EE6604

Personal & Mobile Communications

Weeks 9 & 10

Modulation Techniques

Modulation for Wireless Systems

- To achieve high spectral efficiency, power- and bandwidth-efficient modulation techniques are used that have the following properties:
 - Compact Power Density Spectrum: To minimize the effect of adjacent channel interference, the power radiated into the adjacent band is often limited to be 60 to 80 dB below that in the desired band. This requires modulation techniques having a power spectrum characterized by a narrow main lobe and fast roll-off of side-lobes.
 - Robust Communication: Reliable communication must be achieved in the presence of delay and Doppler spread, adjacent and co-channel interference, and thermal noise. Modulation schemes that promote good power and bandwidth efficiency in the presence of channel impairments are desirable.
 - Envelope Properties: Power amplifier non-linearities will degrade the performance of modulation schemes that transmit information in the amplitude of the carrier and/or have a non-constant envelope. To prevent the regrowth of spectral side-lobes with non-linear amplification, modulation schemes having a relatively constant envelope are desirable. However, modulation schemes having a higher peak-to-average power ratio (PAPR) are generally more spectrally efficient. Linearized power amplifiers may be required with high PAPR waveforms.

Three Representations of Bandpass Signals

- Let s(t) be a bandpass waveform. This is a waveform whose frequency content occupies a narrow range of frequency about a carrier frequency f_c . Let $\tilde{s}(t)$ be its complex envelope.
- The complex envelope representation of a bandpass signal is

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

- Writing $\tilde{s}(t) = \tilde{s}_I(t) + j\tilde{s}_Q(t)$, gives the quadrature respresentation of a bandpass signal $s(t) = \tilde{s}_I(t)\cos 2\pi f_c t - \tilde{s}_Q(t)\sin 2\pi f_c t$
- Finally, the amplitude-phase representation of a bandpass signal is

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

where

$$a(t) = \sqrt{\tilde{s}_I^2(t) + \tilde{s}_Q^2(t)}$$

$$\phi(t) = \operatorname{Tan}^{-1} \left[\frac{\tilde{s}_Q(t)}{\tilde{s}_I(t)} \right]$$

Generalized Representation of Modulated Waveforms

• The complex envelope of a modulated waveform can be expressed in the generalized format

$$\tilde{s}(t) = A \sum_{n} b(t - nT, \mathbf{x}_n)$$

where $b(t, \mathbf{x}_n)$ is a generalized modulation symbol, and A is an amplitude scalar.

• In some cases,

$$\mathbf{x}_n = (x_n, x_{n-1}, \ldots, x_{n-K})$$

where one data symbol is transmitted every T seconds, and K is the modulator memory order.

- For example, with binary phase shift keying (BPSK) and a rectangular amplitude pulse shaping function of duration T:

$$b(t, \mathbf{x}_n) = x_n u_T(t), \quad x_n \in \{-1, +1\}$$

where $u_T(t) = u(t) - u(t - T)$ is a unit amplitude rectangular amplitude shaping pulse of duration T, and K = 0 in this case since the BPSK modulator is memoryless.

• In other cases,

$$\mathbf{x}_n = (x_{n_0}, x_{n_1}, \ldots, x_{n_{K-1}})$$

where K is the number of complex data symbols transmitted every T seconds.

Quadrature Amplitude Modulation (QAM)

• With QAM, the bandpass signal is

$$s(t) = A \sum_{n} \left\{ x_{n}^{I} h_{a}(t - nT) \cos 2\pi f_{c} t - x_{n}^{Q} h_{a}(t - nT) \sin 2\pi f_{c} t \right\}$$

where

$$x_n^I, x_n^Q \in \{\pm 1, \pm 3, \dots, \pm (\hat{M} - 1)\}$$

 $h_a(t) = \text{amplitude shaping pulse}$

- the index of summation "n" in the above expression is a time index.

