

# DIGITAL COMMUNICATION THROUGH BAND-LIMITED CHANNELS

## (DC III)

DIRK DAHLHAUS

COMMUNICATIONS LABORATORY [COMLAB]

Summer Semester 2021

# Organisation

- time: Thursdays 05:00-06:30 pm, Fridays 08:30-10:00 am
- online lecturing via Zoom <https://uni-kassel.zoom.us/j/95576869612>,  
Meeting ID: 955 7686 9612; Moodle password: NTDHD\_DC3\_SS2021
- workload: 60 hours course attendance, 120 hours self-study
- language: English, oral exam (30 minutes either in English or in German)
- exercises: are integrated in the lecture (on demand)
- regular attendance of the lecture *and* the exercises is mandatory to pass the exam
- the lecture is based on the books (well-known in communications)
  - John G. Proakis, „Digital Communications“, McGraw-Hill, 4th ed.,  
ISBN 0-07-118183-0
- upon passing the exam, you obtain **6 credit points for lecture and exercises.**
- Digital Communication Through Band-Limited Channels (4 credits) is part of module *Digital Communications R1b* in addition to Digital Communication Through Band-Limited Channels Lab (2 credits), Simulation of Digital Communication Systems Using Matlab (3 credits) and Seminar of Medium Access Control Protocols in Wireless Communications (3 credits). Overall the module comprises **12 credits.**

# Table of Contents

- 1 Introduction
- 2 Carrier and Symbol Synchronization
  - Carrier Phase ( $\phi$ ) Estimation
  - Symbol Timing ( $\tau$ ) Estimation
  - Joint Estimation of  $\phi$  and  $\tau$
- 3 Signal Design for Band-Limited Channels
  - Characterization of Band-Limited Channels
  - Signal Design for No ISI-Nyquist Criterion
  - Signal Design with Controlled ISI
  - Probability of Error in Detection of PAM
  - Modulation Codes for Spectrum Shaping
- 4 Communication Through Band-Limited Linear Filter Channels
  - Optimum Receiver for AWGN Channels with ISI
  - Linear Equalization
  - Decision-Feedback Equalization
  - Equalization at the Transmitter
- 5 Multicarrier and Multichannel Systems
  - ISI in Single Carrier Systems
  - Design Criteria for Broadband System
  - Basic Principle of OFDM Signaling

# Table of Contents

- 1 Introduction
- 2 Carrier and Symbol Synchronization
  - Carrier Phase ( $\phi$ ) Estimation
  - Symbol Timing ( $\tau$ ) Estimation
  - Joint Estimation of  $\phi$  and  $\tau$
- 3 Signal Design for Band-Limited Channels
  - Characterization of Band-Limited Channels
  - Signal Design for No ISI-Nyquist Criterion
  - Signal Design with Controlled ISI
  - Probability of Error in Detection of PAM
  - Modulation Codes for Spectrum Shaping
- 4 Communication Through Band-Limited Linear Filter Channels
  - Optimum Receiver for AWGN Channels with ISI
  - Linear Equalization
  - Decision-Feedback Equalization
  - Equalization at the Transmitter
- 5 Multicarrier and Multichannel Systems
  - ISI in Single Carrier Systems
  - Design Criteria for Broadband System
  - Basic Principle of OFDM Signaling

# Overview

Scope of the lectures *Introduction to Digital Communications (DCI)* and *Introduction to Information Theory and Coding (DCII)*:

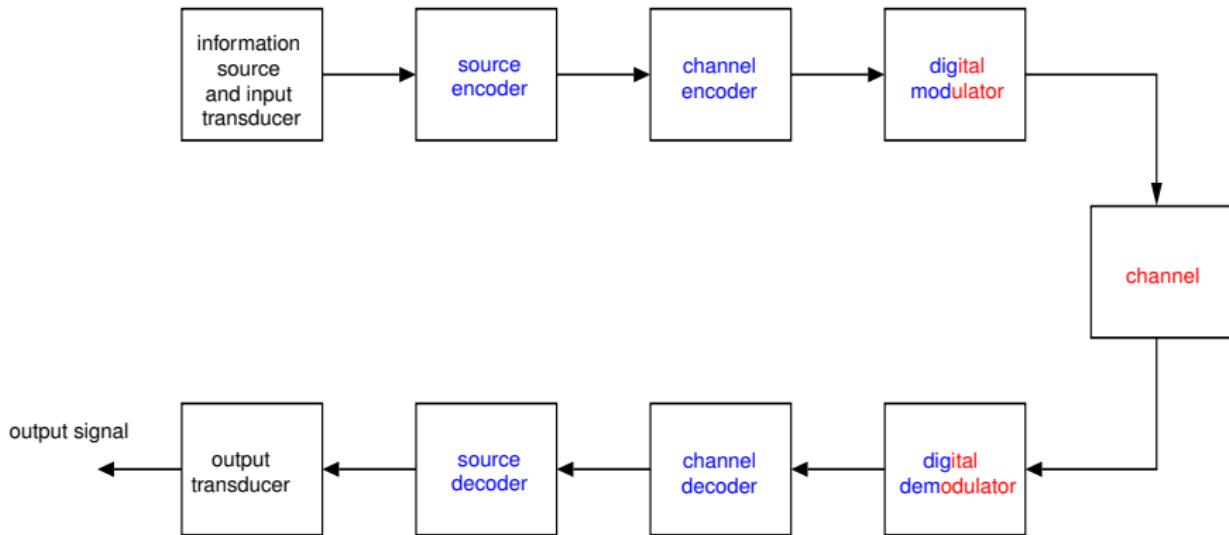


Fig. 1.1-1 Generic model of a digital communication system.

- Before we consider new topics, like e.g. channel parameter estimation and transmission over band-limited channels, let us recap the essentials of the courses *Introduction to Digital Communications (DCI)* and *Introduction to Information Theory and Coding (DCII)*.

- **Introduction to Digital Communications (DCI):**

- ① Overview of digital communication systems and OSI model
- ② Channel models
- ③ Fourier transform
- ④ Basics in random variables and stochastic processes
- ⑤ Signal representation in complex baseband
- ⑥ Memoryless modulation
- ⑦ Optimum binary detection in AWGN

- **Introduction to Information Theory and Coding (DCII):**

- ① Basics in information theory
- ② Channel capacity
- ③ Channel coding: block and convolutional codes
- ④ Hard and soft decision decoding
- ⑤ Source coding: discrete memoryless and analog sources

## Contents of **Digital Communication Through Band-Limited Channels (DCIII)**:

- Introduction
- Carrier and symbol synchronization
- Signal design for band-limited channels
- Communication through band-limited linear filter channels
- Intersymbol interference
- Adaptive equalization
- Multicarrier communications

What is the motivation for this table of contents?

- Until now, we have assumed idealized properties of the transceiver and the channel for the modem part of the physical layer in contrast to real-world situations requiring additional measures to account for the real-world effects.
- Idealized assumptions include
  - perfect channel state information (CSI) at the receiver such as carrier phase and symbol synchronization
    - ⇒ CSI is never perfect (fading, drifts, interference, nonlinearities etc)
    - ⇒ estimation schemes for carrier phase and symbol timing (with/without knowledge of symbol values)

- Idealized assumptions include (cont.)
  - **AWGN channel**, i.e., apart from the thermal noise, the transmission channel is ideal (constant amplitude spectrum, linear phase, infinite bandwidth)
    - ⇒ for sufficiently large bandwidth/time, the channel is always frequency-selective/time-selective
    - ⇒ assume **band-limited channel characteristics**
  - **missing interference and channel fading**
    - ⇒ there are **different forms of interference** (intersymbol interference, nonlinearities (self-interference), multiple-access interference, modulation-specific interference e.g. intercarrier interference) and **time-/frequency selective fading in wireless transmission**
    - ⇒ extend the model of the received signal in order to take into account the most important type of interference, namely **intersymbol interference**

- Idealized assumptions include (cont.)

- **performance metrics like bit-error probabilities being independent of channel estimation or equalizer schemes**
  - ⇒ performance depends critically on channel estimation errors and the properties of equalizers employed
  - ⇒ take into account the estimation errors in the performance characterization in order not to overestimate the achievable performance
- **time-invariant channel characteristics (in real communication systems, in particular in wireless/mobile radio transmission, time-invariance is a substantial oversimplification)**
  - ⇒ there is always a time constant (sometimes called *coherence time of the channel*) limiting the time scale for a stationary operation of the transceiver
  - ⇒ the operation has to be **adapted to the instantaneous channel characteristics**

- Idealized assumptions include (cont.)
  - an oversimplification as far as transceiver complexity is concerned (often we have to find a trade-off between performance and complexity)
    - ⇒ the complexity is limited e.g. by battery lifetime and performance depends critically on the type of implementation (e.g. floating-point vs. fixed-point arithmetics).
    - ⇒ compare solutions of different complexities to each other and select the one providing the best trade-off e.g. **multicarrier transmission**

# Table of Contents

- 1 Introduction
- 2 Carrier and Symbol Synchronization
  - Carrier Phase ( $\phi$ ) Estimation
  - Symbol Timing ( $\tau$ ) Estimation
  - Joint Estimation of  $\phi$  and  $\tau$
- 3 Signal Design for Band-Limited Channels
  - Characterization of Band-Limited Channels
  - Signal Design for No ISI-Nyquist Criterion
  - Signal Design with Controlled ISI
  - Probability of Error in Detection of PAM
  - Modulation Codes for Spectrum Shaping
- 4 Communication Through Band-Limited Linear Filter Channels
  - Optimum Receiver for AWGN Channels with ISI
  - Linear Equalization
  - Decision-Feedback Equalization
  - Equalization at the Transmitter
- 5 Multicarrier and Multichannel Systems
  - ISI in Single Carrier Systems
  - Design Criteria for Broadband System
  - Basic Principle of OFDM Signaling

# Carrier and Symbol Synchronization

- In a digital communication system, there are two fundamental synchronization tasks to be carried out at the receiver before a reliable information retrieval can be achieved:
  - ① The **delay in the propagation channel** results in a **carrier offset** which must be estimated at the receiver in case of **phase-coherent detection** ⇒ **carrier phase estimation**
  - ② In every detector type, some form of **sampling** takes place, in most cases with a rate of one sample per symbol time to obtain **decision variables** for post-processing ⇒ **symbol timing estimation**
- In the case of perfect carrier phase and symbol timing estimation, the resulting overall channel in the complex baseband can be represented as an **additive white Gaussian noise (AWGN)** channel. Therefore, we expect a certain performance loss of a real-world system as compared to the corresponding AWGN scenario.
- Below, we will first discuss the general approach for **signal parameter estimation**, which is treated in greater detail in the lecture *Introduction to Signal Detection and Estimation* in the summer term. Here, we will limit ourselves to **maximum-likelihood (ML) based parameter estimation** without prior information about the statistics of the parameter to be estimated.
- Subsequently, the ML schemes will be used for **carrier phase estimation** and **symbol timing estimation**, where we first present separate estimation schemes and finally a joint scheme for both tasks.

## Signal Parameter Estimation

- In order to study the principal procedure for estimating a parameter like e.g. the carrier phase or the propagation delay, we first consider a simple model of the received signal given by

$$r(t) = s(t - \tau) + n(t) = \Re \left\{ s_\ell(t - \tau) e^{j2\pi f_c(t-\tau)} + z(t) e^{j2\pi f_c t} \right\},$$

where  $s(t) = \Re \{ s_\ell(t) e^{j2\pi f_c t} \}$  is the signal being transmitted,  $s_\ell(t)$  is the equivalent signal in the complex baseband,  $n(t)$  is an AWGN process with spectral power density  $N_0/2$  and complex baseband representation  $z(t)$ ,  $f_c$  denotes the carrier frequency and  $\tau$  is the **propagation delay caused by the channel, which is to be estimated.**

- From the signal model above, it seems that  $\tau$  is the only parameter to be estimated, since we can write

$$r(t) = \Re \left\{ [s_\ell(t - \tau) e^{j\phi} + z(t)] e^{j2\pi f_c t} \right\},$$

where the carrier phase  $\phi$  is a function of  $\tau$  given by  $\phi = -2\pi f_c \tau$ .

- The aforementioned model is an oversimplified one, since, in general, the front-end of a receiver introduces additional phase errors which do not depend solely on the propagation delay. In addition, as can be seen from  $\phi = -2\pi f_c \tau$ , a small error in an estimation of  $\tau$  causes a large error in the estimation of  $\phi$  due to the usually large value of the carrier frequency  $f_c$ .
- As a result, we have to abandon the above simple model and replace it by a model including two independent parameters  $\tau$  and  $\phi \neq -2\pi f_c \tau$  to be estimated independently.
- Upon defining the parameter vector  $\Psi = [\phi, \tau]$ , we have the

$$r(t) = s(t; \Psi) + n(t),$$

and our objective is to calculate an estimate  $\hat{\Psi}$  of the parameter vector  $\Psi$  for an observation  $r(t)$  within an interval of length  $T_0$ .

- To formulate the sought for estimator, we first use a Karhunen-Loève expansion to represent  $r(t)$  in an  $N$ -dimensional signal space spanned by the orthonormal functions  $\{f_n(t)\}$ . We obtain a random vector

$$\mathbf{r} = [r_1, r_2, \dots, r_N]^T,$$

where  $r_n = \langle r(t), f_n(t) \rangle = \int_{T_0} r(t) f_n(t) dt$  is the inner product of the received signal and the  $n$ -th base function.

- In general, there is no statistical model about  $\Psi$ , so that it is usually modelled as a **deterministic unknown parameter vector**. In this case, due to the properties of the AWGN process, the observation vector  $\mathbf{r}$  is a real multivariate Gaussian random vector with

$$\mathbf{r} \sim p(\mathbf{r}; \Psi) = \mathcal{N}\left(\mathbf{s}(\Psi); \frac{N_0}{2} \mathbf{I}_N\right),$$

where  $\mathcal{N}(\mu; \Sigma)$  is a multivariate Gaussian PDF with expectation  $\mu$  and covariance matrix  $\Sigma$ ,  $\mathbf{s}(\Psi)$  is the signal vector in the  $N$ -dimensional signal space and  $\mathbf{I}_N$  is an identity matrix of dimension  $N \times N$ .

- Since we have not modeled  $\Psi$  as a random vector, there is no well-defined notion of probability w.r.t.  $\Psi$ , which we could use to define an objective function to be optimized in statistical terms.
- However, there is an intuitive procedure to find an estimation scheme for  $\Psi$ :
  - Obviously, we have observed  $\mathbf{r}$ , whose component values depend on  $\Psi$  and the values of the noise vector  $\mathbf{n}$  in the signal space.
  - We can argue that the value of  $\Psi$  which maximizes  $p(\mathbf{r}; \Psi)$  is the most likely (not the most probable!) among all possible values that  $\Psi$  might take, since  $\mathbf{r}$  has indeed been observed.
- As a consequence, we can define the maximum-likelihood (ML) estimate of  $\Psi$  in the  $N$ -dimensional signal space as

$$\hat{\Psi}_{\text{ML},N} = \arg \max_{\Psi \in \Omega_\Psi} p(\mathbf{r}; \Psi),$$

where  $\Omega_\Psi$  is the set of all possible values for  $\Psi$ .

- Note that we have

$$p(\mathbf{r}; \Psi) = (\pi N_0)^{-N/2} \exp \left( - \sum_{n=1}^N \frac{(r_n - s_n(\Psi))^2}{N_0} \right).$$

- In order to simplify the expression for the estimation scheme, we can apply any strictly monotonically increasing function to  $p(\mathbf{r}; \Psi)$  prior to the maximization without changing the value of  $\hat{\Psi}_{ML,N}$ . Choosing the natural logarithm  $\ln(\cdot)$  as this function and noting that the factor  $(\pi N_0)^{-N/2}$  does not depend on  $\Psi$ , we have

$$\begin{aligned}\hat{\Psi}_{ML,N} &= \arg \max_{\Psi \in \Omega_\Psi} p(\mathbf{r}; \Psi) = \arg \max_{\Psi \in \Omega_\Psi} \ln(p(\mathbf{r}; \Psi)) \\ &= \arg \max_{\Psi \in \Omega_\Psi} - \sum_{n=1}^N \frac{(r_n - s_n(\Psi))^2}{N_0}.\end{aligned}$$

- Note that  $r_n$  is independent of  $\Psi$ , so that we obtain

$$\hat{\Psi}_{ML,N} = \arg \max_{\Psi \in \Omega_\Psi} \sum_{n=1}^N \frac{2r_n s_n(\Psi) - (s_n(\Psi))^2}{N_0}.$$

- Finally, we define the **maximum-likelihood estimate of  $\Psi$**  as the limit

$$\begin{aligned}\hat{\Psi}_{ML} &= \lim_{N \rightarrow \infty} \hat{\Psi}_{ML,N} = \lim_{N \rightarrow \infty} \arg \max_{\Psi \in \Omega_\Psi} \sum_{n=1}^N \frac{2r_n s_n(\Psi) - (s_n(\Psi))^2}{N_0} \\ &= \arg \max_{\Psi \in \Omega_\Psi} \frac{1}{N_0} \int_{T_0} \left[ 2r(t)s(t; \Psi) - (s(t; \Psi))^2 \right] dt.\end{aligned}$$

- The function

$$\Lambda_L(\Psi) = \frac{1}{N_0} \int_{T_0} \left[ 2r(t)s(t; \Psi) - (s(t; \Psi))^2 \right] dt$$

is the **log-likelihood function (LLF)**, since it results from taking the natural logarithm of the function  $\Lambda(\Psi) = \exp \left( \frac{1}{N_0} \int_{T_0} \left[ 2r(t)s(t; \Psi) - (s(t; \Psi))^2 \right] dt \right)$ , which is also known as the **likelihood function**.

- As a result, we have to implement the estimation rule

$$\hat{\Psi}_{ML} = \arg \max_{\Psi \in \Omega_\Psi} \Lambda_L(\Psi).$$

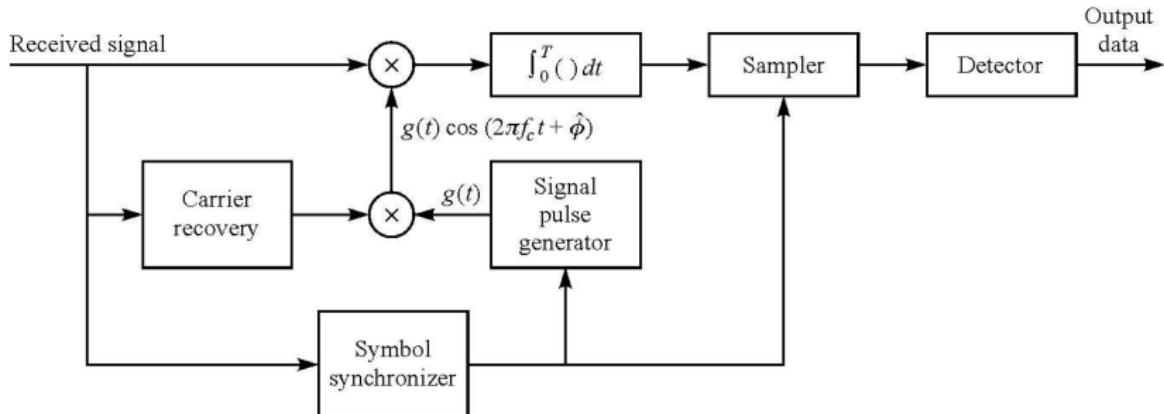
This can be done either **directly by solving the right hand side for the maximum value  $\hat{\Psi}_{ML}$**  or by **implementing a necessary condition to be satisfied by  $\hat{\Psi}_{ML}$** , where in the latter case we usually make assumptions about the properties of  $\Lambda_L(\Psi)$  w.r.t.  $\Psi$ .

- Below, we will derive ML estimators for carrier phase and symbol timing estimation and investigate their performance.

# Carrier Recovery and Symbol Synchronization in Signal Demodulation

Before we study the ML estimation schemes in greater detail, we first consider the overall structure of receivers for different types of modulation. Below we consider

- binary and  $M$ -ary PSK
- $M$ -ary PAM
- QAM.



**Fig. 6.1-1** Block diagram of a **binary PSK receiver**.

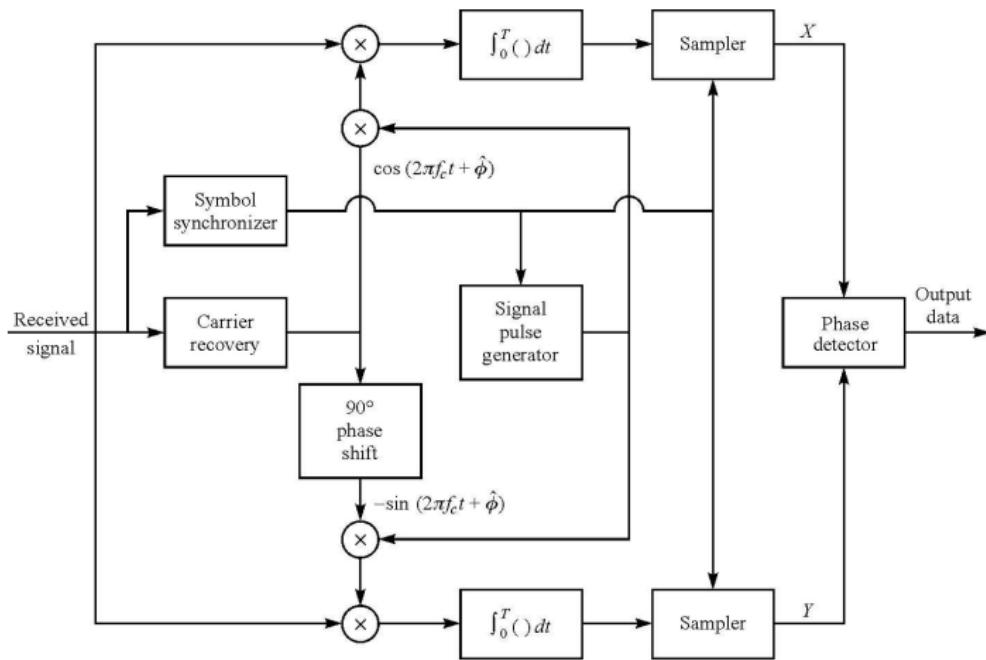


Fig. 6.1-2 Block diagram of an  $M$ -ary PSK receiver.

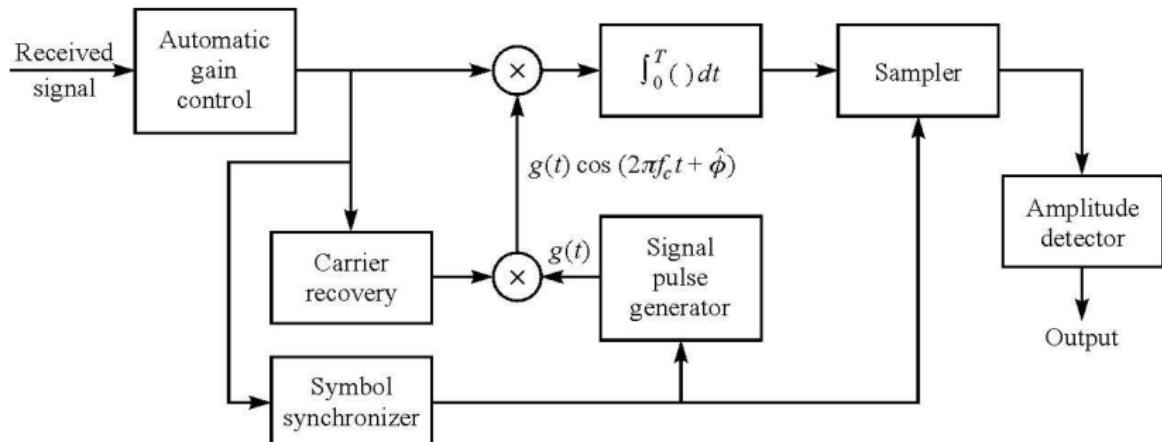
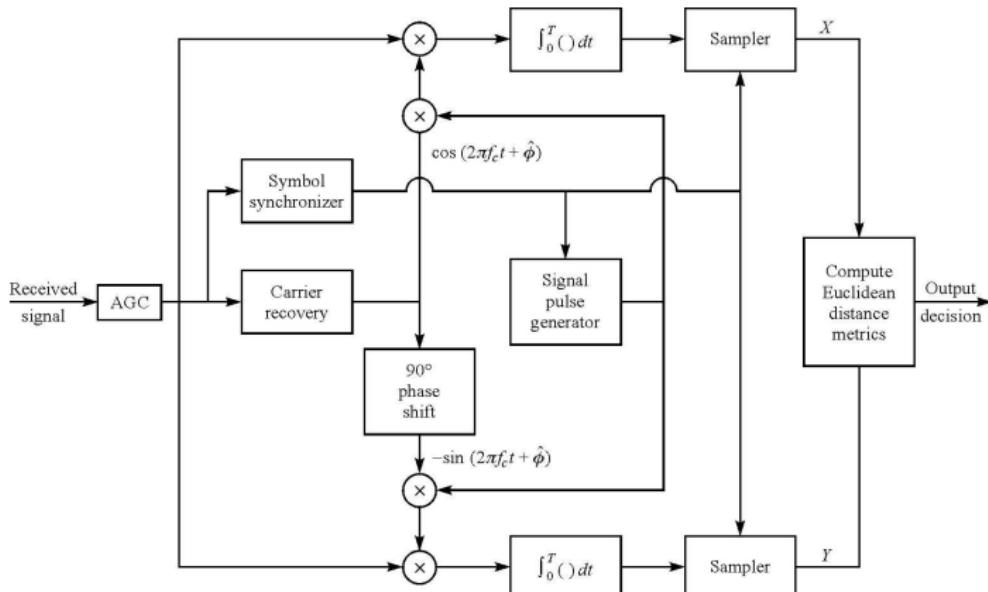


Fig. 6.1-3 Block diagram of an  $M$ -ary PAM receiver.

The automatic gain control is required in order to **eliminate channel gain variations** which would affect the amplitude detector.



**Fig. 6.1-4** Block diagram of a QAM receiver.

As in the case of a PAM receiver, the automatic gain control is required.

## Carrier Phase ( $\phi$ ) Estimation

- In order to estimate the carrier phase at the receiver, one can **transmit an unmodulated carrier (pilot signal) along with the information bearing signal.**
- A so-called **phase-locked loop (PLL)** can then be used in order to **acquire and track the carrier component.**
- The disadvantage of transmitting a pilot carrier signal is its inefficiency, since power is being wasted only for synchronization instead of using it for information transmission.
- Therefore, most digital communication systems do not transmit a pilot carrier, but **estimate the carrier phase directly from the modulated signal.** Subsequently, we will assume a transmission scheme without pilot insertion.
- In order to consider different effects of phase estimation schemes, we will first consider the demodulation of two different types of signals **neglecting the AWGN**, namely a **double-sideband suppressed carrier (DSB/SC) signal**

$$r(t) = A(t) \cos(2\pi f_c t + \phi)$$

and a **quadrature amplitude modulated (QAM) signal**

$$r(t) = A(t) \cos(2\pi f_c t + \phi) - B(t) \sin(2\pi f_c t + \phi).$$

- What happens in the demodulation process if the phase estimate  $\hat{\phi}$  is imperfect, i.e.  $\hat{\phi} \neq \phi$ ?

### Demodulation of a DSB/SC signal

- A product demodulator multiplies  $r(t)$  with a locally generated oscillator signal  $c(t) = 2 \cos(2\pi f_{ct} t + \hat{\phi})$  and uses the output  $y(t)$  of a subsequent low-pass filter (LPF) as an estimate of  $A(t)$ .
- Since we have

$$y(t) = A(t) \cos(\phi - \hat{\phi}),$$

we see that a deviation  $\varepsilon_\phi = \phi - \hat{\phi}$  leads to a signal level reduction by  $\cos \varepsilon_\phi$  and a power reduction by  $\cos^2 \varepsilon_\phi$ .

- **Example:** For  $\varepsilon_\phi = \pi/6$ , we have a signal power loss of  $10 \log_{10} (\cos^2 \varepsilon_\phi) = -1.25$  dB.

### Demodulation of a QAM signal

- Using two product demodulators with locally generated oscillator signals  $c_I(t) = 2 \cos(2\pi f_c t + \hat{\phi})$  and  $c_Q(t) = -2 \sin(2\pi f_c t + \hat{\phi})$  for the in-phase and quadrature components, we obtain two LPF output signals

$$\begin{aligned}y_I(t) &= A(t) \cos(\phi - \hat{\phi}) - B(t) \sin(\phi - \hat{\phi}) \\y_Q(t) &= B(t) \cos(\phi - \hat{\phi}) + A(t) \sin(\phi - \hat{\phi}).\end{aligned}$$

- Obviously, the phase error in QAM (and also in  $M$ -PSK) signals has a much more severe effect than in PAM signals.
- There is a power reduction of the required signal component by a factor  $\cos^2 \varepsilon_\phi$ .
- In addition, there is crosstalk interference from the in-phase and quadrature components. Since the average power levels of  $A(t)$  and  $B(t)$  are similar, a small phase error causes a large degradation in performance.
- As a result, the accuracy requirements for phase estimation schemes for coherent detection of QAM and  $M$ -PSK signals are substantially higher than for DSB/SC PAM signals.

## Maximum-likelihood carrier phase estimation

- Here, we will consider the ML estimate of  $\phi$ , where, for simplicity, we assume the delay  $\tau = 0$  be given at the receiver.
- As a result, we have  $\Psi = \phi$ . Since  $\int_{T_0} (s(t; \phi))^2 dt$  does not depend on  $\phi$ , we obtain as maximum-likelihood phase estimate  $\hat{\phi}_{\text{ML}} = \arg \max_{\phi \in [0, 2\pi)} \int_{T_0} r(t) s(t; \phi) dt$ .
- Example:** For simplicity, consider an unmodulated carrier signal in AWGN

$$r(t) = A \cos(2\pi f_c t + \phi) + n(t).$$

The ML estimate of  $\phi$  is given by  $\hat{\phi}_{\text{ML}} = \arg \max_{\phi \in [0, 2\pi)} A \int_{T_0} r(t) \cos(2\pi f_c t + \phi) dt$ .

The necessary condition

$$\left. \frac{d}{d\phi} \int_{T_0} r(t) \cos(2\pi f_c t + \phi) dt \right|_{\phi=\hat{\phi}_{\text{ML}}} = - \int_{T_0} r(t) \sin(2\pi f_c t + \hat{\phi}_{\text{ML}}) dt = 0$$

first provides

$$-\cos(\hat{\phi}_{\text{ML}}) \int_{T_0} r(t) \sin(2\pi f_c t) dt - \sin(\hat{\phi}_{\text{ML}}) \int_{T_0} r(t) \cos(2\pi f_c t) dt = 0.$$

## Maximum-likelihood carrier phase estimation

- Solving for  $\tan(\hat{\phi}_{\text{ML}})$ , we have

$$\tan(\hat{\phi}_{\text{ML}}) = \frac{- \int r(t) \sin(2\pi f_c t) dt}{\int_{T_0} r(t) \cos(2\pi f_c t) dt}.$$

- Example:** For simplicity, consider an unmodulated carrier signal in AWGN

$$r(t) = A \cos(2\pi f_c t + \phi) + n(t).$$

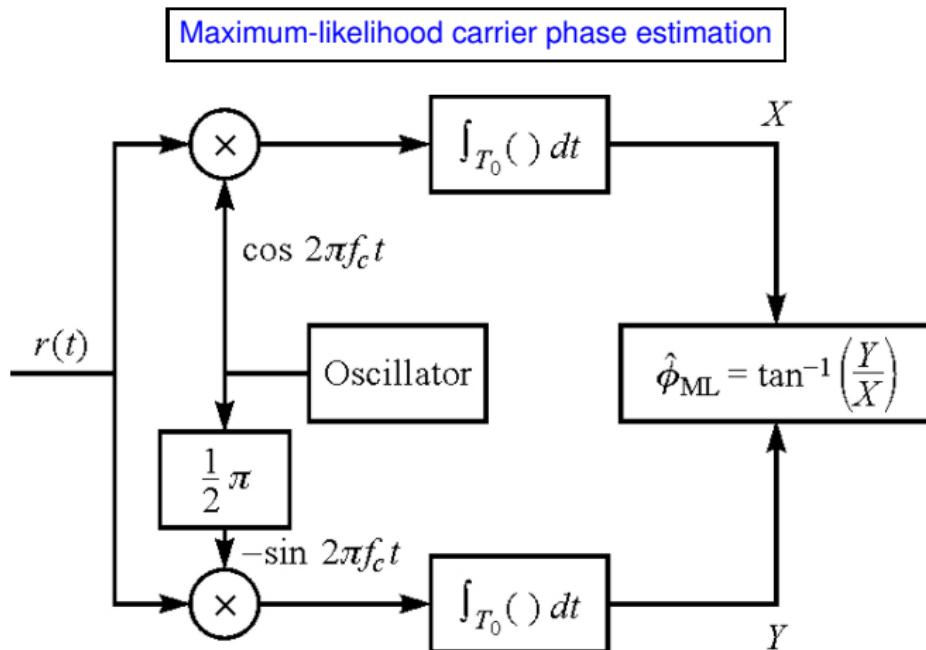
The ML estimate of  $\phi$  is given by

$$\hat{\phi}_{\text{ML}} = \arg \max_{\phi \in [0, 2\pi)} A \int_{T_0} r(t) \cos(2\pi f_c t + \phi) dt.$$

The necessary condition

$$\left. \frac{d}{d\phi} \int_{T_0} r(t) \cos(2\pi f_c t + \phi) dt \right|_{\phi=\hat{\phi}_{\text{ML}}} = - \int_{T_0} r(t) \sin(2\pi f_c t + \hat{\phi}_{\text{ML}}) dt = 0$$

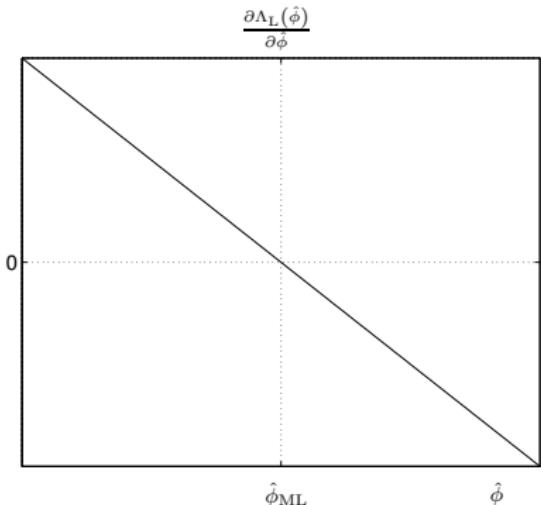
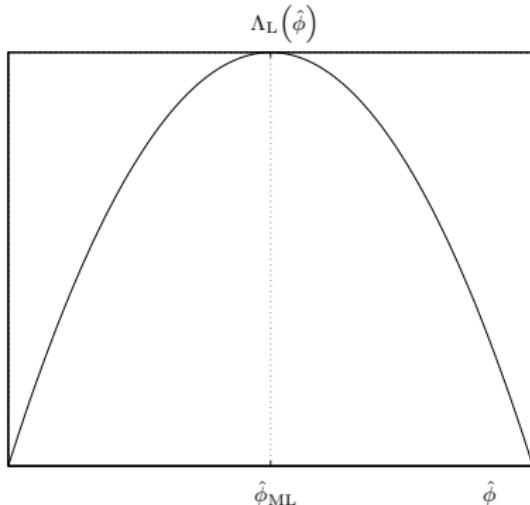
$$\hat{\phi}_{\text{ML}} = -\arctan \left( \frac{\int_{T_0} r(t) \sin(2\pi f_c t) dt}{\int_{T_0} r(t) \cos(2\pi f_c t) dt} \right) = \arctan \left( \frac{\int_{T_0} -r(t) \sin(2\pi f_c t) dt}{\int_{T_0} r(t) \cos(2\pi f_c t) dt} \right).$$



**Fig. 6.2-2** A (one-shot) ML estimate of the phase of an unmodulated carrier.

## Phase locked loop

- The aforementioned estimate  $\hat{\phi}_{\text{ML}}$  results from explicitly solving the necessary condition for  $\phi$ .
- Often, we can use a less complex implementation for calculating the ML estimate by the following heuristic approach: The form of the LLF suggests a **dynamic implementation** of the estimation rule. This can be seen from inspection of  $\Lambda_L(\phi)$  in the low-noise case  $N_0 \ll 1$ :

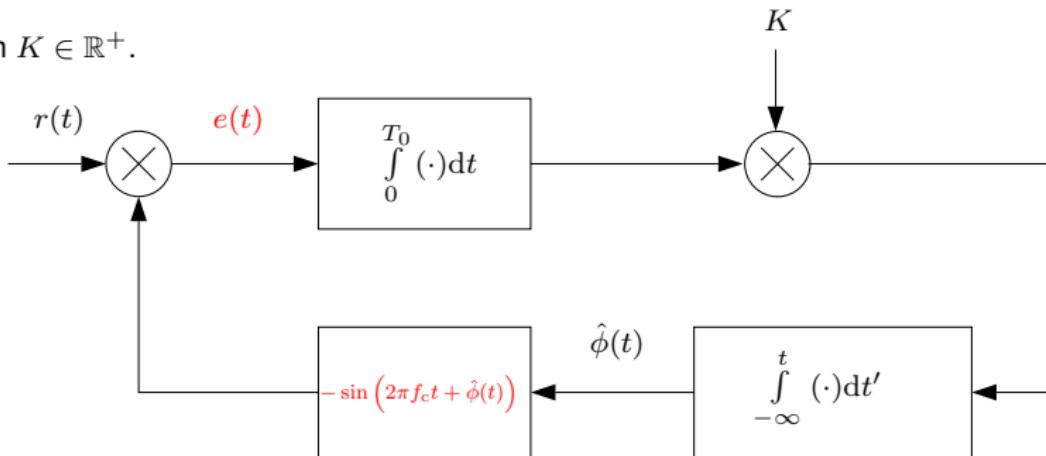


### Phase locked loop

- Example of a **dynamic implementation** of the estimation rule

$$K \frac{\partial \Lambda_L(\hat{\phi})}{\partial \hat{\phi}} = \frac{\partial \hat{\phi}}{\partial t}$$

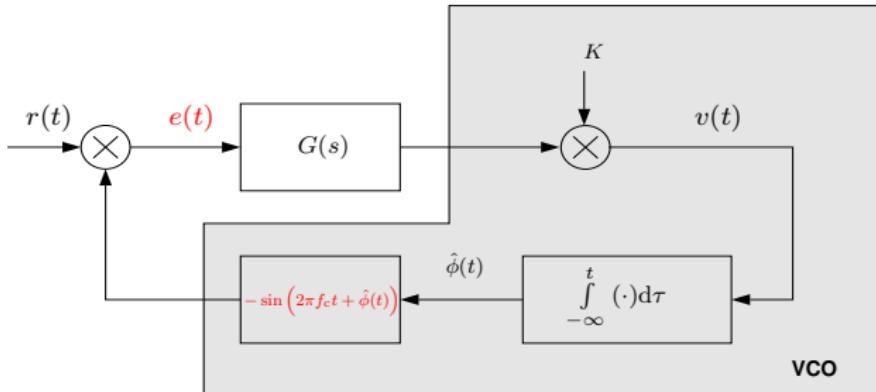
with  $K \in \mathbb{R}^+$ .



