MODULATION AND DETECTION

Sven Erik Nordholm
Western Australian Telecommunications Research Institute, Nedlands, Australia

Hans-Jürgen Zepernick
Blekinge Institute of Technology, Ronneby, Sweden

Keywords: Continuous wave modulation, Amplitude modulation, Frequency modulation, Phase modulation, Pulse amplitude modulation, Pulse modulation, Pulse amplitude modulation, Pulse duration modulation, Pulse position modulation, Pulse code modulation, Digital modulation, Phase shift keying, Frequency shift keying, Amplitude shift keying, Quadrature amplitude modulation, In-phase component, Quadrature component, Continuous-phase modulation, Multiplexing, Additive white Gaussian noise, Signal space, Signal vector, Orthonormal basis functions, Coherent detector, Matched filter, Maximum likelihood detector, Non-coherent detector

UNESCO – EOLSS SAMPLE CHAPTERS

Contents

Summary

This chapter presents principles of modulation techniques for modern communication systems along with the corresponding signal detection schemes and prominent receiver

structures. Modulation is an integral part of every modern communication systems and provides a means of matching an information-carrying signal to the transmission medium or transmission channel. Modulation techniques are commonly classified into continuous wave (CW) modulation and pulse (PU) modulation. While modulation takes place at the transmitting end of a communication system, an inverse operation called detection needs to be performed at the receiving end. In general, detection aims at recovering the information-bearing signal from the received signal as close as possible in terms of a specified optimization criterion. Digital communications, which usually stands for the digitized representation of originally analogue signals by discrete amplitude and discrete time, is common practice with today's communication systems. In digital communication systems, discrete information is conveyed between the communicating entities using signal formats from a finite set of waveforms. This allows for the design of very robust transmission systems with the deployed waveforms constructed in such a manner as to facilitate easy information recovery.

1. Introduction

Communication between people plays a vital role in today's society serving the fundamental human needs of exchanging information and ideas. Since the invention of telegraph and telephone, more sophisticated forms of telecommunication devices and services have been developed over the years such as wireless communication systems and the Internet. In addition, the tremendous advances in integrated circuit technology enabled the production of small and low cost communication devices, which nowadays consume only little energy and are affordable by almost anyone. The demand to communicate from anywhere in the world and at any time has lead to the development of mobile radio systems. Especially, with the deployment of the digital mobile radio systems in the early 1990's, global penetration of mobile phones has reached a penetration of close to 80-90 percent of the population in some regions. A similar rapid increase of users has been observed with the cable-based communication systems.

All these communication systems depend heavily on the engineering efforts that have been put into the design of the underlying transmission system and the associated physical layer functions. Apart from operations such as source encoding, channel encoding, equalization, and the respective decoding functions, modulation and detection are crucial components in the physical layer of modern communication systems. Modulation not only transforms information-bearing signals into a format that is suitable for transmission over a given channel but also takes care that the available transmission resources are efficiently used. Furthermore, modulation facilitates the multiplexing of large numbers of users onto a common transmission channel, which in view of the global user penetration is a challenging task. At the receiver, powerful detectors are required that can recover information-bearing signals out of severely corrupted received signal.

2. Principles of Modulation

The purpose of a communication system is to deliver a communication service from an information source to an information sink. Typically, source and sink are spatially separated and a transmission channel is needed to serve as link between the

UNESCO – EOLSS SAMPLE CHAPTERS

communicating entities. Depending on the particular transmission medium, the original messages as released by the source are mapped onto a different format or signal representation before the actual transmission such that it suits the channel characteristics. This operation is performed at the transmitting end of a communication system and is referred to as modulation. At the receiver, a reverse operation takes place, which is called detection or demodulation aiming at recovering the original message. In the sequel of this chapter, we will mainly concentrate on principles of prominent modulation techniques.

The different types of modulation techniques can be classified into Continuous Wave (CW) modulation and Pulse (PU) modulation. In the case of CW modulation, a sinusoidal wave is used as carrier of the messages. If the amplitude of the carrier wave is varied according to the amplitude of the message signal, amplitude modulation (AM) is obtained. Similarly, angle modulation refers to those CW schemes in which the angle of the carrier is changed according to the message signal. It is noted that angle modulation can be further differentiated into frequency modulation (FM) and phase modulation (PM) schemes indicating that either instantaneous frequency of the carrier or carrier phase is modified by the message signal.