• Every T seconds, one of the following M waveforms are transmitted

$$s_m(t) = A x_m^I h_a(t) \cos 2\pi f_c t - A x_m^Q h_a(t) \sin 2\pi f_c t, \quad m = 1, \dots, M$$

- The data sequence $\{x_m^I\}$ modulates the amplitude of the cosine component of the RF carrier, whereas $\{x_m^Q\}$ modulates the amplitude of the sine component of the RF carrier.
- Since $\cos 2\pi f_c t$ and $\sin 2\pi f_c t$ are orthogonal over the duration of the amplitude shaping pulse $h_a(t)$, hence, the name **quadrature amplitude modulation**.

Vector Space Representation of Modulated Signals

• Consider a set of finite energy complex signals $\{\tilde{s}_m(t)\}_{m=0}^{M-1}$.

- Let $\{\varphi_i(t)\}_{i=0}^{N-1}$ be a set of N complete complex orthonormal basis functions, i.e.,

$$\int_{-\infty}^{\infty} \varphi_i(t) \varphi_j^*(t) dt = \delta_{ij}$$

such that

$$\tilde{s}_m(t) = \sum_{i=0}^{N-1} \tilde{s}_{mi} \varphi_i(t), \quad m = 1, \ldots, M$$

where

$$\tilde{s}_{mi} = \int_{-\infty}^{\infty} \tilde{s}_m(t) \varphi_i^*(t) dt$$

- The complete set of basis functions $\{\varphi_i(t)\}_{i=0}^{N-1}$ can be obtained from $\{\tilde{s}_m(t)\}_{m=0}^{M-1}$ using a Gram-Schmidt orthonormalization procedure; N is the dimension of the signal space.

• The signal $\tilde{s}_m(t)$ is represented by the length-N vector

$$\widetilde{\mathbf{s}}_m = \{\widetilde{s}_{m_1}, \ \widetilde{s}_{m_2}, \ \dots, \ \widetilde{s}_{m_N}\}$$

• Example QAM: In this case N = 1, and

$$\tilde{s}_m(t) = \sqrt{2E_h} x_m \varphi_0(t) , \qquad x_m = x_{I,m} + j x_{Q,m}$$

$$\varphi_0(t) = \sqrt{\frac{A^2}{2E_h}} h_a(t) , \qquad E_h = \frac{A^2}{2} \int_{-\infty}^{\infty} h_a^2(t) dt$$

Square QAM Constellations

• For QAM, $x_{I,m}$ and $x_{Q,m}$ take on odd integer values, i.e.,

$$x_{I,m}, x_{Q,m} \in \{\pm 1, \pm 3, \dots, \pm \hat{M} - 1\}$$

• For a square QAM constellation, $M = 2^k$ for some even k, i.e., $M = \hat{M}^2$.



Complex signal-space diagram for square QAM constellations.

Cross QAM Constellations

• For a cross QAM constellation, $M = 2^k$ for some odd k.



Complex signal-space diagram for cross QAM constellations.

Gray Coding



Signal-space diagram for 16-QAM with Gray coding. Adjacent symbols differ in only one bit position. Gray coding is optimal for uncoded waveforms, where no error control coding is used. It is not optimal when error control coding is used.

8-QAM Constellations

• The cross QAM constellation is not optimal for M = 8.



Alternative 8-QAM signal constellations.

Front-end Processing



• Optimum front-end processing for a digital signaling on an ISI channel.

- Sometimes called Forney's approach.

- In the above diagram, waveforms representing modulated symbols are transmitted using the amplitude shaping pulse $h_a(t)$.
- To maximize the signal-to-noise ratio at the output of the receiver (or detection) filter, the receiver filter in theory is an analog filter that is matched to the received pulse $h(t) = h_a(t) * g(t)$, i.e., the receiver filter is the matched filter h(-t), and the output of the matched filter is sampled at symbol-spaced intervals.
 - Unfortunately, if the channel g(t) is unknown, then so is h(t). Therefore, implementation of the matched filter h(-t) is practically impossible.

