The First Course in Telecommunications, a Top-down Approach

Peter F. Driessen
Department of Electrical and Computer Engineering
University of Victoria, Canada V3W 3P6
driessen@uvic.ca

ABSTRACT— *The traditional first course in telecommunications starts with signals and spectra, linear filtering and then analog modulation AM, FM, random signals and noise, noise in analog modulation. Only after all that does the student see anything digital. In this paper, we outline a top-down approach to teaching the first (required, core) course in telecommunications, starting with baseband digital messages, link budget and general amplitude/phase modulation/demodulation, and leading to AM, FM, QPSK, FSK, QAM etc as special cases.*

I. Introduction

The traditional first course in telecommunications [4] is generally about the concept of a carrier wave with analog AM and FM modulation, and performance of AM and FM in noise. Later courses cover other physical layer parts of communications systems such as digital modulation, coding, spread spectrum, microwave components, fiber optics and antennas. Other communications courses cover wireless (cellular) systems and the higher layers (protocols and networks).

Today, most physical layer communications using carrier waves (fiber, wireless) uses digital modulation and the last bastions of analog modulation (AM and FM broadcast radio and TV carried over cable or wireless channels) are being replaced by digital broadcast systems. Thus it seems appropriate for the first course in communications to start with digital modulation. In this paper, we show a top-down approach which introduces carrier waves and modulation with digital signals first, and analog as a special case.

In section 2, we give the course outline and its motivation. In section 3, we provide details of the top-down approach.

II. Course outline

The course starts with the Webster's definition of telecommunications as "the science and technology of communications at a distance by electronic transmission of impulses, as by telegraph, cable, telephone, ... ".

We then introduce the idea of a carrier wave, and then introduce the link budget as a means to find the distance that can be covered, and it's dependence on power and bandwidth. Via many link budget examples ranging from cellphone to broadcast TV, students get a feeling for what communications systems are and can do.

Only after developing this intuition, do we introduce details of baseband message signals and modulation. We consider baseband messages in the form of sine waves or square waves. Sine waves can be interpreted as spectral components of an analog message (voice, music). Both sine and square waves can be interpreted as representing a digital message with a 101010... data pattern.

Modulation is introduced with the general expression for amplitude/phase modulation of one carrier wave, and then several carrier waves. We show both the real passband and complex baseband forms of the modulated signal. We then show how amplitude-shift keying, phase-shift keying and frequency-shfit keying are all special cases.

Receiver/demodulators are introduced as a general I-Q receiver which recovers the I and Q components, from which the amplitude and phase of the complex envelope are obtained.

Performance calculations, bit error rates and FM threshold effect are left for a later course.

III. Details

A. Definition

The course starts by introducing telecommunications as per the dictionary definition (see Appendix).

Smoke signals, shouting and handwaving are ways of communications at a distance, but telecommunications, as per Webster's, is exclusively electronic. Thus a telecommunications system for people must first convert the message from human-readable form (speech, music, image, video, text, data) into electronic form then transmit this electronic message over the distance via some channel (electronic pathway), and convert it back to human-readable form.

For the free space radio channel a carrier wave (radio wave) is needed to carry the signal across free space using electromagnetic (radio) waves. Similarly, for the optical fiber channel, a carrier wave (light wave) is needed to pass the signal through (inside) the optical fiber. A carrier wave can also be used on the copper wire channel.

In this way, we introduce telecommunications as communications over a distance, requiring a carrier wave to carry the message over the distance. The choice of carrier frequency depends in part on the distance to be covered. We then mention that a message can be sent by turning the carrier on and off in some pattern, as was done with

Thanks to all the students in my courses who put up with my approach and seemed to enjoy it.

Morse code, and is now done on fiber optic links. We use the general overview in [2].

B. Link Budget

To discover just how much distance we can cover with the carrier wave, the link budget is introduced [1]. The link budget specifies the distance in terms of the available power and bandwidth, and the unavoidable noise and interference, and can be used to calculate the range of a cellphone from the base station, or the time it takes to send an image from a space probe near Saturn. The link budget is introduced as follows.