- The dynamic implementation aims at driving  $\hat{\phi}$  towards  $\phi$  or, expressed differently, at driving the estimation error  $\varepsilon(t) = \phi(t) - \hat{\phi}(t)$  towards zero **based on the signal  $e(t)$** .

## Phase locked loop

- The dynamic properties of the process  $\phi(t)$  can be taken into account by a suitable filter with Laplace transform  $G(s)$ , which can be used simultaneously to implement the integration required in the necessary condition.
- Note that the dynamic implementation does not drive the parameter estimate toward  $\hat{\phi}_{ML}$  over the whole range of possible values for  $\hat{\phi}$ . In general, the dynamic implementation works primarily in the vicinity of  $\hat{\phi} = \hat{\phi}_{ML}$ .
- The overall scheme is known as **PLL (phase-locked loop)**.



VCO ... voltage-controlled oscillator

$K$  ... **sensitivity** of the VCO in [rad/V sec]

$v(t)$  ... control voltage of the VCO.

## Phase locked loop

- In order to characterize the estimation performance of the PLL, we consider the signal

$$e(t) = -\cos(2\pi f_c t + \phi(t)) \sin(2\pi f_c t + \hat{\phi}(t)) = \frac{1}{2} \sin(\varepsilon(t)) - \frac{1}{2} \sin(4\pi f_c t + \phi(t) + \hat{\phi}(t))$$

and assume the loop filter  $G(s)$  to have essentially a low-pass characteristics so that the double frequency component of  $e(t)$  is filtered out. The whole operation represents a **phase discrimination**.

- The loop filter  $G(s)$  is usually chosen to have the relatively simple transfer function

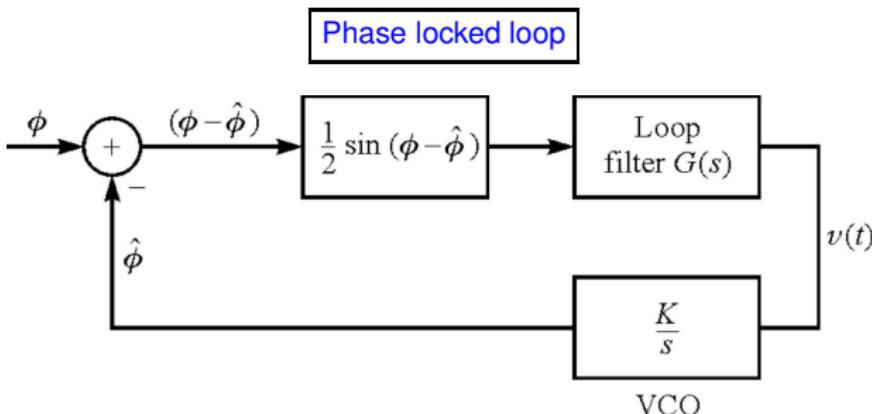
$$G(s) = \frac{1 + \tau_2 s}{1 + \tau_1 s}$$

where  $\tau_1$  and  $\tau_2$  with  $\tau_1 \gg \tau_2$  are design parameters to control the bandwidth of the loop. Higher-order filters are possible to obtain a better loop response.

- In view of the phase estimate

$$\hat{\phi}(t) = K \int_{-\infty}^t v(\tau) d\tau,$$

and the fact that the double frequency contribution in  $e(t)$  is filtered out, the VCO input is a function of  $\varepsilon(t)$  and we can translate the PLL block diagram into an **equivalent base-band model** depending on  $\phi(t)$ ,  $\hat{\phi}(t)$  and  $\varepsilon(t)$  rather than on  $r(t)$ ,  $\sin(2\pi f_c t + \hat{\phi}(t))$  and  $e(t)$ .



**Fig. 6.2-4** Model of phase-locked loop (corrected figure!).

- In the simplest case where we set  $\tau_1 = \tau_2 = 0$ , i.e.  $G(s) \equiv 1$ , and absorb the factor  $1/2$  into the gain parameter  $K$ , we have  $K \sin \varepsilon(t) = \dot{\hat{\phi}}(t) = \dot{\phi}(t) - \dot{\varepsilon}(t)$  or equivalently

$$\dot{\varepsilon}(t) = \dot{\phi}(t) - KS(\varepsilon(t))$$

Here,  $S(\varepsilon) = \sin \varepsilon$  is the so-called ***S*-function** whose name results from the shape (a horizontal ***S***) of the phase discriminator output characteristics.

## Phase locked loop

- In normal operation mode when the loop is tracking the phase of the incoming carrier, the phase error  $\varepsilon(t) = \phi(t) - \hat{\phi}(t)$  is small and hence  $S(\varepsilon(t)) = \sin \varepsilon(t) \approx \varepsilon(t) = \phi(t) - \hat{\phi}(t)$ .
- As a consequence, the **equivalent baseband loop represents a linear network** with a transfer function  $H(s)$  between  $\Phi(s)$  and  $\hat{\Phi}(s)$ .
- Assuming a general low-pass filter  $G(s)$  in the loop,  $H(s)$  can be found from Fig. 6.2-4. We get upon Laplace transformation

$$\frac{KG(s)}{s} (\Phi(s) - \hat{\Phi}(s)) = \hat{\Phi}(s) \quad \Leftrightarrow \quad \hat{\Phi}(s) = \frac{KG(s)}{s + KG(s)} \Phi(s) = H(s)\Phi(s).$$

- Inserting  $G(s) = \frac{1+\tau_2 s}{1+\tau_1 s}$  from above, we obtain

$$H(s) = \frac{1 + \tau_2 s}{1 + (\tau_2 + \frac{1}{K})s + \frac{\tau_1}{K}s^2}.$$

- The closed-loop behavior of the PLL is described a **second-order transfer function** where  $\tau_2$  controls the position of the zero, while  $K$  and  $\tau_1$  are used to control the position of the closed-loop system poles.
- It is customary to express the denominator of  $H(s)$  in the standard form

$$D(s) = s^2 + 2\zeta\omega_n s + \omega_n^2,$$

where  $\zeta$  is the loop damping factor and  $\omega_n$  is the natural frequency of the loop.

## Phase locked loop

- In terms of the loop parameters, we obtain

$$\omega_n = \sqrt{K/\tau_1} \quad \text{and} \quad \zeta = \frac{\omega_n}{2} \left( \tau_2 + \frac{1}{K} \right)$$

and upon insertion

$$H(s) = \frac{(2\zeta\omega_n - \omega_n^2/K)s + \omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}.$$

- The one-sided noise-equivalent bandwidth of the loop can be shown to be

$$B_{\text{eq}} = \frac{\tau_2^2 (1/\tau_2^2 + K/\tau_1)}{4(\tau_2 + 1/K)} = \frac{1 + (\tau_2\omega_n)^2}{8\zeta/\omega_n}.$$

- For  $\tau_1 \gg 1$ , the magnitude response  $20 \log_{10} |H(\omega)|$  can be characterized as a function of the **normalized frequency**  $\omega/\omega_n$  with the **damping parameter  $\zeta$  as a parameter**. For

$\zeta = 1$ , we have a critically damped loop response,  
 $\zeta < 1$  produces an underdamped response, and  
 $\zeta > 1$  yields an overdamped response.

### Phase locked loop

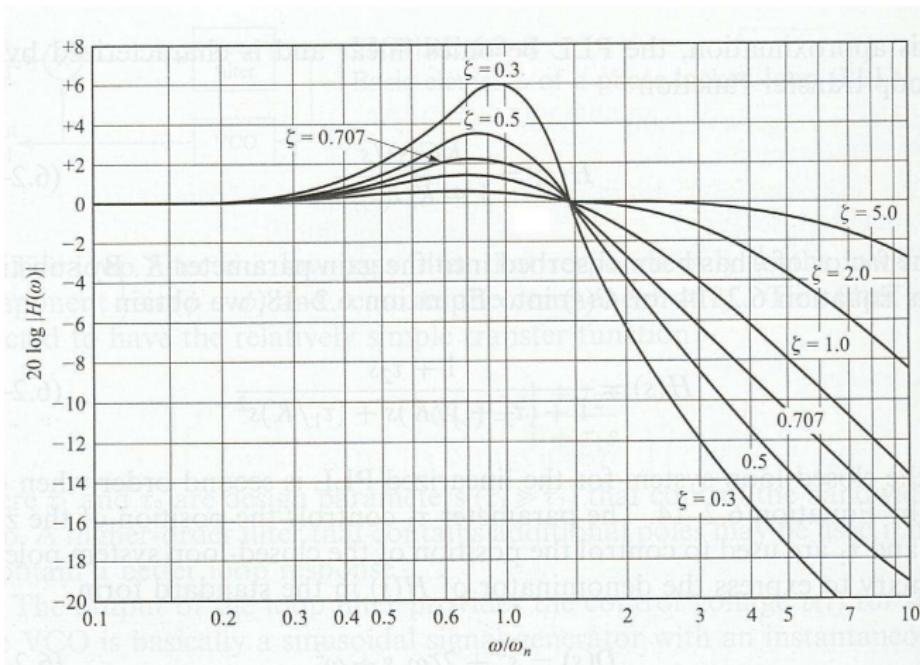


Fig. 6.2-5 Frequency response of a second-order loop.

## Phase locked loop

- Until now, we have not yet considered AWGN in the received signal and its effect on the tracking performance of the PLL. The treatment of the performance degradation is complicated due to the nonlinear mapping of the AWGN by the phase discriminator.
- There exist baseband models which builds on certain assumptions about the signal-to-noise ratio (SNR). If the SNR is sufficiently large, one can linearize the characteristics  $\sin \varepsilon(t)$ , much the same way like in the noise-free case, and obtain a linear model of the loop including the noise.
- In the aforementioned modelling, one concentrates on different scenarios including the transient behavior of the PLL during initial acquisition, the behavior at low SNR values and the mean time it takes to lose lock or to reacquire the phase estimate and reenter the tracking phase again.
- For example, it is known that when the SNR at the input of the PLL drops below a certain value, there is rapid deterioration in the performance of the PLL. The loop begins to lose lock and in impulsive type of noise, characterized as clicks, is generated leading to the performance degradation.
- A detailed treatment of noise in PLLs is beyond the scope of this lecture. The interested reader is referred to the books of Viterbi (1966, *Principles of Coherent Communication*) and Gardner (1979, *Phaselock Techniques*).

### Decision directed loop

- Until now, we have not yet addressed the fact that the signal  $s(t; \Psi)$  carries an information sequence  $\{I_n\}$  being unknown at the receiver.
- To be specific, consider an equivalent low-pass representation of the linearly modulated received signal given by

$$r_\ell(t) = e^{-j\phi} \sum_n I_n g(t - nT) + z(t) = s_\ell(t)e^{-j\phi} + z(t),$$

where  $s_\ell(t)$  is a known signal for a known sequence  $\{I_n\}$ .

- There are two basic approaches to estimate the phase in the case of modulated signals.
  - Either we assume that the sequence  $\{I_n\}$  has been estimated so we can assume it to be known neglecting any errors in it. Since we feed back the decisions  $\{\hat{I}_n\}$  being made about  $\{I_n\}$  to the phase estimation scheme, this approach is called **decision-directed**.
  - The alternative is to treat  $\{I_n\}$  as a random sequence and average the likelihood approach over its statistics where, as a result, we deal with expectations of likelihood and log-likelihood functions. As a result of the averaging which does not require a decision to be made about  $\{I_n\}$  prior to phase estimation, this approach is called **non-decision-directed**
- Here, we will consider **decision-directed phase estimation**.

### Decision directed loop

- As in the maximum-likelihood carrier phase estimation, we can neglect the energy term (which for the considered case of complex-valued signals reads  $\int_{T_0} |s(t; \phi)|^2 dt$ ), since it does not depend on  $\phi$ .
- The log-likelihood function is equivalent to the inner product between the received signal  $r_\ell(t)$  and the locally generated signal hypotheses  $s_\ell(t)e^{-j\phi}$ , where we use  $\{\hat{I}_n\}$  to build  $s_\ell(t)$ .
- Clearly, the inner product in the case of complex-valued signals  $r_\ell(t)$  and  $s_\ell(t)$  is given by  $\Re \left\{ \int_{T_0} r_\ell(t) s_\ell^*(t) dt \right\}$ . Correspondingly, if we neglect the factor 2 and assume an observation interval of length  $T_0 = KT$ , the log-likelihood function is given by

$$\Lambda_L(\phi) = \Re \left\{ \frac{e^{j\phi}}{N_0} \int_{T_0} r_\ell(t) s_\ell^*(t) dt \right\} = \Re \left\{ \frac{e^{j\phi}}{N_0} \sum_{n=0}^{K-1} \hat{I}_n^* y_n \right\},$$

where we have defined  $y_n = \int_{nT}^{(n+1)T} r_\ell(t) g^*(t - nT) dt$ .

### Decision directed loop

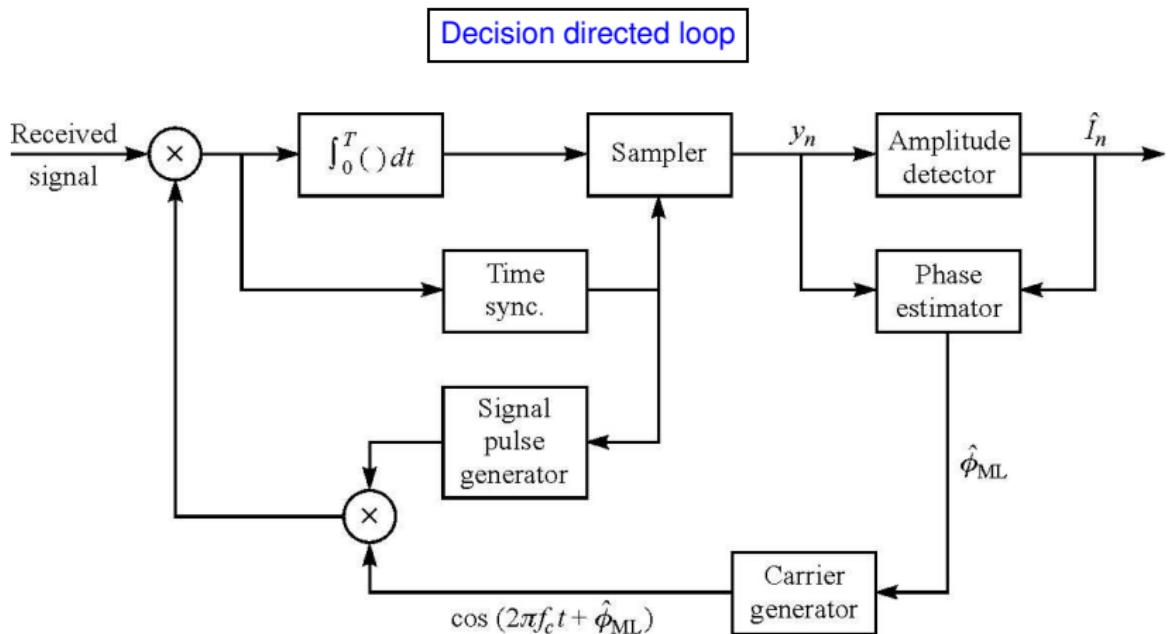
- With the identity  $\Re\{ab\} = \Re\{a\}\Re\{b\} - \Im\{a\}\Im\{b\}$ , where  $\Im\{\cdot\}$  characterizes the imaginary part of the argument, we obtain

$$\Lambda_L(\phi) = \Re \left\{ \frac{1}{N_0} \sum_{n=0}^{K-1} \hat{I}_n^* y_n \right\} \cos \phi - \Im \left\{ \frac{1}{N_0} \sum_{n=0}^{K-1} \hat{I}_n^* y_n \right\} \sin \phi.$$

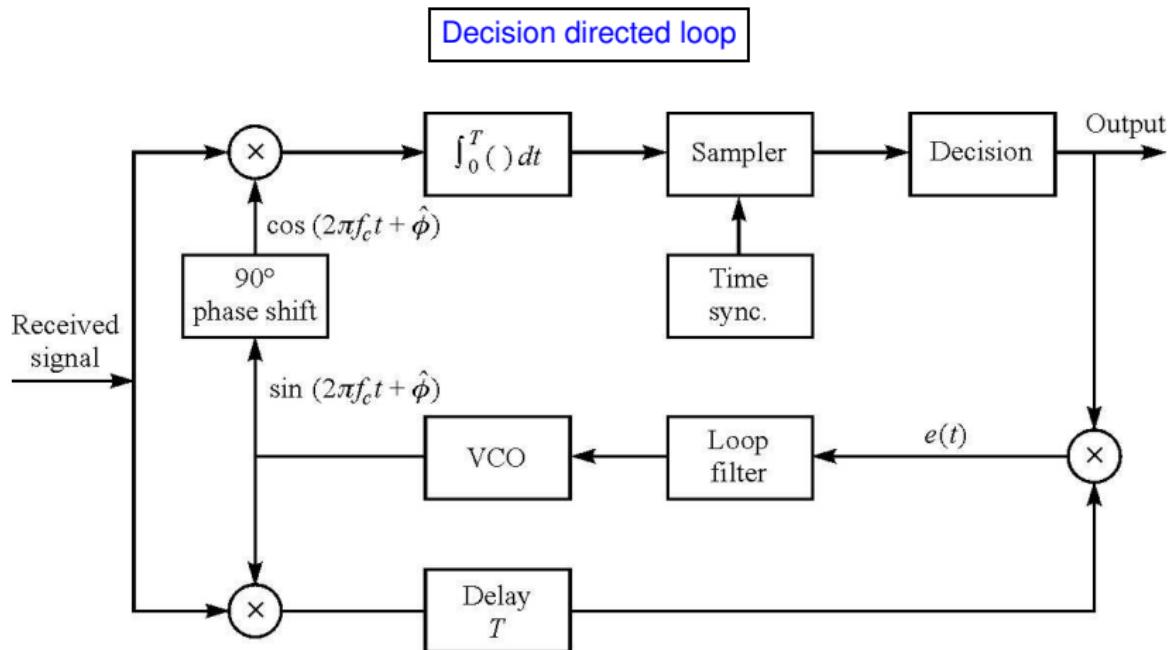
- Setting the derivative of  $\Lambda_L(\phi)$  equal to zero and solving for  $\phi$ , we obtain the **decision-directed (or decision-feedback) carrier phase estimate**

$$\hat{\phi}_{ML} = -\arctan \left( \frac{\Im \left\{ \sum_{n=0}^{K-1} \hat{I}_n^* y_n \right\}}{\Re \left\{ \sum_{n=0}^{K-1} \hat{I}_n^* y_n \right\}} \right).$$

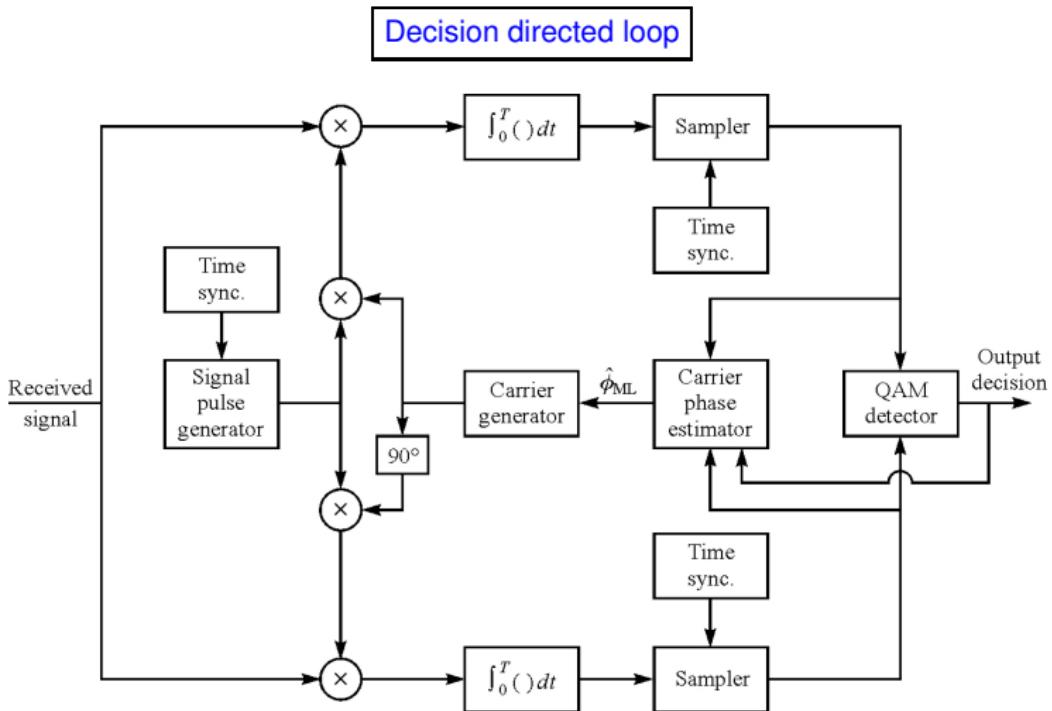
- Since  $r_\ell(t) = s_\ell(t)e^{-j\phi} + z(t)$ , this carrier phase estimation scheme is intuitive.
- It can be shown that  $\hat{\phi}_{ML}$  is an unbiased estimator for  $\hat{I}_n = I_n$ . Also, the PDF of  $\hat{\phi}_{ML}$  can be derived using methods known from the derivation of the probability of error for  $M$ -ary PSK signaling in AWGN.
- The estimator will be investigated for different signaling schemes for building  $\{I_n\}$ .



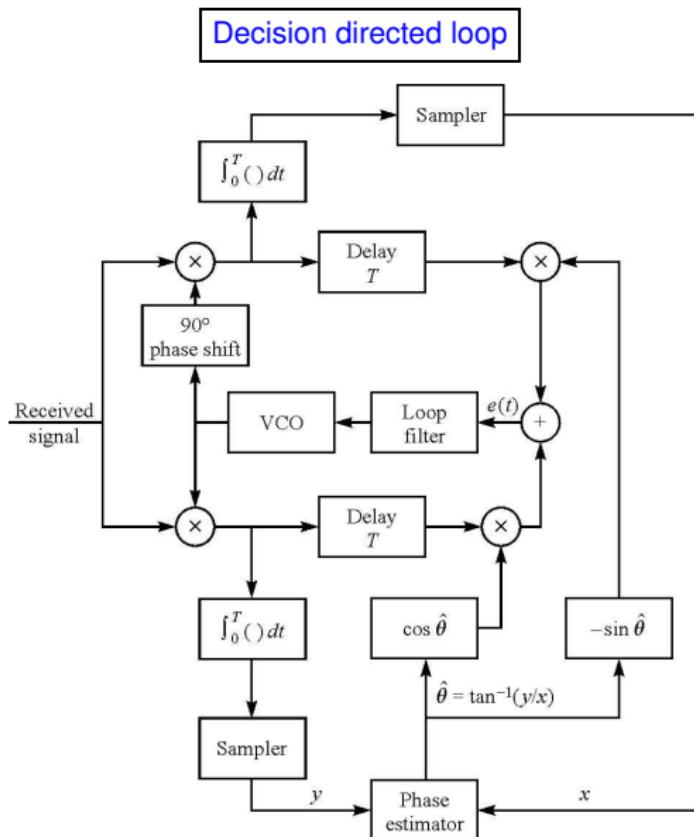
**Fig. 6.2-9** Block diagram of double-sideband PAM signal receiver with **decision-directed carrier phase estimation**.



**Fig. 6.2-10 Carrier recovery with a decision-feedback PLL.**



**Fig. 6.2-11** Block diagram of QAM signal receiver with **decision-directed carrier phase estimation**.



**Fig. 6.2-12** Carrier recovery for  $M$ -ary PSK using a **decision-feedback PLL**.

### Non-decision directed loop

- As we have seen above, the decision-directed receivers require a high effort for implementing suitable phase estimation schemes. **Can we find less complex schemes with a reasonable performance?**
- Instead of assuming, like in decision-directed schemes, that we have a decision on the information symbols available which we can feed back to the phase estimation, we can **treat the information symbols as random variables** and **average the likelihood function over the corresponding PDF** in order to achieve a maximum average performance (i.e., formulated in the language of estimation theory, a minimum average cost).
- Example:** Suppose we have a real-valued signal  $s(t)$  with BPSK modulated symbols  $A$ . Then, in a signal interval, we have

$$s(t) = A \cos(2\pi f_c t + \phi), \quad 0 \leq t \leq T,$$

where  $A = \pm 1$  with equal probability, i.e. the PDF of  $A$  is given by

$$p(A) = \frac{1}{2}\delta(A - 1) + \frac{1}{2}\delta(A + 1).$$

- The likelihood function  $\Lambda(\phi) = \exp\left(\frac{2}{N_0} \int_{T_0} r(t)s(t; \phi)dt\right)$  is now **conditional on a given value  $A$**  and **must be averaged over the PDF for the two values  $A = \pm 1$** .

## Non-decision directed loop

- We obtain the likelihood function

$$\begin{aligned}
 \bar{\Lambda}(\phi) &= E_A[\Lambda(\phi)] \\
 &= \frac{1}{2} \exp \left( \frac{2}{N_0} \int_{T_0} r(t) \cos(2\pi f_c t + \phi) dt \right) \\
 &\quad + \frac{1}{2} \exp \left( -\frac{2}{N_0} \int_{T_0} r(t) \cos(2\pi f_c t + \phi) dt \right) \\
 &= \cosh \left( \frac{2}{N_0} \int_{T_0} r(t) \cos(2\pi f_c t + \phi) dt \right)
 \end{aligned}$$

and the corresponding log-likelihood function

$$\bar{\Lambda}_L(\phi) = \ln \cosh \left( \frac{2}{N_0} \int_{T_0} r(t) \cos(2\pi f_c t + \phi) dt \right).$$

### Non-decision directed loop

- The usual procedure of differentiating  $\bar{\Lambda}_L(\phi)$  w.r.t.  $\phi$ , setting the derivative to zero and implementing the resulting equation leads, unfortunately, to highly non-linear equations. However, we can use the following **approximations** in order to find analytical solutions:

$$\ln \cosh x = \begin{cases} \frac{1}{2}x^2 & \text{for } |x| \ll 1 \\ |x| & \text{for } |x| \gg 1. \end{cases}$$

- As a result of the **discrete approximation**, we would base our approach on the function

$$\bar{\Lambda}_L(\phi) \approx \left[ \frac{2}{N_0} \int_{T_0} r(t) \cos(2\pi f_c t + \phi) dt \right]^2.$$

### Non-decision directed loop

- Clearly, if we have  $M$ -ary modulation with large  $M$ , it might be reasonable to assume a continuous PDF on the modulating signals rather than discrete PDFs leading to highly non-linear equations.
- If we assume in the above example that the values of  $A$  are **standard Gaussian distributed** with

$$p(A) = \frac{1}{\sqrt{2\pi}} e^{-A^2/2},$$

we again obtain a log-likelihood upon averaging as

$$\bar{\Lambda}_L(\phi) = \left[ \frac{2}{N_0} \int_{T_0} r(t) \cos(2\pi f_c t + \phi) dt \right]^2.$$

- Interestingly, the both cases yield similar results so that the assumption of Gaussian symbol distributions does not seem to be completely unreasonable.

### Non-decision directed loop

- With similar arguments, assuming again an observation interval  $T_0 = KT$  and information symbols being statistically independent and identically Gaussian distributed, we obtain a log-likelihood function

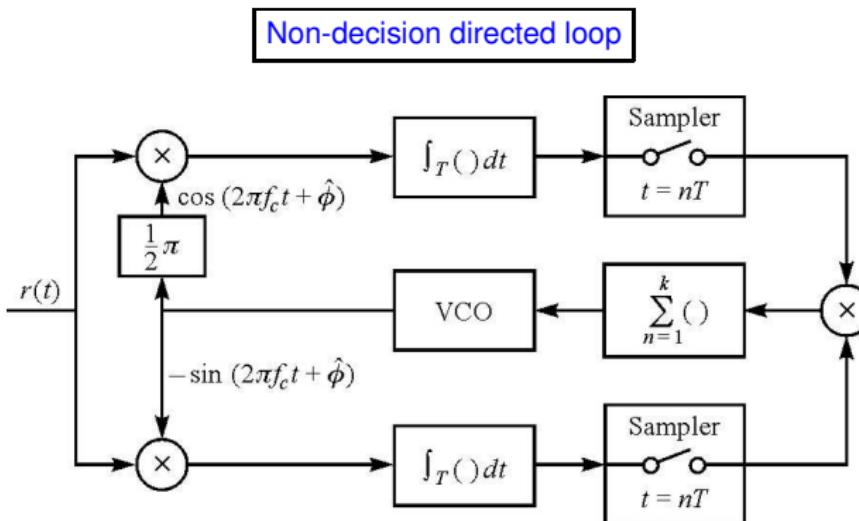
$$\bar{\Lambda}_L(\phi) = \sum_{n=0}^{K-1} \left[ \frac{2}{N_0} \int_{nT}^{(n+1)T} r(t) \cos(2\pi f_c t + \phi) dt \right]^2.$$

- Upon setting the derivative of  $\bar{\Lambda}_L(\phi)$  w.r.t.  $\phi$  equal to zero and using a dynamic implementation, we obtain with a gain constant  $\kappa > 0$

$$-\kappa \sum_{n=0}^{K-1} \int_{nT}^{(n+1)T} r(t) \cos(2\pi f_c t + \hat{\phi}(t)) dt \int_{nT}^{(n+1)T} r(t) \sin(2\pi f_c t + \hat{\phi}(t)) dt = \frac{\partial \hat{\phi}}{\partial t}$$

- Obviously, we have a similar scheme as in the case of known information symbols.

Now, however, the first term  $\int_{nT}^{(n+1)T} r(t) \cos(2\pi f_c t + \hat{\phi}(t)) dt$  is an estimate of the  $n$ -th symbol. This term replaces the otherwise available information symbol (either known or estimated) which would be multiplied by  $\int_{nT}^{(n+1)T} r(t) \sin(2\pi f_c t + \hat{\phi}(t)) dt$ .



**Fig. 6.2-13 Non-decision-directed PLL for carrier phase estimation of PAM signals.**

- The upper branch provides the data symbol estimate, the lower one is the usual PLL branch.
- The summer can be replaced by a LPF (loop filter).

Non-decision directed loop

- Similar to the loop in Fig. 6.2-13 is the so-called **Costas loop**. It has been developed by Costas in 1956.

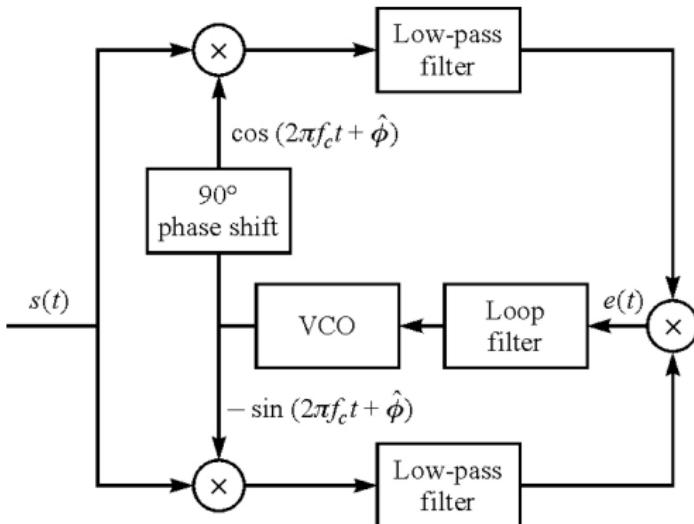


Fig. 6.2-15 Block diagram of **Costas loop**.

- Obviously, the general structure of both loops is the same, only the filtering operations are implemented differently.

### Non-decision directed loop

- Another approach to eliminate estimates of information symbols in phase recovery circuits can be investigated by reconsidering digitally modulated PAM signals of the form

$$s(t) = A(t) \cos(2\pi f_c t + \phi),$$

where  $A(t)$  carries the digital information.

- If  $A(t)$  was a constant with  $A(t) > 0$ , we could directly apply a usual PLL. However, since  $E[A(t)] = 0$ , we will have at the output of the phase discriminator a zero-average signal level. If, on the contrary, we would have  $E[A(t)] > 0$ , we could again use a PLL.
- Since  $A(t)$  is real-valued, we can generate a non-zero average signal level at the double carrier frequency  $2f_c$  by squaring the received signal prior to phase recovery. Since

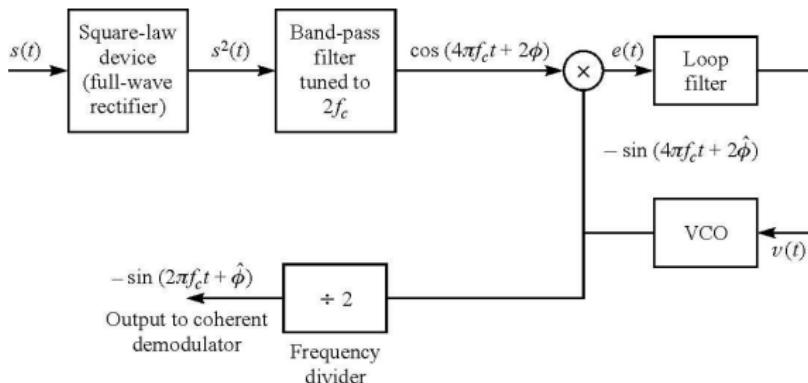
$$s^2(t) = \frac{A^2(t)}{2} (1 + \cos(4\pi f_c t + 2\phi)),$$

we can use a band-pass filter at the frequency  $2f_c$  where the average power is non-zero.

### Non-decision directed loop

- The following items must be taken into account in the design of a squaring loop:

- In order to estimate  $\phi$ , the output of the squaring loop must be frequency-divided by 2, which results in a phase ambiguity of  $\pi$ . For this reason, in the case of binary data, the latter must be differentially encoded prior to transmission and differentially decoded at the receiver.
- Clearly, the squaring of the signal implies a squaring of the noise resulting in a performance loss as compared to a usual PLL.