Practical Matched Filtering



- Choose the receiver filter to be an analog filter that is matched to the transmitted pulse $h_a(t)$, i.e., $h_a(-t)$, over-sample by a factor of 2, and process 2 samples per baud interval.
 - This front end processing is optimal, similar to Forney's receiver. Refer to Chapter 7 for the proof.
- For sequence estimation based equalizers, such as maximum likelihood sequence estimation (MLSE), the branch metrics will consist of 2 additive terms, one for each of the 2 samples per modulated symbol interval.
- For symbol-by-symbol equalizers, such as a decision feedback equalizer (DFE), the feed-forward filter can be a fractionally-spaced filter, i.e., a fractionally-spaced DFE with a T/2-spaced feed-forward filter.
 - The T/2 -spaced feed-forward filter is essentially the same as the T/2 -spaced noise whitening filter.

Matched Filtering and Pulse Shaping

• To design the transmitter filter and receiver matched filter, assume an ideal channel, i.e., set $g(t) = \delta(t)$, so that the overall pulse is

$$p(t) = h_a(t) * g(t) * h_a(-t)$$
$$= h_a(t) * h_a(-t)$$

• Taking the Fourier transform of both sides

$$P(f) = H_a(f)H_a^*(f) = |H_a(f)|^2$$

• Hence

$$|H_a(f)| = \sqrt{|P(f)|}$$

- For many practical amplitude shaping pulses, $h_a(t) = h_a(-t)$, i.e., the pulse is even in t.
- Implementation can be realized as an FIR filter.
 - If $H_a(f)$ is strictly bandlimited, the filter $h_a(t)$ is non-causal.
 - The filter can be approximated as a causal filter by adding delay in the amplitude shaping pulse $h_a(t)$ and truncating it to a finite length. For example, the time shifted pulse and truncated pulse $h_a(t 3T)u_{6T}(t)$ could be implemented. An example of this is shown later.

Conditions for ISI-free Transmission

The Nyquist first criterion for ISI-free transmission is

$$p_k = \delta_{k0} p_0 = \begin{cases} p_0 & k = 0\\ 0 & k \neq 0 \end{cases}$$

That is, the overall pulse $p(t) = h_a(t) * h_a(-t)$ must have equally spaced zero crossings, separated by the symbol interval T seconds.

<u>Theorem</u>: The pulse p(t) satisfies $p_k = \delta_{k0} p_0$ iff

$$P_{\Sigma}(f) \stackrel{\Delta}{=} \frac{1}{T} \sum_{n=-\infty}^{\infty} P(f+n/T) = p_0$$

That is the **folded spectrum** of the pulse p(t), denoted by $P_{\Sigma}(f)$, is flat in the frequency domain.



Nyquist Pulse Design



Construction of pulses satisfying the (first) Nyquist criterion. $P_N(f)$ is an "ideal" Nyquist pulse, and $P_o(f)$ is a transmittence function having skew symmetry about the Nyquist frequency 1/2T. Any transmittence function $P_o(f)$ having skew symmetry will work.

Raised Cosine Pulse

• The overall pulse p(t) is often chosen to be a **raised cosine pulse**, defined by the spectral description:

$$P(f) = \begin{cases} T & , \ 0 \le |f| \le (1-\beta)/2T \\ \frac{T}{2} \left[1 - \sin \frac{\pi T}{\beta} \left(f - \frac{1}{2T} \right) \right] & , \ (1-\beta)/2T \le |f| \le (1+\beta)/2T \\ 0 & , \ |f| \ge (1+\beta)/2T \end{cases}$$

- The bandwidth of the raised cosine pulse is $(1 + \beta)/2T$, where the parameter $\beta, 0 \le \beta \le 1$ is called the roll-off factor and controls the bandwidth expansion.
- The term "raised cosine" comes from the fact that pulse spectrum P(f) with $\beta = 1$ has a "raised cosine" shape, i.e., with $\beta = 1$

$$P(f) = \frac{T}{2} \left[1 + \cos(\pi fT) \right], \quad 0 \le |f| \le 1/T$$

• The corresponding time domain pulse is

$$p(t) = \frac{\sin \pi t/T}{\pi t/T} \frac{\cos \beta \pi t/T}{1 - 4\beta^2 t^2/T^2}$$

• When $\beta = 0$, the raised cosine pulse becomes the ideal Nyquist pulse $p(t) = \operatorname{sinc}(t/T)$.