For any communications system designed to send a message from point A to point B linked via some channel (fiber, coax, radio), the available resources are transmit power and channel bandwidth, and the obstacles to be overcome are noise and interference. Both power and bandwidth cost money; how much will depend on the details of the communications system.

To design a working link with acceptable quality at the lowest possible cost we need to first specify the customer performance requirements, and then compute the necessary power and bandwidth. These performance requirements are usually expressed for digital signals as the data rate at a given error rate, or for analog signals as the fidelity (bandwidth) and signal-to-noise ratio. Another performance requirement is the distance to be covered. Alternately, we may be given the power and/or bandwidth that we can afford, and then need to compute the maximum data rate and distance that can be achieved.

The link budget is presented in a simplified way, and all quantities are expressed in dB. We calculate the received power obtained $P_{r,o}$ for a given link, and the received power needed $P_{r,n}$ for acceptable quality. For the link to work, the parameters must be adjusted so that the link margin

$$M = P_{r,o} - P_{r,n} > 0 \tag{1}$$

We write the power obtained as

$$P_{R,o} = P_T + G_T + G_R - L_0 \tag{2}$$

where P_T, G_T, G_R are transmit power, transmit and receive antenna gains and $L_0 > 0$ is the path loss. Expressions for the antenna gain are given in terms of antenna area. Expressions for L_0 are given for cables of given length and specified attenuation per unit distance. For radio channels, we consider free space $L_0 = 32.4 + 20logd + 20logf$ for d in meters and f in GHz, as well as two-ray flat-earth channels with $L_0 = 40logd - 20logh_1 - 20logh_2$ attenuation where h_1, h_2 are the antenna heights, or the more general two-slope model $L_0 = L_B + 10n_1 logd$ for $d < d_{brk}$ and $L_0 = L_B + 10(n_1 - n_2)log(d_{brk}) + 10n_2 logd$ where L_B depends on f, h_1, h_2, typically $n_1 = 2, n_2 = 4$, and $d_{brk} = 4h_1h_2/\lambda$ is the breakpoint distance where the propagation loss shifts from exponent n_1 to n_2. We write the power needed

$$P_{r,n} = (S/N) + W + F - k \tag{3}$$

where (S/N) is the signal to noise ratio needed for the desired quality of service or bit error rate, W is the system bandwidth, $k = -204$ dBW/Hz represents thermal noise at room temperature 290 degrees Kelvin, and F is the noise figure of the system. Interference can be approximately represented as a larger value of k.

For digital systems, we point out that the relationship between data rate R and bandwidth W depends on the particular system, and are typically in the range from $R = W$ for simple systems to $R = 16W$ for 56Kbit/sec dialup telephone modems. We mention the Shannon theoretical capacity limit $R = C = Wlog_2(1 + S/N)$ and show that for telephone lines with $W = 3.5$KHz, 8 bit samples and a signal-to- quantization noise $S/N = 48$ dB = 63,000 (6 dB/bit), $R = 16W \simeq 56,000$. This example is very popular, since it relates to something the students use (dialup modems), and explains why dialup modems have reached their theoretical speed limit. We also mention that $\frac{S}{N} = \frac{E_b}{N_0}\frac{R}{W}$ so that textbook bit error rate curves can be used in link budget calculations.

We then illustrate the use of the link budget with numerous examples. Examples include:

- find the maximum data rate of a telemetry link to Saturn assuming $R = W$ and free space path loss.
- find the range of a cellphone from a base station given the antenna heights
- find the maximum distance between repeaters on an undersea fiber link
- find the transmit power of a TV broadcast (6 MHz bandwidth) needed to cover a city from a single 1000 foot tower, and compare to the power needed for FM radio broadcast (200 KHz bandwidth).
- how much power and antenna gain is needed to bounce a 10 bit/sec signal at 900 MHz from the moon and correctly receive the echo ?
- what is the range of an 11 Mbits/sec wireless local area network using 100 millwatts transmit power ?