In pulse modulation, on the other hand, a pulse train is used as the carrier wave and a suitable pulse parameter is modified in accordance with the message signal rather than modifying a sinusoidal carrier wave. Especially, amplitudes, durations, or positions of the pulses in a pulse train are suitable for representing the message signal. The corresponding pulse modulation techniques are referred to as Pulse Amplitude Modulation (PAM), Pulse Duration Modulation (PDM) also known as Pulse Width Modulation (PWM), and Pulse Position Modulation (PPM). These three PM techniques are classified as analog pulse modulation, which accounts for the fact that the respective pulse characteristics are modified with the sample value of the message signal. A second class of PM techniques is known as digital pulse modulation in which not only time but also amplitude of the message is represented and processed in discrete form. For example, Pulse Code Modulation (PCM) is a prominent digital pulse modulation technique. In a first processing step, PCM applies an analog-to-digital conversion to the analog message signal resulting in a PAM signal. Then, the obtained pulse amplitudes are quantized, thus, giving a finite set of possible amplitude values. In a final processing step, each quantized amplitude value is represented by a binary code word. This type of discrete time and encoded discrete amplitude modulation schemes are used in many digital communication systems.

Apart from matching messages to the characteristics of a given transmission medium, the concept of modulation also provides a framework for multiplexing messages of different users onto a common but shared channel. The three major types of multiplexing are as follows:

1. Frequency-Division Multiplexing (FDM). This technique is based on CW modulation and translates the message signals of the different users into different frequency bands where each band is identified by a distinct carrier frequency.
2. Time-Division Multiplexing (TDM). This technique exploits PM techniques and as such positions the message signals of the different users into non-overlapping time slots.

UNESCO – EOLSS SAMPLE CHAPTERS

3. Code-Division Multiplexing (CDM). This technique uses different spreading sequences or signatures to separate the messages of the various users. In that way, all users can use the total available frequency band simultaneously while the individual user can still be identified by its unique spreading sequence.

2.1. Continuous Wave Modulation

The continuous wave modulation schemes can be described mathematically by way of a modulated sinusoidal signal

$$s(t) = a(t)\cos\left[\omega_c t + \varphi(t)\right] \tag{1}$$

where $a(t)$ denotes instantaneous signal amplitude, $\omega_c t+\varphi(t)$ represents instantaneous phase, $\omega_c=2\pi f_c$ is referred to as angular velocity with f_c being the carrier frequency, and $\varphi(t)$ denotes a time-varying phase. In CW modulation the message signal $m(t)$ is used to modify either amplitude or the phase of a high frequency carrier signal.

2.1.1. Linear Modulation

Linear modulation techniques vary the amplitude of a high frequency carrier signal

$$c(t) = a_c \cos(\omega_c t) \tag{2}$$

linearly with the instantaneous amplitude of the baseband message signal $m(t)$. Accordingly, the modulated signal is given by

$$s(t) = a(t)c(t) \tag{3}$$

In particular, amplitude modulation schemes impose the following modification to the carrier amplitude:

$$a(t) = 1 + \mu m(t) \tag{4}$$

where μ is referred to as modulation index. This type of modulation causes the envelope of the carrier waveform to change proportionally with the instantaneous amplitude of the baseband message signal. In order to avoid an over-modulation of the carrier signal, which would result in undesirable carrier phase reversals, it is necessary to keep $a(t)$ always positive by requiring

$$\left|\mu m(t)\right| < 1 \tag{5}$$

If this condition is fulfilled, then an envelope detector may be deployed for demodulation purposes, which would simply track the slowly changing envelope of the high frequency carrier signal. A frequency domain representation of an AM signal can be obtained by Fourier transform and is given by

UNESCO – EOLSS SAMPLE CHAPTERS

$$S(f) = \frac{a_c}{2}\left[\delta(f - f_c) + \delta(f + f_c)\right] + \frac{a_c \mu}{2}\left[M(f - f_c) + M(f + f_c)\right] \tag{6}$$

where $\delta(f)$ denotes the Dirac delta function. It can be concluded from Eq. (6) that the spectrum $S(f)$ of the modulated signal $s(t)$ consists of two discrete spectral lines which are located at the positive and virtual negative carrier frequency. In addition, the baseband spectrum $M(f)$ of the modulating message signal $m(t)$ appears now shifted in frequency and is centered around the discrete spectral lines of the carrier signal. This constitutes a lower and an upper sideband of the spectrum $M(f \pm f_c)$ with respect to the carrier frequency $\pm f_c$ giving rise to call such an approach a Double Sideband (DSB) technique.