Root Raised Cosine Pulse Shaping

• The amplitude shaping pulse $H_a(f)$ is often chosen to be the **root raised cosine pulse**, defined by the spectral description:

$$H_a(f) = \begin{cases} \sqrt{T} & 0 \le |f| \le (1-\beta)/2T \\ \sqrt{\frac{T}{2}} \left[1 - \sin \frac{\pi T}{\beta} \left(f - \frac{1}{2T} \right) \right]^{1/2} & (1-\beta)/2T \le |f| \le (1+\beta)/2T \\ 0 & |f| \ge (1+\beta)/2T \end{cases}$$

where, again, β is the *roll-off* factor.

• The corresponding time domain pulse is

$$h_{a}(t) = \begin{cases} 1 - \beta + 4\beta/\pi &, t = 0\\ (\beta/\sqrt{2})\left((1 + 2/\pi)\sin(\pi/4\beta) + (1 - 2/\pi)\cos(\pi/4\beta)\right) &, t = \pm T/4\beta\\ \frac{4\beta(t/T)\cos((1+\beta)\pi t/T) + \sin((1-\beta)\pi t/T)}{\pi(t/T)(1 - (4\beta t/T)^{2})} &, \text{elsewhere} \end{cases}$$

- The pulse $h_a(t)$ is non-causal, so the time-shifted pulse and length-L truncated pulse $h_a(t (L/2)T)u_{LT}(t)$ is implemented instead.
- For $\beta = 0$, the root raised cosine pulse reduces to the ideal Nyquist pulse

$$h_a(t) = \operatorname{sinc}(t/T)$$
.



Raised cosine and root raised cosine pulses with roll-off factor $\beta = 0.5$. The pulses are truncated to length 6T and time shifted by 3T to yield causal pulses.

PSK and $\pi/4$ -**DQPSK**

• **PSK:** The PSK bandpass signal is

$$s(t) = \sum_{n} h_a(t - nT) \cos \left\{ 2\pi f_c t + \theta_n \right\}$$

 $-\theta_n = 2\pi x_m/M$, $x_m \in \{0, 1, 2, \dots, M-1\}.$

- $-h_a(t)$ is an amplitude shaping pulse, e.g., square-root raised cosine pulse.
- $\pi/4$ -**DQPSK:** was used in the IS-54/136 and PDC systems.
 - Let $\Delta \theta(n) = \theta(n) \theta(n-1)$ be the differential carrier phase. The differential phase is related to the quaternary data sequence $\{x_n\}, x_n \in \{\pm 1, \pm 3\}$ through the mapping (this may be different from IS-54) $\Delta \theta(n) = x_n \pi/4$.
- The $\pi/4$ -DQPSK bandpass signal is

$$s(t) = \sum_{n} h_a(t - nT) \left[\cos\left(\theta_{n-1} + \frac{\pi}{4}x_n\right) \cos 2\pi f_c t - \sin\left(\theta_{n-1} + \frac{\pi}{4}x_n\right) \sin 2\pi f_c t \right]$$
$$\theta_{n-1} = \frac{\pi}{4} \sum_{k=-\infty}^{n-1} x_k$$



8PSK Signal-space constellation.



Signal-space constellations for QPSK and $\pi/4$ -DQPSK.



Phaser diagram, $\tilde{s}_Q(t)$ vs. $\tilde{s}_I(t)$, for $\pi/4$ -DQPSK with root raised cosine amplitude pulse shaping; $\beta = 0.5$.



Phaser diagram of $\pi/4$ -DQPSK with raised cosine pulse, $\beta = 0.5$. The raised cosine pulse assumes the transmit and receiver filters are square root raised cosine pulses.

