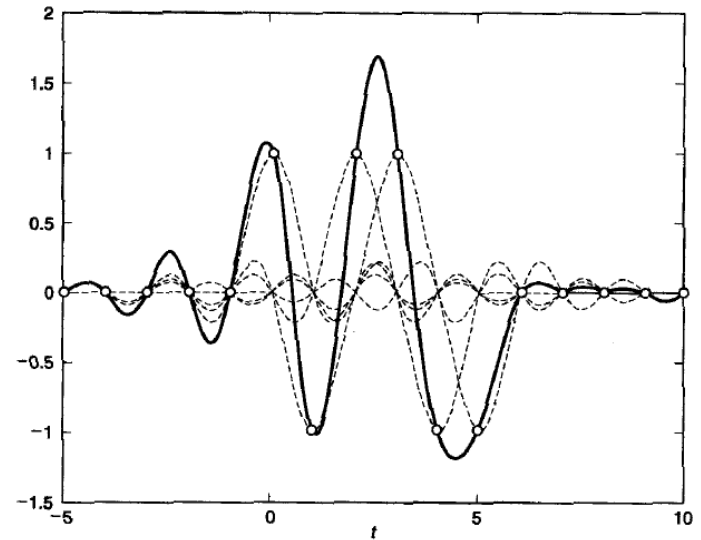
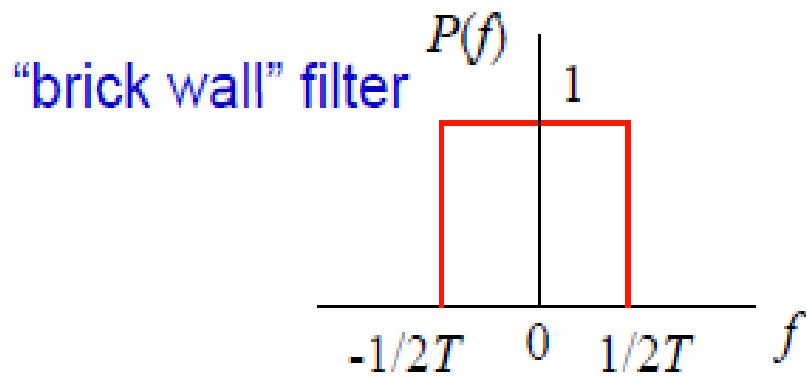
## Not Nyquist



## Nyquist with minimum bandwidth







PAM filters in practice often have **raised cosine** transforms. The raised cosine frequency function, for any given rolloff  $\alpha$  between 0 and 1, is defined by

$$\hat{g}_\alpha(f) = \begin{cases} T, & 0 \leq |f| \leq \frac{1-\alpha}{2T}; \\ T \cos^2 \left[ \frac{\pi T}{2\alpha} \left( |f| - \frac{1-\alpha}{2T} \right) \right], & \frac{1-\alpha}{2T} \leq |f| \leq \frac{1+\alpha}{2T}; \\ 0, & |f| \geq \frac{1+\alpha}{2T}. \end{cases} \quad (2.3)$$

The inverse transform of  $\hat{g}_\alpha(f)$  is

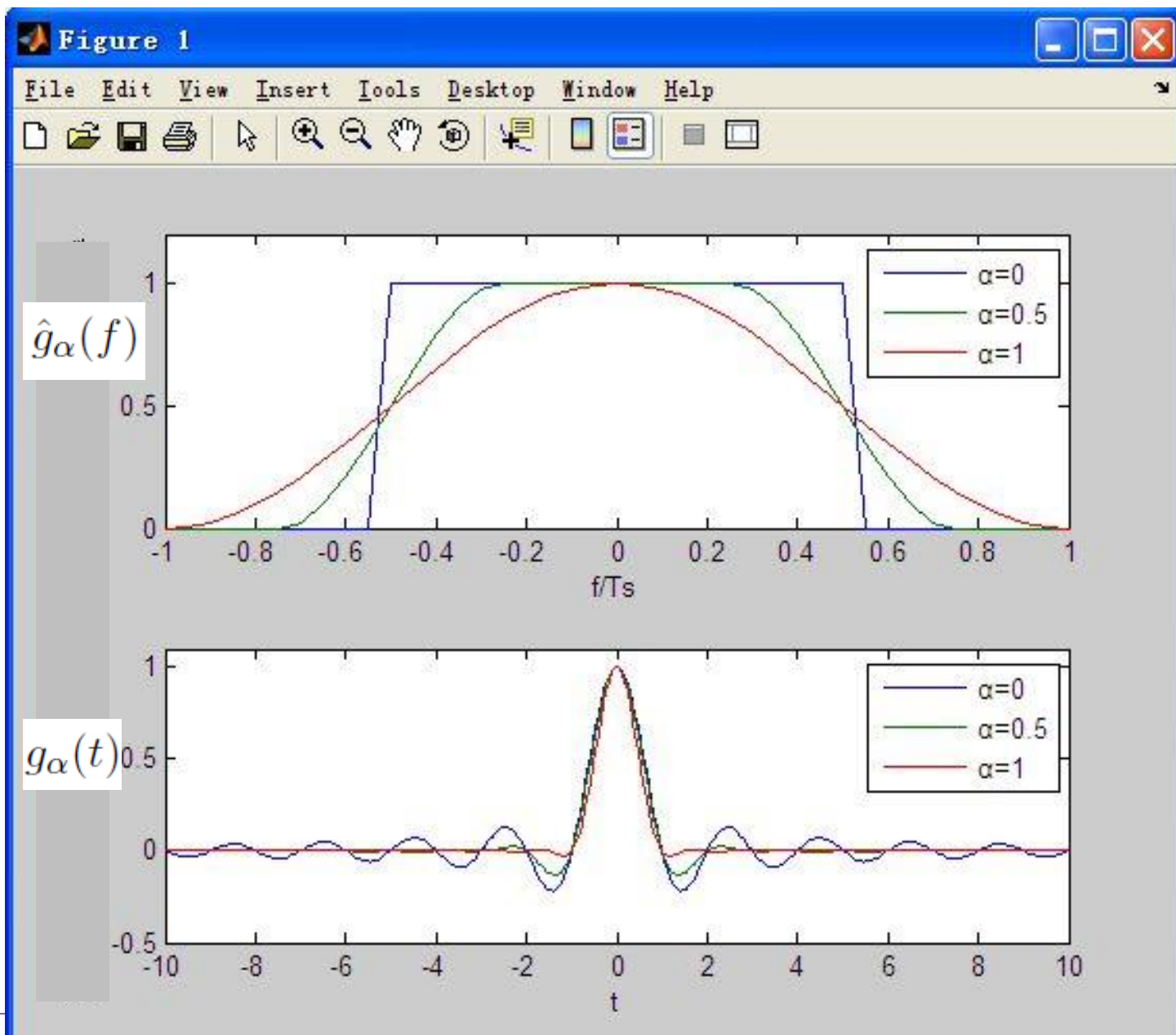
$$g_\alpha(t) = \text{sinc}\left(\frac{t}{T}\right) \frac{\cos(\pi\alpha t/T)}{1 - 4\alpha^2 t^2/T^2}, \quad (2.4)$$