DSB techniques with Suppressed Carrier (DSB-SC) aim at conserving transmission power by not transmitting the carrier waveform $a(t)=a_c m(t)$ but only the information-bearing signal components within the sidebands. The benefit of DSB-SC being more power efficient than the conventional AM schemes comes at the expense of the required more involved coherent detector at the receiver (see Section 3).

Single Sideband (SSB) techniques address the second shortcoming of DSB techniques, which is their inefficiency in the use of bandwidth. Since the information associated with the message signal $m(t)$ is completely contained within a single sideband of the modulated signal, either upper sideband or lower sideband may be suppressed before transmission. In particular, an SSB signal containing only the lower sideband (+) or the upper sideband (–) is given in the time domain by

$$s(t) = \frac{a_c}{2} m(t) \cos(\omega_c t) \pm \frac{a_c}{2} \hat{m}(t) \sin(\omega_c t) \tag{7}$$

where the Hilbert transform $\hat{m}(t)$ of the band-limited message signal $m(t)$ is defined as

$$\hat{m}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{m(\tau)}{t - \tau} d\tau \tag{8}$$

The corresponding frequency domain representation of the Hilbert transform is given by

$$\hat{M}(f) = -j\, sgn(f) M(f) \tag{9}$$

where sgn(f) denotes the signum function and is defined as

$$sgn(f) = \begin{cases} 1, & f > 0 \\ 0, & f = 0\,1 \\ -1 & f < 0 \end{cases} \tag{10}$$

UNESCO – EOLSS SAMPLE CHAPTERS

2.1.2. Frequency and Phase Modulation

Instead of using the message signal to modify the amplitude of a sinusoidal carrier wave, the message signal may be utilized to vary the angle of the carrier while keeping its amplitude constant. Either varying the frequency or the phase of the carrier signal in accordance with the message signal may be applied. The numerous advantages of frequency and phase modulation techniques over the above described linear modulation schemes include:

1. Better discrimination of noise and interference.
2. Less stringent requirements on the linearity of the front-end power amplification enable the use of more efficient class C power amplifiers.
3. Trade-off possibilities between bandwidth and noise suppression.

Given the general formulation of the continuous wave modulation schemes as defined in Eq. (1), an angle-modulated signal can be described as

$$s(t) = a_c \cos[\omega_c t + \varphi(t)] \tag{11}$$

Then, the instantaneous angle $\theta_c(t)$ of the angle-modulated signal $s(t)$ is given by

$$\theta_c(t) = \omega_c t + \varphi(t) \tag{12}$$

with the instantaneous frequency $f_c(t)$ being the derivative of the instantaneous angle $\theta_c(t)$, i.e.

$$f_c(t) = \frac{1}{2\pi}\frac{d\theta_c(t)}{dt} \tag{13}$$

Among the many means of varying the instantaneous angle of the carrier wave in accordance with the message signal, we will consider in the sequel the two commonly used methods of phase modulation and frequency modulation. In the case of a PM scheme, the message signal $m(t)$ linearly varies the instantaneous angle $\theta_c(t)$ of the carrier signal. In view of Eq. (12), we have

$$\theta_c(t) = \omega_c t + k_p m(t) \tag{14}$$

where k_p denotes the phase sensitivity of the modulator. Similarly and in view of Eq. (13), an FM scheme varies the instantaneous frequency $f_c(t)$ of the carrier signal according to

$$f_c(t) = f_c t + k_f m(t) \tag{15}$$

where k_f denotes the frequency sensitivity of the modulator. The instantaneous angle $\theta_c(t)$ is then obtained by simply integrating the instantaneous frequency $f_c(t)$, that is

UNESCO – EOLSS SAMPLE CHAPTERS

$$\theta_{\mathrm{c}}(t) = 2\pi f_{\mathrm{c}} t + 2\pi k_{\mathrm{f}} \int_{0}^{t} m(\tau) d\tau \tag{16}$$

Therefore, phase-modulated and frequency-modulated signal $s(t)$, respectively, can be described in the time domain as

$$s(t) = \begin{cases} a_{\mathrm{c}} \cos\left[2\pi f_{\mathrm{c}} t + k_p m(t)\right] & \text{for PM} \\ a_{\mathrm{c}} \cos\left[2\pi f_{\mathrm{c}} t + 2\pi k_f \int_{0}^{t} m(\tau) d\tau\right] & \text{for FM} \end{cases} \tag{17}$$