Multi-Carrier Modulation



Block transmission through the principle of multicarrier modulation

Multi-Carrier Demodulation

• Optimal demodulation/detection for an AWGN channel



Orthogonal Frequency Division Multiplexing

• The OFDM complex envelope is given by

$$\tilde{s}(t) = A \sum_{n} b(t - nT, \mathbf{x}_{n})$$
(1)

where

$$b(t, \mathbf{x}_n) = u_T(t) \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi kt}{T}}$$
(2)

where

$$n = \text{block index}$$

 $k = \text{sub-carrier index}$
 $\mathbf{x}_n = \{x_{n,0}, x_{n,1}, \dots, x_{n,N-1}\}$ is the data symbol block at epoch n
 $N = \text{block size, usually a power of } 2$
 $u_T(t) = u(t) - u(t - T) = \text{unit amplitude length-} T$ rectangular pulse
 $\Delta_f = 1/T$

- The subcarrier phase on sub-carrier *n* during block *k* is $\phi_n = \operatorname{Tan}^{-1}(x_{k,n}^Q/x_{k,n}^I)$.
- The sub-carrier frequency separation of 1/T ensures that the sub-carriers are orthogonal in time regardless of the random sub-carrier phase that is introduced by the data symbols.

OFDM with Cyclic Extension

- A cyclic extension (or guard interval) is usually added to the OFDM waveform in (1) and (2) to combat ISI.
 - This can be a cyclic prefix, cyclic suffix, or both.
- With a cyclic suffix, the OFDM complex envelope becomes

$$\tilde{s}_g(t) = \begin{cases} \tilde{s}(t), & 0 \le t \le T \\ \tilde{s}(t-T), & T \le t \le (1+\alpha_g)T \end{cases},$$

where $\alpha_g T$ is the length of the guard interval and $\tilde{s}(t)$ is defined in (1) and (2).

• The OFDM waveform with cyclic suffix guard interval can be rewritten in the standard form

$$\tilde{s}(t) = A \sum_{n} b(t - nT_g, \mathbf{x}_n)$$

where

$$b(t, \mathbf{x}_n) = u_T(t) \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi kt}{T}} + u_{\alpha_g T}(t-T) \sum_{k=0}^{N-1} x_{n,k} e^{j\frac{2\pi k(t-T)}{T}} ,$$

and $T_g = (1 + \alpha_g)T$ is the **OFDM symbol** period with guard interval.

IFFT Implementation

• The OFDM complex envelope (without a cyclic guard) on the interval $mT \le t \le (m+1)T$ has the form

$$\tilde{s}(t) = A \sum_{k=0}^{N-1} x_{m,k} e^{\frac{j2\pi k(t-mT)}{T}} u_T(t-mT) = A \sum_{k=0}^{N-1} x_{m,k} e^{\frac{j2\pi kt}{NT_s}} u_T(t-mT) , \quad mT \le t \le (m+1)T .$$
(3)

• Sample the complex envelope at synchronized T_s -second intervals to yield the sequence

$$X_{m,n} = \tilde{s}(nT_s) = A \sum_{k=0}^{N-1} x_{m,k} e^{\frac{j2\pi kn}{N}}, \quad n = 0, 1, \dots, N-1$$
.

- The sampled envelope sequence has duration $T = NT_s$
- The vector $\mathbf{X}_m = \{X_{m,n}\}_{n=0}^{N-1}$ is the IDFT/IFFT of the vector $A\mathbf{x}_m = A\{x_{m,k}\}_{k=0}^{N-1}$. Contrary to conventional usage, the lower case vector $A\mathbf{x}_m$ is used to represent the frequency domain coefficients, while the upper case vector \mathbf{X}_m is used to represent the time domain coefficients.
- The OFDM modulator is just an IFFT operation followed by a pair of digital-to-analog converters (DACs) each with clock rate $R_s = 1/T_s$.

IFFT Implementation with Cyclic Guard

• In a typical OFDM transmitter each IDFT/IFFT output vector $\mathbf{X}_m = \{X_{m,n}\}_{n=0}^{N-1}$ is appended with a cyclic suffix to yield the vector $\mathbf{X}_m^g = \{X_{m,n}^g\}_{n=0}^{N+G-1}$, where

$$X_{m,n}^{g} = X_{m,(n)_{N}}$$

= $A \sum_{k=0}^{N-1} x_{m,k} e^{\frac{j2\pi kn}{N}}$, $n = 0, 1, ..., N + G - 1$

G is the length of the guard interval in samples, and $(n)_N$ is the residue of n modulo N.