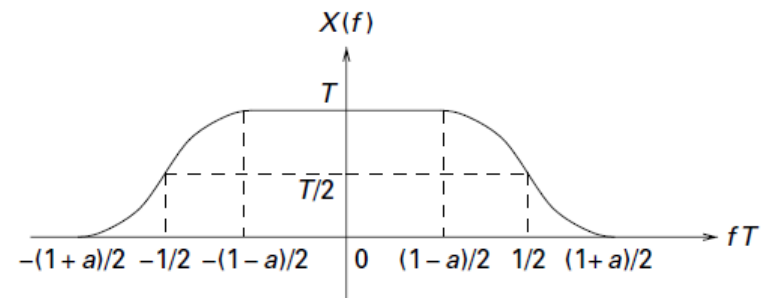
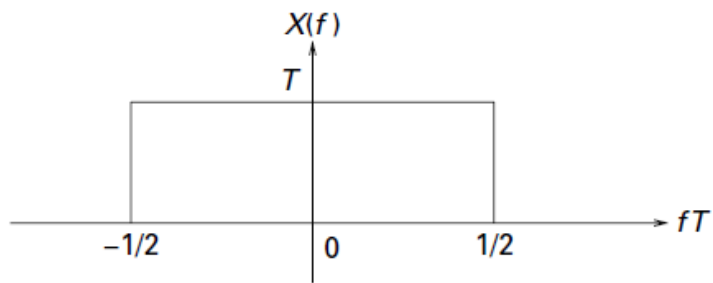
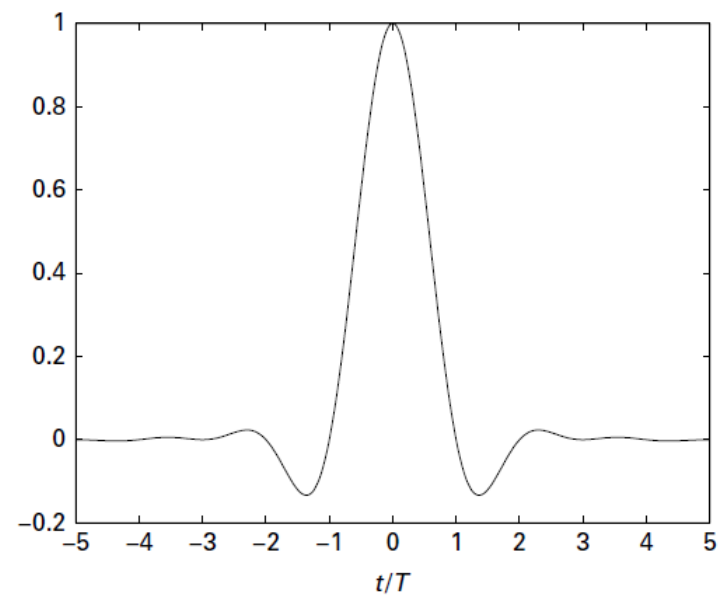
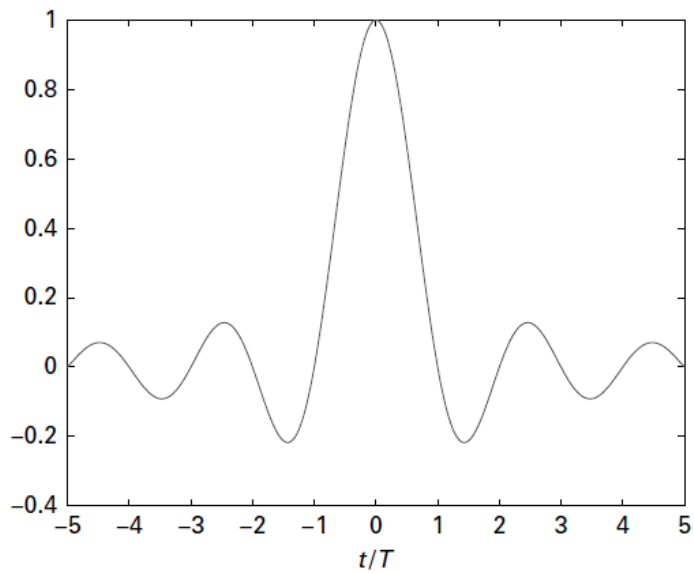
which decays asymptotically as  $1/t^3$ , compared to  $1/t$  for  $\text{sinc}(\frac{t}{T})$

In particular, for a rolloff  $\alpha=1$ ,  $\hat{g}_\alpha(f)$  is nonzero from  $-2W_b = -1/T$  to  $2W_b = 1/T$  and  $g_\alpha(t)$  has most of its energy between  $-T$  and  $T$ .

Rolloffs as sharp as 5–10% are used in current practice







After choosing  $\hat{g}(f) \geq 0$ , the question remains of choosing the transmit filter  $p(t)$  and the receive filter  $q(t)$  subject to  $\hat{p}(f)\hat{q}(f) = \hat{g}(f)$

$$\hat{p}(f)\hat{q}(f) = \hat{g}(f) \geq 0$$

When studying white Gaussian noise later, we will find that  $\hat{q}(f)$  should be chosen to equal  $\hat{p}^*(f)$ . Thus

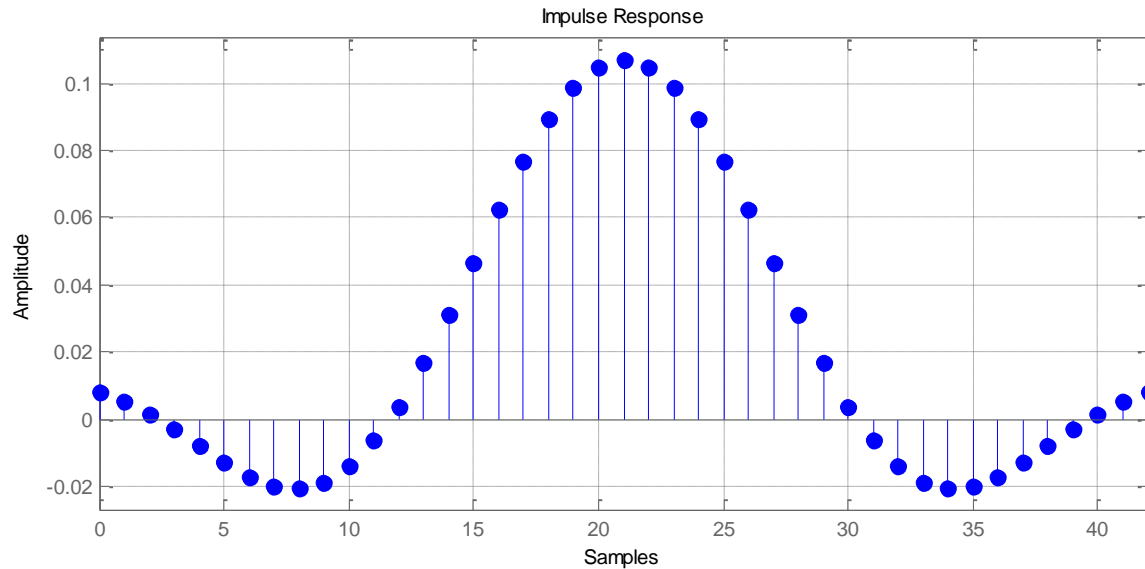
$$|\hat{p}(f)| = |\hat{q}(f)| = \sqrt{\hat{g}(f)}. \quad \hat{q}(f) = \hat{p}^*(f).$$

The phase of  $\hat{p}(f)$  can be chosen in an arbitrary way, but this determines the phase of  $\hat{q}(f)$ .

The filter  $q(t)$  is called the **matched filter** to  $p(t)$ . The matched filter is the optimal linear filter for maximizing the signal-to-noise ratio in the presence of additive white Gaussian noise.