In many of these examples, reasonable assumptions must be made. These examples provide motivation and context for the detailed studies of modulation to follow, and will help students understand how we use power and bandwidth to overcome noise and interference.

C. Analog and digital messages

We then focus on the nature of the message signal. For a digital message, the sequence of 1's and 0's needs to be converted to an analog voltage waveform $m(t)$. This is done by assigning a pulse shape to each bit. Pulses may be square (a constant voltage for the bit time T) or rounded (a half cycle of a sine wave of length T). We also mention that pulses may be spread out over a time longer than T. We send positive pulses for 1 bits, and negative pulses for 0 bits. A one-zero (10) message sequence is seen to be one cycle of a square wave or sine wave with period $2T$. A repeating 101010... message sequence can be a square wave or a sine wave with period $2T$. We mention frequency domain raised-cosine pulses which extend over many symbol periods.

An analog message such as voice, music, image or video, can be decomposed into a sum of sine waves. Thus, a sine wave message $m(t)$ can represent both an analog and a digital message. We point out that for digital messages, the square or rounded pulses can have multiple amplitudes, e.g 4 levels $\pm 1, \pm 3$. thus allowing the transmission of 2 bits of information (4 combinations 00 10 11 10) per symbol time.

We briefly show how an analog message can be digitized (sampled, quantized) to create bits, and how these bits are formatted (segmentation into blocks, source coding, encryption, channel coding) to make different bits. This digital message, representing the original analog mesage, is then modulated onto the carrier.

D. Modulation

At this point, we introduce modulation of a carrier wave as general amplitude/phase modulation, and show show how AM, PM and other modulations arise as a special cases. We also show the waveforms for $m(t)$ both a square wave and a sine wave.

We start with a general phasor $a(t)exp[j\theta(t)]$ shown with a polar diagram, and recall that the sum of two phasors rotating in opposite directions yields

$$s(t) = a(t)cos[\theta(t)] \tag{4}$$

If the phase changes linearly with time $\theta(t) = 2\pi f_c t + \phi$ with f_c rotations of 2π in $t = 1$ second, then we have a carrier wave

$$s(t) = acos[2\pi f_c t + \phi] \tag{5}$$

We then show that we can vary both the amplitude and the phase with time, giving rise to a modulated signal

$$s(t) = a(t)cos[2\pi f_c t + \phi(t)] \tag{6}$$

where there is a relationship (to be defined) between $a(t), \phi(t)$ and the message signal $m(t)$. [1] [2] We introduce the polar diagram of $s(t)$ as a 2-D vector with time varying amplitude $a(t)$ and phase $\phi(t)$, and its corresponding rectangular form

$$s(t) = x(t)cos2\pi f_c t - y(t)sin2\pi f_c t \tag{10}$$

with relationships $x(t) = a(t)cos\phi(t), y(t) = a(t)sin\phi(t)$.

Using these polar and rectangular form expressions for $s(t)$ as general amplitude/phase modulation, we show how amplitude and phase modulation arise as special cases. Different modulations simply use different mappings between $m(t)$ and $x(t), y(t)$, or equivalently between $m(t)$ and $a(t), \phi(t)$. We draw a diagram of the general modulator with a mapper block with input $m(t)$ and outputs $x(t), y(t)$, followed by a diagram to implement (10) yielding the modulated signal $s(t)$.