**Fig. 6.2-14 Carrier recovery for PAM signals using a square-law device.**

### Non-decision directed loop

- The approach of the squaring loop can be generalized to  $M$ -ary PSK, where the transmitted signals are defined by

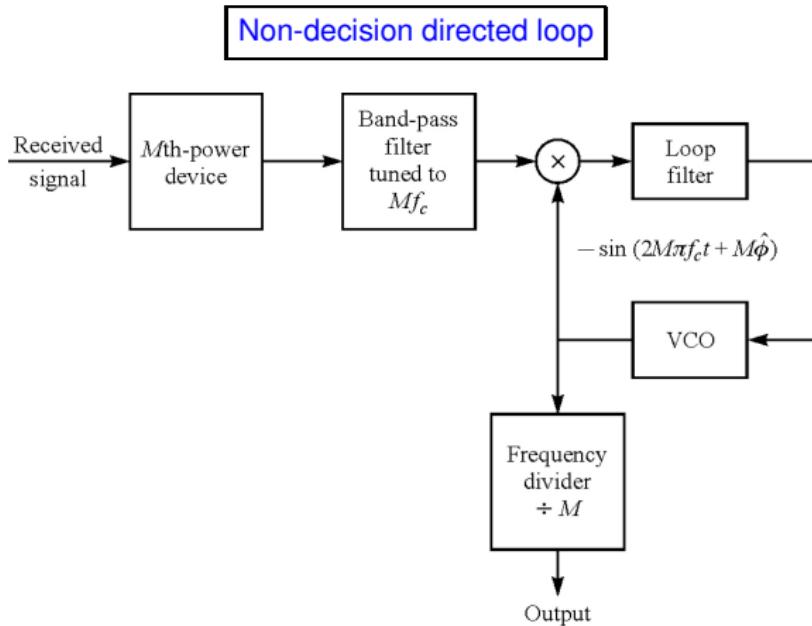
$$s(t) = A \cos \left( 2\pi f_c t + \phi + \frac{2\pi}{M} (m - 1) \right), \quad m = 1, 2, \dots, M,$$

and  $m$  represents the information to be transmitted.

- If we pass the received signal through an  $M$ -th-power-law device and use a suitable band-pass filter, we can select the harmonics  $\cos(2\pi M f_c t + M\phi)$  to drive the PLL, since the information has been eliminated due to the relation

$$\frac{2\pi}{M} (m - 1) M = 2\pi(m - 1) = 0 \bmod 2\pi.$$

- Again, a frequency divider is necessary and a differential encoding/decoding required in order to overcome the phase ambiguity of  $2\pi/M$ .



**Fig. 6.2-16** Carrier recovery with  $M$ -th-power-law device for  $M$ -ary PSK.

## Symbol Timing ( $\tau$ ) Estimation

- In a digital communication system, the output of the demodulator must be sampled periodically at the symbol rate, at the precise sampling time instants  $t_m = mT + \tau$ , where  $T$  is the symbol interval and  $\tau$  is the nominal time delay due to signal propagation.
- The process of extracting a corresponding periodic clock signal is usually called **symbol synchronization** or **timing recovery**.
- Basically, the timing requires two steps, namely the estimation of  $T$  and of  $\tau$ . The latter is sometimes called the **timing phase**.
- There are several ways to deal with the timing recovery problem which are similar to the corresponding approaches for carrier phase recovery. We can, e.g., transmit a clock signal in parallel to the information bearing signal with the aforementioned disadvantages.
- Here, we will focus on timing recovery in systems where the clock signal is to be extracted from a received data signal. Again, we can distinguish **decision-directed** and **non-decision-directed** methods.
- As shown below, the resulting loop structures are analogous to those for carrier phase recovery.

## Maximum-likelihood timing estimation for decision-directed loops

- If the signal is a baseband PAM waveform, it is represented as

$$r(t) = s(t; \tau) + n(t) = \sum_n I_n g(t - nT - \tau).$$

For **decision-directed schemes**, we simply replace the information symbols  $I_n$  by estimates  $\hat{I}_n$  which are assumed to be perfect, i.e.  $\hat{I}_n = I_n$ .

- We assume the observation interval to be large enough so that the signal energy contained in it does not depend on  $\tau$ . The LLF for this problem is given by

$$\Lambda_L(\tau) = \int_{T_0} r(t)s(t; \tau)dt = \sum_n \hat{I}_n \int_{T_0} r(t)g(t - nT - \tau)dt = \sum_n \hat{I}_n y_n(\tau)$$

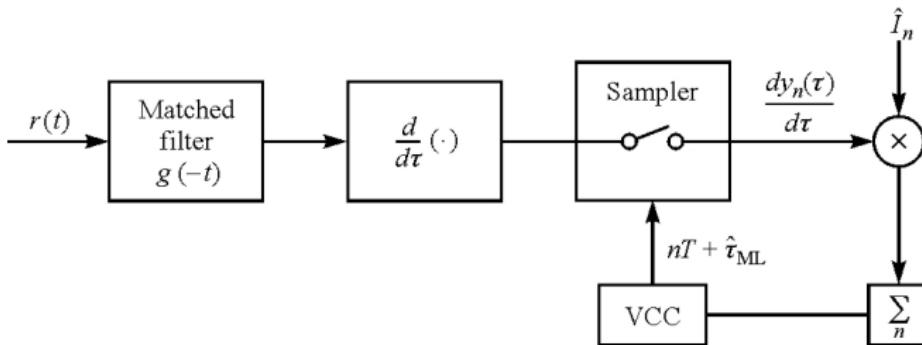
with  $y_n(\tau) = \int_{T_0} r(t)g(t - nT - \tau)dt$ .

- The dynamic implementation of the ML estimation provides

$$\frac{d\Lambda_L(\tau)}{d\tau} \Big|_{\tau=\hat{\tau}} = \sum_n \hat{I}_n \frac{dy_n(\tau)}{d\tau} \Big|_{\tau=\hat{\tau}} = \frac{d\hat{\tau}}{dt}.$$

- The role of the VCO is taken by a so-called **voltage-controlled clock** controlling the sampling times for the input to the loop.

Maximum-likelihood timing estimation for decision-directed loops



**Fig. 6.3-1** Decision-directed ML estimation of timing for baseband PAM.

## Maximum-likelihood timing estimation for decision-directed loops

- Sometimes, the calculation of the derivative  $\left. \frac{dy_n(\tau)}{d\tau} \right|_{\tau=\hat{\tau}}$  is implemented by a structure commonly known as **early-late gate symbol synchronizer**. Its functional principle should be demonstrated for the example of a **coherent spread-spectrum system** where in view of the decision direction, we neglect the information symbols  $\{I_n\}$  again.
- Consider a **real-valued spread spectrum baseband signal in AWGN**

$$r(t) = \sqrt{P}c(t - \tau) + n(t)$$

with a periodic pseudo-noise sequence  $c(t) = \sum_{i=-\infty}^{\infty} \tilde{c}(t - iT)$  and

$$\tilde{c}(t) = \sum_{\nu=0}^{N_c-1} c_\nu \text{rect}_{T_c}(t - \nu T_c).$$

- For simplicity, we assume that the observation time is  $T_0 = T$ . The part of the LLF depending on the hypothesis  $\tau$  is given with  $T = N_c T_c$  by

$$\Lambda_L(\tau) = P \int_{t-T}^t r(t') c(t' - \tau) dt'.$$

## Maximum-likelihood timing estimation for decision-directed loops

- We can distinguish a **coherent** and a **non-coherent** implementation. Here, we discuss the **coherent case**.
- The derivative of  $\Lambda_L(\tau)$  w.r.t.  $\tau$  is approximated by the difference quotient

$$\frac{\partial \Lambda_L(\tau)}{\partial \tau} \Big|_{\tau=\hat{\tau}} \approx \frac{\Lambda\left(\hat{\tau} + \frac{\Delta}{2}\right) - \Lambda\left(\hat{\tau} - \frac{\Delta}{2}\right)}{\Delta}$$

where  $\Delta$  is a positive time parameter, typically  $\Delta = T_c$ .

- As in the PLL, let us define the **deviation**  $\varepsilon = \tau - \hat{\tau}$  and consider again the **noise-free case**, where  $R_{cc}(\tau)$  is the **periodic autocorrelation function** of  $c(t)$ :

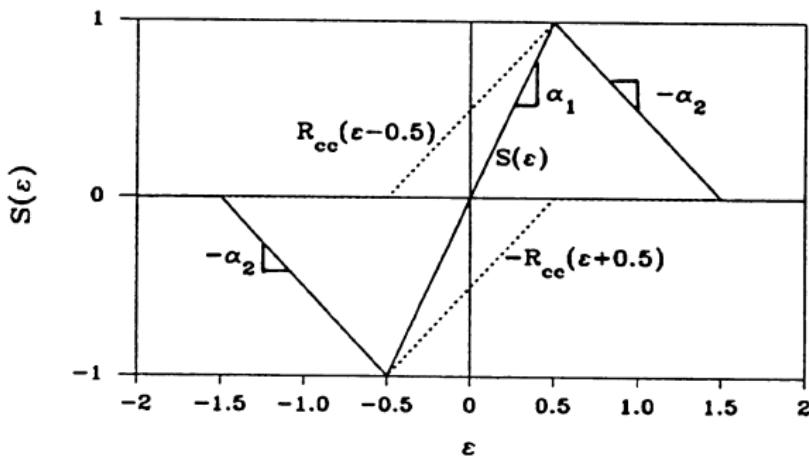
$$\frac{\partial \Lambda_L(\tau)}{\partial \tau} \Big|_{\tau=\hat{\tau}} \sim S(\varepsilon) = R_{cc}\left(\varepsilon - \frac{\Delta}{2}\right) - R_{cc}\left(\varepsilon + \frac{\Delta}{2}\right).$$

Again, we have an ***S*-function**  $S(\varepsilon)$  which corresponds to  $S(\varepsilon) = \sin \varepsilon$  in the phase recovery PLL. Also, for the dynamic implementation  $\frac{\partial \Lambda_L(\tau)}{\partial \tau} \Big|_{\tau=\hat{\tau}} = \frac{\partial \hat{\tau}}{\partial t}$ , we have  $\dot{\varepsilon} = \dot{\tau} - KS(\varepsilon)$ .

- Note that the difference quotient can be built using an **early channel correlator** using  $c_+(t) = c(t - \tau + \Delta/2)$  and a **late channel correlator** using  $c_-(t) = c(t - \tau - \Delta/2)$  whose outputs are subtracted from each other to provide the discriminator output.

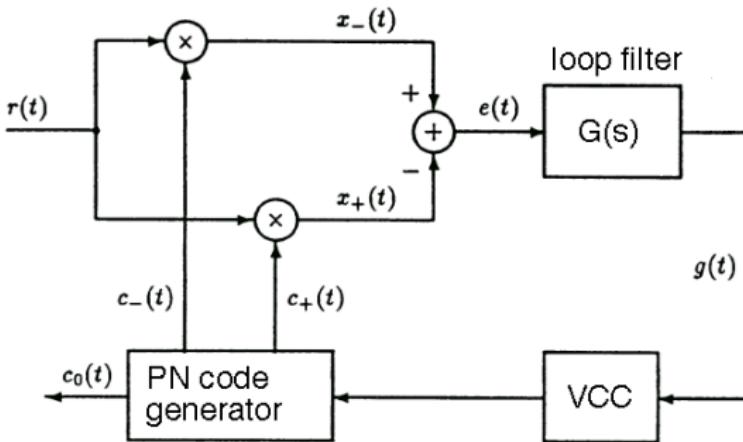
**S-curve**

Form of the **S-curve** in the coherent case and  $\Delta = T_c$ :



### Delay-locked loop (DLL)

Using again a **dynamic implementation** of the estimation rule as in the PLL, we obtain the **coherent delay-locked loop (DLL)**.



## Non-decision-directed timing estimation

- If no symbol estimates  $\hat{I}_n$  are available, we can **average the likelihood function  $\Lambda(\tau)$  over the PDF of  $I_n$**  to obtain  $\bar{\Lambda}(\tau)$  and  $\bar{\Lambda}_L(\tau)$ , respectively. The further procedure for deriving the estimate  $\hat{\tau}$  is completely analogous to the case of phase recovery.
- The averaging of binary baseband PAM, where  $I_n = \pm 1$  with equal probabilities, provides an averaged LLF given by

$$\bar{\Lambda}_L(\tau) = \sum_n \ln \cosh Cy_n(\tau).$$

with some constant  $C$ . Again, we obtain the well-known cosh-characteristics and a suitable loop implements the following equation in a dynamic way:

$$\frac{d}{d\tau} \sum_n \ln \cosh Cy_n(\tau) \Bigg|_{\tau=\hat{\tau}} = 0.$$

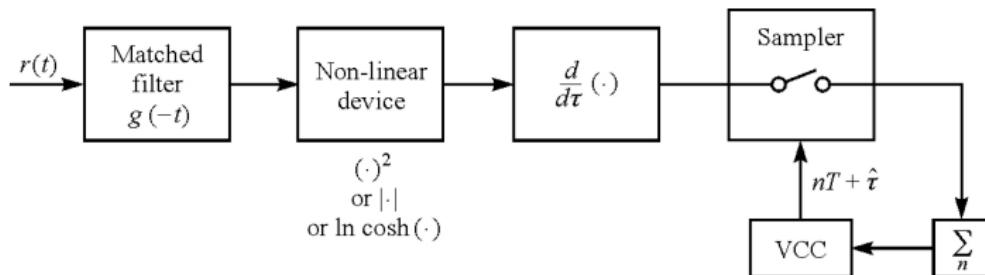
- Using the aforementioned approximation  $\ln \cosh x \approx x^2/2$  for small values of  $|x|$ , we obtain an alternative formulation

$$\bar{\Lambda}_L(\tau) \approx \frac{C^2}{2} \sum_n y_n^2(\tau).$$

and we have to implement

$$\frac{d}{d\tau} \sum_n y_n^2(\tau) \Bigg|_{\tau=\hat{\tau}} = 2 \sum_n y_n(\tau) \frac{d}{d\tau} y_n(\tau) \Bigg|_{\tau=\hat{\tau}} = 0.$$

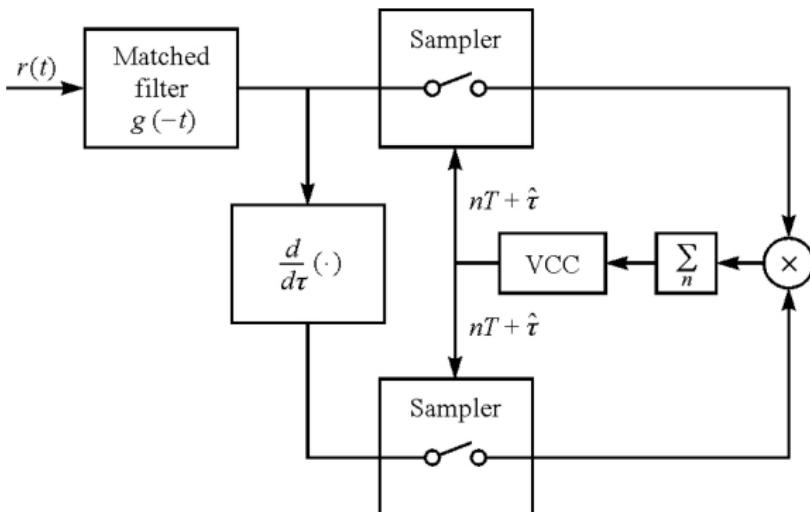
### Non-decision-directed timing estimation



**Fig. 6.3-2** Non-decision directed estimation of timing for binary baseband PAM implementing

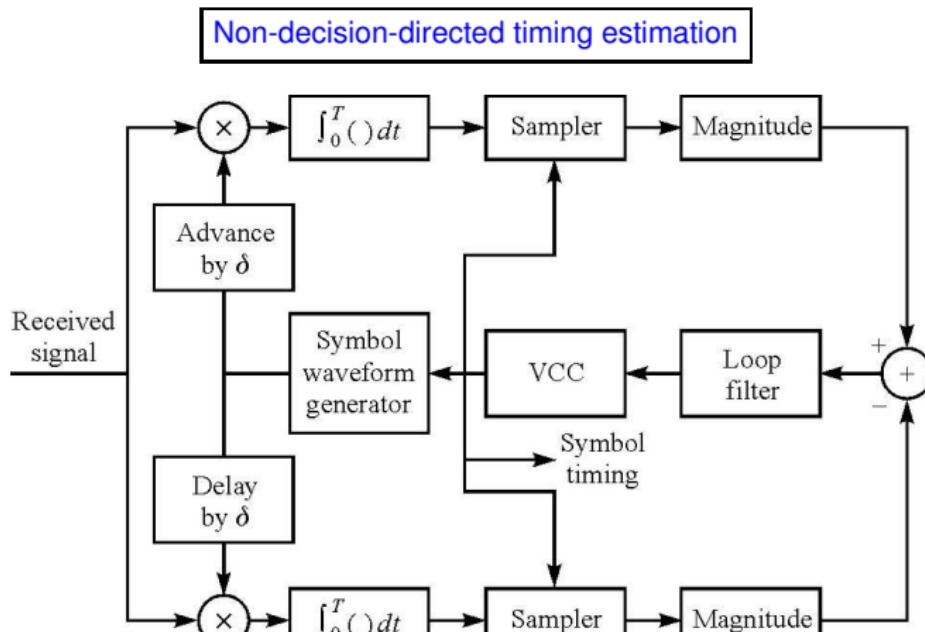
$$\frac{d}{d\tau} \sum_n \ln \cosh C y_n(\tau) \Big|_{\tau=\hat{\tau}} = 0.$$

### Non-decision-directed timing estimation



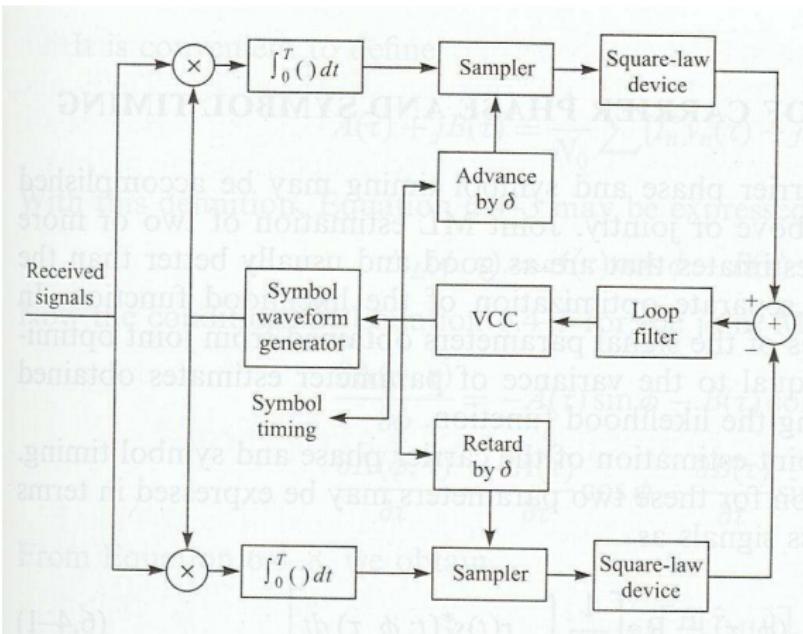
**Fig. 6.3-3** Non-decision directed estimation of timing for baseband PAM implementing

$$\sum_n y_n(\tau) \frac{d}{d\tau} y_n(\tau) \Big|_{\tau=\hat{\tau}} = 0.$$



**Fig. 6.3-5 Block diagram of an early-late gate synchronizer.**

### Non-decision-directed timing estimation



**Fig. 6.3-6** Block diagram of an **early-late gate synchronizer** - an alternative form.

## Joint Estimation of Carrier Phase and Symbol Timing

- Instead of estimating the carrier phase and timing separately, we can estimate both jointly. Typically, a joint scheme is more complex, but also provides estimates with improved performance, such as a reduced variance.
- Assuming that the energy in the observation interval does not depend on the parameter vector  $\Psi = [\phi, \tau]$ , we have the LLF

$$\Lambda_L(\Psi) = \Re \left\{ \frac{1}{N_0} \int_{T_0} r(t) s_\ell^*(t; \Psi) dt \right\},$$

where  $s_\ell(t; \Psi) = s_\ell(t; \phi, \tau)$  is the equivalent lo-pass signal, which is assumed to have the general form

$$s_\ell(t; \phi, \tau) = e^{-j\phi} \sum_n [I_n g(t - nT - \tau) + j J_n w(t - nT - \tau)],$$

where  $\{I_n\}$  and  $\{J_n\}$  are the two quadrature information sequences.

- Again, we assume decision-direction, i.e., we have estimates  $\{\hat{I}_n\}$  and  $\{\hat{J}_n\}$  available at the receiver.

- The LLF becomes

$$\Lambda_L(\Psi) = \Re \left\{ \frac{e^{j\phi}}{N_0} \sum_n \left[ \hat{I}_n^* y_n(\tau) + j \hat{J}_n^* x_n(\tau) \right] \right\}$$

with  $y_n(\tau) = \int_{T_0} r(t)g^*(t - nT - \tau)dt$ ,  $x_n(\tau) = \int_{T_0} r(t)w^*(t - nT - \tau)dt$ .

- Necessary conditions for the estimates  $\hat{\phi}$  and  $\hat{\tau}$  are

$$\frac{d\Lambda_L(\Psi)}{d\phi} \Big|_{\Psi=\hat{\Psi}} = 0 \quad \text{and} \quad \frac{d\Lambda_L(\Psi)}{d\tau} \Big|_{\Psi=\hat{\Psi}} = 0.$$

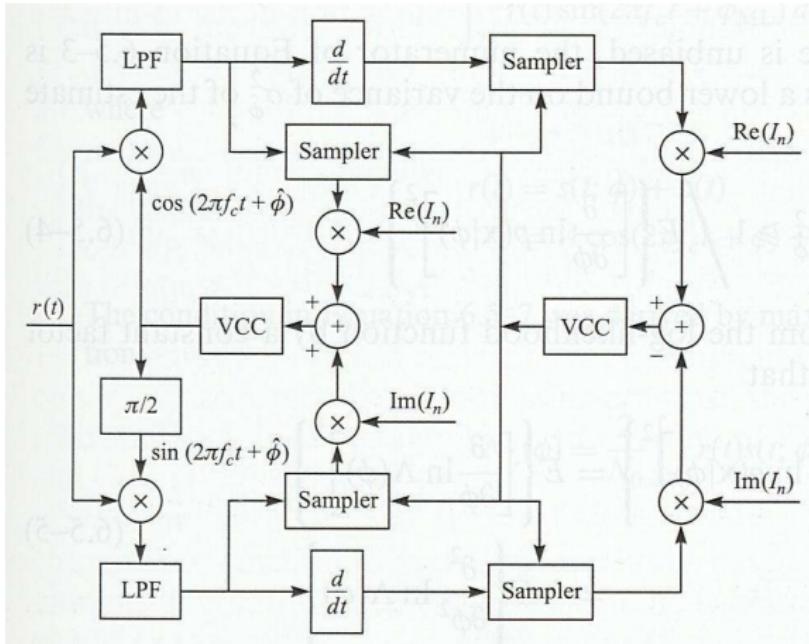
- It is convenient to define

$$A(\tau) + jB(\tau) = \frac{1}{N_0} \sum_n \left[ \hat{I}_n^* y_n(\tau) + j \hat{J}_n^* x_n(\tau) \right].$$

- Problem** Derive the equations defining implicitly the ML estimate  $\hat{\Psi}_{ML}$ . Show that they are given by

$$-\arctan \left( \frac{B(\hat{\tau}_{ML})}{A(\hat{\tau}_{ML})} \right) = \hat{\phi}_{ML}$$

$$\left[ A(\tau) \frac{dA(\tau)}{d\tau} + B(\tau) \frac{dB(\tau)}{d\tau} \right] \Big|_{\tau=\hat{\tau}_{ML}} = 0.$$



**Fig. 6.4-1** Decision-directed joint tracking loop for carrier phase and symbol timing in QAM and PSK.

# Table of Contents

- 1 Introduction
- 2 Carrier and Symbol Synchronization
  - Carrier Phase ( $\phi$ ) Estimation
  - Symbol Timing ( $\tau$ ) Estimation
  - Joint Estimation of  $\phi$  and  $\tau$
- 3 Signal Design for Band-Limited Channels
  - Characterization of Band-Limited Channels
  - Signal Design for No ISI-Nyquist Criterion
  - Signal Design with Controlled ISI
  - Probability of Error in Detection of PAM
  - Modulation Codes for Spectrum Shaping
- 4 Communication Through Band-Limited Linear Filter Channels
  - Optimum Receiver for AWGN Channels with ISI
  - Linear Equalization
  - Decision-Feedback Equalization
  - Equalization at the Transmitter
- 5 Multicarrier and Multichannel Systems
  - ISI in Single Carrier Systems
  - Design Criteria for Broadband System
  - Basic Principle of OFDM Signaling

# Signal Design for Band-Limited Channels

- In the case of AWGN channels, no bandwidth constraint has to be imposed on the signal design and the communication system design.
- Assume from now on that the channel is band-limited to some specified bandwidth of  $W$  Hz. The channel is modeled as a linear filter having an equivalent low-pass frequency response  $C(f)$  that is zero for  $|f| > W$ .
- Consider a linearly modulated signal represented as

$$v(t) = \sum_n I_n g(t - nT)$$

that efficiently utilizes the total available channel bandwidth  $W$ .

- We shall see that
  - when the channel is ideal for  $|f| \leq W$ , a signal pulse can be designed that allows us to transmit at symbol rates comparable to or exceeding the channel bandwidth  $W$
  - when the channel is not ideal, signal transmission at a symbol rate equal or exceeding  $W$  results in inter symbol interference (ISI) among a number of adjacent symbols.
  - the use of coding to shape the spectrum of the transmitted signal can reduce the problem of ISI.

## Characterization of Band-Limited Channels

- Consider a band-limited channel (e.g. a telephone channel) to be characterized as a linear filter with low-pass frequency-response characteristic  $C(f)$  and low-pass impulse response  $c(t)$ .
- If a signal of the form

$$s(t) = \Re \left\{ v(t) e^{j2\pi f_c t} \right\}$$

is transmitted over a band-pass telephone channel, the equivalent low-pass received signal is

$$r_\ell(t) = \int_{-\infty}^{\infty} v(\tau) c(t - \tau) d\tau + z(t),$$

where the integral expression contains the signal information and  $z(t)$  represents the additive noise.

- Since the channel is band-limited to  $W$  Hz, we assume correspondingly  $V(f) = 0$  for  $|f| > W$ .

- Within the channel bandwidth, we may express the frequency response  $C(f)$  as

$$C(f) = |C(f)| e^{j\theta(f)},$$

where  $|C(f)|$  is the amplitude-response characteristic and  $\theta(f)$  is the phase-response characteristic. The **envelope delay characteristic** (sometimes also called **group delay**) is defined as

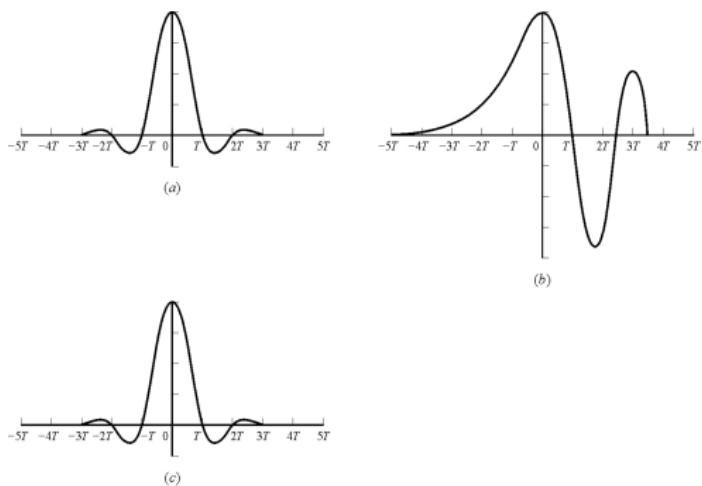
$$\tau(f) = -\frac{1}{2\pi} \frac{d\theta(f)}{df}.$$

- A channel is **non-distorting** or **ideal** if  $|C(f)| = \text{const.}$  for all  $|f| \leq W$  **AND**  $\tau(f) = \text{const.}$  for all  $|f| \leq W$ , i.e.,  $\theta(f)$  is a linear function of  $f$ .
- If  $|C(f)| \neq \text{const.}$  for all  $|f| \leq W$ , the channel **distorts the transmitted signal  $V(f)$  in amplitude**.
- If  $\tau(f) \neq \text{const.}$  for all  $|f| \leq W$ , the channel **distorts the transmitted signal  $V(f)$  in delay**.

**Problem**

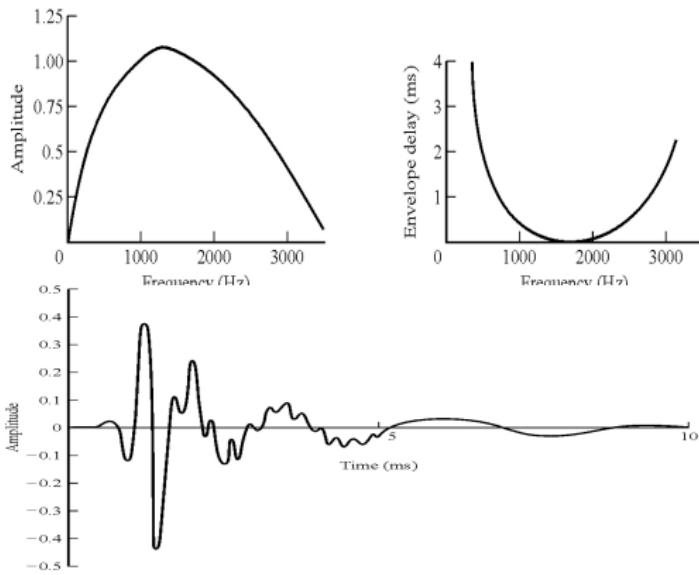
Proof that the conditions required for a channel to be non-distorting indeed guarantee that a band-limited signal can be received without distortion at the output of the channel.

- If a train of successive pulses is transmitted over a distorting channel at rates comparable to the bandwidth  $W$ , the pulses are *smeared* to the point that they are no longer distinguishable as single pulses. Instead, they overlap and thus cause **intersymbol interference**.
- Example (distorted pulse train being **filtered** at the receiver to provide an equalized **output signal**):



**Fig. 9.1-1** Effect of channel distortion: a) channel input signal; b) **channel output signal**; c) **equalizer output**.

### Example of medium-range telephone channel

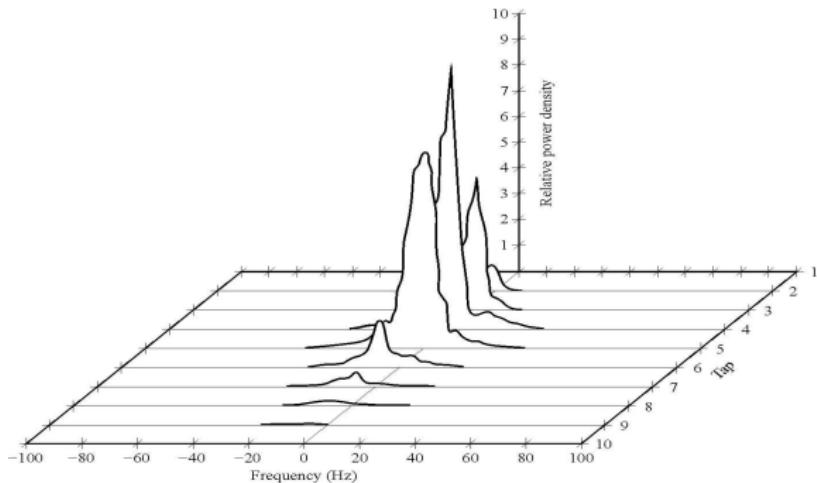


**Fig. 9.1-2/3** Top: average amplitude and delay characteristics of medium-range telephone channel; bottom: impulse response of average channel with amplitude and delay shown on top; **for a rate of 2500 baud  $\Rightarrow$  ISI over 20-30 symbols**

Other sources of distortion:

- non-linearities in amplifiers and compandors
  - ⇒ usually small and difficult to correct
- frequency offset of local oscillators
  - ⇒ cannot be tolerated in broadband systems with phase-coherent demodulation; carrier recovery loop to be used
- phase jitter, i.e. low-index FM, caused by power lines (50-60 Hz)
  - ⇒ serious problem in digital broadband transmission; use compensation device
- impulse noise caused by switching equipment in the telephone network
  - ⇒ usual approach includes some form of clipping, i.e., non-linear post-processing
- thermal noise caused by the power amplifier
  - ⇒ not a serious problem in properly designed transmission systems (including channel coding)

Linear distortions in **linear time-variant transmission channels** (e.g. in cellular mobile radio, tropospheric scatter channel):



**Fig. 9.1-4** Scattering function of a medium-range tropospheric scatter channel.

⇒ here, we exclusively consider channels which can be described as a **linear time-invariant system**.

- In the following, we consider the transmitted signal

$$v(t) = \sum_{n=0}^{\infty} I_n g(t - nT),$$

where  $\{I_n\}$  is the set of information-bearing symbols and the pulse  $g(t)$  is bandlimited with  $G(f) = 0$  for  $|f| > W$ .

- At the output of the bandlimited channel, we obtain in view of the linear modulation

$$r_\ell(t) = \sum_{n=0}^{\infty} I_n h(t - nT) + z(t) \quad \text{with} \quad h(t) = \int_{-\infty}^{\infty} g(\tau) c(t - \tau) d\tau.$$

Thus,  $h(t)$  is the response of the channel to the input pulse  $g(t)$ .

- Let us suppose that the received signal  $r_\ell(t)$  is passed through a filter whose output is sampled at a rate  $1/T$ . The optimum filter is one matched to the received pulse  $h(t)$  so that the filter's frequency response is  $H^*(f)$ .

- The output of the receiving filter is given by

$$y(t) = \sum_{n=0}^{\infty} I_n x(t - nT) + \nu(t),$$

where  $x(t)$  is the pulse representing the response of the receiving filter to the input pulse  $h(t)$  and  $\nu(t)$  is the response of the filter to the noise  $z(t)$ .

- We further assume that  $y(t)$  is sampled at times  $t = kT + \tau_0$ ,  $k = 0, 1, \dots$ , with  $\tau_0$  denoting the transmission delay through the channel. Then,

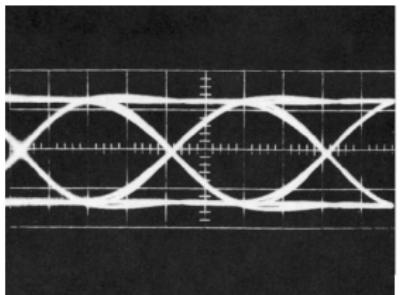
$$y(kT + \tau_0) \equiv y_k = \sum_{n=0}^{\infty} I_n x(kT - nT + \tau_0) + \nu(kT + \tau_0),$$

or, equivalently,

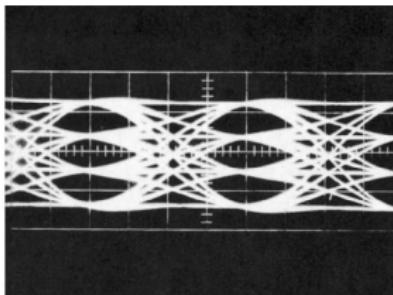
$$\begin{aligned} y_k &= \sum_{n=0}^{\infty} I_n x_{k-n} + \nu_k, \quad k = 0, 1, \dots, \\ &= x_0 \left( I_k + \frac{1}{x_0} \sum_{n=0, n \neq k}^{\infty} I_n x_{k-n} \right) + \nu_k, \quad k = 0, 1, \dots. \end{aligned}$$

- Since we are interested in the value of  $I_k$ , we assume  $x_0$  to be a non-zero scaling factor which we set equal to unity for convenience. Then, at the  $k^{\text{th}}$  sampling instant

$$y_k = \underbrace{I_k}_{\text{desired}} + \underbrace{\sum_{n=0, n \neq k}^{\infty} I_n x_{k-n}}_{\text{undesired ISI}} + \underbrace{\nu_k}_{\text{additive Gaussian noise}}.$$



(a) Binary



(b) Quaternary

**Fig. 9.2-1** Examples of **eye patterns** for binary and quaternary amplitude-shift keying (or PAM).

- The effect of ISI is to **cause the eye to close**, thereby reducing the margin for additive noise to cause errors.