$$|\hat{p}(f)| = |\hat{q}(f)| = \sqrt{\hat{g}(f)}.$$



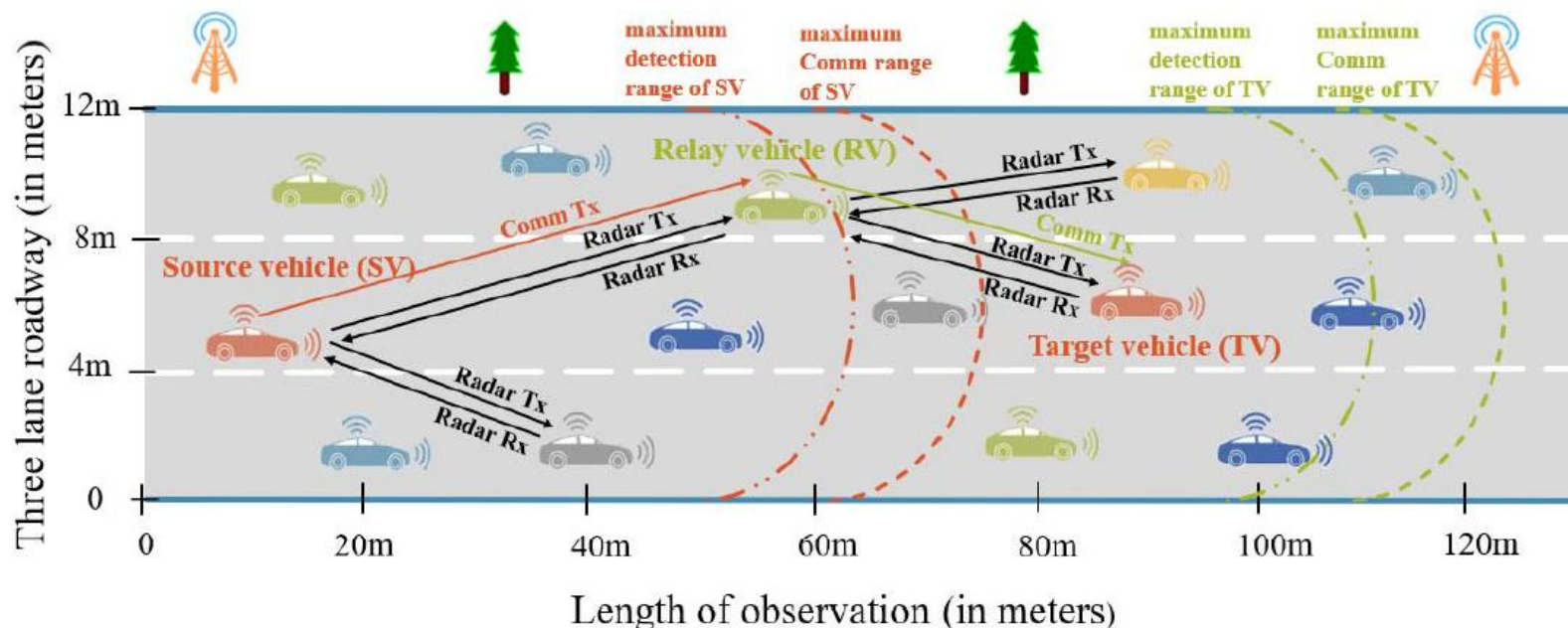
$$p(t) = \begin{cases} \frac{1}{\sqrt{T}} \frac{\sin[\pi(1-\alpha)t/T] + (4\alpha t/T) \cos[\pi(1+\alpha)t/T]}{(\pi t/T) [1 - (4\alpha t/T)^2]}, & t \neq 0, \neq \pm \frac{T}{4\alpha} \\ \frac{1}{\sqrt{T}} \left[ 1 - \alpha + \frac{4\alpha}{\pi} \right], & t = 0 \\ \frac{\alpha}{\sqrt{T}} \left[ \left( 1 + \frac{2}{\pi} \right) \sin\left(\frac{\pi}{4\alpha}\right) + \left( 1 - \frac{2}{\pi} \right) \cos\left(\frac{\pi}{4\alpha}\right) \right], & t = \pm \frac{T}{4\alpha} \end{cases}$$



# Q1: How to design the waveform $p(t)$ ?—additional discussions

## Integrated Sensing and Communications (ISAC) for Vehicular Communication Networks (VCN)

Xiang Cheng<sup>ID</sup>, *Fellow, IEEE*, Dongliang Duan<sup>ID</sup>, *Senior Member, IEEE*, Shijian Gao<sup>ID</sup>, *Member, IEEE*, and Liuqing Yang<sup>ID</sup>, *Fellow, IEEE*

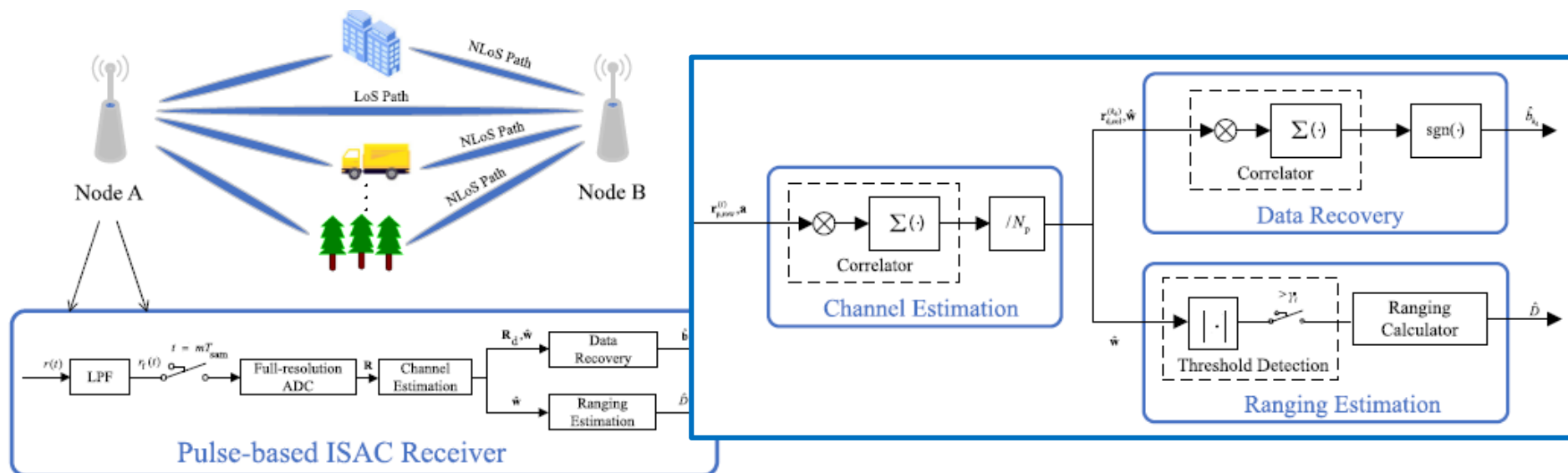


Schematic of a vehicular network where each vehicle can serve sensing and communications simultaneously.