For example, the mapper block for AM arises from the definition of the idea of AM as an instantaneous amplitude varying about a constant amplitude in step with the modulation $a(t) = A_c + A_c k_a m(t), \phi(t) = 0$ assuming $max|m(t)| = 1$ and k_a is the modulation index, and $A_c^2/2$ is the carrier power. Thus the mapping from $m(t)$ to $x(t), y(t)$ is $x(t) = A_c[1 + k_a m(t)], y(t) = 0$. For double sideband suppressed carrier, we delete the constant term. For single sideband suppressed carrier $x(t) = m(t), y(t) = \pm\hat{m}(t)$ where $\hat{m}(t)$ is the Hilbert transform of $m(t)$, with $-$ for upper sideband, $+$ for lower sideband. The mapper block for phase modulation arises from $a(t) = A_c, \phi(t) = k_p m(t)$ where k_p is the phase modulation index, so that the mapping from $m(t)$ to $x(t), y(t)$ is $x(t) = A_c cos[k_p m(t)], y(t) = A_c sin[k_p m(t)]$. With $k_p = \pi/2$ and $m(t)$ a square wave with $max|m(t)| = 1$, we obtain binary phase shift keying. For frequency modulation [3] we obtain the mapping

$$x(t) = A_c cos\left[2\pi \triangle f \int_{\tau=0}^{t} m(\tau)d\tau\right], \tag{15}$$

$$y(t) = A_c sin\left[2\pi \triangle f \int_{\tau=0}^{t} m(\tau)d\tau\right] \tag{16}$$

Different modulations simply use different mappings between $m(t)$ and $x(t), y(t)$. In this way, we present a unified theory of modulation that shows the relationship between the special case expressions for AM, DSB, SSB, PM, FM. We also point out that the mapping block can have two independent inputs $m_1(t), m_2(t)$ with simple mapping $x(t) = m_1(t), x(t) = m_2(t)$, so that we can send two independent messages on a single carrier wave.

We then consider a digital message 101010... and show the time-domain waveforms $s(t)$ for messages $m(t)$ both a square wave and a sine wave with period $2T$. For square waves, we obtain amplitude, frequency and phase shift keying (ASK,FSK,PSK), and show how they can be sampled at the appropriate times $t = nT$ to retrieve the bits.

[1] We also recall the idea that

$$s(t) = a(t)\{exp[j2\pi f_c t + \phi(t)] + exp[-j2\pi f_c t - \phi(t)]\}/2 \tag{7}$$

is made of of two phasors rotating in opposite directions, thus giving rise to positive and negative frequencies.

[2] Frequency modulation is shown later as arising from a time-varying phase $s(t) = a(t)cos[\theta(t)]$ where the phase $\theta(t) = 2\pi f_c t + \phi(t)$ gives rise to a time-varying frequency

$$f_i(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt} \tag{8}$$

$$= f_c + \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{9}$$

[3] The idea of frequency modulation is an instantaneous frequency varying about a constant frequency in step with the modulation

$$f_i(t) = f_c + \triangle f m(t) \tag{11}$$

with frequency deviation $\triangle f$ for $max|m(t)| = 1$ so that from (9)

$$\triangle f m(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{12}$$

or

$$\phi(t) = 2\pi \triangle f \int_{\tau=0}^{t} m(\tau)d\tau \tag{13}$$

For $m(t) = cos[2\pi f_m t]$

$$\phi(t) = (\triangle f/f_m)sin[2\pi f_m t] \tag{14}$$

where $\beta = \triangle f/f_m$ is the modulation index.

Using the unified modulation theory, we point out that the mapping block can have two independent inputs $m_1(t), m_2(t)$ with simple mapping $x(t) = m_1(t), x(t) = m_2(t)$. Thus we show how we can send two independent messages on a single carrier wave, one on $x(t)$ and one on $y(t)$. For 2-level (± 1) pulse amplitudes in each of $m_1(t), m_2(t)$, we can send 2 bits per symbol time T (Quadrature Phase shift keying), and for 4-level ($\pm 1, \pm 3$) pulse amplitudes in each of $m_1(t), m_2(t)$, we can send 4 bits per symbol time T (Quadrature Amplitude Modulation).

We show the 3-D representation of signals plotting $x(t), y(t), t$, and the corresponding 2-D 'side view' of the I component $x(t), t$ and Q component $y(t), t$. We also show the 2-D 'end view' of $x(t), y(t)$ (where t goes perpendicular into the page), which we call the signal constellation. We show both the continuous time constellation for all times t, and also the traditional sampled constellation with points at times $t = nT$.