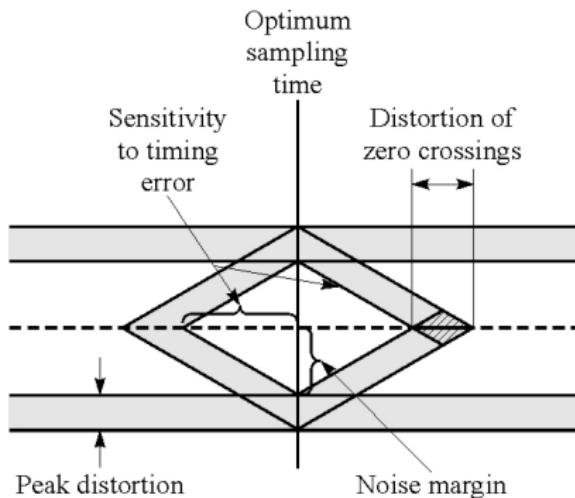
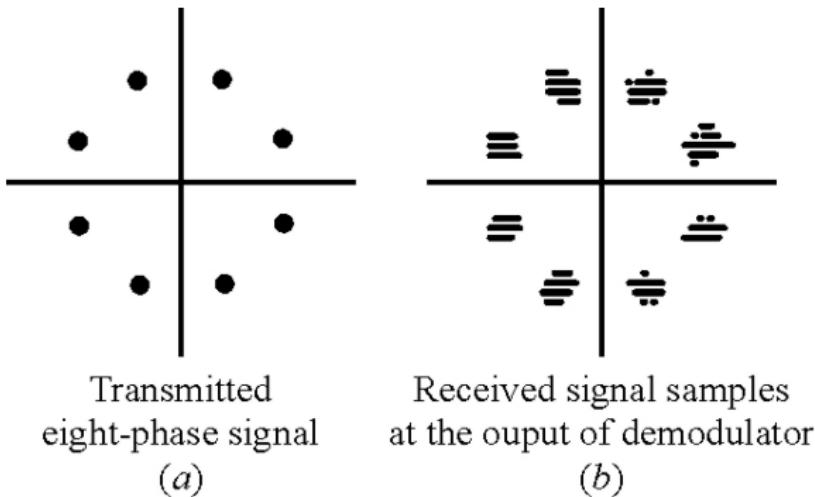


Fig. 9.2-2 Effect of intersymbol interference on eye opening.

- Another way to display the effect of ISI is in the complex plane where the decision variables  $y_k$  are displayed using quadrature components.



**Fig. 9.2-3** Two-dimensional scatter plots (corresponding to eye patterns in both quadrature components at sampling instants).

# Signal Design for No ISI-Nyquist Criterion

- First, we assume an **ideal channel**, i.e.,  $C(f) = 1$  for  $|f| \leq W$ . Then, the pulse  $x(t)$  has a spectral characteristic  $X(f) = |G(f)|^2$ , where

$$x(t) = \int_{-W}^W X(f) e^{j2\pi f t} df.$$

- We want to determine the spectral properties of the pulse  $x(t)$  and, as a consequence, of the transmitted pulse  $g(t)$  **that results in no ISI**. Since

$$y_k = \underbrace{I_k}_{\text{desired}} + \underbrace{\sum_{n=0, n \neq k}^{\infty} I_n x_{k-n}}_{\text{undesired ISI}} + \underbrace{\nu_k}_{\text{additive Gaussian noise}},$$

the condition for no ISI, in view of  $\tau_0 = 0$ , is

$$x(t = kT) \equiv x_k = \begin{cases} 1 & \text{for } k = 0 \\ 0 & \text{for } k \neq 0. \end{cases}$$

Subsequently, we derive the necessary and sufficient condition on  $X(f)$  in order for  $x(t)$  to satisfy the above relation. This condition is known as **Nyquist pulse shaping criterion** or **Nyquist condition for zero ISI**.

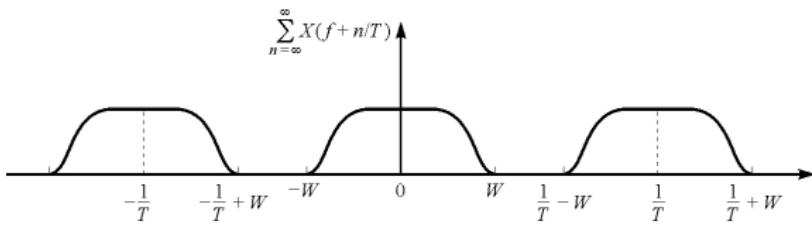
**Theorem (Nyquist):** The necessary and sufficient condition for  $x(t)$  to satisfy

$$x(t = kT) \equiv x_k = \begin{cases} 1 & \text{for } k = 0 \\ 0 & \text{for } k \neq 0 \end{cases}$$

is that its Fourier transform  $X(f)$  satisfies

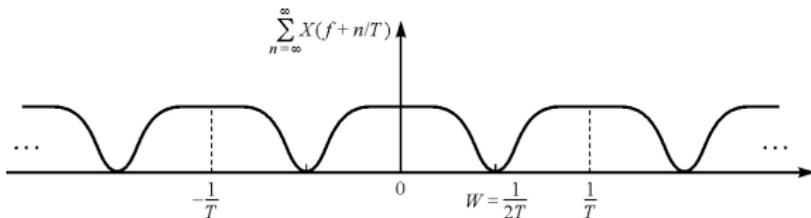
$$\sum_{m=-\infty}^{\infty} X\left(f + \frac{m}{T}\right) = T.$$

**Problem** Proof the Nyquist theorem.



**Fig. 9.2-4** Plot of  $\sum_{n=-\infty}^{\infty} X\left(f + \frac{n}{T}\right)$  for the case  $T < 1/2W$ .

- When  $T < 1/2W$ , or, equivalently,  $1/T > 2W$ , there is no choice for  $X(f)$  to ensure  $\sum_{n=-\infty}^{\infty} X\left(f + \frac{n}{T}\right) = T$ .
- As a result, **for a symbol rate  $1/T$  chosen too large**, there is no way to design a system without ISI.

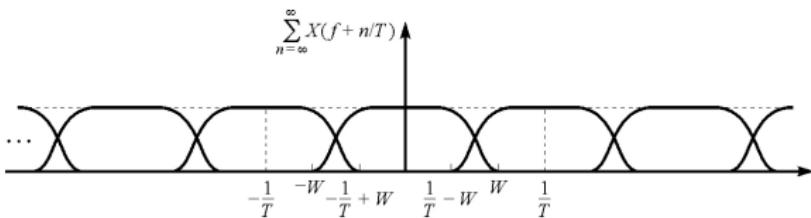


**Fig. 9.2-5** Plot of  $\sum_{n=-\infty}^{\infty} X\left(f + \frac{n}{T}\right)$  for the case  $T = 1/2W$ .

- When  $T = 1/2W$ , or, equivalently,  $1/T = 2W$ , there is only one possible  $X(f)$  to ensure  $\sum_{n=-\infty}^{\infty} X\left(f + \frac{n}{T}\right) = T$ , namely a rectangular frequency characteristic resulting in a time-domain pulse  $x(t) = \frac{\sin(\pi t/T)}{\pi t/T} \equiv \text{sinc}(\frac{\pi t}{T})$ .

### Problems:

- ① non-causal behavior of  $x(t)$  and thus  $x(t)$  not realizable
- ② small timing error results in infinite ISI due to the slow decay of the sinc-pulse (decays with a rate of  $1/t$ ).



**Fig. 9.2-6** Plot of  $\sum_{n=-\infty}^{\infty} X\left(f + \frac{n}{T}\right)$  for the case  $T > 1/2W$ .

- When  $T > 1/2W$ , or, equivalently,  $1/T < 2W$ , there are numerous  $X(f)$  to ensure  $\sum_{n=-\infty}^{\infty} X\left(f + \frac{n}{T}\right) = T$ .
- A pulse which is chosen often in practical systems is one with a **raised cosine frequency characteristic**.

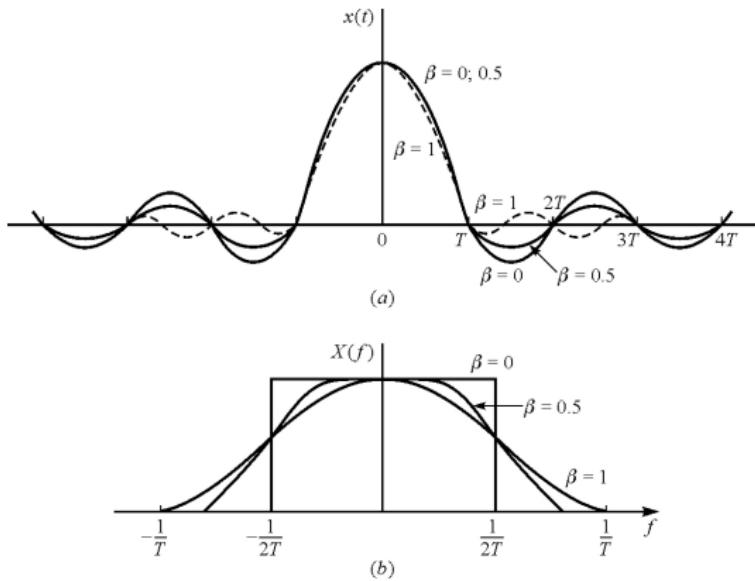


Fig. 9.2-7 Pulses having a raised cosine spectrum.

- The raised cosine frequency characteristic is given as

$$X_{\text{rc}}(f) = \begin{cases} T & \text{for } 0 \leq |f| \leq \frac{1-\beta}{2T} \\ \frac{T}{2} \left( 1 + \cos \left( \frac{\pi T}{\beta} \left[ |f| - \frac{1-\beta}{2T} \right] \right) \right) & \text{for } \frac{1-\beta}{2T} \leq |f| \leq \frac{1+\beta}{2T} \\ 0 & \text{for } |f| > \frac{1+\beta}{2T} \end{cases}$$

where  $\beta$  is the roll-off factor in the range  $0 \leq \beta \leq 1$ .

- The bandwidth occupied by the signal beyond the Nyquist frequency  $1/2T$  equals  $\beta/2T$  and is called the **excess bandwidth**. It is usually expressed as a percentage of the Nyquist frequency and in this case equals  $\beta$ .
- The pulse  $x(t)$  in the time domain is given by

$$x(t) = \text{sinc}(\pi t/T) \frac{\cos \pi \beta t/T}{1 - 4\beta^2 t^2/T^2}.$$

Note that  $x(t)$  is normalized to  $x(0) = 1$ . Furthermore, the tails of  $x(t)$  decay as  $1/t^3$  for  $\beta > 0$ . Thus, a **mistiming error in sampling leads to a series of ISI components converging to a finite value**.

- Problem** Proof that a raised-cosine spectrum pulse  $X_{\text{rc}}(f)$  satisfies the Nyquist criterion for any value of the roll-off factor  $\beta$

- Since  $X_{rc}(f)$  satisfies the Nyquist criterion in any case, it can serve for the design of practical filters for the transmitter and the receiver.
- In the special case  $C(f) = 1$  for  $|f| \leq W$ , we have to guarantee that

$$X_{rc}(f) = G_T(f)G_R(f),$$

where  $G_T(f)$  and  $G_R(f)$  are the frequency responses of the two filters. If the receiver is matched to the transmitter filter, we have

$$X_{rc}(f) = G_T(f)G_R(f) = |G_T(f)|^2,$$

so that ideally we have

$$G_T(f) = \sqrt{|X_{rc}(f)|}e^{-j2\pi f t_0} = G_R^*(f)$$

where  $t_0$  is some nominal delay required to ensure physical realizability of the filter. Thus, the overall raised cosine spectral characteristic is split evenly between the transmitting and receiving filters, where in the latter an additional delay is necessary to ensure realizability.

- For zero ISI, we have to reduce the symbol rate  $1/T$  below the Nyquist rate of  $2W$  symbols/s to realize practical transmitting and receiving filters. By abandoning the zero ISI requirement, we can achieve the Nyquist rate with practical filters as shown next.

# Signal Design with Controlled ISI - Partial Response Signals

- One way to control the ISI is to require, instead of zero ISI, an ISI from one symbol:

$$x(nT) = \begin{cases} 1 & \text{for } n = 0, 1 \\ 0 & \text{otherwise.} \end{cases}$$

- Then, from the proof of the Nyquist theorem, it follows that

$$B(f) = \sum_{m \in \mathbb{Z}} X_{f+m/T} = T \left( 1 + e^{-j2\pi fT} \right)$$

which **cannot be satisfied for  $T < 1/2W$** . However, for  $T = 1/2W$ , we obtain

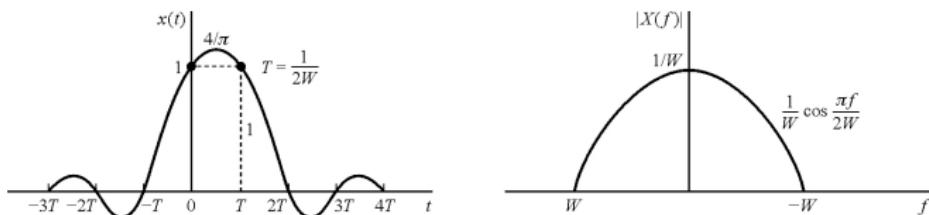
$$X(f) = \begin{cases} \frac{1}{2W} (1 + e^{-j\pi f/W}) & \text{for } |f| < W \\ 0 & \text{otherwise.} \end{cases}$$

- The time-domain pulse results to

$$x(t) = \operatorname{sinc}(2\pi Wt) + \operatorname{sinc}\left(2\pi\left(Wt - \frac{1}{2}\right)\right).$$

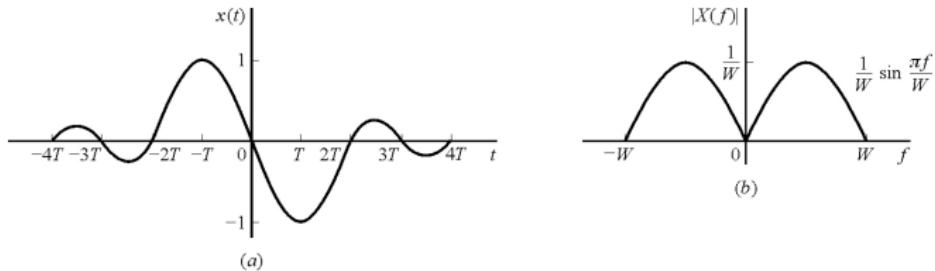
This pulse is called a **duobinary signal pulse**.

- The spectrum decays to zero smoothly so that physically realizable filters can be designed that approximate the spectrum very closely.



**Fig. 9.2-8** Time-domain and frequency-domain characteristics of a duobinary signal.

The approach can be extended to control ISI from two samples resulting in a modified duobinary signal.



**Fig. 9.2-9** Time-domain and frequency-domain characteristics of a modified duobinary signal.

## Data detection for controlled ISI - symbol-by-symbol suboptimum detection

- The special property of received signals with controlled ISI can be exploited in the detection process. Here, we will discuss two approaches for the detection, namely a **simple symbol-by-symbol detection process** and a **complex maximum-likelihood sequence estimation scheme**.
- We consider the detection of the **duobinary** and the **modified duobinary** partial-response signals. In both cases, we assume that the desired spectral characteristic  $X(f)$  for the partial-response signal is split evenly between transmitting and receiving filters, i.e.  $|G_T(f)| = |G_R(f)| = \sqrt{|X(f)|}$ .
- For simplicity, we consider schemes for PAM signals, but the schemes can easily be extended to QAM and PSK.
- We first consider the duobinary pulse with  $x(nT) = 1$  for  $n = 0, 1$  and zero otherwise. Hence, we have the samples at the demodulator output

$$y_m = B_m + \nu_m = I_m + I_{m-1} + \nu_m,$$

where  $\{I_m\}$  is the transmitted sequence of amplitudes and  $\{\nu_m\}$  is a sequence of additive Gaussian noise samples.

## Data detection for controlled ISI - symbol-by-symbol suboptimum detection

- Here, we neglect the noise and consider exclusively the ISI as the main noise contribution in the decision variable upon demodulation.
- Consider the binary case where  $I_m = \pm 1$  with equal probability. Our objective is to construct a signaling scheme which is tailored to the duobinary pulse and the form of  $y_m$ .
- Obviously,  $B_m$  can take the values  $-2, 0, 2$  with probabilities  $1/4, 1/2, 1/4$ , respectively.
- A simple, yet non-robust approach is to detect  $I_{m-1}$  and, given that it is correct, subtract it from  $B_m$  in order to obtain  $I_m$ . This process can be repeated sequentially for every received symbol. Clearly, the main problem of this interference cancellation approach is error propagation. How can error propagation be eliminated?
- One way to avoid error propagation is a special kind of differential encoding usually called precoding. The precoding is performed on the binary data sequence prior to modulation

## Data detection for controlled ISI - symbol-by-symbol suboptimum detection

- From the data sequence  $\{D_m\}$  of 1s and 0s to be transmitted, a new sequence  $\{P_m\}$ , called the **precoded sequence**, is generated. For the duobinary signal, the precoded sequence is defined as

$$P_m = D_m \ominus P_{m-1}, \quad m = 1, 2, \dots,$$

where  $\ominus$  denotes modulo-2 subtraction. Then we set  $I_m = 2P_m - 1$ .

- The noise-free samples at the output of the receiving filter are given by

$$B_m = I_m + I_{m-1} = 2(P_m + P_{m-1} - 1)$$

and thus

$$P_m + P_{m-1} = \frac{1}{2}B_m + 1.$$

## Data detection for controlled ISI - symbol-by-symbol suboptimum detection

- Clearly, for  $P_m + P_{m-1} = 0$  or  $P_m + P_{m-1} = 2$ , we have  $D_m = 0$ . Otherwise,  $P_m + P_{m-1} = 1$  and  $D_m = 1$ . Thus, we have

$$D_m = \left( \frac{1}{2}B_m + 1 \right) \bmod 2.$$

- An example is given in the following

Data sequence $D_n$	1	1	1	0	1	0	0	1	0	0	0	1	1	0	1
Precoded sequence $P_n$	0	1	0	1	1	0	0	0	1	1	1	1	0	1	1
Transmitted sequence $I_n$	-1	1	-1	1	1	-1	-1	-1	1	1	1	1	-1	1	1
Received sequence $B_n$	0	0	0	2	0	-2	-2	0	2	2	2	2	0	0	2
Decoded sequence $D_n$	1	1	1	0	1	0	0	1	0	0	0	1	1	0	1

Table 9.2-1 Binary signaling with duobinary pulses

- In the presence of additive noise, we have  $y_m = B_m + \nu_m$  which is compared to the thresholds  $\pm 1$  and the data sequence  $\{D_m\}$  is obtained using the detection rule

$$D_m = \begin{cases} 1 & \text{for } |y_m| < 1 \\ 0 & \text{for } |y_m| \geq 1. \end{cases}$$

## Data detection for controlled ISI - symbol-by-symbol suboptimum detection

- The extension to  $M$ -level PAM  $\{I_m\}$  using duobinary pulses is straightforward. Now,  $B_m = I_m + I_{m-1}$  has  $2M - 1$  possible equally spaced levels and the amplitude levels are given by  $I_m = 2P_m - (M - 1)$  with the precoded sequence  $\{P_m\}$ , an  $M$ -level data sequence  $\{D_m\}$  with  $D_m \in \{0, 1, \dots, M - 1\}$  and

$$P_m = (D_m \oplus P_{m-1}) \bmod M.$$

- In analogy with the case  $M = 2$ , we have in the absence of noise

$$B_m = I_m + I_{m-1} = 2[P_m + P_{m-1} - (M - 1)].$$

- Hence,  $P_m + P_{m-1} = \frac{1}{2}B_m + (M - 1)$ . Since  $D_m = (P_m + P_{m-1}) \bmod M$ , it follows that

$$D_m = \left( \frac{1}{2}B_m + (M - 1) \right) \bmod M.$$

- An example is given in the following

<b>Data sequence <math>D_n</math></b>	0	0	1	3	1	2	0	3	3	2	0	1	0
Precoded sequence $P_n$	0	0	0	1	2	3	3	1	2	1	1	3	2
Transmitted sequence $I_n$	-3	-3	-3	-1	1	3	3	-1	1	-1	-1	3	1
Received sequence $B_n$	-6	-6	-4	0	4	6	2	0	0	-2	2	4	2
Decoded sequence $D_n$	0	0	1	3	1	2	0	3	3	2	0	1	0

**Table 9.2-2** Four-level signal transmission with duobinary pulses

## Data detection for controlled ISI - symbol-by-symbol suboptimum detection

- In the presence of additive noise, the noisy signal  $y_m = B_m + \nu_m$  is quantized to the nearest of the possible signal levels and then the above rule is used on the quantized values to recover the data sequence.
- Modified duobinary pulses** can be treated similarly. Here, the controlled ISI is specified by the values  $x(nT) = -1$  for  $n = 1$ ,  $x(nT) = 1$  for  $n = -1$  and zero otherwise. Consequently, we have in the noise-free case

$$B_m = I_m - I_{m-2}.$$

- The mapping of the precoded sequence  $\{P_m\}$  is as before, where

$$P_m = (D_m \oplus P_{m-2}) \bmod M.$$

- It can be shown that the detection rule for recovering the data sequence  $\{D_m\}$  from  $\{B_m\}$  in the absence of noise is

$$D_m = \left(\frac{1}{2}B_m\right) \bmod M.$$

- Summarizing the symbol-by-symbol detection**, it is a simple scheme which can be implemented without a memory element to store the previous symbols. It is, in general, a suboptimum scheme since we have not taken into account the noise characteristics in the derivation of the detection schemes. The signaling and detection is applicable to duobinary and modified duobinary pulse signals which makes them attractive for many applications.

## Data detection for controlled ISI - maximum-likelihood sequence detection

- Unlike the symbol-by-symbol detection, the optimum maximum-likelihood (ML) sequence detector takes into account the overall memory of the received signal waveforms. This memory can be represented in a trellis, similarly to a trellis in convolutional coding.
- The number of states for  $M$ -ary modulation and a signal with memory length  $L$  is given by  $M^L$ . Thus, the trellis for the duobinary partial-response signal with memory length  $L = 1$  for binary data transmission ( $M = 2$ ) illustrated below has  $2^1 = 2$  states.

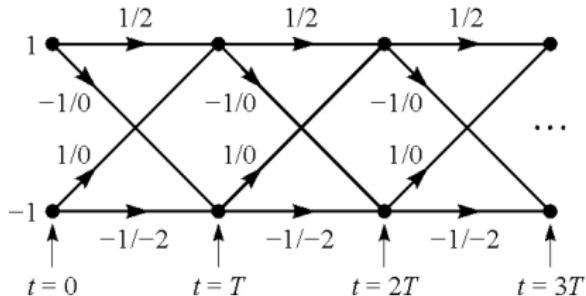


Fig. 9.2-11 Trellis for duobinary partial-response signal.

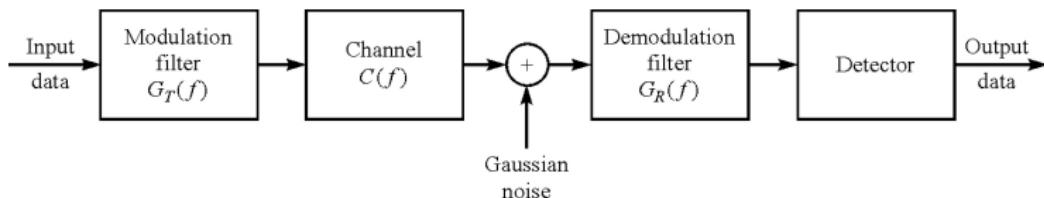
- Each branch in the trellis is labeled by two numbers. The **first number is the new data bit, i.e.  $I_{m+1} = \pm 1$**  and **determines the transition to the new state**. The **second number is the received signal level**.

## Data detection for controlled ISI - maximum-likelihood sequence detection

- In general, each node of the trellis has  $M$  incoming paths and  $M$  corresponding metrics. One out of the  $M$  incoming paths is selected at the most probable, based on the values of the metrics and the other  $M - 1$  paths and their metrics are discarded. The surviving path at each node is extended to  $M$  new paths, one for each of the  $M$  possible input symbols, and the search continues. The algorithm is basically a **Viterbi algorithm (VA)** used also for decoding of convolutionally encoded data streams.
- The metric computations in the trellis search are complicated by the **correlation of noise samples at the output of the matched filter for the partial-response signal**. The correlation properties are reflected by the covariance matrix of the multivariate Gaussian PDF of the received sequence  $\{y_m, 1 \leq m \leq N\}$ . The memory length  $L$  determines the number of non-zero off-diagonal elements in the covariance matrix.
- Substantial **simplifications of the VA are achievable by assuming a diagonal covariance matrix**. In this way, the inversion of the matrix is circumvented and the inner product for metric calculation is a simple norm of an error vector. Clearly, in general, performance losses result from these assumptions.
- The usual properties of the VA have to be taken into account in a practical implementation, in particular the **need for truncating the surviving sequences in order to achieve a fixed detection delay**.

# Signal Design for Channels with Distortion

- In the signal design criteria for the modulation filter at the transmitter and the demodulation filter at the receiver, we have assumed that the channel is ideal. Now, we consider the more realistic case of **a channel distorting the transmitted signal** where the **channel frequency-response  $C(f)$**  is assumed known for  $|f| \leq W$  and that  $C(f) = 0$  for  $|f| > W$ .
- The filter responses  $G_T(f)$  and  $G_R(f)$  may be selected to minimize the error probability at the detector. The additive channel noise is assumed to be Gaussian with power spectral density  $\Phi_{nn}(f)$ .



**Fig. 9.2-12** System model for the design of the modulation and demodulation filters.

- For the transceiver chain, we require

$$G_T(f)C(f)G_R(f) = X_d(f)e^{-j2\pi f t_0} \quad \text{for } |f| \leq W,$$

where  $X_d(f)$  is the desired frequency response of the cascade of the modulator, channel and demodulator, and  $t_0$  is a time delay that is necessary to ensure the physical realizability of the modulation and demodulation filters.

- As above, we might select  $X_d(f)$  to yield either zero or controlled ISI at the sampling instants. For simplicity, we consider the case of zero ISI by selecting  $X_d(f) = X_{rc}(f)$  with  $X_{rc}(f)$  denoting the frequency response of the raised cosine spectrum with an arbitrary roll-off factor  $\beta$ .
- The noise process at the output of the demodulation filter is given by  $\nu(t) = n(t) * g_R(t)$ , where  $n(t)$  is the input to the filter. Since  $n(t)$  is zero-mean Gaussian,  $\nu(t)$  is also zero-mean Gaussian with a spectral power density

$$\Phi_{\nu\nu}(f) = \Phi_{nn}(f) |G_R(f)|^2.$$

- We consider binary PAM transmission with a matched filter output

$$y_m = x_0 I_m + \nu_m = I_m + \nu_m,$$

where  $x_0 = 1$ ,  $I_m = \pm d$  and  $\nu_m$  represents the noise term which is zero-mean Gaussian with variance

$$\sigma_\nu^2 = \int_{\mathbb{R}} \Phi_{nn}(f) |G_R(f)|^2 df.$$

- Since we assume the data symbols  $I_m = \pm d$  to be equally probable a priori, the probability of error is

$$P_2 = \frac{1}{\sqrt{2\pi}} \int_{d/\sigma_\nu}^{\infty} e^{-y^2/2} dy = Q\left(\sqrt{\frac{d^2}{\sigma_\nu^2}}\right).$$

Clearly, the probability of error is minimized by maximizing the signal-to-noise ratio  $d^2/\sigma_\nu^2$ .

- Let us consider two possible solutions for the case in which the additive Gaussian noise is white, so that  $\Phi_{nn}(f) = N_0/2$ . In case 1, we assume that we pre-compensate for the total channel distortion at the transmitter. In this case, we have the magnitude characteristics for  $|f| \leq W$  given by

$$|G_T(f)| = \frac{\sqrt{X_{rc}(f)}}{|C(f)|} \quad |G_R(f)| = \sqrt{X_{rc}(f)},$$

where the phase characteristic of  $C(f)$  may also be compensated at the transmitter filter.

- The average transmitted power is

$$P_{\text{av}} = \frac{E[I_m^2]}{T} \int_{\mathbb{R}} g_T(t) dt = \frac{d^2}{T} \int_{-W}^W |G_T(f)|^2 df = \frac{d^2}{T} \int_{-W}^W \frac{|X_{\text{rc}}(f)|}{|C(f)|^2} df$$

and thus  $d^2 = P_{\text{av}} T \int_{-W}^W \frac{|X_{\text{rc}}(f)|}{|C(f)|^2} df$ .

- Finally, the noise at the output of the receiver filter is  $\sigma_\nu^2 = N_0/2$  and the SNR at the detector for case 1 results to

$$\frac{d^2}{\sigma_\nu^2} = \frac{2P_{\text{av}}T}{N_0} \left[ \int_{-W}^W \frac{|X_{\text{rc}}(f)|}{|C(f)|^2} df \right]^{-1}.$$

- Now consider **case 2** where we split the channel compensation equally between the transmitter and receiver filters, i.e. for  $|f| \leq W$  we have

$$|G_T(f)| = |G_R(f)| = \sqrt{\frac{X_{rc}(f)}{|C(f)|}},$$

where the phase characteristic of  $C(f)$  may also be split equally between the transmitter and receiver filters. In this case, the **average transmitter power** is

$$P_{av} = \frac{d^2}{T} \int_{-W}^W \frac{X_{rc}(f)}{|C(f)|} df$$

and the noise variance at the output of the receiver filter is

$$\sigma_\nu^2 = \frac{N_0}{2} \int_{-W}^W \frac{X_{rc}(f)}{|C(f)|} df.$$

- Hence, the SNR at the detector for case 2 results to

$$\frac{d^2}{\sigma_\nu^2} = \frac{2P_{av}T}{N_0} \left[ \int_{-W}^W \frac{X_{rc}(f)}{|C(f)|} df \right]^{-2}.$$

- In the case of an ideal channel, we have  $C(f) = 1$  for  $|f| \leq W$  and  $\int_{-W}^W X_{rc}(f) df = 1$ .

As a consequence, we obtain that for the ideal channel the SNR values for both cases are identical and given by  $2P_{av}T/N_0$ .

- If, however, there is an amplitude distortion  $|C(f)| < 1$  for some range of frequencies in the band  $|f| \leq W$ , a loss results for both cases. These losses are

$$10 \log \int_{-W}^W \frac{X_{rc}(f)}{|C(f)|^2} df \quad \text{for case 1}$$

$$10 \log \left[ \int_{-W}^W \frac{X_{rc}(f)}{|C(f)|} df \right]^2 \quad \text{for case 2.}$$

- **Problem** Use the Cauchy-Schwarz inequality to show that case 2 minimizes the noise-to-signal ratio given by  $\sigma_\nu^2/d^2$ .
- **Example 9.2-1** Let us determine the transmitting and receiving filters given by case 2 for a binary communication system that transmits data at a rate of 4800 bit/s over a channel with frequency (magnitude) response

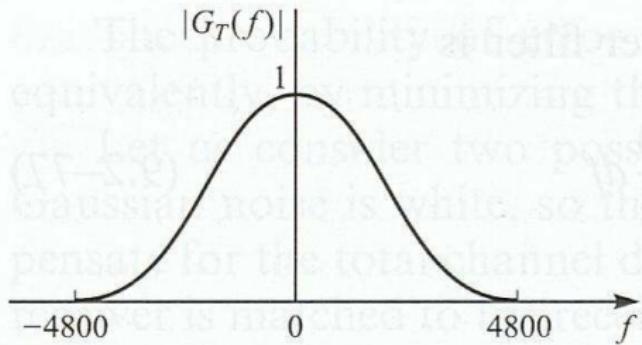
$$|C(f)| = \frac{1}{\sqrt{1 + (f/W)^2}}, \quad |f| \leq W$$

where  $W = 4800$  Hz. The additive noise is zero-mean white Gaussian with spectral density  $N_0/2 = 10^{-15}$  W/Hz.

Since  $W = 1/T$ , we use a signal pulse with a raised cosine spectrum and  $\beta = 1$  since for  $\beta < 1$ , we would have  $W < 1/T$ . Thus,  $X_{rc}(f) = T(1 + \cos(\pi T |f|))/2 = T \cos^2(\pi |f|/9600)$ .

Then,  $|G_T(f)| = |G_R(f)| = 0$  for  $|f| > 4800$  and

$$|G_T(f)| = |G_R(f)| = \left[ 1 + \left( \frac{f}{4800} \right)^2 \right]^{1/4} \cos\left(\frac{\pi |f|}{9600}\right) \quad \text{for } |f| \leq 4800.$$



**Fig. 9.2-13** Frequency response of an optimum transmitter filter.

# Probability of Error in Detection of PAM

zero ISI

- Subsequently, we will investigate the probability-of-error performance of both zero-ISI transceivers as well as of partial-response signals, namely duobinary and modified duobinary signals.
- In the absence of ISI, the output samples at the receiving filter are given by

where  $x_0 = \int_{-W}^W |G_T(f)|^2 df = \mathcal{E}_g$  and  $\nu_m$  is the additive Gaussian noise with zero-mean and variance  $\sigma_\nu^2 = \mathcal{E}_g N_0 / 2$ .

- In general,  $I_m$  takes one of  $M$  possible equally spaced amplitude values with equal probability. Due to the fact that there is no ISI, we can directly use the formula for the probability of error for  $M$ -ary PAM in AWGN. We obtain (cf. Sect. 5.2 in Proakis' book) the **probability of a symbol error**

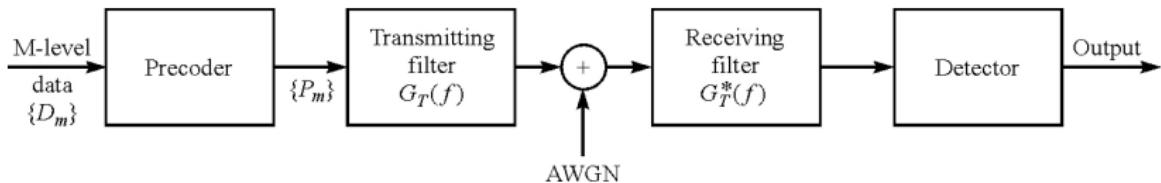
$$P_M = \frac{2(M-1)}{M} Q\left(\sqrt{\frac{6 \ln(M) \mathcal{E}_{b,\text{av}}}{(M^2 - 1) N_0}}\right),$$

where  $\mathcal{E}_{b,\text{av}}$  denotes the average energy per bit.

- Note that although the probability-of-error expression is the same like in the case of transmission in AWGN, we have imposed here the additional constraint that the transmitted signal is band-limited to the bandwidth allocated for the channel. Consequently, the transmitted signal pulses were designed to be band-limited and to have zero ISI.

### Partial-response signal

- We now turn to the case of **partial-response signals**, where the channel is assumed ideal band-limited with AWGN and ***M*-ary PAM transmission** is considered.



**Fig. 9.3-1** Block diagram of modulator and demodulator for partial-response signals.

- Let us first consider a **symbol-by-symbol detector**. The  $M$ -level data sequence  $\{D_m\}$  is precoded as described above. The receiving filter  $G_R(f)$  satisfies  $G_R(f) = G_T^*(f)$  and we have  $|G_T(f)G_R(f)| = |X(f)|$ .
- We have

$$y_m = I_m + I_{m-1} + \nu_m = B_m + \nu_m \quad \text{for the duobinary signal}$$

$$y_m = I_m - I_{m-2} + \nu_m = B_m + \nu_m \quad \text{for the modified duobinary signal}$$

### Partial-response signal

- For  $M$ -ary PAM signal transmission, where  $I_m = \pm d, \pm 3d, \dots, \pm(M-1)d$ , the received signal levels are  $B_m = 0, \pm 2d, \pm 4d, \dots, \pm 2(M-1)d$ . Hence, the number of received levels is  $2M - 1$  and the scale factor  $d$  is equivalent to  $x_0 = \mathcal{E}_g$ .
- The input transmitted symbols  $\{I_m\}$  are assumed to be equally probable. Then, for both duobinary and modified duobinary signals, it is easily demonstrated that in the noise-free case, we have a distribution of the received signal levels given by

$$P(B = 2md) = \frac{M - |m|}{M^2}, \quad m = 0, \pm 1, \pm 2, \dots, \pm(M-1),$$

where  $B$  denotes the noise-free received level and  $2d$  is the distance between any two adjacent signal levels.

- The power spectral density of the AWGN equals  $N_0/2$ , so that the noise component  $\nu_m$  is zero-mean Gaussian with variance

$$\sigma_\nu^2 = \frac{N_0}{2} \int_{-W}^W |G_R(f)|^2 df = \frac{N_0}{2} \int_{-W}^W |X(f)| df.$$

For both duobinary and modified duobinary signals, we have  $\int_{-W}^W |X(f)| df = 4/\pi$ , so that  $\sigma_\nu^2 = 2N_0/\pi$ .