# Q1: How to design the waveform $p(t)$ ?—additional discussions

## Pulse-Based ISAC: Data Recovery and Ranging Estimation for Multi-Path Fading Channels

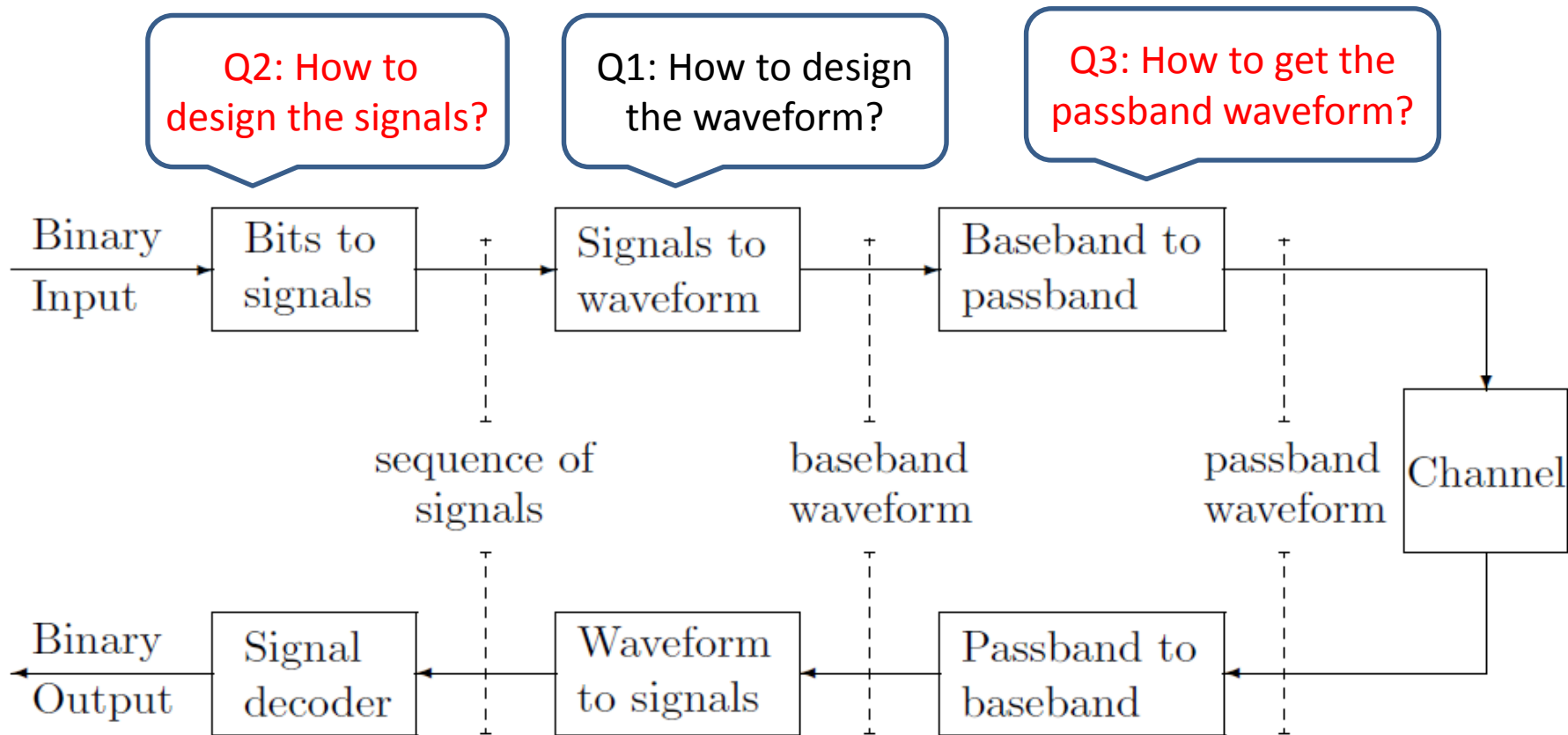
Shusen Cai, Li Chen<sup>ID</sup>, *Senior Member, IEEE*, Yunfei Chen<sup>ID</sup>, *Senior Member, IEEE*,  
Huarui Yin<sup>ID</sup>, *Member, IEEE*, and Weidong Wang<sup>ID</sup>



System Model: a node-to-node ISAC system based on pulsed signals.





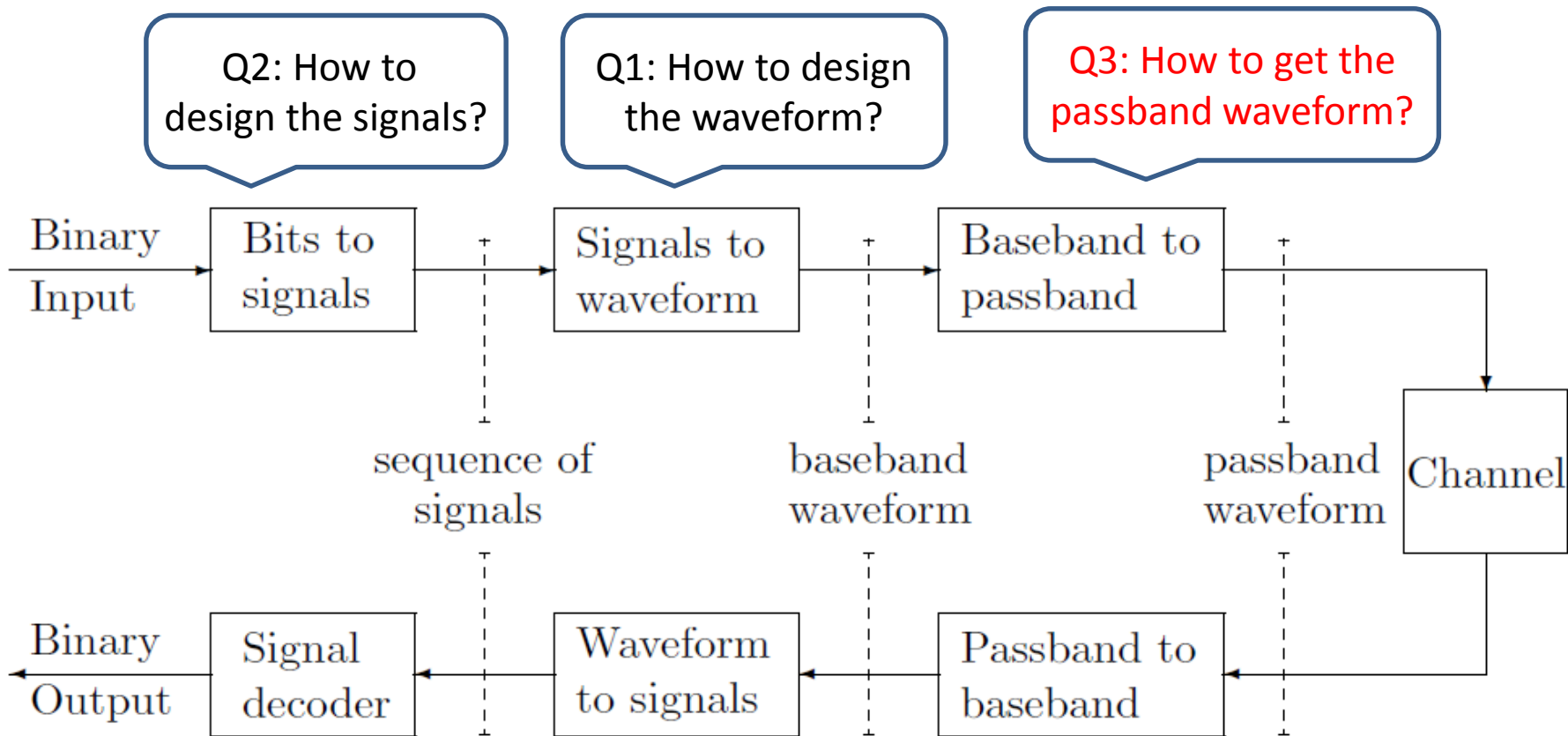


## Why multilevel PAM systems use a standard M-PAM signal set?

Because the Gaussian density drops off so fast with increasing distance, the error probability due to confusion of nearest neighbors drops off equally fast.

Thus error probability is dominated by the points in the constellation that are closest. If the signal points are constrained to have some minimum distance  $d$  between points, it can be seen that the minimum energy  $E_s$  for a given number of points  $M$  is achieved by the standard M-PAM set.

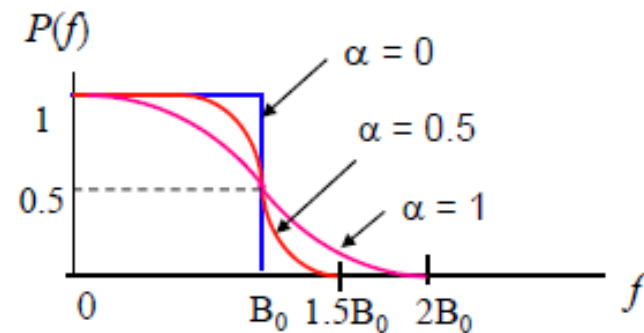
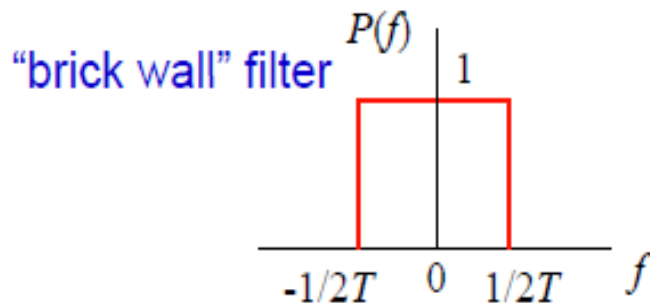




## Baseband to passband and back

The discussion in the previous sections focused on converting a T-spaced sequence of real signals into a real waveform of bandwidth  $B_b$  slightly larger than the Nyquist bandwidth  $W_b = \frac{1}{2T}$ .