- the cyclic guard can be implemented as a cyclic prefix, cyclic suffix, or both. Here, we consider a cyclic suffix.
- The purpose of this guard interval is to eliminate the effects of inter-symbol interference (ISI), or interblock interference (IBI) from one OFDM symbol to the next. Any residual ISI or IBI will cause large error floors.
- To maintain the same input data rate $R_s = 1/T_s$ with a cyclic guard, the DACs must be clocked with increased rate $R_s^g = \frac{N+G}{N}R_s$, due to the insertion of the cyclic guard interval.



FFT-based OFDM Transmitter

Discrete-time, Continuous-time Equivalence?

- The waveform generated by using the IDFT/IFFT implementation is *NOT* exactly the same as the waveform generated by (3).
- The waveform in (3) uses an amplitude shaping pulse $u_T(t)$ that is strictly time-limited to T seconds.
 - The resulting OFDM waveform is not strictly band-limited, so sampling $\tilde{s}(t)$ in (3) at any finite sampling rate will lead to aliasing and imperfect reconstruction.
- With the IDFT/IFFT implementation, we apply the IDFT/IFFT of the vector $A\mathbf{x}_m$ to a pair of balanced DACs as explained earlier.
 - The perfect reconstruction filter is an ideal low pass filter with cutoff frequency $1/(2T_s^g)$; in the time-domain the perfect reconstruction filter has the non-causal impulse response $h(t) = \operatorname{sinc}(t/T_s^g)$.
 - In practice, a perfect reconstruction filter is non-realizable and a causal, finite-length, reconstruction filter must used instead. However, such a finite-length filter will generate a waveform $\tilde{s}(t) = \tilde{s}_I(t) + j\tilde{s}_Q(t)$ that is not strictly bandlimited, i.e., side-lobes will appear in the power spectrum.

Orthogonal Modulation with Binary Orthogonal Codes

• Construct a Hadamard matrix using the recursion

$$\mathbf{H}_M = \left[egin{array}{cc} \mathbf{H}_{M/2} & \mathbf{H}_{M/2} \ \mathbf{H}_{M/2} & -\mathbf{H}_{M/2} \end{array}
ight]$$

where $\mathbf{H}_1 = [1]$.

• For example,

• Notice that the rows of the Hadamard matrix are mutually orthogonal.

Orthogonal Modulation with Binary Orthogonal Codes

• A set of M equal energy orthogonal waveforms can be constructed according to

$$\tilde{s}_m(t) = A \sum_{k=1}^M h_{m_k} h_c(t - kT_c) , \quad m = 0, \dots, M - 1$$

- $-h_{m_k}$ is the kth co-ordinate in the mth row of the Hadamard matrix
- $-T = MT_c$ is the symbol duration
- $-h_c(t)$ is a "chip" shaping pulse
- This type of orthogonal modulation is used with CDMA cellular, e.g., IS-95A/B, cdma2000, reverse link.
- The same waveforms can be used for synchronous CDMA multi-access. One waveform is assigned to each user, e.g., IS-95A/B, cdma2000 forward link.

Orthogonal Multipulse Modulation

• A more bandwidth efficient scheme can be obtained by using the rows of the Hadamard matrix \mathbf{H}_N to define N orthogonal amplitude shaping pulses

$$h_k(t) = A \sum_{\ell=0}^{N-1} h_{k\ell} h_c(t - \ell T_c) , \quad k = 0, \dots, N-1$$

- The block of N information symbols is transmitted in parallel by using the N orthogonal amplitude shaping pulses.
- The transmitted complex envelope is

$$\tilde{s}(t) = \sum_{n} b(t - nT, \mathbf{x}_n)$$

where

$$b(t, \mathbf{x}_n) = \sum_{k=0}^{N-1} x_{n_k} h_k(t)$$

 $T = NT_c$, and $\mathbf{x}_n = (x_{n_0}, x_{n_1}, \dots, x_{n_{N-1}})$ is the block of N data symbols transmitted at epoch n. With OFDM, from (2)

$$h_k(t) = e^{j\frac{2\pi kt}{T}} u_T(t)$$