### Partial-response signal

- We assume that a symbol error occurs whenever the magnitude of the additive noise exceeds the distance  $d$ . This assumption, in view of the modulo-operation at the detector, neglects the rare event that a large noise component with magnitude exceeding  $d$  may result in a signal level yielding a correct symbol decision.
- The upper bound based on the above reasoning can be constructed according to

$$\begin{aligned}
 P_M &< \sum_{m=-(M-2)}^{M-2} P(|y - 2md| > d | B = 2md) P(B = 2md) \\
 &\quad + 2P(y + 2(M-1)d > d | B = -2(M-1)d) P(B = -2(M-1)d) \\
 &= P(|y| > d | B = 0) \left[ 2 \sum_{m=0}^{M-1} P(B = 2md) - P(B = 0) - P(B = -2(M-1)d) \right] \\
 &= (1 - M^{-2}) P(|y| > d | B = 0),
 \end{aligned}$$

where we have used in the last equation the expression for  $P(B = 2md)$  from the previous slide.

- The probability in the last equation is given by

$$P(|y| > d | B = 0) = \frac{2}{\sqrt{2\pi}\sigma_\nu} \int_d^\infty e^{-x^2/2\sigma_\nu^2} dx = 2Q\left(\sqrt{\frac{\pi d^2}{2N_0}}\right).$$

### Partial-response signal

- Thus, we have the **upper bound on the symbol error** as

$$P_M < 2(1 - M^{-2})Q\left(\sqrt{\frac{\pi d^2}{2N_0}}\right).$$

- We want to express the scale factor  $d$  by the average power  $P_{av}$ . To this end, we note that

$$P_{av} = \frac{E[I_m^2]}{T} \int_{-W}^W |G_T(f)|^2 df = \frac{E[I_m^2]}{T} \int_{-W}^W |X(f)| df = \frac{4}{\pi T} E[I_m^2],$$

where  $E[I_m^2]$  is the mean square value of the  $M$  signal levels, which is  $E[I_m^2] = d^2(M^2 - 1)/3$ .

- As a result, we have

$$d^2 = \frac{3\pi P_{av} T}{4(M^2 - 1)}$$

and obtain the final expression for the upper bound

$$P_M < 2(1 - M^{-2})Q\left(\sqrt{\left(\frac{\pi}{4}\right)^2 \frac{6}{M^2 - 1} \frac{\text{ld}(M)\mathcal{E}_{b,av}}{N_0}}\right).$$

### Partial-response signal

- If we compare the bound for the partial-response signals with the exact expression for  $P_M$  in the case of zero ISI, we see that the partial-response signal detection has a loss of  $(\frac{\pi}{4})^2$  or 2.1 dB. This loss results from the suboptimum symbol-by-symbol detection which ignores the inherent memory carrying information which has to be exploited in the optimum detection scheme.
- For the optimum maximum-likelihood sequence detection scheme, it can be shown (without proof) for the duobinary and modified duobinary signals that the aforementioned loss of 2.1 dB can be completely recovered.

## Modulation Codes for Spectrum Shaping

- Coding can be employed to achieve different effects in a digital communication system. Apart from **source coding** and **channel coding** treated in the lecture *Digital Communications II*, there is another application of codes in the context of channels with ISI. Here, coding can be employed to shape the spectrum of the transmitted signal. These codes are called **modulation codes, line codes** or **data translation codes**.
- To explain the use of these codes, we consider the example of a magnetic recording which can be cast in the framework of a usual transmission channel in a digital communication system. Here, the modulation code is used to increase the distance between transitions in the recorded waveform and, thus, **to reduce intersymbol interference**.

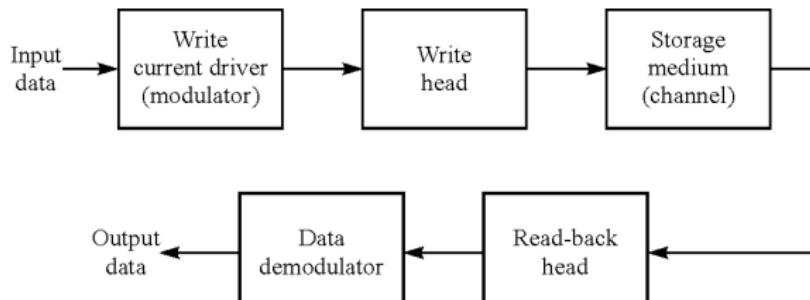


Fig. 9.4-1 Block diagram of magnetic storage read/write system.

- The most common write-waveform is NRZI, where a transition from one amplitude to another (i.e.  $A$  to  $-A$  or  $-A$  to  $A$ ) occurs only when the information bit is a 1. The positive amplitude pulse results in magnetizing the medium in one polarity and the negative pulse magnetizes the medium in the opposite polarity.
- The readback signal for a positive transition from  $-A$  to  $A$  is a pulse that is well-modeled mathematically as

$$g(t) = \frac{1}{1 + (2t/T_{50})^2},$$

where  $T_{50}$  is defined as the width of the pulse at its 50 percent amplitude level. A negative transition from  $A$  to  $-A$  results in a negative pulse  $-g(t)$ .

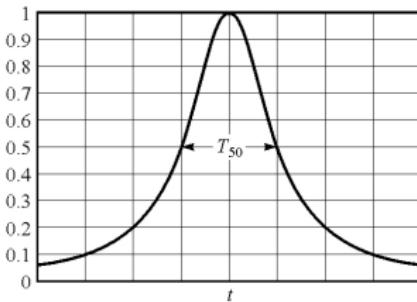
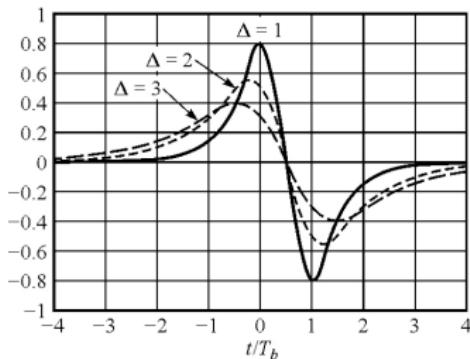


Fig. 9.4-2 Readback pulse in magnetic recording system.

- Consider the transition signal  $g(t) - g(t - T_b)$  where  $T_b$  denotes the bit interval. Varying the bit interval is equivalent to varying the so-called **normalized density**  $\Delta = T_{50}/T_b$ . The closer the bit transitions, the larger will be the packing density and  $\Delta$ . As  $\Delta$  is increased, the **peak amplitudes of the readback signal is reduced** thus limiting the density with which we can write.

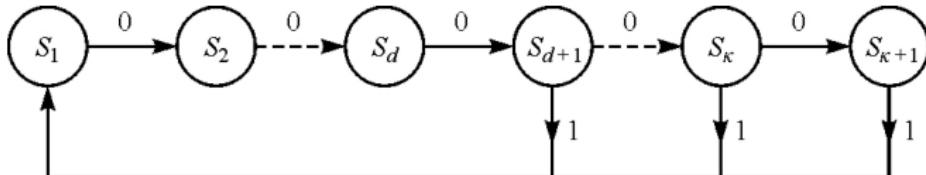


**Fig. 9.4-3** Readback signal response to a pulse for different values of  $\Delta$ .

- This problem serves as a motivation to design **modulation codes** that take the original data sequence and transform (encode) it into another sequence that results in a write waveform in which amplitude transitions are spaced farther apart (**data translation codes**). As an example, if we use NRZI, the encoded sequence into the modulator must contain one or more zeros between ones.

## Runlength-limited codes

- Codes having a restriction on the number of consecutive ones or zeros in sequence are generally called runlength-limited codes.
- Clearly, we can classify such codes by two numbers, namely the minimum and maximum numbers  $d$  and  $\kappa$ , resp., of zeros between two ones in a sequence. When used, for example, with NRZI modulation, the effect of
  - placing  $d$  zeros between successive ones is to spread the transitions farther apart
  - upper limiting  $\kappa$  ensures that transitions occur frequently enough to facilitate symbol timing estimation.
- Runlength-limited codes are usually called  $(d, \kappa)$  codes.
- The  $(d, \kappa)$  code sequence constraints may be represented by a finite-state sequential machine with  $\kappa + 1$  states, denoted as  $S_i$ ,  $1 \leq i \leq \kappa + 1$ .



**Fig. 9.4-4** Finite-state sequential machine for a  $(d, \kappa)$ -coded sequence.

## Runlength-limited codes

- We see that
  - an output bit 0 takes the sequence from state  $S_i$  to state  $S_{i+1}$
  - the output bit 1 takes the sequence to state  $S_1$
  - the output bit may be 1 only when the sequence is in state  $S_i$  for  $d + 1 \leq i \leq \kappa + 1$
  - when the sequence is in state  $S_{\kappa+1}$ , the output bit is always 1.
- These properties of the finite-state machine may also be represented in form of a  $(\kappa + 1) \times (\kappa + 1)$ -dimensional **state transition matrix  $D = \{d_{ij}\}$** , where

$$d_{ij} = \begin{cases} 1 & \text{for a possible transition from state } S_i \text{ to state } S_j \\ 0 & \text{otherwise.} \end{cases}$$

We obtain  $d_{i1} = 1$  for  $i \geq d + 1$  and  $d_{ij} = \delta_{j-(i+1)}$  with  $\delta_i$  denoting the Kronecker delta.

- The  $(d, \kappa)$  code sequence constraints may be represented by a finite-state sequential machine with  $\kappa + 1$  states, denoted as  $S_i$ ,  $1 \leq i \leq \kappa + 1$ .

### Runlength-limited codes

- **Example 9.4-1** For a  $(d, \kappa) = (1, 3)$  code with its four states, we obtain a state transition matrix

$$\mathbf{D} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

- An important parameter of any  $(d, \kappa)$  code is the number of sequences of a certain length  $n$  satisfying the  $(d, \kappa)$  constraints.
- For increasing  $n$ , the number of sequences  $N(n)$  that satisfy the  $(d, \kappa)$  constraints also increases. Clearly, the number  $k$  of information bits that can be uniquely represented with  $N(n)$  code sequences (in the sense that there is a one-to-one mapping between the  $k$  bits and the  $N(n)$  sequences) is given by

$$k = \lfloor \log_2(N(n)) \rfloor.$$

with  $\lfloor x \rfloor$  denoting the largest integer being less or equal to  $x$ . The maximum achievable code rate is then  $R = k/n$ .

### Runlength-limited codes

- Usually, a channel can be assigned a **Shannon capacity** in bits/use. Since the code design is based on the assumption of an error-free transmission of the uniquely decodable code words through a channel, we can (with a slight abuse of terms) define a **capacity of a  $(d, \kappa)$  code** as

$$C(d, \kappa) = \lim_{n \rightarrow \infty} \frac{\text{ld}(N(n))}{n}.$$

$C(d, \kappa)$  is the **maximum possible rate that can be achieved with the  $(d, \kappa)$  constraints**. Shannon showed that the capacity is given by

$$C(d, \kappa) = \text{ld}(\lambda_{\max}),$$

where  $\lambda_{\max}$  is the largest eigenvalue of the state transition matrix  $\mathbf{D}$ .

### Runlength-limited codes

- **Example 9.4-2** For the  $(d, \kappa) = (1, 3)$  code in Example 9.4-1, we have

$$\det \mathbf{D} - \lambda \mathbf{I} = \begin{vmatrix} -\lambda & 1 & 0 & 0 \\ 1 & -\lambda & 1 & 0 \\ 1 & 0 & -\lambda & 1 \\ 1 & 0 & 0 & -\lambda \end{vmatrix} = \lambda^4 - \lambda^2 - \lambda - 1 = 0,$$

from which we obtain the capacity  $C(1, 3) = \text{ld}(\lambda_{\max}) = \text{ld}(1.4656) = 0.5515$ .

$\kappa$	$d = 0$	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$d = 6$
2	.8791	.4057					
3	.9468	.5515	.2878				
4	.9752	.6174	.4057	.2232			
5	.9881	.6509	.4650	.3218	.1823		
6	.9942	.6690	.4979	.3746	.2269	.1542	
7	.9971	.6793	.5174	.4057	.3142	.2281	.1335
8	.9986	.6853	.5293	.4251	.3432	.2709	.1993
9	.9993	.6888	.5369	.4376	.3620	.2979	.2382
10	.9996	.6909	.5418	.4460	.3746	.3158	.2633
11	.9998	.6922	.5450	.4516	.3833	.3285	.2804
12	.9999	.6930	.5471	.4555	.3894	.3369	.2924
13	.9999	.6935	.5485	.4583	.3937	.3432	.3011
14	.9999	.6938	.5495	.4602	.3968	.3478	.3074
15	.9999	.6939	.5501	.4615	.3991	.3513	.3122
$\infty$	1.000	.6942	.5515	.4650	.4057	.3620	.3282

**Table 9.4-1** Capacity  $C(d, \kappa)$  versus runlength parameter  $d$  and  $\kappa$ .

### Runlength-limited codes

- We can construct  $(d, \kappa)$  codes either as **fixed-length codes** or as **variable-length codes**. In a fixed-length code, each block of  $k$  bits is encoded into a block of  $n > k$  bits.
- Principle of a fixed-length code design:
  - For a given block length  $n$ , we select the subset of the  $2^n$  code words that satisfy the specified runlength constraints.
  - From this subset, we eliminate the code words not satisfying the runlength constraints when concatenated.
  - The encoding and decoding operations can be performed by use of a lookup table.

### Runlength-limited codes

- **Example 9.4-3** Let us construct a  $(d, \kappa) = (0, 2)$  code of length  $n = 3$ , and determine its efficiency. Upon listing all possible code words, we first obtain the following code words satisfying the  $(0, 2)$  constraint:  $(010), (011), (101), (110), (111)$ . We may **select any four of these code words** and use them to encode the pairs of data bits  $(00, 01, 10, 11)$ . Thus, we have a **rate  $k/n = 2/3$  code** that **satisfies the  $(0,2)$  constraint**.

The **efficiency is not very high**. From Table 9.4-1, the capacity is  $C(0, 2) = 0.8791$  and thus, we obtain an **efficiency** of

$$\frac{R_c}{C(d, \kappa)} = \frac{2/3}{0.8791} = 0.76.$$

Surely, **better  $(0,2)$  codes can be constructed by increasing the block length  $n$** .

- A possible **variable-length code** will be discussed for

**Example 9.4-6** We want to construct a  $(2,7)$  variable block size code, which is neither unique nor trivial. The example chosen here is the  $(2,7)$  code which has been widely used by IBM in many of its disk storage systems.

### Runlength-limited codes

Input data bits	Output coded sequence
1 0	1 0 0 0
1 1	0 1 0 0
0 1 1	0 0 0 1 0 0
0 1 0	0 0 1 0 0 0
0 0 0	1 0 0 1 0 0
0 0 1 1	0 0 1 0 0 1 0 0
0 0 1 0	0 0 0 0 1 0 0 0

**Table 9.4-3** Code book for variable-length (2,7) code.

- The input data blocks of 2,3 and 4 bits are mapped into data blocks of 4,6 and 8 bits, respectively. Hence, the code rate is  $R_c = 1/2$ . Since this is the **code rate for all code words**, the code is called a **fixed-rate code**. It has an efficiency of  $0.5/0.5174 = 0.966$  and satisfies the prefix condition which ensures instantaneous decodability.

## Problems

- Show that, for any value of  $\beta$ , the raised cosine spectrum satisfies  $\int_{\mathbb{R}} X_{\text{rc}}(f) df = 1$ .
- A voice-band telephone channel has a passband characteristic in the frequency range  $300\text{Hz} < f < 3000\text{Hz}$ .
  - Select a symbol rate and a power efficient constellation size to achieve a 9600 bit/s signal transmission.
  - If a square-root raised cosine pulse is used for the transmitter pulse  $g(t)$ , select the roll-off factor. Assume that the channel has an ideal frequency-response characteristic.
- The binary sequence 10010110010 is the input to a precoder whose output is used to modulate a duobinary transmitting filter. Construct a table showing the precoded sequence, the transmitted amplitude levels, the received signal levels, and the decoded sequence. Rewrite the table if a modified duobinary signal pulse is used instead.
- A 4-PAM modulation is used for transmitting at a bit rate of 9600 bit/s on a channel having a frequency response

$$C(f) = \frac{1}{1 + \gamma f / 2400}.$$

for  $|f| \leq 2400$  and  $C(f) = 0$  otherwise. The additive noise is zero-mean white Gaussian with power spectral density  $N_0/2$  W/Hz. Determine the magnitude frequency-response characteristic of the optimum transmitting and receiving filters.

# Table of Contents

- 1 Introduction
- 2 Carrier and Symbol Synchronization
  - Carrier Phase ( $\phi$ ) Estimation
  - Symbol Timing ( $\tau$ ) Estimation
  - Joint Estimation of  $\phi$  and  $\tau$
- 3 Signal Design for Band-Limited Channels
  - Characterization of Band-Limited Channels
  - Signal Design for No ISI-Nyquist Criterion
  - Signal Design with Controlled ISI
  - Probability of Error in Detection of PAM
  - Modulation Codes for Spectrum Shaping
- 4 Communication Through Band-Limited Linear Filter Channels
  - Optimum Receiver for AWGN Channels with ISI
  - Linear Equalization
  - Decision-Feedback Equalization
  - Equalization at the Transmitter
- 5 Multicarrier and Multichannel Systems
  - ISI in Single Carrier Systems
  - Design Criteria for Broadband System
  - Basic Principle of OFDM Signaling

# Communication Through Band-Limited Linear Filter Channels

- The signal design for band-limited channels discussed above is based on the assumption of a channel response characteristic  $C(f)$  being known a priori.
- However, in practical digital communication systems,  $C(f)$  is not known with sufficient precision to design optimum filters for the modulator and demodulator. Examples are dial-up telephone networks with channel routes depending on the dialed number, wireless communications with time-variant propagation characteristics or underwater acoustic channels used for signal transmission by moving submarines.
- For these channels, it is impossible to design optimum fixed demodulation filters.
- Here, we will consider the problem of a transmission over a linear filter channel embedded in AWGN. In general, we will be confronted with intersymbol interference (ISI) and the problem is to design a receiver that employs a means for compensating or reducing the ISI in the received signal.
- The compensator for the ISI is called an **equalizer** since it aims at equalizing the frequency-response of the overall channel characteristic (including the equalizer) to resemble the form of a non-dispersive channel. The latter has an equal channel gain for all frequencies  $f \in \mathbb{R}$ .

- Several types of equalization methods can be employed:
  - maximum-likelihood (ML) sequence estimation (MLSE) being optimum from a probability-of-error viewpoint
  - linear equalization representing a simple suboptimum method where filter coefficients are adapted to satisfy a certain design objective
  - decision-feedback equalization using previously detected symbols to suppress ISI
  - reduced complexity ML detection methods aiming at achieving the performance of the MLSE with reduced complexity.
- Before we discuss the different equalizer concepts, we have to formulate the problem and introduce a so-called equivalent discrete-time channel model. The latter serves for deriving the equalizers.
- Note that we will assume that the characteristics describing the discrete-time channel model, e.g. the channel impulse response or the transfer function to be known at the receiver. The topic of estimating these characteristics will be discussed later on.

# Optimum Receiver for AWGN Channels with ISI

MLSE

- We consider the following model of the received signal in the equivalent low-pass representation

$$r_\ell(t) = \sum_n I_n h(t - nT) + z(t),$$

where  $h(t)$  represents the channel impulse response to the input signal pulse  $g(t)$  and  $z(t)$  is the AWGN process.

- First, we demonstrate that the optimum demodulator can be realized as
  - a filter matched to  $h(t)$
  - followed by a sampler operating at the symbol rate  $1/T$  and
  - a subsequent processing algorithm for estimating the information sequence  $\{I_n\}$  from the sample values.

Consequently, we can say that the samples at the output of the matched filter are sufficient for the estimation of the sequence  $\{I_n\}$ . We say that the matched filter outputs form a sufficient statistics for the detection of  $\{I_n\}$ , since we do not lose any information by limiting ourselves to these values in the detection process.

## MLSE

- We expand  $r_\ell(t)$  in the series

$$r_\ell(t) = \lim_{N \rightarrow \infty} \sum_{k=1}^N r_k f_k(t),$$

where  $f_k(t)$  is a complete set of orthonormal functions and  $\{r_k\}$  are the observable random variables obtained by projecting  $r_\ell(t)$  onto the set  $\{f_k(t)\}$ .

- Clearly, the coefficients  $r_k$  satisfy

$$r_k = \sum_n I_n h_{kn} + z_k, \quad k = 1, 2, \dots,$$

where  $h_{kn}$  is the value obtained from projecting  $h(t - nT)$  onto  $f_k(t)$  and  $z_k$  is the value obtained from projecting  $z(t)$  onto  $f_k(t)$ .

- The assumption of an AWGN channel results in a covariance function of  $z_k$  given by

$$\frac{1}{2} E[z_k^* z_m] = N_0 \delta_{k-m}.$$

- The joint probability density function of the random variables  $\mathbf{r}_N \equiv [r_1 \ r_2 \ \dots \ r_N]$  for a given transmitted sequence  $\mathbf{I}_p \equiv [I_1 \ I_2 \ \dots \ I_p]$  with  $p \leq N$  is

$$p(\mathbf{r}_N | \mathbf{I}_p) = \left( \frac{1}{2\pi N_0} \right)^N \exp \left( -\frac{1}{2N_0} \sum_{k=1}^N \left| r_k - \sum_n I_n h_{kn} \right|^2 \right).$$

## MLSE

- As the number  $N$  approaches infinity, the contributions in the log-likelihood function can be, similar to the derivation in the carrier phase estimation section, expressed by time integral expression.
- Neglecting the term  $\int_{\mathbb{R}} |r_\ell(t)|^2 dt$  being independent of the information symbols to be detected, we obtain the metrics

$$CM(\mathbf{I}_p) = 2\Re \left\{ \sum_n I_n^* y_n \right\} - \sum_n \sum_m I_n^* I_m x_{n-m},$$

where  $y_n = \int_{\mathbb{R}} r_\ell(t) h^*(t - nT) dt$  and  $x_n = \int_{\mathbb{R}} h^*(t) h(t + nT) dt$ .

- If we substitute the signal model for  $r_\ell(t)$ , we obtain

$$y_k = \sum_n I_n x_{k-n} + \nu_k,$$

where  $\nu_k$  denotes the additive noise sequence of the output of the matched filter, i.e.

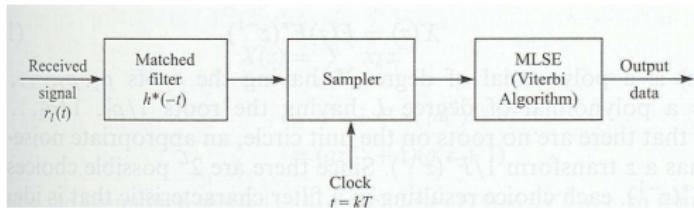
$$\nu_k = \int_{\mathbb{R}} z(t) h^*(t - kT) dt.$$

Obviously, the output of the demodulator (matched filter) is corrupted by ISI.

## MLSE

- In any system, the ISI spans over a limited number of symbols. Therefore, we may assume that  $x_n = 0$  for  $|n| > L$ . Consequently, the ISI observed at the output of the demodulator may be viewed as the output of a finite state machine. This implies that the channel output with ISI may be represented by a trellis diagram and the MLSE provides a means for estimating the sequence with minimum probability of error. Clearly, the Viterbi algorithm can be used to find this sequence.
- From the form of the metrics  $CM(\mathbf{I}_p) = 2\Re \left\{ \sum_n I_n^* y_n \right\} - \sum_n \sum_m I_n^* I_m x_{n-m}$ , it can be seen that  $CM(\mathbf{I}_p)$  can be calculated recursively in the Viterbi algorithm according to

$$CM_n(\mathbf{I}_n) = CM_{n-1}(\mathbf{I}_{n-1}) + \Re \left\{ I_n^* \left( 2y_n - x_0 I_n - 2 \sum_{m=1}^L x_{n-m} I_{n-m} \right) \right\}.$$

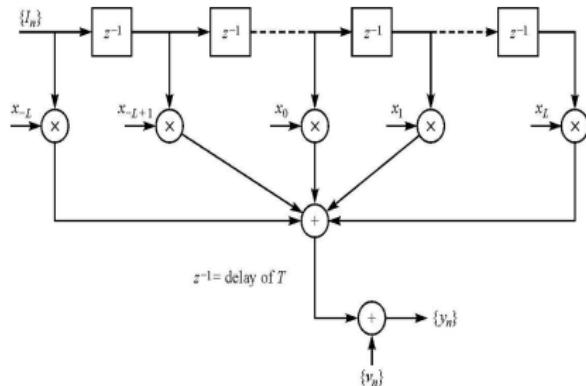


**Fig. 10.1-1** Optimum receiver for an AWGN channel with ISI.

## A discrete-time model for a channel with ISI

- Since the transmitter sends symbols at a rate  $1/T$  and the matched filter at the receiver is sampled also with a rate  $1/T$ , the cascade of
  - analog filter at the transmitter with impulse response  $g(t)$
  - the channel with impulse response  $c(t)$
  - the matched filter at the receiver with impulse response  $h^*(t)$  and
  - the sampler

can be represented by an equivalent discrete-time transversal filter that spans a time interval of  $2LT$  seconds.



**Fig. 10.1-2** Equivalent discrete-time model of channel with intersymbol interference.

A discrete-time model for a channel with ISI

- From the fact that the matched filter at the receiver has, in general, a memory of  $L$  non-zero taps, it follows that the covariance function of the complex baseband noise process  $\nu_k$  is given by

$$\begin{aligned}
 \frac{1}{2}E[\nu_{k+m}\nu_k^*] &= \frac{1}{2}E\left[\int_{\mathbb{R}} \int_{\mathbb{R}} z(t)h^*(t-(k+m)T)z^*(\tau)h(t-kT)dtd\tau\right] \\
 &= \frac{1}{2}\int_{\mathbb{R}} \int_{\mathbb{R}} h^*(t-(k+m)T)E[z(t)z^*(\tau)]h(\tau-kT)dtd\tau \\
 &= \int_{\mathbb{R}} \int_{\mathbb{R}} h^*(t-(k+m)T)N_0\delta(t-\tau)h(\tau-kT)dtd\tau \\
 &= N_0 \int_{\mathbb{R}} h^*(t)h(t+mT)dtd\tau \\
 &= N_0x_m.
 \end{aligned}$$

A discrete-time model for a channel with ISI

- Hence, the random vector

$$\mathbf{y} = [y_1, y_2, \dots, y_N]^T = \begin{bmatrix} \sum_n I_n x_{1-n} \\ \sum_n I_n x_{2-n} \\ \vdots \\ \sum_n I_n x_{N-n} \end{bmatrix} + \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_N \end{bmatrix} = \bar{\mathbf{y}} + \boldsymbol{\nu}$$

is Gaussian distributed with  $\mathbf{y} \sim \mathcal{N}(\bar{\mathbf{y}}; \Sigma_y)$  and a covariance matrix  $\Sigma_y = \{n_{kj}\} = \{\frac{1}{2}E[\nu_j \nu_k^*]\} = \{N_0 x_{j-k}\} = N_0 \mathbf{X}$  which is, in general, non-diagonal.

- We know that any linear transformation  $\mathbf{Gy}$  of a Gaussian vector  $\mathbf{y}$  results in a Gaussian vector. If the transformation matrix  $\mathbf{G}$  is one-to-one, the information contained in the vectors  $\mathbf{y}$  and  $\mathbf{Gy}$  is the same.
- Since it is much easier to describe the properties of the ISI channel, in particular the performance of detection schemes, if the involved multivariate Gaussian vectors contain white noise, we look for a transformation  $\mathbf{G}$  which diagonalizes the covariance matrix of the transformed vector.

### A discrete-time model for a channel with ISI

- Indeed, such a transformation exists and can be obtained from the following

**PROPOSITION:** For every hermitian positive definite matrix  $\Sigma$ , there exists a unique lower invertible triangular matrix  $\mathbf{A}$ , i.e.  $F_{kj} = 0$  for  $k < j$ , with positive diagonal elements such that

$$\Sigma = \mathbf{A}\mathbf{A}^H.$$

Consequently, if we set  $\mathbf{X} = \mathbf{A}\mathbf{A}^H$  and choose  $\mathbf{G} = \mathbf{A}^{-1}$ , we obtain a transformed vector  $\mathbf{v} = [v_1, v_2, \dots, v_N]^T = \mathbf{G}\mathbf{y} \sim \mathcal{N}(\bar{\mathbf{v}}; \Sigma_{\mathbf{v}})$  with  $\bar{\mathbf{v}} = \mathbf{G}\bar{\mathbf{y}}$  and covariance matrix  $\Sigma_{\mathbf{v}} = \mathbf{G}\Sigma_{\mathbf{y}}\mathbf{G}^H = N_0\mathbf{G}\mathbf{X}\mathbf{G}^H = N_0\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^H(\mathbf{A}^{-1})^H = N_0\mathbf{A}^{-1}\mathbf{A}(\mathbf{A}^{-1}\mathbf{A})^H = N_0\mathbf{I}$ .

- Obviously, after the transformation, the resulting Gaussian vector  $\mathbf{v}$  has a white spectrum where the filtering expressed by  $\mathbf{G}$  is the so-called **noise-whitening filter**.
- Interestingly, the lower triangularity of  $\mathbf{A}$  implies that  $\mathbf{G} = \mathbf{A}^{-1}$  is also lower triangular. As a consequence, we can write

$$v_k = \sum_{\ell=1}^k a_{kl} y_\ell,$$

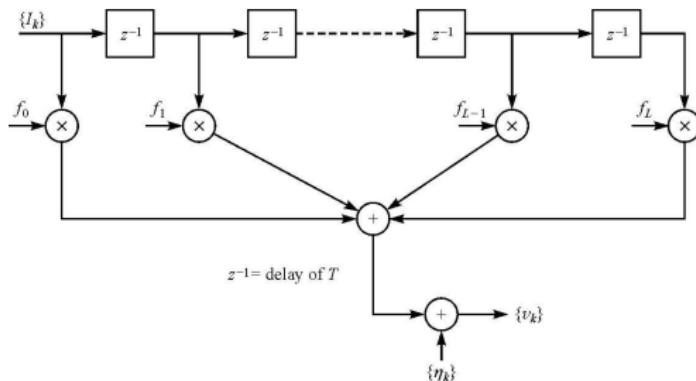
where  $\mathbf{G} = \{a_{kl}\}$ . Note that this operation represents a **causal, but possibly time-varying linear filtering**.

A discrete-time model for a channel with ISI

- We say that the cascade of the matched filter, the sampler and the noise-whitening filter constitutes the **whitened matched filter (WMF)**  $f_0, \dots, f_L$  with

$$v_k = \sum_{n=0}^L f_n I_{k-n} + \eta_k,$$

and a white noise sequence  $\{\eta_k\}$  with variance  $N_0$ .



**Fig. 10.1-3** Equivalent discrete-time model of intersymbol interference channel with AWGN.

### A discrete-time model for a channel with ISI

- The discrete-time white noise linear filter model for ISI effects arising in high-speed digital transmission over non-ideal band-limited channels will be used subsequently in the discussion of compensation techniques for the interference.
- In general, the compensation methods are called **equalization techniques** or simply **equalizers**. The reason for that is the fact that conceptually, all the techniques aim at providing a flat channel gain characteristic with ideally a linear phase over the transmission band as it exists in an ideal channel and which has an equal gain value within the band.
- We may consider the following approaches for constructing the equalizer:
  - Viterbi algorithm for implementing an MLSE
  - linear equalization
  - decision-feedback equalization
  - reduced complexity ML detectors
  - iterative equalization and decoding: turbo equalization.

## Viterbi algorithm

- An MLSE of the information sequence  $\{I_n\}$  is most easily described in terms of the received sequence  $\{v_k\}$  at the output of the whitening filter.
- In the presence of ISI spanning  $L + 1$  symbols, i.e.  $L$  interfering components, the MLSE criterion is equivalent to the problem of estimating the state of a discrete-time finite-state machine.
- The finite-state machine in this case is the equivalent discrete-time channel with coefficients  $\{f_k\}$  and its state at any instant in time is given by the  $L$  most recent inputs, i.e. the state at time  $k$  is

$$S_k = (I_{k-1}, \dots, I_{k-L}),$$

where  $I_k = 0$  for  $k \leq 0$ .

- Hence, if the information symbols are  $M$ -ary, the channel filter has  $M^L$  states and the channel is described by an  $M^L$ -state trellis. As in the case of a convolutional code, the Viterbi algorithm can be used to determine the most likely path through the trellis.
- Starting from samples  $v_1, v_2, \dots, v_{L+1}$ , we compute the  $M^{L+1}$  metrics

$$\sum_{k=1}^{L+1} \ln p(v_k | I_k, \dots, I_{k-L}).$$

- Then, the  $M^{L+1}$  possible sequences  $I_{L+1}, \dots, I_1$  are subdivided into  $M^L$  groups corresponding to the  $M^L$  states  $(I_{L+1}, \dots, I_2)$ .

### Viterbi algorithm

- Note that the  $M$  sequences in each group (state) differ in  $I_1$  and correspond to the paths through the trellis that merge at a single node.
- From the  $M$  sequences in each of the  $M^L$  groups, we **select the sequence with the largest likelihood (with respect to  $I_1$ )** and **assign to the surviving sequence the metric**

$$PM_1(\mathbf{I}_{L+1}) \equiv PM_1(I_{L+1}, \dots, I_2) = \max_{I_1} \sum_{k=1}^{L+1} \ln p(v_k | I_k, \dots, I_{k-L}).$$

The  $M - 1$  remaining sequences from each of the  $M^L$  groups are discarded. Thus, we are left with  $M^L$  surviving sequences and their metrics.

- Upon reception of  $v_{L+2}$ , the  $M^L$  surviving sequences are extended by one stage and the corresponding  $M^{L+1}$  likelihoods for the extended sequences are computed using the previous metrics and the new increment, which is  $\ln p(v_{L+2} | I_{L+2}, \dots, I_2)$ .
- Again, the  $M^{L+1}$  sequences are subdivided into  $M^L$  groups corresponding to the  $M^L$  possible states  $(I_{L+2}, \dots, I_3)$  and the most likely sequence from the group is selected while the other  $M - 1$  sequences are discarded.
- The procedure continues with the reception of subsequent signal samples.

### Viterbi algorithm

- In general, upon reception of  $v_{L+k}$ , the metrics

$$PM_k(\mathbf{I}_{L+k}) = \max_{I_k} [\ln(p(v_{L+k} | I_{L+k}, \dots, I_k)) + PM_{k-1}(\mathbf{I}_{L+k-1})].$$

which have to be computed give the likelihoods of the  $M^L$  surviving sequences.

- Thus, as each signal sample is received, the Viterbi algorithm involves

- first the computation of the  $M^{L+1}$  likelihoods

$$\ln(p(v_{L+k} | I_{L+k}, \dots, I_k)) + PM_{k-1}(\mathbf{I}_{L+k-1})$$

corresponding to the  $M^{L+1}$  sequences forming the continuation of the  $M^L$  surviving sequences from the previous stage of the process

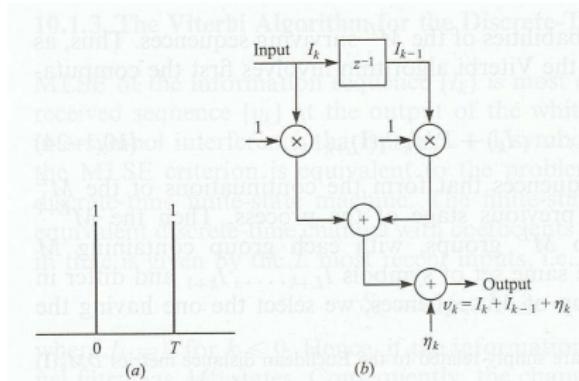
- the subdivision of the  $M^{L+1}$  sequences into  $M^L$  groups terminating in the same set of symbols  $I_{L+k}, \dots, I_{k+1}$  and differing in symbol  $I_k$
  - selection from the  $M$  sequences in each group the one with the largest likelihood
  - discarding the residual  $M - 1$  sequences in each group
  - storing the  $M^L$  sequences and the corresponding metrics  $PM_k(\mathbf{I}_{L+k})$ .

### Viterbi algorithm

- As indicated in the context of decoding convolutional codes, the delay in detecting each information symbol is variable since the  $M^L$  surviving sequences at time  $k$  might disagree on the symbol  $I_{k-q}$  with  $q \gg L$ .
- In practice, the variable delay is avoided by truncating the surviving sequences to the  $q$  most recent symbols where  $q \gg L$ , thus achieving a fixed delay.
- In the case where the  $M^L$  surviving sequences at time  $k$  disagree on the symbol  $I_{k-q}$ , the symbol in the most probable sequence may be chosen. Still, we keep all surviving sequences independently of the decided symbol, since it might turn out later on that we have erroneously chosen a symbol which is not lying in the sequence (extended in length compared to the time instant of the final choice of the symbol) with highest likelihood.
- However, the loss of performance resulting from the suboptimum decision procedure is negligible if  $q \gg 5L$ .