Clearly, Eq. (17) indicates that PM and FM signals are very closely related. For example, an FM signal may be considered as a PM signal in which the message signal was replaced by the integral over the message signal:

$$m(t) \rightarrow \int_{0}^{t} m(\tau) d\tau \tag{18}$$

As a consequence, the characteristics and properties of PM signals can be deduced from FM signals. On the other hand, it is noted that the properties of FM signals, such as their spectrum and required transmission bandwidth, are more complicated to derive analytically compared to the case of linear modulation. This is due to the fact that an FM signal is a nonlinear function of the message signal. An estimation of the bandwidth that is occupied by an FM signal when the message signal is sinusoidal is given by the so-called Carson's rule (see *Communication Systems*):

$$B_T \approx 2k_f a_c + 2 f_m \tag{19}$$

where f_m denotes the frequency of the sinusoidal message signal $m(t)$.

2.2. Pulse Modulation

A fundamental operation associated with all pulse modulation techniques is the sampling process. To be more specific, sampling converts an analog signal into a corresponding sequence of samples, which in view of applications in practical systems are spaced uniformly in time. Given an arbitrary signal $m(t)$ of finite energy, then the instantaneously sampled version of the signal can be formulated mathematically as

$$s(t) = \sum_{n=-\infty}^{\infty} m(nT_{\mathrm{S}})\delta(t - nT_{\mathrm{S}}) \tag{20}$$

where T_{S} is the sampling period, $m(nT_{\mathrm{S}})$ is the sample value of $m(t)$, and $\delta(t)$ denotes the Dirac delta function. This operation is called ideal sampling. Let the analog signal $m(t)$ be strictly band-limited to a bandwidth B. Then, the instantaneously sampled version $s(t)$ uniquely defines $m(t)$ if the sampling rate $f_{\mathrm{S}}=1/T_{\mathrm{S}}$ is chosen as

$$f_{\mathrm{S}} \geq 2B \tag{21}$$

UNESCO – EOLSS SAMPLE CHAPTERS

where $2B$ is called the Nyquist rate and $1/2B$ is called the Nyquist interval. As long as the sampling rate is equal or greater than the Nyquist rate, the original signal $m(t)$ can be completely reconstructed from its instantaneously sampled version $s(t)$ (Shannon-Nyquist sampling Theorem).

-
-
-

TO ACCESS ALL THE **25 PAGES** OF THIS CHAPTER,
Visit: http://www.eolss.net/Eolss-sampleAllChapter.aspx

Bibliography

Anderson J. B., Aulin T., and Sundberg C.-E. (1986) *Digital Phase Modulation*, New York: Plenum Press. [This book offers a comprehensive coverage of digital phase modulation techniques on an advanced level]

Haykin S. (2001). *Communication Systems*, New York: John Wiley & Sons. [This book presents an overview of communication systems on an accessible level]

Papoulis A. (1991). *Probability, Random Variables, and Stochastic Processes*, New York: McGraw-Hill. [This book presents a background on stochastic processes]

Proakis J. G. (2001). *Digital Communications*, New York: McGraw-Hill. [This book gives a detailed coverage of the theoretical aspects of digital communications on an advanced level]

Sklar B.(1988) *Digital Communications: Fundamentals and Applications* , Englewood Cliffs: Prentice Hall. [This book presents an overview of digital communication systems on an accessible level]

Biographical Sketches

Sven Erik Nordholm received his MSc. EE, Lic. Eng. and PhD degree from the University of Lund, Sweden, in 1983, 1989 and 1992. From 1983 to 1986, he was a Research and Development Engineer with Gambro AB a biotechnical company.
From 1990-1999 he hold various positions at Blekinge Institute of Technology starting from lecturer to full Professor. Currently, Dr. Nordholm holds the positions of a Professor in Telecommunications at Curtin University of Technology, Director of the Signal Processing Laboratory of the Western Australian Telecommunications Research Institute (WATRI), and Research Executive of the Australian Telecommunications Cooperative Research Centre (ATcrc), Perth, Australia. His research interests include speech enhancement, array signal processing, multi-rate processing and signal processing for communication systems.

Hans-Jürgen Zepernick received his MSc degree from the University of Siegen, Germany, in 1987, and his Ph.D. degree from the University of Hagen, Germany, in 1994, both in electrical engineering. Currently, he holds the position of Professor in Radio Communications at the Department of Signal Processing of the Blekinge Institute of Technology, Ronneby, Sweden. His research interests include radio channel characterization and modeling, coding and modulation, equalization, spread-spectrum systems, wireless multimedia communications, and future generation wireless systems.

UNESCO – EOLSS SAMPLE CHAPTERS