This section focuses on converting that baseband waveform into a passband waveform appropriate for the physical medium, regulatory constraints, and avoiding other transmission bands.



## Double-sideband amplitude modulation

The objective of modulating a baseband PAM waveform  $u(t)$  to some high frequency passband around some carrier  $f_c$  is to simply shift  $\hat{u}(f)$  up in frequency to  $\hat{u}(f)e^{2\pi i f_c t}$ .

**Since only real waveforms can actually be transmitted**,  $u(t)$  is also multiplied by the complex conjugate of  $e^{2\pi i f_c t}$ , i.e.,  $e^{-2\pi i f_c t}$ , resulting in the following passband waveform:

$$x(t) = u(t)[e^{2\pi i f_c t} + e^{-2\pi i f_c t}] = 2u(t) \cos(2\pi f_c t), \quad (2.5)$$

$$\hat{x}(f) = \hat{u}(f - f_c) + \hat{u}(f + f_c). \quad (2.6)$$

$u(t)$  is both translated up in frequency by  $f_c$  and also translated down by  $f_c$ .



Note that the entire set of frequencies in  $[-B_b, B_b]$  is both translated up to  $[-B_b + f_c, B_b + f_c]$  and down to  $[-B_b - f_c, B_b - f_c]$ . Thus (assuming  $f_c > B_b$ ) the range of nonzero frequencies occupied by  $x(t)$  is twice as large as that occupied by  $u(t)$ .

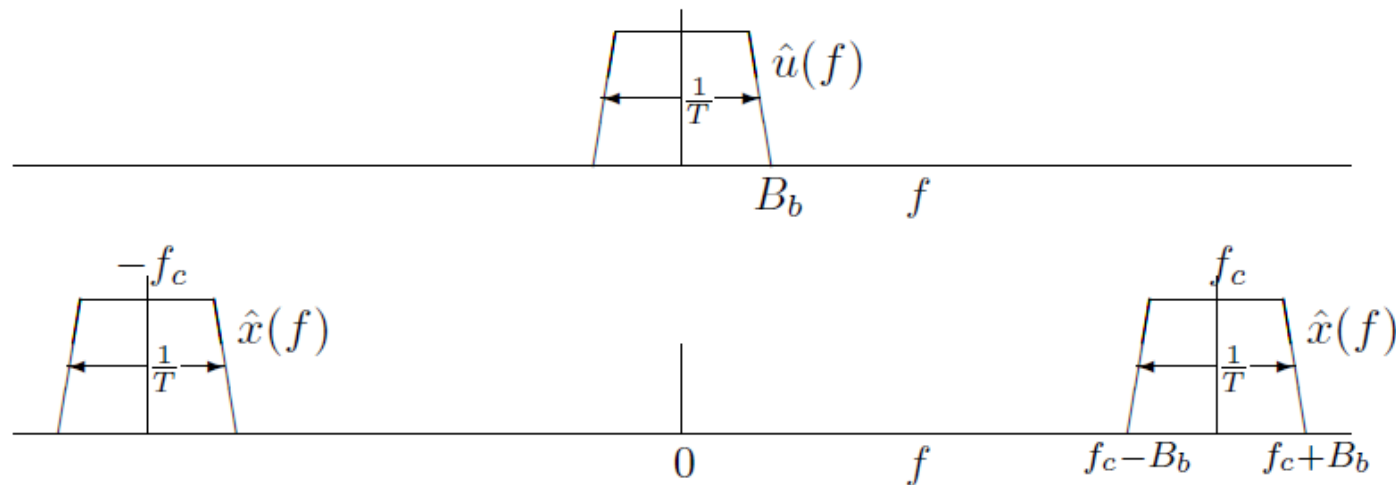
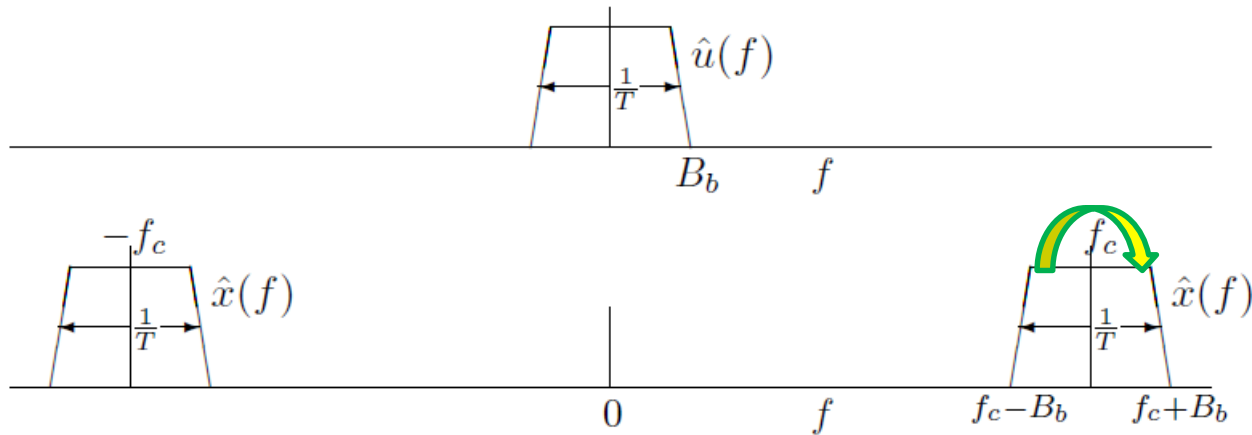


Fig. 2.4: Frequency domain representation of a baseband waveform  $u(t)$  shifted up to a passband around the carrier  $f_c$ .





The passband modulation scheme is called **double-sideband** amplitude modulation.

**Why?** Viewing  $[f_c - B_b, f_c + B_b]$  as two sidebands, the upper,  $[f_c, f_c + B_b]$ , coming from the positive frequency components of  $u(t)$  and the lower,  $[f_c - B_b, f_c]$  from its negative components. If  $u(t)$  is real,  $u(f) = u^*(-f)$ , these two bands are redundant.

→ quite waste of bandwidth

→ quadrature amplitude modulation (QAM) ( $u(t)$  is complex)



## Quadrature amplitude modulation (QAM)

A QAM modulator (see figure 2.5) has the same 3 layers as a PAM modulator. The demodulator performs the inverse of these operations in reverse order.

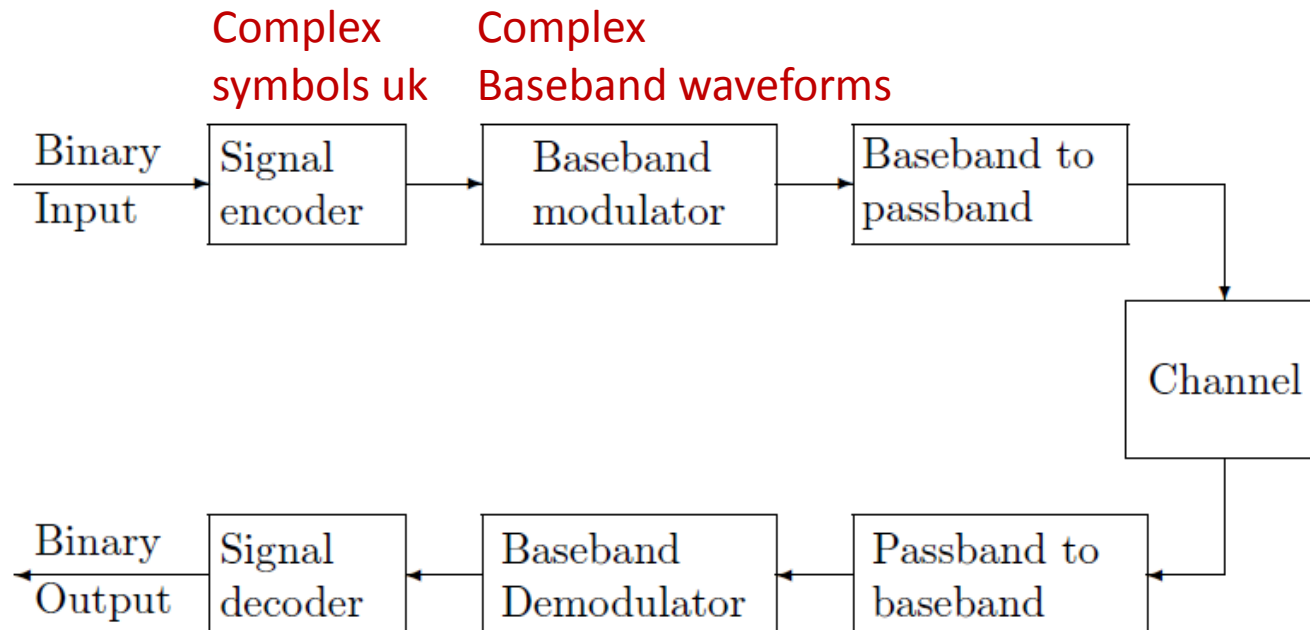


Fig. 2.5: QAM modulator and demodulator.