**Example 10.1-2** Suppose we use a duobinary pulse for transmission of a four-level PAM. Thus,  $L = 1$  and  $M = 4$ , where  $v_k = I_k + I_{k-1} + \eta_k$ .

## Viterbi algorithm



**Fig. 10.1-4** Equivalent discrete-time model for ISI resulting from a duobinary pulse.

- In the first step, we compute the 16 metrics

$$PM_1(I_2, I_1) = - \sum_{k=1}^2 \left( v_k - \sum_{j=0}^1 I_{k-j} \right)^2, \quad I_1, I_2 = \pm 1, \pm 3,$$

where  $I_k = 0$  for  $k \leq 0$ .

### Viterbi algorithm

- From these 16 metrics, we form 4 groups, namely

$$\{(I_1, I_2) | I_1 \in \{-3, -1, 1, 3\}, I_2 = 3\}$$

$$\{(I_1, I_2) | I_1 \in \{-3, -1, 1, 3\}, I_2 = 1\}$$

$$\{(I_1, I_2) | I_1 \in \{-3, -1, 1, 3\}, I_2 = -1\}$$

$$\{(I_1, I_2) | I_1 \in \{-3, -1, 1, 3\}, I_2 = -3\}.$$

- Then, we select the surviving sequences by maximizing the expressions  $PM_1(I_2, I_1)$  within each group. For the first group above, where  $I_2 = 3$ , we select for example the root branch which satisfies  $PM_1(I_2 = 3, I_1) = \max_{I_1} \left( - \sum_{k=1}^2 \left( v_k - \sum_{j=0}^1 I_{k-j} \right)^2 \right)$ .
- When  $v_3$  is received, the four surviving paths are extended to yield again 16 paths with corresponding metrics

$$PM_2(I_3, I_2, I_1) = PM_1(I_2, I_1) - \left( v_3 - \sum_{j=0}^1 I_{3-j} \right)^2$$

and select again the surviving paths in the new four groups.

## Viterbi algorithm

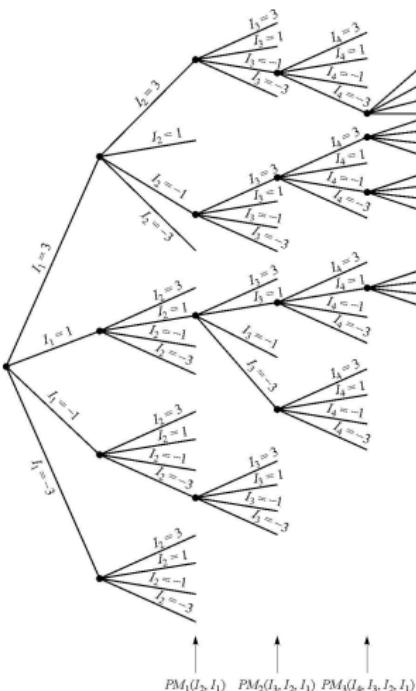


Fig. 10.1-4 Tree diagram for Viterbi decoding of the duobinary pulse.

### Viterbi algorithm

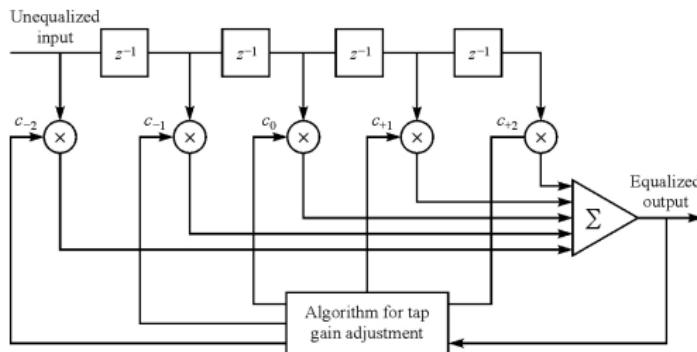
- The performance of the Viterbi algorithm in a given channel can be characterized in form of an upper bound on the loss in SNR caused by the channel. The characterization which has been proposed by Forney in 1972 is based on the calculation of the probability of an error event in the trellis.
- Some examples of the maximum loss are given in

Channel length $L + 1$	Performance loss $-10 \log \delta_{\min}^2 \text{dB}$	Minimum-distance channel
3	2.3	0.50, 0.71, 0.50
4	4.2	0.38, 0.60, 0.60, 0.38
5	5.7	0.29, 0.50, 0.58, 0.50, 0.29
6	7.0	0.23, 0.42, 0.52, 0.52, 0.42, 0.23

**Table 10.1-1** Maximum performance loss and corresponding channel characteristics.

## Linear Equalization

- The MLSE approach requires, in general, a large computational effort that grows exponentially with the length of the channel time dispersion  $L$ .
- If the size of the symbol alphabet is  $M$  and the number of interfering symbols contributing to ISI is  $L$ , the Viterbi algorithm computes  $M^{L+1}$  metrics for each new received symbol. In most practical systems, in particular hand-held devices being battery powered, such a large computational effort is prohibitive.
- Hence, we look for a sub-optimum approach to channel equalization with substantially reduced complexity compared to MLSE. One possible scheme employs a linear transversal filter whose computational complexity grows only linearly with the channel dispersion length  $L$ .



**Fig. 10.2-1** Linear transversal filter for channel equalization.

- The input to the equalizer is the output of the whitening filter  $\{v_k\}$ . At the **equalizer output**, we observe an **estimated sequence**  $\{\hat{I}_k\}$  of the information sequence  $\{I_k\}$ . We have

$$\hat{I}_k = \sum_{j=-K}^K c_j v_{k-j},$$

where  $\{c_j\}$  are the  $2K + 1$  complex-valued tap weight coefficients of the filter.

- The estimate  $\hat{I}_k$  is quantized to the nearest (in distance) information symbol to form the **decision**  $\tilde{I}_k$ . If  $\tilde{I}_k \neq I_k$ , and **error has been made**.
- Clearly, the most straightforward approach for choosing the filter coefficients would be to minimize the symbol-error probability as a function of  $\{c_j\}$ .
- Unfortunately, the symbol-error probability is a **highly non-linear function of  $\{c_j\}$**  not allowing a computationally efficient implementation of the equalization rule.
- In the following we will discuss two criteria for optimizing  $\{c_j\}$  which have been adopted in many practical systems, namely the **peak distortion criterion** and the **mean-square-error criterion**.

### Peak distortion criterion

- The peak distortion is simply defined as the worst-case intersymbol interference at the output of the equalizer. The minimization of this performance index is called the peak distortion criterion.
- The cascade of the discrete-time linear filter model with impulse response  $\{f_n\}$  and an equalizer with impulse response  $\{c_n\}$  can be represented by a single equivalent filter having the overall impulse response

$$q_n = \sum_{j=-\infty}^{\infty} c_j f_{n-j}.$$

- Note that the equalizer is allowed to have an infinite number of taps. Its output at the  $k$ th sampling instant can be expressed in the form

$$\hat{I}_k = q_0 I_k + \sum_{n \neq k} I_n q_{k-n} + \sum_{j=-\infty}^{\infty} c_j \eta_{k-j}.$$

- The first term represents the desired signal where, for convenience, we normalize  $q_0$  to unity.
- The second term is the ISI. The peak value of this interference is called peak distortion and is given by

$$\mathcal{D}(\mathbf{c}) = \sum_{n \in \mathbb{Z}/\{0\}} |q_n| = \sum_{n \in \mathbb{Z}/\{0\}} \left| \sum_{j \in \mathbb{Z}} c_j f_{n-j} \right|.$$

### Peak distortion criterion

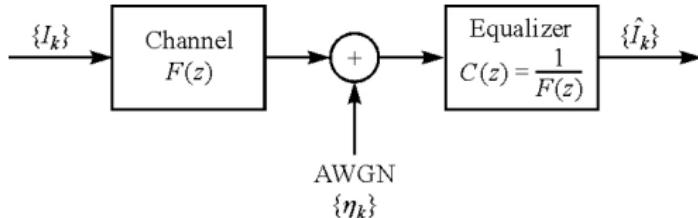
- Clearly, the peak distortion is a function of the equalizer tap weights  $\{c_n\}$ . With an equalizer having an infinite number of taps, it is possible to select the tap weights such that  $\mathcal{D}(\mathbf{c}) = 0$ , i.e.  $q_n = \delta_n$ .
- Hence, the ISI can be completely eliminated. Taking the  $z$ -transform of the equation

$$q_n = \sum_{j=-\infty}^{\infty} c_j f_{n-j} = \delta_n,$$

we immediately obtain

$$Q(z) = C(z)F(z) = 1 \quad \Leftrightarrow \quad C(z) = \frac{1}{F(z)}.$$

- Obviously, the **complete elimination of the ISI requires a filter which is the inverse of the linear filter model  $F(z)$** . We call such a filter a **zero-forcing filter**, since it forces the ISI to zero.



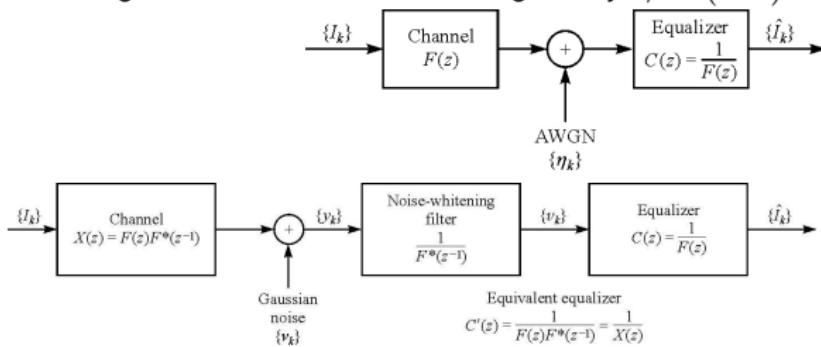
**Fig. 10.2-2** Block diagram of channel with zero-forcing equalizer.

### Peak distortion criterion

- Using from the  $z$ -transform  $X(z)$  of the sampled autocorrelation function  $\{x_k\}$  of the received signal pulse  $h(t)$ , i.e. the channel with intersymbol interference, it can be shown that

$$X(z) = F(z)F^*(z^{-1}),$$

and that the whitening filter has a transfer function given by  $1/F^*(z^{-1})$ .



**Fig. 10.2-2/3** Equivalent block diagrams of channel with zero-forcing equalizer.

- Note that the overall filter characteristic at the receiver results from the **cascade of the whitening filter and the zero-forcer** according to

$$C'(z) = \frac{1}{F(z)F^*(z^{-1})} = \frac{1}{X(z)}.$$

### Peak distortion criterion

- The input to the filter  $C'(z)$  is the sequence  $\{y_k\}$ , i.e. the samples from the matched filter output.
- The impulse response of the combined filter can be achieved from the inverse  $z$ -transform according to

$$c'_k = \frac{1}{2\pi j} \oint C'(z) z^{k-1} dz = \frac{1}{2\pi j} \oint \frac{z^{k-1}}{X(z)} dz,$$

where the integration is performed on a closed contour that lies within the region of convergence of  $C'(z)$  which includes the unit circle.

- The performance of the infinite-tap equalizer that completely eliminates the ISI can be expressed in terms of the SNR at its output.
- For mathematical convenience, we normalize the received signal energy to unity which implies that  $q_0 = 1$  and  $E[|I_k|^2] = 1$ . Then, the SNR is simply the reciprocal of the noise variance  $\sigma_n^2$  at the output of the equalizer.
- We first observe that the noise sequence  $\{\nu_k\}$  at the input to the equivalent zero-forcing equalizer  $C'(z)$  has zero-mean and a power spectral density

$$\Phi_{\nu\nu}(\omega) = N_0 X\left(e^{j\omega T}\right), \quad |\omega| \leq \frac{\pi}{T},$$

where  $X(e^{j\omega T}) = X(z = e^{j\omega T})$ .

### Peak distortion criterion

- Since  $C'(z) = 1/X(z)$ , it follows that the noise sequence at the output of the equalizer has a power spectral density

$$\Phi_{nn}(\omega) = \Phi_{\nu\nu}(\omega) \frac{1}{|X(e^{j\omega T})|^2} = \frac{N_0}{X(e^{j\omega T})}, \quad |\omega| \leq \frac{\pi}{T},$$

since  $|X(e^{j\omega T})| = X(e^{j\omega T})$ .

- As a consequence, the noise variance at the equalizer output is

$$\sigma_n^2 = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \Phi_{nn}(\omega) d\omega = \frac{TN_0}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{d\omega}{X(e^{j\omega T})}$$

and the SNR results to

$$\gamma_\infty = \frac{1}{\sigma_n^2} = \left[ \frac{TN_0}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{d\omega}{X(e^{j\omega T})} \right]^{-1},$$

where the subscript  $\infty$  indicates that the equalizer has an infinite number of taps.

- What are the properties of  $\gamma_\infty$  in terms of the channel characteristics?

### Peak distortion criterion

- To answer that question, we have to establish a relationship between  $X(e^{j\omega T})$  and  $H(\omega)$ , the latter being the Fourier transform of the channel impulse response  $h(t)$ .
- Using Parseval's theorem, we obtain

$$x_k = \int_{\mathbb{R}} h^*(t)h(t+kT)dt = \frac{1}{2\pi} \int_{\mathbb{R}} |H(\omega)|^2 e^{j\omega kT} d\omega.$$

- Similarly to the proof of the Nyquist theorem, we can split the integral into

$$x_k = \frac{1}{2\pi} \int_{-\pi/T}^{\pi/T} \left[ \sum_{n \in \mathbb{Z}} \left| H\left(\omega + \frac{2\pi n}{T}\right) \right|^2 \right] e^{j\omega kT} d\omega.$$

Now, the Fourier transform of  $\{x_k\}$  is

$$X\left(e^{j\omega T}\right) = \sum_{k \in \mathbb{Z}} x_k e^{-j\omega kT}$$

Peak distortion criterion

and the inverse transform yields

$$x_k = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} X(e^{j\omega T}) e^{j\omega kT} d\omega.$$

- Thus, equating the both expressions for  $x_k$ , we obtain the sought for relationship between  $X(e^{j\omega T})$  and  $H(\omega)$  according to

$$X(e^{j\omega T}) = \frac{1}{T} \sum_{n \in \mathbb{Z}} \left| H\left(\omega + \frac{2\pi n}{T}\right) \right|^2 = \frac{1}{T} \sum_{n \in \mathbb{Z}} X\left(\omega + \frac{2\pi n}{T}\right), \quad |\omega| \leq \frac{\pi}{T}.$$

The middle term is the so-called **folded spectrum** of  $|H(\omega)|^2 = X(\omega)$ .

- Finally, we substitute for  $X(e^{j\omega T})$  in the SNR expression and obtain

$$\gamma_\infty = \frac{1}{\sigma_n^2} = \left[ \frac{T^2 N_0}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{d\omega}{\sum_{n \in \mathbb{Z}} \left| H\left(\omega + \frac{2\pi n}{T}\right) \right|^2} \right]^{-1}.$$

### Peak distortion criterion

- **Observations:**

- If the folded spectrum of  $|H(\omega)|^2$  possesses any zeros, the integrand becomes infinite and the SNR is zero. Thus, the performance of the equalizer will be poor whenever the folded spectral characteristic possesses nulls or takes on small values.
- This effect is called **noise enhancement** since the equalizer, in eliminating the ISI, enhances the additive noise.
- In the case of an **ideal channel** coupled with an appropriate signal design that results in no ISI will have a folded spectrum satisfying the condition

$$\sum_{n \in \mathbb{Z}} \left| H\left(\omega + \frac{2\pi n}{T}\right) \right|^2 = T, \quad |\omega| \leq \frac{\pi}{T}.$$

In this case, the SNR achieves its maximum value, namely  $\gamma_\infty = 1/N_0$ .

- The noise enhancement results from the fact that no information about the distortion resulting from the noise is taken into account in the equalizer design. Clearly, there are other ways to design the equalizer. The most important approach is the so-called **mean-square-error (MSE) equalizer** discussed next.

### Mean-square-error (MSE) criterion

- In the mean-square-error (MSE) equalizer, the tap weight coefficients  $\{c_j\}$  of the equalizer are adjusted to minimize the mean square value of error

$$\varepsilon_k = I_k - \hat{I}_k.$$

- When the information symbols  $\{I_k\}$  are complex-valued, the performance index for the MSE criterion, denoted by  $J$ , is defined as

$$J = E[|\varepsilon_k|^2] = E[|I_k - \hat{I}_k|^2].$$

- When the information symbols  $\{I_k\}$  are real-valued, the performance index for the MSE criterion, denoted by  $J$ , is defined as

$$J = E[\Re\{\varepsilon_k\}^2] = E[\Re\{I_k - \hat{I}_k\}^2].$$

- Hence, in either case,  $J$  is a quadratic function of the equalizer coefficients  $\{c_j\}$ .
- Below, we consider the case of complex-valued  $I_k$ .

### Mean-square-error (MSE) criterion

- We will assume that the **equalizer has infinitely many taps**. In this case, the symbol estimate reads

$$\hat{I}_k = \sum_{j \in \mathbb{Z}} c_j v_{k-j}.$$

- Upon inserting  $\hat{I}_k = \sum_{j \in \mathbb{Z}} c_j v_{k-j}$  into  $J$  and setting the derivatives w.r.t.  $c_j$ ,  $j \in \mathbb{Z}$ , equal to zero, we obtain the conditions

$$E[\varepsilon_k v_{k-\ell}^*] = 0, \quad \ell \in \mathbb{Z}.$$

This is the well-known **orthogonality principle of mean-square-estimation schemes**.

- Substitution for  $\varepsilon_k$  yields

$$E \left[ \left( I_k - \sum_{j \in \mathbb{Z}} c_j v_{k-j} \right) v_{k-\ell}^* \right] = 0$$

or, equivalently,

$$\sum_{j \in \mathbb{Z}} c_j E[v_{k-j} v_{k-\ell}^*] = E[I_k v_{k-\ell}^*], \quad \ell \in \mathbb{Z}.$$

### Mean-square-error (MSE) criterion

- Since the signal  $v_k$  is given by

$$v_k = \sum_{n=0}^L f_n I_{k-n} + \eta_k,$$

we obtain

$$E[v_{k-j} v_{k-\ell}^*] = \sum_{n=0}^L f_n^* f_{n+\ell-j} + N_0 \delta_{ij} = \begin{cases} x_{\ell-j} + N_0 \delta_{ij} & \text{for } |\ell - j| \leq L \\ 0 & \text{otherwise.} \end{cases}$$

and

$$E[I_k v_{k-\ell}^*] = \begin{cases} f_{-\ell}^* & \text{for } -L \leq \ell \leq 0 \\ 0 & \text{otherwise.} \end{cases}$$

- Inserting these expressions into

$$\sum_{j \in \mathbb{Z}} c_j E[v_{k-j} v_{k-\ell}^*] = E[I_k v_{k-\ell}^*], \quad \ell \in \mathbb{Z}.$$

and taking the  $z$ -transform, we obtain

$$C(z)(F(z)F^*(z^{-1}) + N_0) = F^*(z^{-1}).$$

### Mean-square-error (MSE) criterion

- As a result, the transfer function of the equalizer based on the MSE criterion is given by

$$C(z) = \frac{F^*(z^{-1})}{F(z)F^*(z^{-1}) + N_0}$$

and the equivalent equalizer including the noise-whitening filter has a transfer function

$$C'(z) = \frac{1}{F(z)F^*(z^{-1}) + N_0} = \frac{1}{X(f) + N_0}.$$

- If we compare the transfer functions derived from the peak distortion and MSE criteria, we observe that the only difference is the **noise spectral density factor  $N_0$**  in the denominator of  $C'(z)$  for the MSE criterion.
- Clearly, when  $N_0$  is small in comparison with the signal, we have practically the coefficients from the peak distortion criterion. In this case, noise enhancement does not degrade the performance too much in the case of a peak-distortion based design.
- In the **other extreme case, where  $N_0 \gg X(f)$** , the MSE equalizer just scales the signal, i.e. there is no equalization at all since the noise is the dominant source of distortion, not the ISI.
- We want to **characterize the performance of the MSE equalizer** in terms of the **minimum value of  $J$**  when the transfer function is given by  $C(z) = \frac{F^*(z^{-1})}{F(z)F^*(z^{-1}) + N_0}$ .

## Mean-square-error (MSE) criterion

- Since

$$J = E[|\varepsilon_k|^2] = E\left[\left|I_k - \hat{I}_k\right|^2\right] = E\left[\left(I_k - \hat{I}_k\right)\left(I_k^* - \hat{I}_k^*\right)\right] = E[\varepsilon_k I_k^*] - E\left[\varepsilon_k \hat{I}_k^*\right]$$

and  $E\left[\varepsilon_k \hat{I}_k^*\right] = 0$  by virtue of the orthogonality principle, it follows that

$$J_{\min} = E[\varepsilon_k I_k^*] = E\left[|I_k|^2\right] - \sum_{j \in \mathbb{Z}} c_j E[v_{k-j} I_k^*] = 1 - \sum_{j \in \mathbb{Z}} c_j f_{-j}.$$

- The term  $\sum_{j \in \mathbb{Z}} c_j f_{-j}$  is the convolution of the sequences  $\{c_k\}$  and  $\{f_k\}$  denoted as  $b_k = c_k \star f_k$  evaluated at time instant  $k = 0$ .
- In order to quantify the value  $b_0$ , we consider the  $z$ -transform of  $\{b_k\}$  given by

$$B(z) = C(z)F(z) = \frac{F(z)F^*(z^{-1})}{F(z)F^*(z^{-1}) + N_0} = \frac{X(z)}{X(z) + N_0}.$$

- Taking the inverse  $z$ -transform at  $k = 0$ , we obtain

$$b_0 = \frac{1}{2\pi j} \oint \frac{B(z)}{z} dz = \frac{1}{2\pi j} \oint \frac{X(z)z^{-1}}{X(z) + N_0} dz = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{X(e^{j\omega T})}{X(e^{j\omega T}) + N_0} d\omega.$$

### Mean-square-error (MSE) criterion

- Hence, we obtain

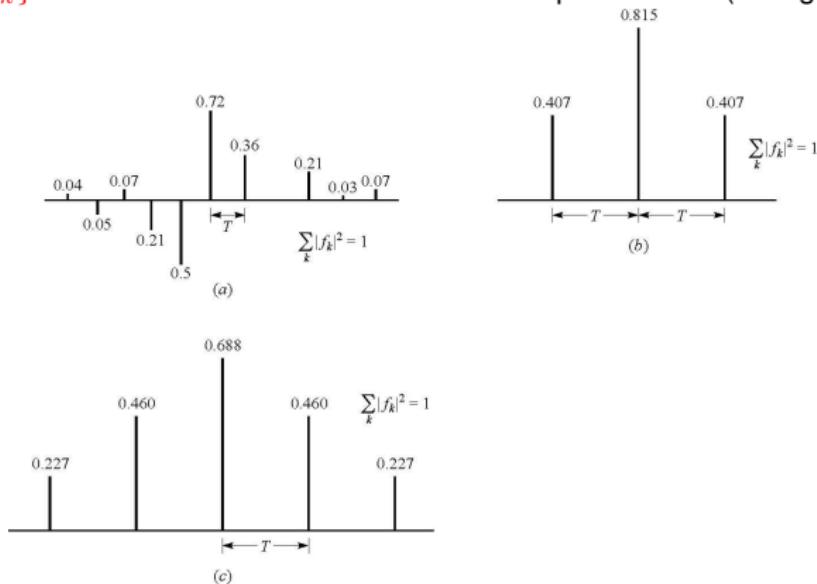
$$\begin{aligned}
 J_{\min} = 1 - b_0 &= 1 - \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{X(e^{j\omega T})}{X(e^{j\omega T}) + N_0} d\omega = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{N_0}{X(e^{j\omega T}) + N_0} d\omega \\
 &= \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \frac{N_0}{\frac{1}{T} \sum_{n \in \mathbb{Z}} |H(\omega + \frac{2\pi n}{T})|^2 + N_0} d\omega.
 \end{aligned}$$

- Observations:

- Since  $X(e^{j\omega T}) \geq 0$ , we have  $0 \leq J_{\min} \leq 1$ .
- In the absence of ISI,  $X(e^{j\omega T}) = 1$  and, hence,  $J_{\min} = \frac{N_0}{1+N_0}$ . For increasing  $N_0$ , the impact of the noise on  $J_{\min}$  increases up to the point (for  $N_0 \rightarrow \infty$ ) where  $J_{\min}$  reaches its maximum value. Clearly, in this case, no equalization takes place anymore since  $C'(z) = 1/N_0$ .
- In a practical system, the MSE equalizer will be implemented with a finite number of coefficients so that  $J_{\min}$  given above represents a lower bound for a given value of  $N_0$ .
- From the last integral expression, we see that if the folded spectrum of  $|H(\omega)|^2$  possesses any zeros we will have a large contribution from the corresponding intervals in the integral and  $J_{\min}$  tends to one.

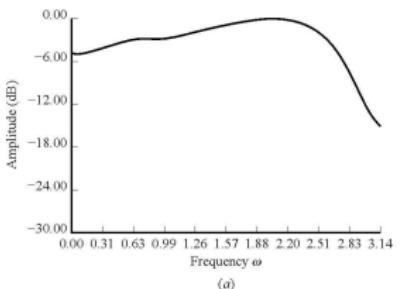
### Mean-square-error (MSE) criterion

The performance of the MSE equalizer will be discussed for **three examples of channels  $\{f_k\}$** . The second and third channels have spectral nulls (cf. Fig. 10.2-5).

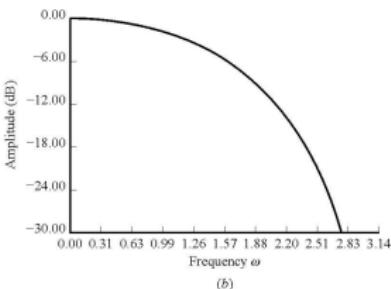


**Fig. 10.2-5** Three discrete-time channel characteristics

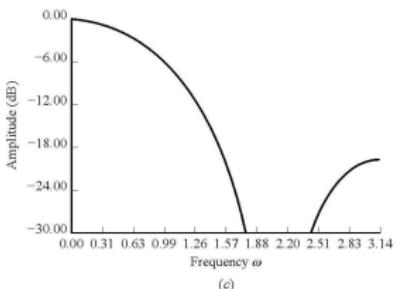
### Mean-square-error (MSE) criterion



(a)



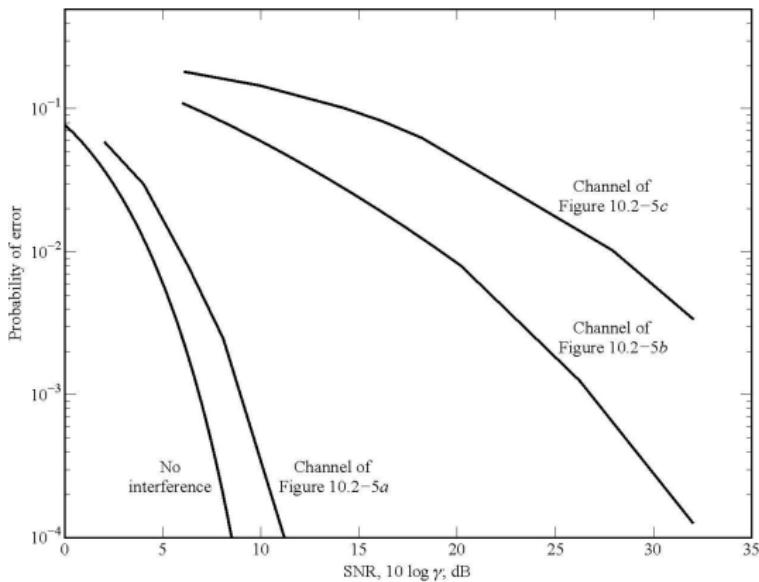
(b)



(c)

**Fig. 10.2-6** Amplitude spectra for the three channels shown in Fig. 10.2-5; note the spectral nulls of channels b) and c).

### Mean-square-error (MSE) criterion



**Fig. 10.2-4** Error rate performance of linear MSE equalizer with 31 taps and  $\gamma = \frac{1}{N_0} \sum_k |f_k|^2$  for the three channels shown in Fig. 10.2-5.

## Decision-Feedback Equalization

- The decision-feedback equalizer (DFE) consists of two parts, a feedforward filter and a feedback filter.
- Both filters have taps spaced at the symbol interval  $T$ .
- The feedforward filter with  $K_1$  taps and input sequence  $\{v_k\}$  is the same like in the linear equalizer. The feedback filter has as its input the sequence of decisions on previously detected symbols. Hence, functionally, the feedback filter is used to remove that part of the ISI caused by previously detected symbols from the present estimate.
- Based on the assumption of correct symbol estimates, the coefficients of the feedback filter results from the convolution of the equivalent channel filter taps  $\{f_k\}$  and the coefficients of the feedforward filter  $\{c_k\}$ . Clearly, if the assumption is correct and the length of the feedback filter  $K_2 \geq L$  is sufficient, the ISI is eliminated completely.

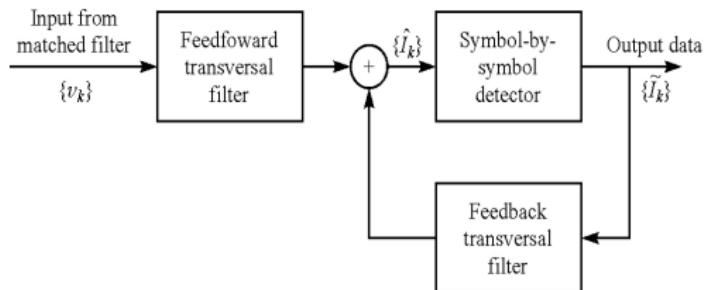
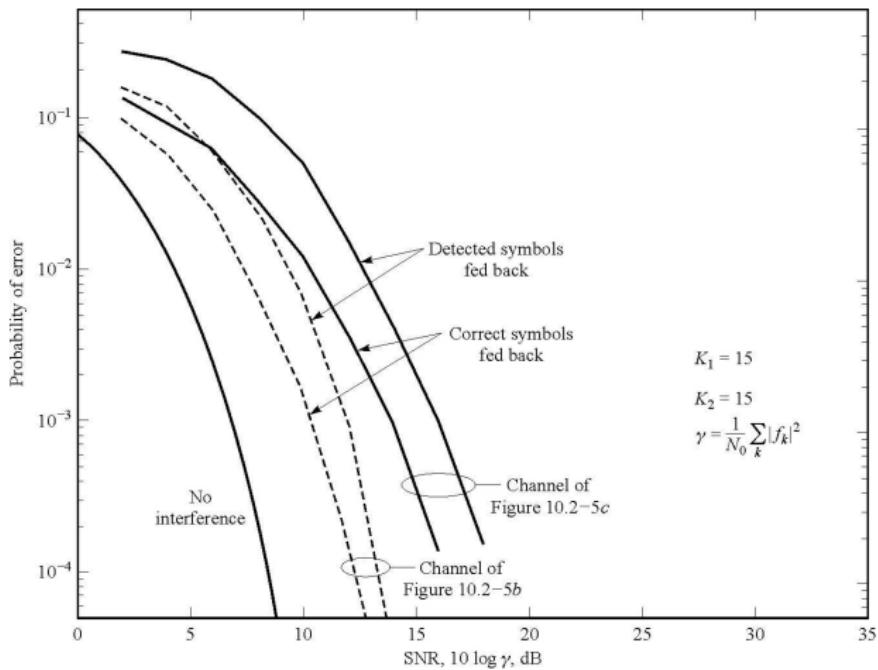
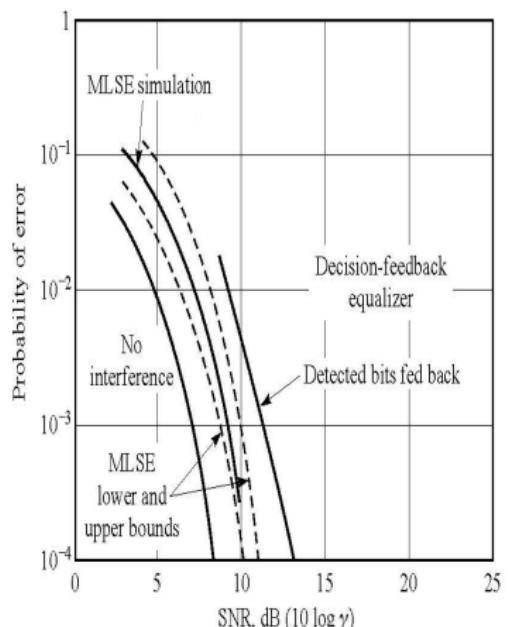


Fig. 10.3-1 Structure of decision-feedback equalizer.

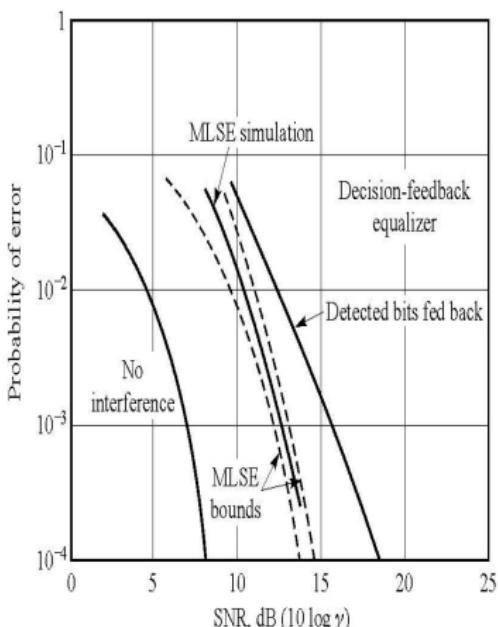
The performance of the DFE will be discussed again for the three channels in Fig. 10.2-5.



**Fig. 10.3-2 Performance of DFE with and without error propagation for the three channels in Fig. 10.2-5.**



(a) Channel of Figure 10.2-5b



(b) Channel of Figure 10.2-5c

**Fig. 10.3-2** Performance comparison between MLSE and DFE for (a) channel in Fig 10.2-5 b) and (b) channel in Fig 10.2-5 c).

# Equalization at the Transmitter

## Tomlinson-Harashima precoding

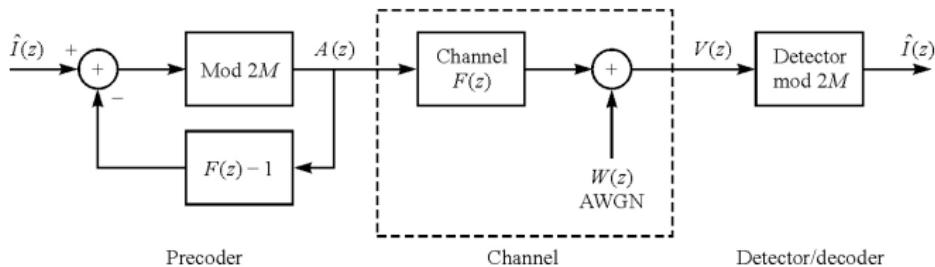
- The main problem of the equalizer at the receiver is the noise enhancement resulting from low channel gain of the overall channel. Since we always observe the sum of signal and noise, an amplification of the signal inevitably also increases the noise level.
- If the **channel response is known to the transmitter**, the **equalizer can be placed at the transmitter**. Thus, the **noise enhancement** being inherent to both linear equalizers and DFEs at the receiver **is avoided**. Clearly, in practical systems it might turn out difficult to place the entire equalizer at the transmitter in view of varying channel characteristics.
- A simple yet efficient way of implementing such an equalizer has been proposed by **Tomlinson and Harashima (Tomlinson-Harashima-precoding)** which will be sketched below for the case of  $M$ -ary PAM where the information symbols  $\{I_k\}$  are assumed to take the values  $\{\pm 1, \pm 3, \dots, \pm ((M-1)\}$ .
- The main idea in the design of the equalizer is based on the fact that the first tap  $f_0$  of the channel is equal to one. Thus, instead of transmitting  $I_k$ , we could, in principle, transmit  $a_k = I_k - \sum_{j=1}^L f_j a_{k-j}$ . Neglecting the noise, we would observe at the receiver the sequence  $\sum_{j=0}^L f_j a_{k-j} = a_k + \sum_{j=1}^L f_j a_{k-j} = I_k$ .