The complex QAM waveform  $u(t)$  is then shifted up to passband as  $u(t)e^{2\pi if_c t}$ . This waveform is complex and is converted into a real waveform for transmission by adding its complex conjugate. The resulting real passband waveform is then

$$x(t) = u(t)e^{2\pi if_c t} + u^*(t)e^{-2\pi if_c t} \quad (2.7)$$

Note that the passband waveform for PAM is a special case of this in which  $u(t)$  is real.

The passband waveform  $x(t)$  in (2.7) can also be written in the following equivalent ways:

$$\begin{aligned} x(t) &= 2\Re\{u(t)e^{2\pi if_c t}\} \\ &= 2\Re\{u(t)\} \cos(2\pi f_c t) - 2\Im\{u(t)\} \sin(2\pi f_c t). \end{aligned} \quad (2.8)$$



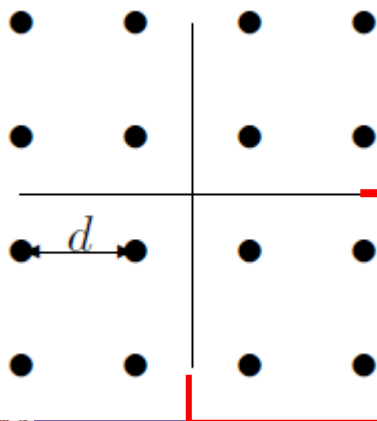
## QAM signal set

The input bit sequence arrives at a rate of  $R$  bps and is converted into a sequence of complex signals  $u_k$  chosen from a signal set  $\mathcal{A}$  of size  $M = |\mathcal{A}| = 2^b$ . Each signal has  $b$  bits. The signal rate is thus  $R_s = R/b$  signals per second, and the signal interval is  $T = 1/R_s = b/R$  sec.

In the case of QAM, the transmitted signals  $u_k$  are complex numbers, rather than real numbers. Alternatively, we may think of each signal as a real 2-tuple in  $\mathbb{R}^2$ . A standard  $(M' \times M')$ -QAM signal set

$$\mathcal{A} = \{(a' + ia'') \mid a' \in \mathcal{A}', a'' \in \mathcal{A}'\},$$

$$M = (M')^2 = 2^b$$



$$\mathcal{A}' = \{-d(M' - 1)/2, \dots, -d/2, d/2, \dots, d(M' - 1)/2\}.$$



The minimum distance between the two-dimensional points is denoted by  $d$ . The average energy per two-dimensional signal, which is denoted as  $E_s$ , is

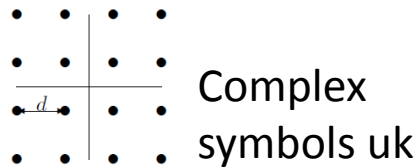
$$E_s = \frac{d^2[(M')^2 - 1]}{6} = \frac{d^2[M - 1]}{6}.$$

In the case of QAM there are many ways to arrange the signal points other than on a square grid as above.

However, finding the optimal signal set to minimize  $E_s$  for practical values of  $M$  and a given  $d$  is a messy and ugly problem.

Now the standard QAM signal set is almost universally used in practice and will be assumed in what follows.

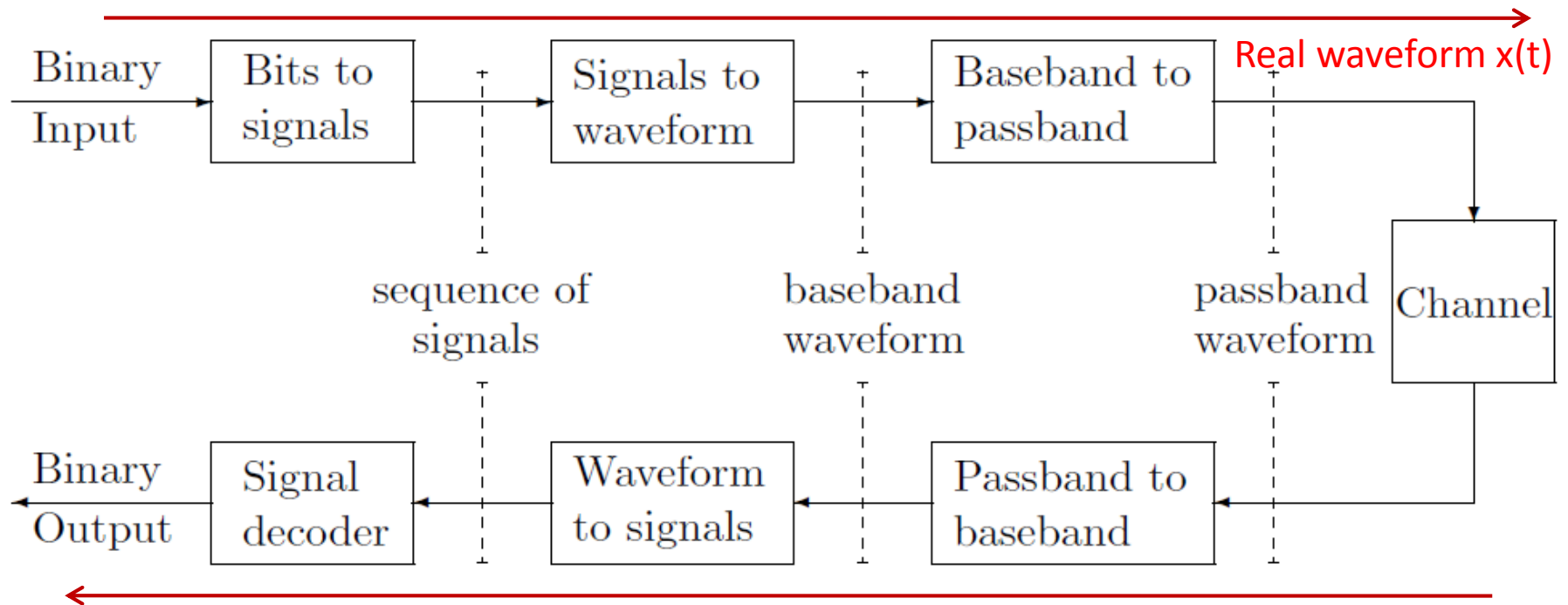




$$u(t) = \sum_{k \in \mathbb{Z}} u_k p(t - kT).$$

$$1. x^+(t) = u(t)e^{2\pi i f_c t}$$

$$2. x(t) = [x^+(t)]^* + x^+(t).$$



## QAM: baseband to passband and back

The expressions for  $x(t)$  and the frequency representation are given as follows

$$u(t) = \sum_{k \in \mathbb{Z}} u_k p(t - kT).$$

$$x(t) = u(t)e^{2\pi i f_c t} + u^*(t)e^{-2\pi i f_c t}$$

$$\begin{aligned} x(t) &= 2\Re\{u(t)e^{2\pi i f_c t}\} \\ &= 2\Re\{u(t)\} \cos(2\pi f_c t) - 2\Im\{u(t)\} \sin(2\pi f_c t). \end{aligned}$$

As with PAM,  $u(t)$  has a nominal baseband bandwidth  $W_b = \frac{1}{2T}$ . The actual baseband bandwidth  $B_b$  exceeds  $W_b$  by some small rolloff factor.



## Implementation of QAM

$$x(t) = u(t)e^{2\pi i f_c t} + u^*(t)e^{-2\pi i f_c t}$$

The positive frequency expression  $x^+(t) = u(t)e^{2\pi i f_c t}$  is a complex multiplication of complex waveforms which requires **4 real multiplications**.

$$\begin{aligned} x(t) &= 2\Re\{u(t)e^{2\pi i f_c t}\} \\ &= 2\Re\{u(t)\} \cos(2\pi f_c t) - 2\Im\{u(t)\} \sin(2\pi f_c t). \end{aligned}$$

From an implementation standpoint, the baseband waveform  $u(t)$  is usually **implemented as two real waveforms**,  $\Re\{u(t)\}$  and  $\Im\{u(t)\}$ . These are then modulated up to passband using multiplication by in-phase and quadrature-phase carriers. (**2 real multiplications**)



For real baseband waveform  $p(t)$

$$\Re\{u(t)\} = \sum_k \Re\{u_k\} p\left(\frac{t}{T} - k\right),$$

$$\Im\{u(t)\} = \sum_k \Im\{u_k\} p\left(\frac{t}{T} - k\right).$$

Letting  $u'_k = \Re\{u_k\}$  and  $u''_k = \Im\{u_k\}$ , the transmitted passband waveform becomes

$$x(t) = 2 \cos(2\pi f_c t) \left( \sum_k u'_k p(t - kT) \right) - 2 \sin(2\pi f_c t) \left( \sum_k u''_k p(t - kT) \right) \quad (2.9)$$



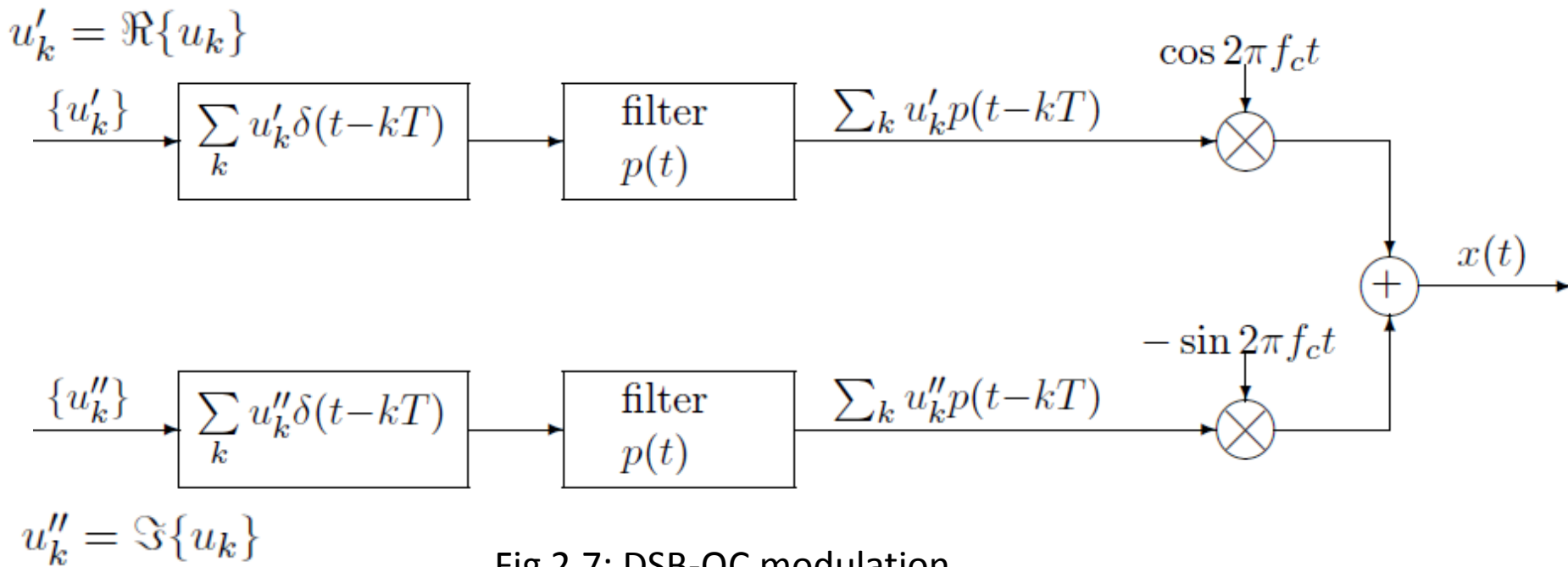


Fig 2.7: DSB-QC modulation

$$x(t) = 2 \cos(2\pi f_c t) \left( \sum_k u'_k p(t-kT) \right) - 2 \sin(2\pi f_c t) \left( \sum_k u''_k p(t-kT) \right).$$

This realization of QAM is called double-sideband quadrature-carrier (DSB-QC) modulation.





Using the trigonometric identities  $2 \cos^2(\alpha) = 1 + \cos(2\alpha)$ ,  $2 \sin(\alpha)\cos(\alpha) = \sin(2\alpha)$ , and  $2\sin^2(\alpha) = 1 - \cos(2\alpha)$ , these terms can be written as

$$\begin{aligned} x(t) \cos(2\pi f_c t) &= \boxed{\Re\{u(t)\}} + \Re\{u(t)\} \cos(4\pi f_c t) + \Im\{u(t)\} \sin(4\pi f_c t), \\ -x(t) \sin(2\pi f_c t) &= \boxed{\Im\{u(t)\}} - \Re\{u(t)\} \sin(4\pi f_c t) + \Im\{u(t)\} \cos(4\pi f_c t). \end{aligned}$$

Filtering out the double frequency terms then yields  $\Re\{u(t)\}$  and  $\Im\{u(t)\}$ .

**How?**



$$\begin{aligned}
 x(t) \cos(2\pi f_c t) &= \Re\{u(t)\} + \Re\{u(t)\} \cos(4\pi f_c t) + \Im\{u(t)\} \sin(4\pi f_c t), \\
 -x(t) \sin(2\pi f_c t) &= \Im\{u(t)\} - \Re\{u(t)\} \sin(4\pi f_c t) + \Im\{u(t)\} \cos(4\pi f_c t).
 \end{aligned}$$