### Tomlinson-Harashima precoding

- The idea sketched above requires, unfortunately, a larger transmit power than the original sequence  $\{I_k\}$ . This increase in transmit power can be avoided by using a modulo- $2M$  operation at the receiver. The latter is implemented in the form of the transmitted symbols

$$a_k = I_k - \sum_{j=1}^L f_j a_{k-j} + 2M b_k,$$

where  $\{b_k\}$  denotes the appropriate integer that brings  $\{a_k\}$  to the desired range  $(M, M]$ .



**Fig. 10.3-5 Tomlinson-Harashima precoding with DFE at the transmitter.**

- Using the  $z$ -transform to describe the operation of the precoder, we obtain with  $f_0 = 1$

$$A(z) = I(z) - [F(z) - 1]A(z) + 2MB(z).$$

### Tomlinson-Harashima precoding

- Solving for  $A(z)$ , we obtain

$$A(z) = \frac{I(z) + 2MB(z)}{F(z)}.$$

- Since the channel response is  $F(z)$ , the spectrum of the received signal is given by

$$V(z) = A(z)F(z) + W(z) = I(z) + 2MB(z) + W(z),$$

where  $W(z)$  represents the AWGN term in the frequency domain.

- Thus, the received data sequence term  $I(z) + 2MB(z)$  at the input to the detector is **free of ISI** and  $I(z)$  can be recovered from  $V(z)$  by using a symbol-by-symbol detector that decodes the symbols modulo- $2M$ .
- Further methods for equalization in so-called single-carrier systems, where the complex baseband signal design is carried out in the time domain, include **reduced complexity ML detectors** and **turbo equalizers**. However, these topics are beyond the scope of this course.

**Problem 1**

In a binary PAM system, the input to the detector is  $y_m = a_m + n_m + i_m$ , where  $a_m = \pm 1$  is the desired signal,  $n_m$  is a zero-mean Gaussian random variable with variance  $\sigma_n^2$ , and  $i_m$  represents the ISI due to channel distortion. The ISI term is a random variable that takes the values  $-1/2$ ,  $0$ , and  $1/2$  with probabilities  $1/4$ ,  $1/2$  and  $1/4$ , respectively. Determine the average probability of error as a function of  $\sigma_n^2$ .

**Problem 2**

Consider the use of a square-root cosine signal pulse with a roll-off factor of unity for transmission of binary PAM over an ideal band-limited channel that passes the pulse without distortion. The transmitted signal is  $v(t) = \sum_{k \in \mathbb{Z}} I_k g_T(Pt - kT_b)$ , where the bit interval, however, is  $T_b = T/2$ . Thus, the symbol rate is double of that for no ISI.

- Determine the ISI values at the output of a matched filter demodulator.
- Sketch the trellis for the MLSE detector and label the states.

**Problem 3**

Binary PAM is used to transmit information over an unequalized linear filter channel. When  $a = 1$  is transmitted, the noise-free output of the demodulator is

$$x_m = \begin{cases} 0.3 & \text{for } m = 1 \\ 0.9 & \text{for } m = 0 \\ 0.3 & \text{for } m = -1 \\ 0 & \text{otherwise.} \end{cases}$$

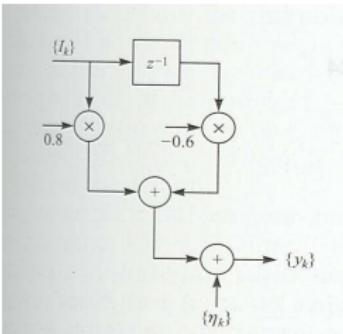
- a) Design a three-tap linear equalizer so that the output is

$$q_m = \begin{cases} 1 & \text{for } m = 0 \\ 0 & \text{for } m = \pm 1. \end{cases}$$

- b) Determine  $q_m$  for  $m = \pm 2$  and  $m = \pm 3$  by convolving the impulse response of the equalizer with the channel response.

**Problem 4**

A time-dispersive channel having an impulse response  $h(t)$  is used to transmit 4-PSK at a rate  $R = 1/T$  symbols/s. The equivalent discrete-time channel described by the transfer function  $F(z)$  is shown in the figure below. The sequence  $\{\eta_k\}$  is a zero-mean AWGN sequence with variance  $\sigma^2 = N_0$ .



**Fig. P10.23** Equivalent discrete-time channel.

- a) What is the sampled autocorrelation function sequence  $\{x_k\}$  defined by

$$x_k = \int h^*(t)h(t + kT)dt$$

for this channel?

- b) The minimum MSE performance of a linear equalizer and a DFE having an infinite number of taps depends on the folded spectrum of the channel

$$\frac{1}{T} \sum_{n \in \mathbb{Z}} \left| H\left(\omega + \frac{2\pi n}{T}\right) \right|^2,$$

where  $H(\omega)$  is the Fourier transform of  $h(t)$ . Determine the folded spectrum of the channel given above.

- c) Use your answer in b) to express the minimum MSE of a linear equalizer in terms of the folded spectrum of the channel. You may leave your answer in integral form, in case you do not find a good formulary :-)

# Table of Contents

- 1 Introduction
- 2 Carrier and Symbol Synchronization
  - Carrier Phase ( $\phi$ ) Estimation
  - Symbol Timing ( $\tau$ ) Estimation
  - Joint Estimation of  $\phi$  and  $\tau$
- 3 Signal Design for Band-Limited Channels
  - Characterization of Band-Limited Channels
  - Signal Design for No ISI-Nyquist Criterion
  - Signal Design with Controlled ISI
  - Probability of Error in Detection of PAM
  - Modulation Codes for Spectrum Shaping
- 4 Communication Through Band-Limited Linear Filter Channels
  - Optimum Receiver for AWGN Channels with ISI
  - Linear Equalization
  - Decision-Feedback Equalization
  - Equalization at the Transmitter
- 5 Multicarrier and Multichannel Systems
  - ISI in Single Carrier Systems
  - Design Criteria for Broadband System
  - Basic Principle of OFDM Signaling

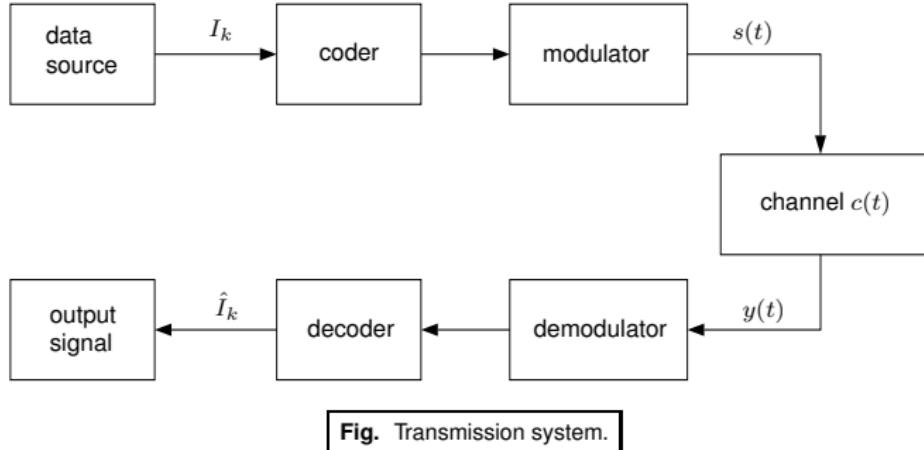
# Multicarrier and Multichannel Systems

- As can be concluded from the previous sections, signaling over band-limited channels in the form considered above has a number of implications for the overall transceiver:
  - First, the transmitting and receiving filters have to be chosen properly. Based on the assumption of an equalized channel, we can optimize the filter coefficients given that we know the form of the autocorrelation function  $x(t)$  both at the transmitter and the receiver.
  - The optimum detection scheme, the MLSE, has large complexity for increasing values of the channel dispersion  $L$  and the size of the modulation alphabet  $M$ .
  - Sub-optimum schemes like e.g. linear equalizers or a DFE show a reduced complexity compared to the MLSE, but still require the estimation of a number of coefficients whose number grows with  $L$  and whose performance depends critically on the existence of spectral nulls of the channel transfer function.
  - On top of the detectors and equalizers, powerful estimation schemes have to be implemented in order to meet the specifications suggested by the use of equalizers and other approaches for ISI based on the assumption of perfect channel state information.
- In a practical system implementation, the mutual dependency of the estimation and detection schemes makes the design of a bandwidth-efficient system a complicated task.
- Hence, the question arises: Can we find bandwidth-efficient schemes which can be used over dispersive channels with reasonable complexity? This question becomes more relevant in view of today's communication systems with increasing data rates and the requirement of high spectral efficiency at an affordable price of the transceivers used.

- The basic limitation of the previously investigated approaches for signaling, equalization and detection is the fact that the **signal design is carried out in the time domain** and that we always aim at reducing the ISI.
- Below we will consider a different approach which is called **orthogonal frequency-division multiplexing (OFDM)**. This modulation scheme is increasingly being used in today's broadband communication systems including
  - digital subscriber lines (DSL)
  - digital video broadcasting (DVB)
  - wireless local area networks (WLANs) according to the standard IEEE 802.11a
  - ultra-wideband (UWB) transmission systems according to the multiband OFDM alliance (MBOA) WiMedia standardization process,to name a few.
- We will first introduce OFDM in a general set-up including
  - dispersive channels and noise
  - the structure of the transceiver
  - implementation aspects based on the use of fast Fourier transforms (FFTs) and
  - a rough complexity comparison with single carrier transceivers.

# ISI in Single Carrier Systems

- Before we characterize the OFDM signaling method, we revisit the **problem of ISI in single-carrier systems** in order to contrast it later on with the **multi-carrier signaling**.



- We still assume the transmission channel to be linear and time-invariant. We obtain in the complex baseband description

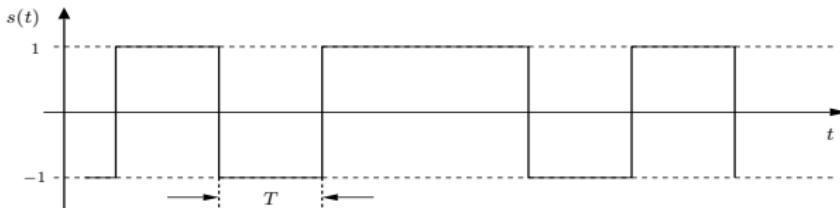
$$r_l(t) = \int_{\mathbb{R}} c(\tau) s(t - \tau) d\tau + z(t) = y(t) + z(t),$$

where we have for the AWGN process  $z(t)$  that  $E[z(t)] = E[z(t)z(t')] = 0$  and  $E[z(t)z^*(t')] = N_0\delta(t - t')$  for arbitrary  $t, t'$ . The noisy demodulator provides a mapping of the received signal to decision variables for subsequent symbol detection.

- The transmitted signal is given by

$$s(t) = \sum_k I_k g(t - kT),$$

where  $g(t)$  is again an arbitrary pulse and  $T = R_T^{-1}$  is the symbol duration with  $R_T$  denoting the symbol rate.



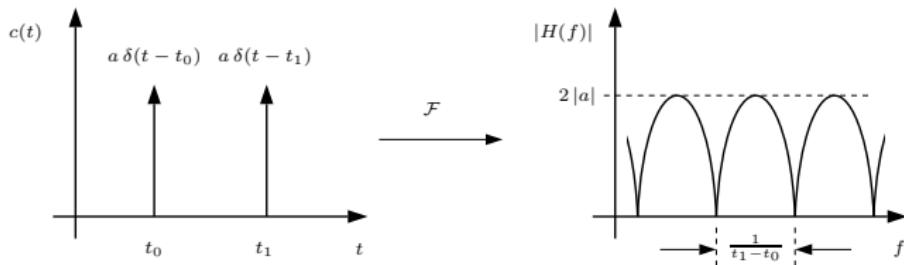
**Fig.** Transmission signal for BPSK signaling with rectangular pulses.

- The received signal is characterized as

$$r_l(t) = \sum_k I_k h_k(t) + z(t),$$

where  $h_k(t) = \int_{\mathbb{R}} c(\tau)g(t - kT - \tau)d\tau$  denotes the channel response to the pulse  $g(t - kT)$  in the  $k$ -th signalling interval.

- Consider the delay-dispersive channel below consisting of two paths of equal strength which are also called **multipath components**. Such a channel typically arises in cellular radio where the signal is reflected from scatterers and reflectors which might have a delay difference of up to  $10 \mu\text{s}$  in urban environments.



**Fig. Two-path channel:** channel impulse response and amplitude response in the frequency domain; obviously, delay dispersion leads to frequency selectivity of the amplitude response.

- As discussed in the previous section, the optimum receiver correlates the received signal  $r_l(t)$  with a replica of the distorted pulses  $h_k(t)$ , i.e. we have to calculate the **sufficient statistics**

$$u_k = \int_{\mathbb{R}} r_l(t) h_k^*(t) dt = \langle r_l, h_k \rangle.$$

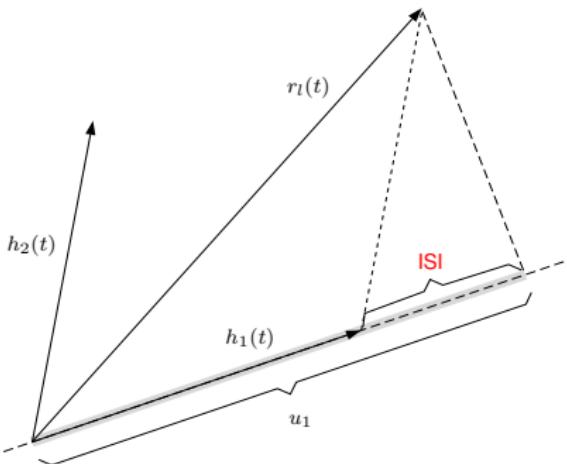
- The sufficient statistics can be fed directly to a symbol detector which may employ a hard decision  $\hat{I}_k = \text{sign}(u_k)$  for estimating  $I_k$  with  $\text{sign}(\cdot)$  denoting the sign function.

- Inserting the model of the received signal provides

$$u_k = I_k \|h_k\|^2 + \sum_{k' \neq k} I_{k'} \langle h_{k'}, h_k \rangle + \langle z, h_k \rangle,$$

where

- $\|h_k\|^2 = \langle h_k, h_k \rangle$  is the energy of the signal  $h_k(t)$
- $I_k \|h_k\|^2$  denotes the information signal to be detected
- $\sum_{k' \neq k} I_{k'} \langle h_{k'}, h_k \rangle$  represents the ISI
- $\langle z, h_k \rangle$  is the noise contribution within  $u_k$ .
- As before, we interpret  $r_l(t)$  as an element of a vector space with inner product  $\langle \cdot, \cdot \rangle$ .
- Below, we consider the example of a two-dimensional vector space with base vectors  $h_k(t)$  satisfying  $\int_{\mathbb{R}} |h_k(t)|^2 dt = 1$ ,  $k = 1, 2$ .
- Note that we only require  $h_1$  and  $h_2$  to be linearly independent, they do not have to be orthogonal.



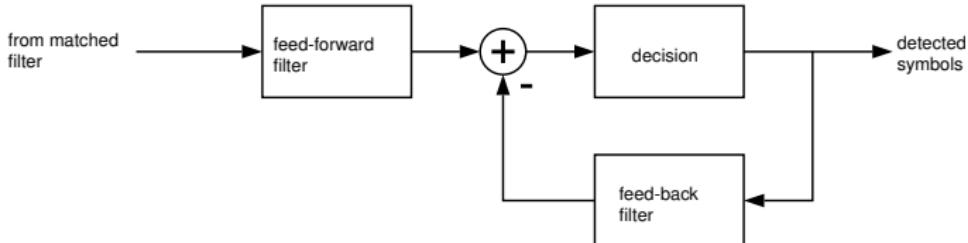
**Fig. Vector space representation of a matched filtering operation.**

- Note that  $u_k$  are the coordinates of  $r_l(t)$  w.r.t. the basis spanned by the base vectors  $\{h_k(t)\}$ .
- For a given transmitted pulse  $g(t)$ , the correlation properties of the vectors  $\{h_k(t)\}$  depend on the characteristics of the propagation channel  $h(t)$ . Furthermore, if we increase the data rate, i.e. reduce the value of  $T$ , the correlation and thus the ISI usually increase.

## Design Criteria for Broadband System

- What are the essential criteria for the design of broadband transmission systems?
- We assume a predefined bandwidth  $W$  of the system which has to be strictly adhered to. In neighbored frequency bands, there are often other systems operating there which are not to be interfered.
- Depending on the application (internet access, mobile radio, WLANs), there are bit rates  $R_{\text{Bit}}$  whose value may differ substantially, in particular for multimedia applications (voice, data, video).
- The regulation authorities prescribe the maximum average transmission power  $P_{\text{av}}$ . This limit might take into account technical and health related aspects.
- Manufacturers are interested in achieving a low complexity of the transmitter and the receiver. Although any manufacturer is free to design the transceivers, the standards defining corresponding wired and wireless communication interfaces have the largest impact on the overall system complexity.
- These criteria are partly contradicting. As an example, the specification of a low complexity air interface usually requires some limitation on the size of the modulation alphabet and lead thus to a usually low spectral efficiency. Furthermore, more and more applications are available over air interfaces, both for residential and mobile users. Apart from the requirement for a broadband transmission, a proper transceiver design must satisfy a low battery consumption in order to keep up the stand-by time of the handheld terminals. As a result, the design of a transmission scheme essentially satisfying all these requirements is a non-trivial task.

- In the context of single-carrier systems, a DFE seems to meet at least some of the requirements since it is a compromise between affordable complexity and achievable performance



**Fig. DFE:** The feedforward filter serves for eliminating the interference by future symbols (**precursor ISI**), the feed-back filter for eliminating the interference by past symbols (**postcursor ISI**).

- Clearly, there are some drawbacks in using the DFE. One is its non-linear behavior: In case where the SNR drops below a critical value, the decisions fed back via the feed-back filter are wrong and the error rate increases substantially.
- For a given pulse  $g(t)$ , the efficiency of the transmission may be very low. As an example, assume that the transmitter is aware of the spectral channel characteristics  $C(f)$ . If the pulse with spectrum  $G(f)$  has a large amplitude in the range of small channel gains, a very low efficiency results from using  $G(f)$ . On the other hand, an optimization of the pulse, in turn, increases the complexity of the scheme tremendously.

## Basic Principle of OFDM Signaling

- As shown on the previous slides, single-carrier schemes seem to be suitable for channels with a small dispersion and a sufficiently high value of the SNR. The former channel property guarantees that the complexity is not too high, while the latter guarantees a sufficient performance or, equivalently, a sufficiently high spectral efficiency.
- OFDM operates, as single-carrier systems, in channels with small dispersion and/or high SNR, but also provides a means for operating in broadband scenarios with high spectral efficiency and moderate implementation complexity.
- Coming back to the representation of ISI in vector spaces, we observe that in a single-carrier signaling scheme, the transmitted pulses  $g(t - kT)$  can be designed such that they are mutually orthogonal. While non-dispersive channels preserve the orthogonality at the receiver where the incoming signal is correlated with the distorted pulses  $h_k(t)$ , dispersive (frequency-selective) channels destroy the orthogonality and require equalizers to undo the distortion. The degree of distortion and thus the ISI depend on the channel under consideration.
- Therefore, a central question in the design of suitable signaling schemes is:

Are there orthogonal pulse waveforms  $g_k(t)$  which preserve their orthogonality at the receiver upon passing through a linear time-invariant channel with arbitrary transfer function?

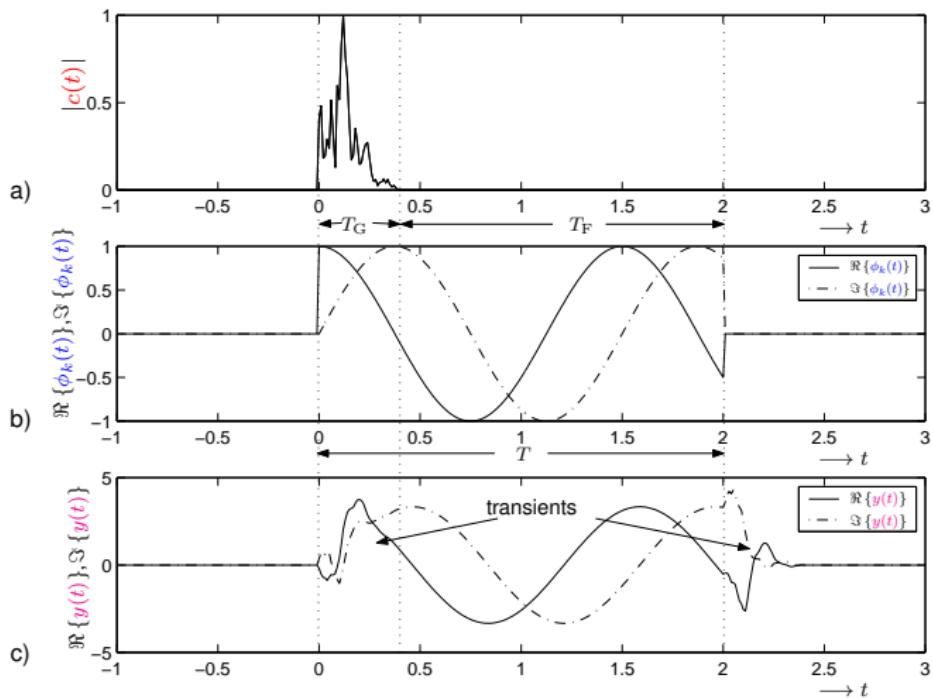
- A suitable candidate for finding such pulses is the concept of **orthogonal eigenfunctions of linear time-invariant (LTI) systems**.
- Such an eigenfunction has the special property that, in analogy with eigenvectors of matrices, if applied as an input signal to a linear system, it creates an output signal which is a scaled version of the input signal. The scaling parameter is called **eigenvalue of the system**.
- Therefore, we just have to **choose pulses from the set of eigenfunctions of LTI systems which have the additional property of being mutually orthogonal**. This is the key idea behind OFDM.
- It is well known that the eigenfunctions of LTI systems are given by

$$\psi_k(t) = \exp(j\omega_n t), \quad t \in \mathbb{R},$$

where  $\psi_k(t)$  and  $\psi_{k'}(t)$  in  $\mathbb{R}$  are mutually orthogonal for  $\omega_n \neq \omega_{n'}$ . Note that the aforementioned scaling (i.e. the eigenvalue) for the function  $\psi_k(t)$  is the transfer function  $H(f)$  of the system at  $f = \omega_n/2\pi$ .

- It is intuitive that the narrow-band signals  $\exp(j\omega_n t)$  with  $t \in \mathbb{R}$  represent a suitable class of base functions in dispersive channels. A **direct application** is, however, **not possible**, since  $\exp(j\omega_n t)$  with  $t \in \mathbb{R}$  **has infinite energy** and it takes infinite time to wait for the end of the signal  $\psi_k(t)$  before we can carry out the projection onto  $\psi_k(t)$  at the receiver.
- There is, however, a simple way to **construct orthogonal eigenfunctions based on**  $\psi_k(t) = \exp(j\omega_n t)$  **for**  $t \in \mathbb{R}$  as explained next.

## OFDM as an efficient multicarrier signaling scheme



**Fig.** Response of a dispersive channel to windowed eigenfunctions: a) absolute value of channel impulse response  $|c(t)|$ , b) real and imaginary parts of windowed eigenfunctions  $\phi_k(t) = \psi_k(t)\text{rect}_T(t)$ , c) real and imaginary parts of response  $y(t)$ .

## OFDM as an efficient multicarrier signaling scheme

- We assume that the **support of the channel impulse response  $c(t)$  is less or equal  $T_G$** .  
The eigenfunction  $\psi_k(t)$  is multiplied by a window of length  $T = T_G + T_F$ . If the system is excited by the function  $\phi_k(t) = \psi_k(t)\text{rect}_T(t)$  with

$$\text{rect}_T(t) = \begin{cases} 1 & \text{für } t \in [0, T) \\ 0 & \text{sonst} \end{cases},$$

we obtain the same output signal  $y(t)$  within  $\mathcal{T} = [T_G, T]$  as for an excitation signal  $\psi_k(t)$ .

- The intervals  $\mathcal{T}_E = [0, T_G]$  and  $\mathcal{T}_A = [T, T + T_G]$  contain the transients.
- Thus, if we transmit signals  $\phi_k(t)$  and **use the interval  $\mathcal{T}$  for the projection of the received signal onto the individual exponential functions**, the latter should be **mutually orthogonal within  $\mathcal{T}$** .
- In view of the observation length  $T_F$ , the **orthogonality holds** (different from the case of the functions  $\psi_k(t)$ ) no more for arbitrary values  $\omega_n \neq \omega_{n'}$ , but **only for  $\omega_n = \omega_0 + 2\pi k/T_F$  with  $k \in \mathbb{Z}$  and  $\omega_0 \in \mathbb{R}$** .
- Note that the **different signals overlap substantially in the frequency domain** so that this signalling is **far more bandwidth efficient than usual frequency multiplexing** in disjoint frequency bands.

## OFDM as an efficient multicarrier signaling scheme

- The OFDM signal in the complex baseband is given by

$$s(t) = \sqrt{S_0} \sum_{k=0}^{N_B-1} \sum_{n=0}^{N_{sc}-1} x_{k,n} g_{k,n}(t),$$

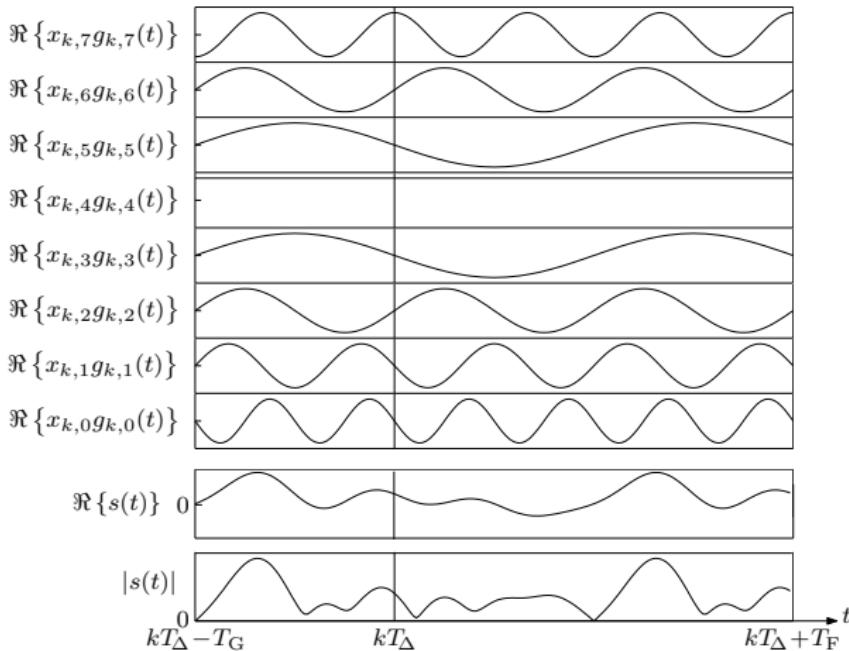
where  $S_0$  is a positive constant,  $N_{sc}$  is the number of subcarrier signals and  $N_B$  is the number of transmitted OFDM symbols.

- The variables  $x_{k,n}$  represent complex-valued data symbols being transmitted using the pulse  $g_{k,n}(t)$ . The pulse for transmission of the symbol  $x_{k,n}$  in the  $k$ -th time slot at subcarrier  $n$  is given by

$$g_{k,n}(t) = \begin{cases} T_F^{-\frac{1}{2}} e^{j2\pi(n-N_{sc}/2)F_\Delta(t-kT_\Delta)}, & t \in [kT_\Delta - T_G, kT_\Delta + T_F) \\ 0, & t \notin [kT_\Delta - T_G, kT_\Delta + T_F) \end{cases}.$$

- The signal  $s(t)$  is the superposition of  $N_{sc}$  linearly modulated **subcarrier signals** with a rate  $1/T_\Delta$ , frequency separation  $F_\Delta = 1/T_F$  and overlapping  $(\sin x/x)^2$ -shaped power spectral densities.
- The data symbols, typically elements of  $M$ -ary PSK or QAM signal sets, are modeled as independent random variables with  $E[x_{k,n}] = 0$  and  $E[|x_{k,n}|^2] = 1$  for all indices  $(k, n)$ .
- For sufficiently large  $N_{sc}$ , the signal  $s(t)$  has a flat power spectrum within the occupied bandwidth with power spectral density  $S_0$ .

## OFDM as an efficient multicarrier signaling scheme



**Fig.** Structure of an OFDM transmitted signal with  $N_{sc} = 8$  subcarriers.

## OFDM as an efficient multicarrier signaling scheme

- In the interval  $[kT_\Delta, kT_\Delta + T_F)$ , the harmonics differ by one period  $T_F$  such that the orthogonality results according to

$$\int_{kT_\Delta}^{kT_\Delta + T_F} x_{k,n} g_{k,n}(t) x_{k,m}^* g_{k,m}^*(t) dt = \delta_{n,m} |x_{k,n}|^2,$$

where  $\delta_{n,m}$  is the Kronecker delta.

- At the receiver, the signal  $r_l(t)$  is projected onto the windowed harmonics

$$h_{k,n}(t) = \begin{cases} T_F^{-\frac{1}{2}} e^{j2\pi(n-N_{sc}/2)F_\Delta(t-kT_\Delta)}, & t \in [kT_\Delta, kT_\Delta + T_F) \\ 0, & t \notin [kT_\Delta, kT_\Delta + T_F) \end{cases}$$

We obtain the decision variables

$$u_{k,n} = \int_{-\infty}^{\infty} r_l(t) h_{k,n}^*(t) dt, \quad k = 0, \dots, N_B - 1, \quad n = 0, \dots, N_{sc} - 1.$$

- Upon insertion of  $r_l(t)$ , we obtain

$$u_{k,n} = \int_{-\infty}^{\infty} \int_0^{T_G} h(\tau) s(t - \tau) d\tau h_{k,n}^*(t) dt + \int_{-\infty}^{\infty} v(t) h_{k,n}^*(t) dt.$$

## OFDM as an efficient multicarrier signaling scheme

- Inserting  $s(t)$ , we obtain

$$u_{k,n} = \sqrt{S_0} \sum_{\ell=0}^{N_B-1} \sum_{m=0}^{N_{sc}-1} x_{\ell,m} \int_0^{T_G} h(\tau) \int_{-\infty}^{\infty} g_{\ell,m}(t-\tau) h_{k,n}^*(t) dt d\tau + \int_{-\infty}^{\infty} v(t) h_{k,n}^*(t) dt.$$

- Since we have

$$\int_{-\infty}^{\infty} g_{\ell,m}(t-\tau) h_{k,n}^*(t) dt = \delta_{k,\ell} \delta_{n,m} e^{-j2\pi(m-N_{sc}/2)F_{\Delta}\tau}$$

for  $\tau \in [0, T_G]$ , we can express the coefficients  $u_{k,n}$  by

$$u_{k,n} = \sqrt{S_0} \alpha_n x_{k,n} + v_{k,n}$$

with the complex weights

$$\alpha_n = \int_{-\infty}^{\infty} h(\tau) e^{-j2\pi(n-N_{sc}/2)F_{\Delta}\tau} d\tau$$

and the noise term

$$v_{k,n} = \int_{-\infty}^{\infty} v(t) h_{k,n}^*(t) dt.$$

## OFDM as an efficient multicarrier signaling scheme

- The weights  $\alpha_n$  are the Fourier transforms of  $h(\tau)$  and thus the transfer function of the channel at  $f_n = \left(n - \frac{N_{sc}}{2}\right) F_\Delta$ , i.e.

$$\alpha_n = H \left( \left( n - \frac{N_{sc}}{2} \right) F_\Delta \right).$$

- The orthogonality of the subcarrier signals suggests a representation of the individual channels as components of vectors. The column vector  $\mathbf{u}_k = (u_{k,0}, \dots, u_{k,N_{sc}-1})^T$  is defined as

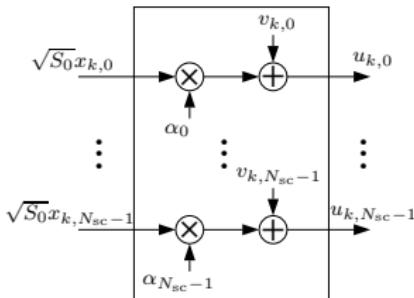
$$\mathbf{u}_k = \sqrt{S_0} \mathbf{C} \mathbf{x}_k + \mathbf{v}_k,$$

where  $\mathbf{x}_k = (x_{k,0}, \dots, x_{k,N_{sc}-1})^T$  und  $(\cdot)^T$  denotes transposition.

- The matrix  $\mathbf{C}$  is diagonal with elements  $\alpha_0, \dots, \alpha_{N_{sc}-1}$  and  $\mathbf{v}_k = (v_{k,0}, \dots, v_{k,N_{sc}-1})^T$  is a zero-mean multivariate complex Gaussian random vector with covariance matrix  $N_0 \mathbf{I}_{N_{sc}}$ .
- Since  $\mathbf{C}$  is diagonal, obviously the different subcarrier signals do not interfere with each other, i.e. they are orthogonal. Together with the fact that the individual data streams are multiplexed on different subcarrier signals, the whole signalling scheme is called **orthogonal frequency-division multiplexing (OFDM)**.

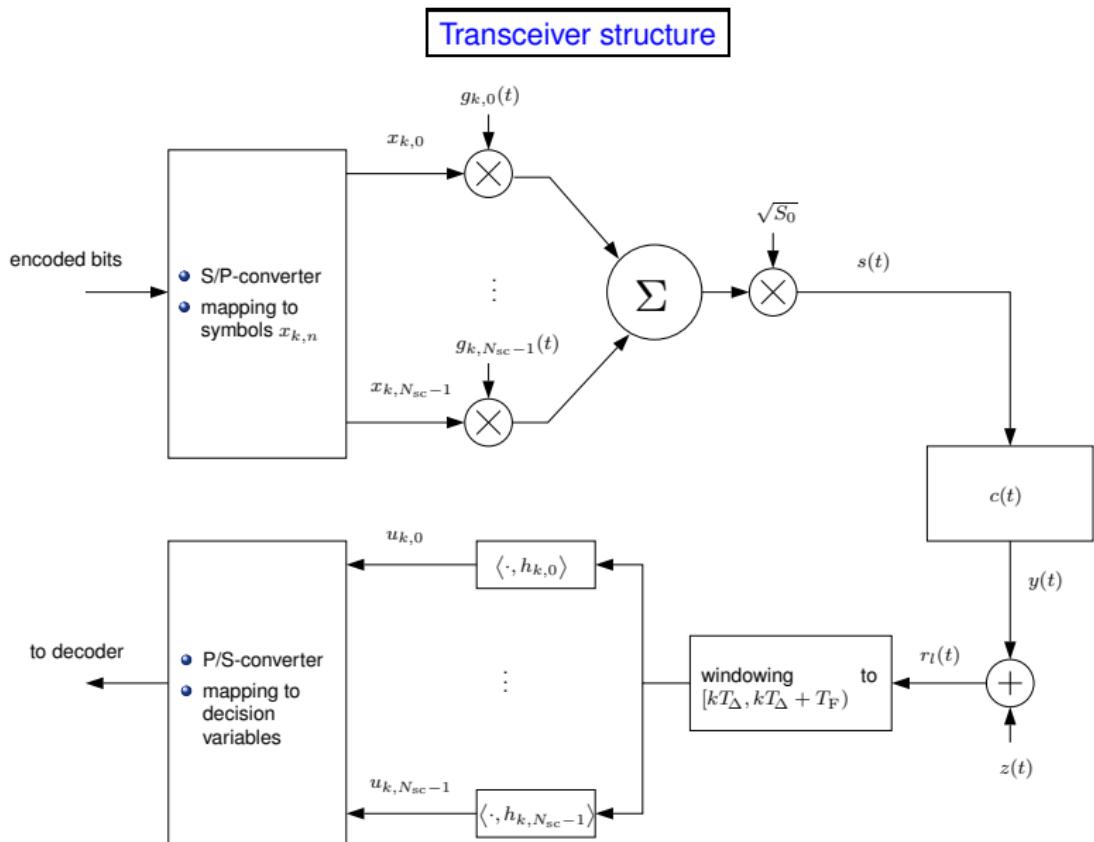
## OFDM as an efficient multicarrier signaling scheme

- The mapping of the symbol vector  $\sqrt{S_0}x_k$  to  $\mathbf{u}_k$  can be represented by the equivalent channel model below.



**Fig.** Equivalent representation of an OFDM system.

- The model contains  $N_{sc}$  parallel narrowband channels with AWGN and an attenuation defined by  $\alpha_n$ .
- The transmitted symbol  $x_{k,n}$  is multiplied by  $\alpha_n = |\alpha_n| \exp(j \arg \{\alpha_n\})$ . Therefore, an **equalization of the channel is quite simple**: We only have to multiply each subcarrier output signal  $u_{k,N_{sc}-1}$  by a complex weight. This gives rise to a so-called **one-tap equalizer**.
- In view of the AWGN noise vector, we can use standard detectors known from AWGN without any performance loss.