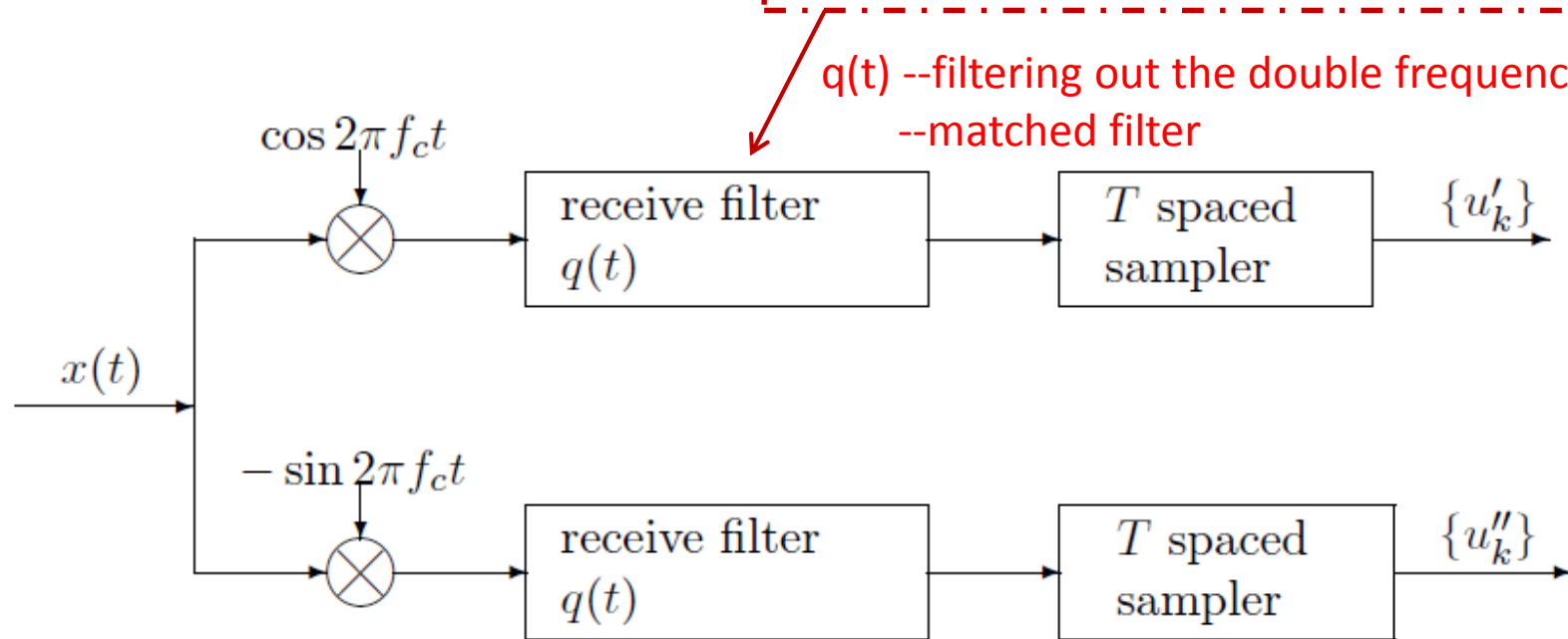


Fig. 3.8: DSB-QC demodulation

Since  $p(t)$  is chosen so that  $|\hat{p}(f)|^2$  satisfies the Nyquist criterion, then the receiver filter should satisfy  $q(t) = p^*(-t)$ .



## Why we deal with the signals in baseband instead of passband?

Is it equivalent to deal with the baseband and the passband?

- The sampling rate in the baseband is much lower. It is easier to perform complicated DSP algorithms in complex baseband.
- Software defined radio : baseband processing + up/down frequency conversion. It allows using cheap DSP and FPGA devices in the baseband without considering the working frequencies.



Why we deal with the signals in baseband instead of passband?  
**Is it equivalent to deal with the baseband and the passband?**

The relationship of the passband signal's and basedband signal's spectrum:

$$S_p(f) = \frac{S(f - f_c) + S^*(-f - f_c)}{\sqrt{2}}$$

$$\text{where } S(f) = \sqrt{2}S_p^+(f + f_c) \quad S_p^+(f) = S_p(f)1_{\{f>0\}}$$



Why we deal with the signals in baseband instead of passband?  
**Is it equivalent to deal with the baseband and the passband?**

The relationship of the passband signal's and basedband signal's spectrum:

$$S_p(f) = \frac{S(f - f_c) + S^*(-f - f_c)}{\sqrt{2}}$$

$$\text{where } S(f) = \sqrt{2}S_p^+(f + f_c) \quad S_p^+(f) = S_p(f)1_{\{f>0\}}$$

**Proof:** let  $v(t) = \sqrt{2}s(t)e^{j2\pi f_c t} \longrightarrow V(f) = \sqrt{2}S(f - f_c)$

The real passband signal

$$s_p(t) = \text{Re}\{v(t)\} = \frac{v(t) + v^*(t)}{2}$$

$$S_p(f) = \frac{V(f) + V^*(-f)}{2} = \frac{S(f - f_c) + S^*(-f - f_c)}{\sqrt{2}}$$



## The relationship of the passband signal's and basedband signal's spectrum:

$$S_p(f) = \frac{S(f - f_c) + S^*(-f - f_c)}{\sqrt{2}}$$

$$\text{where } S(f) = \sqrt{2}S_p^+(f + f_c) \quad S_p^+(f) = S_p(f)1_{\{f>0\}}$$

---

**Proof: continue--**

We already have  $S_p(f) = \frac{S(f - f_c) + S^*(-f - f_c)}{\sqrt{2}}$

$$\begin{aligned} S_p^+(f) &= \frac{1}{\sqrt{2}} S(f - f_c)1_{\{f>0\}} + \frac{1}{\sqrt{2}} S^*(-f - f_c)1_{\{f>0\}} \\ &= \frac{1}{\sqrt{2}} S(f - f_c)1_{\{f>0\}} \end{aligned}$$

=0

$$S_p^+(f + f_c) = \frac{1}{\sqrt{2}} S(f)1_{\{f+f_c>0\}} = \frac{1}{\sqrt{2}} S(f)$$



Why we deal with the signals in baseband instead of passband?  
**Is it equivalent to deal with the baseband and the passband?**

## 1. Filtering in baseband is equivalent to the filtering in passband.

$$y_p(t) = s_p(t) * h_p(t) \quad \Rightarrow \quad Y_p(f) = S_p(f)H_p(f) \quad \Rightarrow \quad Y_p^+(f) = S_p^+(f)H_p^+(f)$$

$$Y(f) = \sqrt{2}Y_p^+(f + f_c)$$

$$= \sqrt{2}S_p^+(f + f_c)H_p^+(f + f_c)$$

$$= \frac{1}{\sqrt{2}} S(f)H(f)$$

$$S(f) = \sqrt{2}S_p^+(f + f_c)$$

$$y(t) = \frac{1}{\sqrt{2}} s(t) * h(t)$$



Why we deal with the signals in baseband instead of passband?  
**Is it equivalent to deal with the baseband and the passband?**

$$\langle u_p, v_p \rangle = \langle u_c, v_c \rangle + \langle u_s, v_s \rangle = \text{Re}(\langle u, v \rangle)$$

Let  $u_p = v_p = s_p$ ,  $u = v = s$ , we can find

$$\text{Re}(\langle s, s \rangle) = \text{Re}(\|s\|^2) = \|s\|^2$$

$$\Rightarrow \|s_p\|^2 = \|s\|^2$$

$$\begin{aligned} n_p(t) &= \sqrt{2}n_c(t)\cos 2\pi f_c t - \sqrt{2}n_s(t)\sin 2\pi f_c t \\ &= \text{Re}\left\{\sqrt{2}(n_c(t) + jn_s(t))e^{j2\pi f_c t}\right\} \end{aligned}$$

$$\Rightarrow P_{n_p} = E\{|n_c(t) + jn_s(t)|^2\} = P_n$$

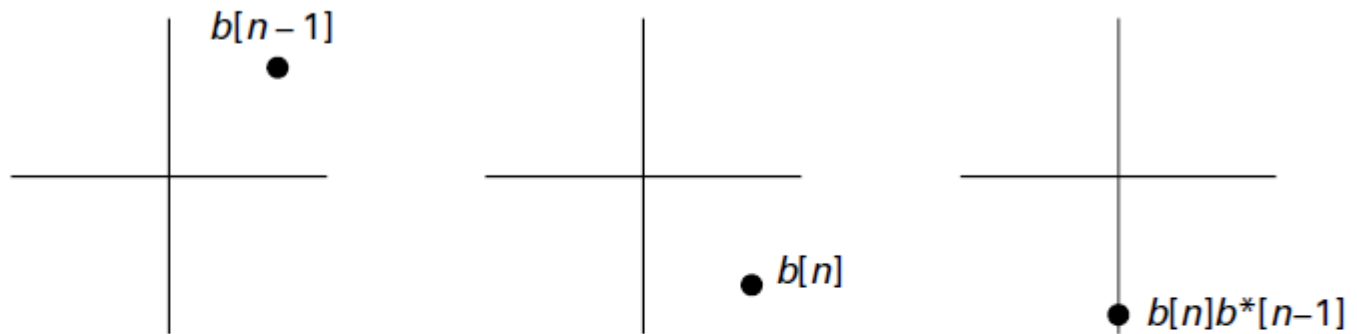
**2. Singal-to-noise ratio (SNR) is the same**





## Differential Modulation (brief introduction)

Transmitted symbols



Noiseless received samples