**Fig.** Structure of OFDM transceiver.

## Use of IFFT/FFT

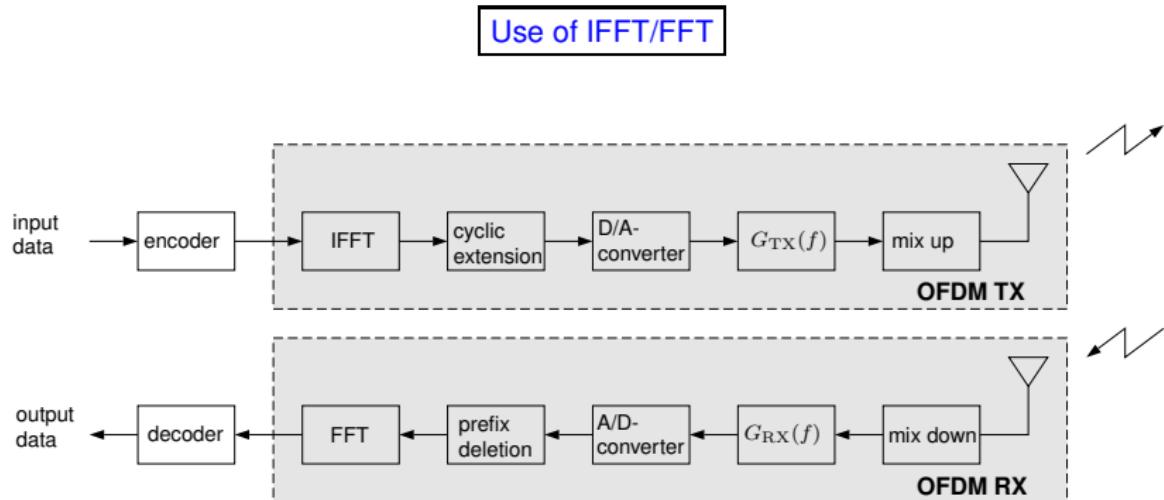
- The modulation and demodulation based on the exponential functions  $g_{k,n}(t)$  and  $h_{k,n}(t)$  are suited for an implementation using the discrete Fourier transform (DFT) and its inverse (IDFT).
- The DFT/IDFT can be implemented efficiently using a so-called **fast Fourier transform (FFT) and its inverse, the IFFT**. For the derivation of the corresponding schemes, we consider the signal  $s(t)$  in the  $k$ -th symbol interval during  $t \in [kT_\Delta, kT_\Delta + T_F)$ . If we take samples at time instants  $kT_\Delta + iT_F/N_{\text{FFT}}$  with  $i = 0, \dots, N_{\text{FFT}} - 1$  and consider a power of 2, namely  $N_{\text{FFT}} = 2^\nu$ ,  $\nu \in \mathbb{N}$ , we obtain the samples

$$s_{k,i} = \sqrt{\frac{S_0}{T_F}} \sum_{n=-N_{\text{sc}}/2}^{N_{\text{sc}}/2-1} x_{k,n+N_{\text{sc}}/2} e^{j2\pi ni/N_{\text{FFT}}}, \quad i = 0, \dots, N_{\text{FFT}} - 1.$$

- If we put the samples  $s_{k,i}$  into a vector  $\mathbf{s}_k = [s_{k,0}, \dots, s_{k,N_{\text{FFT}}-1}]$ , then  $\mathbf{s}_{k,i}$  is, except for a scaling factor  $\sqrt{S_0/T_F}$ , the IDFT of the vector  $[x_{k,N_{\text{sc}}/2}, \dots, x_{k,N_{\text{sc}}-1}, x_{k,0}, \dots, x_{k,N_{\text{sc}}/2-1}]$  for  $N_{\text{FFT}} = N_{\text{sc}}$ .
- If we further assume a guard interval length  $T_G = N_G T_F / N_{\text{FFT}}$ , the periodic extension of the signal  $s(t)$  to the interval  $t \in [kT_\Delta - T_G, kT_\Delta)$  corresponds to the **cyclic extension of the vector  $\mathbf{s}_k$**  by the  $N_G$  components, which are also known as the **cyclic prefix**.

### Use of IFFT/FFT

- Thus, the time-continuous signal  $s(t)$  within the interval  $[kT_\Delta - T_G, kT_\Delta + T_F)$  corresponds to the sequence  $s_{k,N_{\text{FFT}}-N_G}, \dots, s_{k,N_{\text{FFT}}-1}, s_{k,0}, \dots, s_{k,N_{\text{FFT}}-1}$ . For generating the time-continuous transmission signal, this sequence is fed to a digital/analog (D/A) converter, whose output signal is fed to a low-pass filter (LFP)  $G_{\text{TX}}(f)$  and transmitted subsequently.
- The signal processing at the receiver is completely analogous. After an antialiasing filter  $G_{\text{RX}}(f)$ , the receiver signal is fed to an analog/digital (A/D) converter. The windowing corresponds to **deleting the sequence samples corresponding to the cyclic prefix**. The remaining  $N_{\text{FFT}}$  samples are used in a DFT providing the vector of the sought for decision variables  $\mathbf{u}_k$ .
- In a practical OFDM system, we typically have  $N_{\text{FFT}} > N_{\text{sc}}$  for compensating the aliasing arising in the D/A conversion. Here, nulls are placed in the middle of the input vector of the IDFT characterizing the frequency range at half of the sampling frequency.
- Below the transceiver structure is shown for a typical WLAN application (IEEE 802.11a).



**Fig.** OFDM transceiver scheme for IEEE 802.11a: transmitter (OFDM TX) and receiver (OFDM RX).

### Use of IFFT/FFT

- Concerning the **complexity**, it turns out the complex additions are much easier to implement than complex multiplications. Therefore, we only count the number  $N_M$  of complex multiplications.
- The direct calculation of the  $k$ -th component of the vector  $s_k$  requires  $N_{\text{FFT}}$  complex multiplications. Hence, the calculation of the whole vector results in  $N_M = N_{\text{FFT}}^2$  complex multiplications.
- An implementation of a DFT in form of an **FFT requires**, however, **less than**  $N_M = \frac{N_{\text{FFT}}}{2} \log(N_{\text{FFT}})$  complex multiplications, which, for large sizes  $N_{\text{FFT}}$ , leads to relevant savings in the implementation of the transmitter and the receiver. Subsequently, the principle of the FFT is discussed.
- We consider the vector  $Z = [Z_0, \dots, Z_{N_{\text{FFT}}-1}]$  with

$$Z_k = \sum_{n=0}^{N_{\text{FFT}}-1} z_n W_{N_{\text{FFT}}}^{nk}, \quad k = 0, \dots, N_{\text{FFT}} - 1,$$

which defines the DFT of the vector  $z = [z_0, \dots, z_{N_{\text{FFT}}-1}]$ , where  $W_{N_{\text{FFT}}} = \exp(-j2\pi/N_{\text{FFT}})$  is a complex constant.

### Use of IFFT/FFT

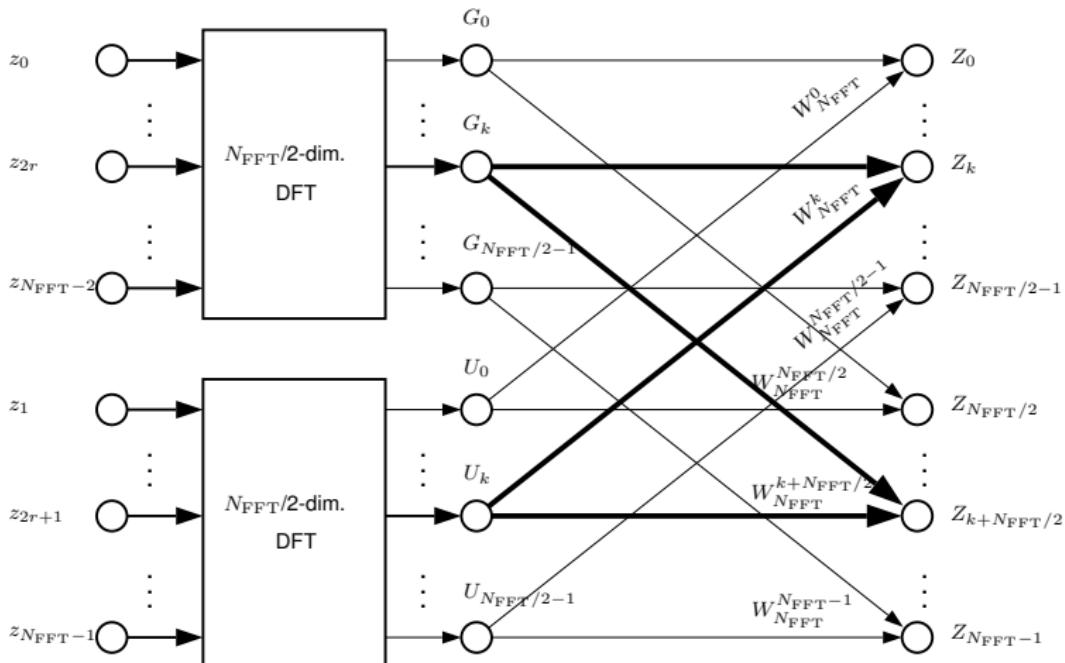
- If we consider in the sum expression the contributions of the terms  $z_n$  with even and odd indices ( $n = 2r$  and  $n = 2r + 1$ , resp.) separately and exploit  $W_{N_{\text{FFT}}}^2 = W_{N_{\text{FFT}}/2}$ , we obtain

$$\begin{aligned}
 Z_k &= \sum_{r=0}^{\frac{N_{\text{FFT}}}{2}-1} z_{2r} W_{N_{\text{FFT}}}^{2rk} + \sum_{r=0}^{\frac{N_{\text{FFT}}}{2}-1} z_{2r+1} W_{N_{\text{FFT}}}^{(2r+1)k} \\
 &= \sum_{r=0}^{\frac{N_{\text{FFT}}}{2}-1} z_{2r} W_{N_{\text{FFT}}/2}^{rk} + W_{N_{\text{FFT}}}^k \sum_{r=0}^{\frac{N_{\text{FFT}}}{2}-1} z_{2r+1} W_{N_{\text{FFT}}/2}^{rk} \\
 &= G_k + W_{N_{\text{FFT}}}^k U_k \\
 &= \begin{cases} G_k + W_{N_{\text{FFT}}}^k U_k & \text{for } k \in \left\{0, 1, \dots, \frac{N_{\text{FFT}}}{2} - 1\right\} \\ G_{k-\frac{N_{\text{FFT}}}{2}} + W_{N_{\text{FFT}}}^k U_{k-\frac{N_{\text{FFT}}}{2}} & \text{for } k \in \left\{\frac{N_{\text{FFT}}}{2}, \dots, N_{\text{FFT}} - 1\right\}. \end{cases}.
 \end{aligned}$$

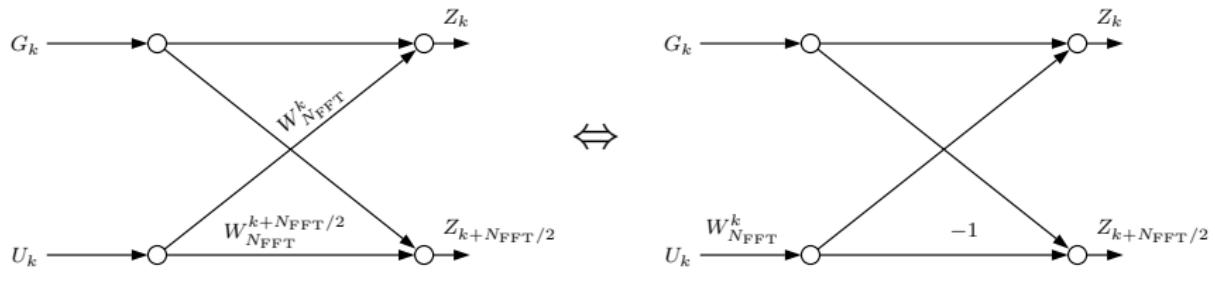
- We make two fundamental observations:

- Obviously,  $G_k$  and  $U_k$  represent the  $N_{\text{FFT}}/2$ -dimensional DFTs of the even and odd, resp., components of  $\mathbf{z}$ .
- Any value  $Z_k$  results from a weighted sum of  $G_k$  and  $U_k$ , where  $G_k$  and  $U_k$  are  $N_{\text{FFT}}/2$ -periodical.

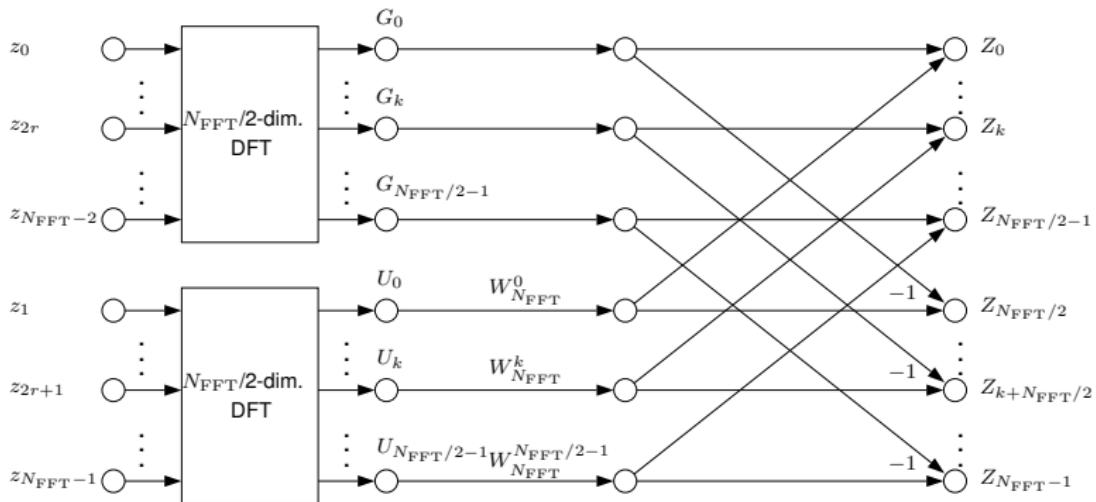
Use of IFFT/FFT



**Fig.** Signal flow graph for calculating  $Z$ .

**Use of IFFT/FFT****Fig. Equivalent butterflies.**

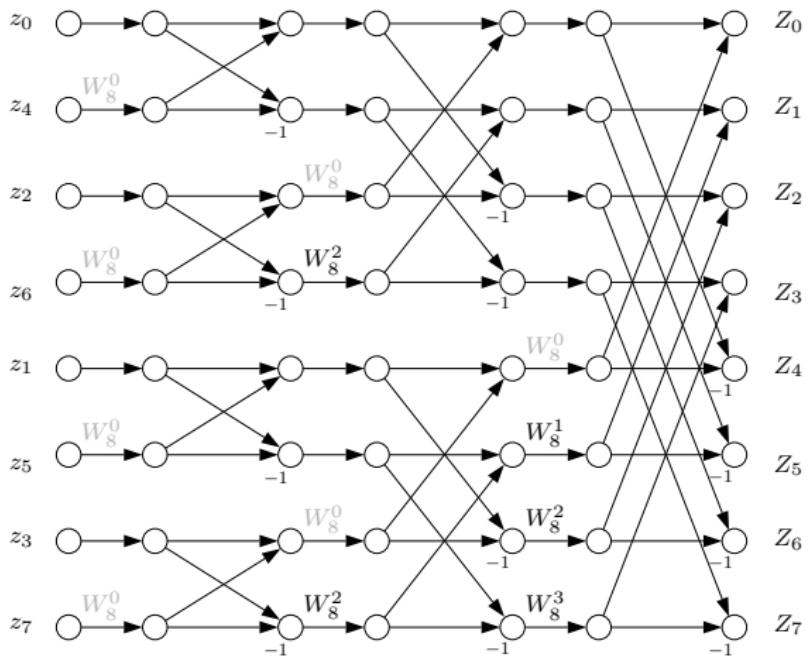
### Use of IFFT/FFT



**Fig.** Simplified signal flow graph for calculating  $\mathbf{Z}$ .

- Both  $N_{\text{FFT}}/2$ -dimensional DFTs can be replaced by four  $N_{\text{FFT}}/4$ -dimensional DFTs etc.
- In each stage, we have  $N_{\text{FFT}}/2$  butterflies with one complex multiplication each. Since for  $N_{\text{FFT}} = 2^\nu$  we have a total of  $\nu = \text{ld}(N_{\text{FFT}})$  stages, the overall number of required complex multiplications is  $N_M = \frac{N_{\text{FFT}}}{2} \text{ld}(N_{\text{FFT}})$ .

### Use of IFFT/FFT



**Fig.** FFT signal flow graph for calculating  $Z$  and  $N_{\text{FFT}} = 8$ ; the coefficients in grey  $W_8^0 = 1$  do not require any multiplication.

### Use of IFFT/FFT

- The comparison of complexities of two different transceivers is a complex task since the absolute effort depends on the implementation at hand.
- Here, we thus compare the **complexity increment** differences for increasing data rates between an OFDM system and a single-carrier system.
- We consider for a given data rate and bandwidth a single-carrier system  $S_{\text{Ein}}$  and an OFDM system  $S_{\text{OFDM}}$ , whose receivers have a complexity of  $C_{\text{Ein}}$  and  $C_{\text{OFDM}}$ , resp., complex multiplications per second. Complex additions and other operations like memory access etc. are neglected here for simplicity.
- Now we want to **double the data rate**, where the bandwidth for the system is doubled as well. In  $S_{\text{Ein}}$ , the sampling rate must be doubled. Since, however, now the number of equalizer coefficients have doubled as well, the complexity increases from  $C_{\text{Ein}}$  to  $4C_{\text{Ein}}$ : Compared to the original data rate, double as many samples must be processed per time unit with a filter of double length.
- For the OFDM system, we have to increase the number of subcarriers from  $N_{\text{sc}}$  to  $2N_{\text{sc}}$ . In view of the complexity formula of the FFT, namely  $\frac{N_{\text{sc}}}{2} \text{ld}(N_{\text{sc}})$ , we obtain a complexity increase by a factor

$$\frac{N_{\text{sc}} \text{ld}(2N_{\text{sc}})}{\frac{N_{\text{sc}}}{2} \text{ld}(N_{\text{sc}})} = 2 \left( 1 + \frac{1}{\text{ld}(N_{\text{sc}})} \right) \approx 2 \quad \text{for } N_{\text{sc}} \gg 1.$$

### Use of IFFT/FFT

- Since, at the same time, double as many coefficients for the channel estimation are required, the effort of which adds to the complexity of the FFT, the complexity increases for sufficiently high values  $N_{sc}$  of  $C_{OFDM}$  to about  $2C_{OFDM}$ . Put differently, the **effort per subcarrier is practically constant, independently of  $N_{sc}$** . This **linear complexity increase for increasing bandwidth** and given channel dispersion is one of the essential advantages of OFDM compared to single-carrier schemes.
- Finally, we want to demonstrate the choice of suitable parameters in an OFDM scheme for a **sample specification of a transmission system**. We assume the following specifications hold:

information bit rate	20 Mbit/s
channel dispersion	800 ns
bandwidth	15 MHz

- First, the ratio of the OFDM symbol duration and the channel dispersion should be sufficiently high in order to limit the relative energy in the guard interval  $T_G$ , which is not used by the receiver. If, for example, we choose  $T = 6T_G$ , we obtain a relative loss of energy of only

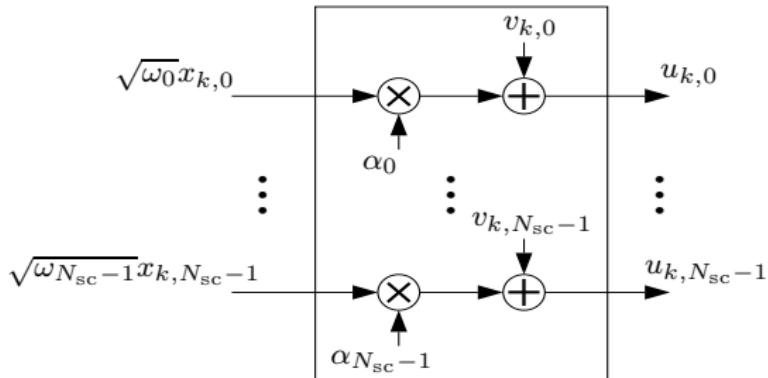
$$10 \log_{10} (T_F/T) \approx -0.8 \text{ dB},$$

an acceptable value in a practical system.

### Design example

- From the numbers above, we have an integration time of the FFT given by  $T_F = 5 T_G = 4 \mu\text{s}$  and the frequency separation of neighbored subcarriers is  $F_\Delta = 1/T_F = 250 \text{ kHz}$ . As a result, we have a maximum number of  $N_{sc,\max} = 15 \text{ MHz}/250 \text{ kHz} = 60$  subcarrier signals.
- 16-QAM signaling:** the required number of information bits to be transmitted per OFDM symbol is given by the product of the OFDM symbol duration and the required information bit rate, i.e.  $4.8 \mu\text{s} \times 20 \text{ Mbit/s} = 96 \text{ bit per OFDM symbol}$  have to be transmitted. We have a transmission of 1d(16) Bit = 4 Bit per subcarrier. Assuming a rate of 1/2 of the channel encoding scheme, we have a transmission of 2 bit per QAM symbol. Thus, in total, we have  $96 \text{ bit}/2 \text{ bit per subcarrier} = 48$  subcarrier, which is compatible with the aforementioned maximum number of  $N_{sc,\max} = 60$  subcarriers. An efficient FFT for this case uses a size of  $N_{FFT} = 64$ , where the **16 unused subcarriers** are placed at the band edges to avoid aliasing.
- QPSK signaling:** i.e. 2 bit per subcarrier, and a code rate of 3/4 leads to a transmission of  $2 \text{ bit} \times 3/4 = 1.5 \text{ bit per subcarrier}$ , leading to  $96 \text{ bit}/(1.5 \text{ bit per subcarrier}) = 64$  subcarriers. In this case, since  $N_{sc,\max} < 64$ , the occupied bandwidth is too large, so that we cannot use the QPSK scheme here.

## Channel capacity and coding in OFDM systems: water-filling



- A fundamental question arises in the design of an OFDM system: How should we choose the different subcarrier signals in order to maximize the throughput with simultaneously optimum performance?

⇒ Approach: maximize the number of correctly received bits (i.e. the capacity) for a

given bandwidth  $N_{sc}/T_F$  and average symbol energy  $\sum_{n=0}^{N_{sc}-1} \omega_n = N_{sc}S_0$  of the transmitted signal

$$s(t) = \sum_{k=0}^{N_B-1} \sum_{n=0}^{N_{sc}-1} \sqrt{\omega_n} x_{k,n} g_{k,n}(t)$$

## Channel capacity and coding in OFDM systems: water-filling

- Capacity for channel coefficients being known at the transmitter in bits per OFDM symbol:

$$C_{\text{OFDM}} = \max_{\omega_0 + \dots + \omega_{N_{\text{sc}}-1} \leq N_{\text{sc}} S_0} \sum_{n=0}^{N_{\text{sc}}-1} \text{ld} \left( 1 + \frac{\omega_n + |\alpha_n|^2}{N_0} \right)$$

with

$$\gamma_n = \omega_n + |\alpha_n|^2 / N_0 \quad \dots \text{ signal-to-noise ratio (SNR) at subcarrier } n$$

$$C_{\text{AWGN}} = \text{ld}(1 + \gamma) \quad \dots \text{ capacity of an AWGN channel with SNR } \gamma$$

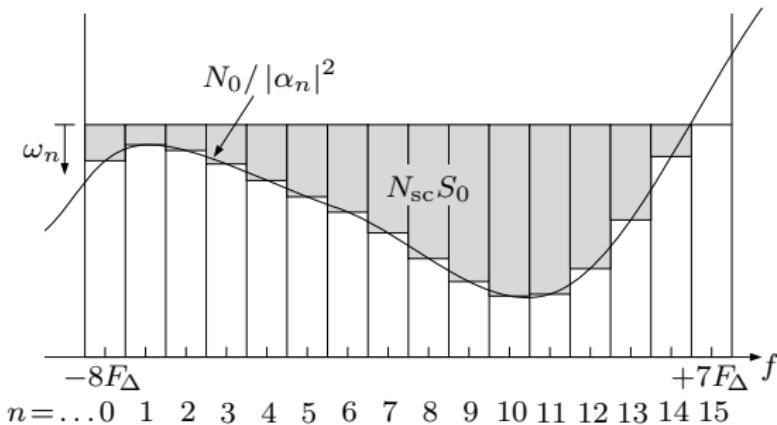
- Apply Langrange's method with  $\lambda \in \mathbb{R}$  and objective function (to be maximized)

$$J = (\omega_0 + \dots + \omega_{N_{\text{sc}}-1}, \lambda) = \sum_{n=0}^{N_{\text{sc}}-1} \text{ld} \left( 1 + \frac{\omega_n + |\alpha_n|^2}{N_0} \right) - \lambda \left( \sum_{n=0}^{N_{\text{sc}}-1} \omega_n - N_{\text{sc}} S_0 \right)$$

- We obtain

$$\omega_n = \max \left\{ (\lambda \text{ld} 2)^{-1} - N_0 / |\alpha_n|^2, 0 \right\}, \quad n = 0, \dots, N_{\text{sc}} - 1,$$

## Channel capacity and coding in OFDM systems: water-filling

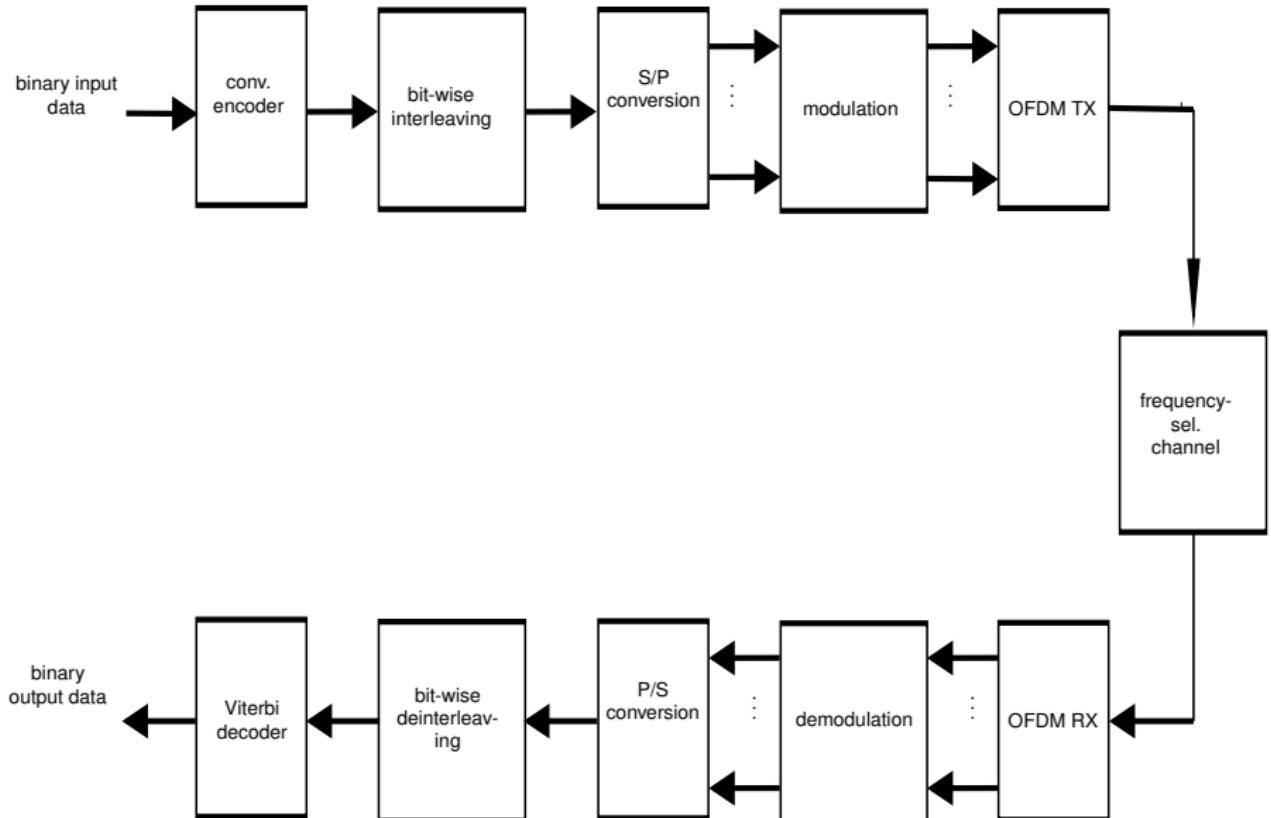


**Fig.** Optimum power allocation for  $N_{sc} = 16$  subcarriers in a frequency-selective channel

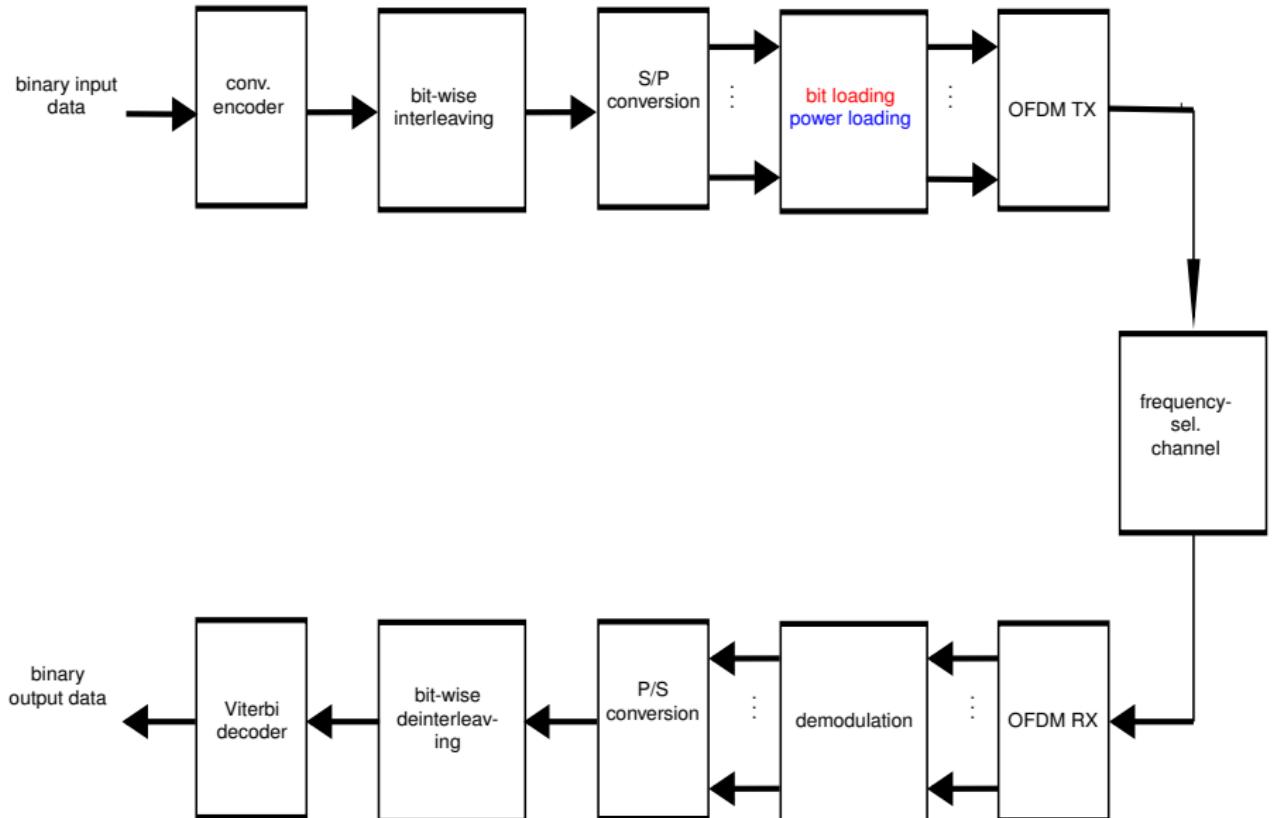
Explanation:

- high gain  $|\alpha_n| \Rightarrow$  high energy  $\omega_n$
- small gain  $|\alpha_n| \Rightarrow$  small energy  $\omega_n$
- for  $|\alpha_n|^2 \leq N_0 \lambda \ln 2 \Rightarrow$  no energy  $\omega_n$

### Example: encoding in a practical OFDM system



### Adaptation to channel characteristics



### Standardized systems using OFDM

#### Digital Video Broadcasting Terrestrial (DVB-T)

application	broadcasting of MPEG-2 encoded video sequences
carrier frequencies	> 470 MHz
bandwidth	8 MHz
max. bit rate	30 MBit/s
number of subcarriers	1705 (2k mode) or 6817 (8k mode)
subcarrier spacing	4.5 KHz (2k mode) or 1.1 KHz (8k mode)
features	four different values of guard periods $T_G$ possible

#### IEEE 802.11a

application	data transmission in wireless networks
carrier frequencies	5.2 ... 5.8 GHz
bandwidth	20 MHz
max. bit rate	54 MBit/s
number of subcarriers	52
subcarrier spacing	312.5 KHz
features	two different values of guard periods $T_G$ possible

## Standardized systems using OFDM

### Asymmetric Digital Subscriber Loop (ADSL)

application	wired data transmission
carrier frequencies	64 ... 1100 KHz
bandwidth	1 MHz
max. bit rate	8 MBit/s (backbone to end user) 640 KBit/s (end user to backbone)
number of subcarriers	256
subcarrier spacing	4 KHz
features	usual analog phone transmission in spectrum 0 ... 4 kHz (in parallel) possible due to filtering

## Interference phenomena in OFDM systems

interference = disturbances caused by signals

in transmitter

in channel

in receiver

### • Interference in transmitter

- for  $Nsc \gg 1 \Rightarrow$  large dynamic changes of signal; characterization by so-called **crest factor (CF)**

$$\kappa = \max \frac{|s(t)|}{s_{\text{eff}}}$$

- for large values of  $\kappa \Rightarrow$  complex analog-digital (A/D) converters and reduced amplifier efficiency
- for limiting of  $\kappa$  : use clipping (nonlinear amplitude limiter)  $\Rightarrow$  intercarrier interference (ICI) as well as interference outside the transmission band

## Interference phenomena in OFDM systems

### ● Interference in channel

- interference by signals of other systems in the same frequency band
- transmission over power lines: interference during switching processes in the power network (impulse noise) or impinging radio waves on transmission lines (narrowband interference)

### ● Interference in receiver

- imperfect frequency synchronization by phase noise  $\Rightarrow$  ICI
- imperfect timing synchronization  $\Rightarrow$  ISI

## Summary

### Basic elements of broadband transmission systems

- Broadband channels are delay dispersive ( $\Rightarrow$  ISI) or, put differently, frequency selective ( $\Rightarrow$  fading).
- The optimum receiver projects the received signal onto distorted signal pulses leading to corresponding high ISI values in the case of single-carrier systems.
- The required equalizer is complex and shows a nonlinear behavior for small signal to noise ratios.

### Properties of OFDM signaling

- OFDM consists of a parallel transmission of mutually orthogonal narrowband subcarrier signals which overlap in the frequency domain and show a high bandwidth efficiency.
- The guard interval preserves the orthogonality of the subcarrier signals for arbitrary channel characteristics with limited dispersion.
- Fast Fourier transforms can be employed to achieve a receiver complexity which increases only linearly with the data rate.
- The equalization can be implemented per subcarrier by a so-called **one-tap equalizer** which consists of a simple multiplication per subcarrier.

## Summary

### Channel capacity and coding in OFDM

- Maximizing the channel capacity (water-filling) assigns high energy to subcarriers with high channel gains; if the gain falls below a critical values, no signal power is assigned to the corresponding subcarriers.
- Practical coding schemes exploit the frequency diversity offered by the channel using bit-interleaved coded modulation.
- In case of CSI available at the transmitter, adaptive bit/power loading can be used to increase the performance of OFDM. Furthermore, adaptive coding (not discussed here) can be employed.

### Aspects in real-world systems

- OFDM is being used in current broadband transmission systems , e.g. DVB-T, IEEE 802.11a and ADSL. Future systems like WiMAX or WiMEDIA will also use OFDM (see corresponding internet sites for further information).
- Interference phenomena, i.e. disturbances caused by signals, arise in the transmitter, the channel and the receiver and require additional counter measures. The interference can be taken into account in generalized multicarrier transmission schemes which operate without a guard period.