We mention the idea of the complex signal representation

$$s(t) = Re\{a(t)exp[j\phi(t)]exp[j2\pi f_c t]\} \tag{17}$$

which reduces to the real form (10). The message signal $m(t)$ is mapped to the complex envelope $a(t)exp[j\phi(t)] = x(t)+jy(t)$ in a manner depending on the modulation type.

E. Channels

We introduce general multipath channel models

$$r(t) = \sum_k h_k(t)s(t - k\tau(t)) + n(t) \tag{18}$$

where h_k is an FIR filter with complex tap coefficients, which may have time-varying gains and spacings. In the link budget $h_0(t) = 1/L_0$ is constant with time and other $h_k = 0$. The channel models are a review of basic FIR digital filter theory, extended to the time-varying case.

F. Demodulation/Receivers

We introduce the general receiver/demodulator which recovers $x(t), y(t)$ from $r(t)$. The received signal $r(t)$ is split into two branches, which are multiplied by $cos(2\pi f_c t + \Phi), sin(2\pi f_c t + \Phi)$ respectively. After multiplication, both branches are low pass filtered. Assuming $\Phi = 0$, the two outputs will be good estimates of $x(t), y(t)$. We also show that we can obtain the envelope $a(t)$ by computing $a(t) = \sqrt{x^2(t) + y^2(t)}$ and obtain the phase from $\phi(t) = arctan[y(t)/x(t)]$. Thus the receiver recovers the complex envelope. The demapping block then combines $x(t), y(t)$ in a manner according to the modulation type, to produce the estimate of $m(t)$. For example, the demapping for AM is $m(t) = a(t) - \overline{a(t)}$, and the demapping for FM is $m(t) = \frac{d\phi(t)}{dt}$. We also show that by multiplying $x(t)$ by $y(t)$ and low pass filtering, we can obtain an estimate of the phase offset $sin2\Phi$. We observe that, provided the phase offset $\Phi = 0$, that $x(t)$ and $y(t)$ are received independenty of one another.

G. Matched filters and orthogonal signalling

For digital messages, we introduce the idea of replacing the low pass filters in the I and Q branches with a matched filter whose impulse response matches the pulse shape (square or half-sine). For a sampled system, I-Q low pass filters (implemented as an FIR filter) are replaced with another FIR filter with coefficients matching the pulse shape. We then show how on the I branch, the sin component of the signal (10) integrates to zero in the matched filter, so that the output is only $x(t)$ with no crosstalk from $y(t)$.

We also introduce the idea of using a second carrier wave with carrier frequency $f_{c2} = f_c + 1/T$ where T is the symbol time. This second carrier is also modulated to produce $s_2(t)$. We show that the I and Q components of $s_2(t)$ are orthogonal to those of $s(t)$. This leads naturally to FSK and multitone signalling. The second carrier can be used for two additional independent data streams.

We then present an elegant general structure to represent multitone signalling with Quadrature Amplitude Modulation (QAM) on each of 4 carriers spaced by $1/T$. There is a mapper which takes in 1, 2, 4, 8 or 16 bits per symbol time. With 1 bit, we can have binary ASK, binary PSK or binary FSK (choosing one of the two carriers f_1, f_2 per symbol time). With 2 bits, we can have 4-level ASK, 4-PSK (QPSK), binary ASK or PSK on two carriers simultaneously, FSK (choosing either f_1, f_2 for one bit, and one of f_3, f_4 for the other bit) or choose-one-out-of-4-carriers FSK. With 4 or more bits, there are many combinations. We point out that this is an 8-dimensional signalling space, where in each dimension, we can send 0, 0/1, -1/+1, or multilevel -3/-1/+1/+3 during each symbol time T. This structure can be used to demodulate all signal types, and unifies, clarifies and distinguishes all types of digital signalling and the number of bits per symbol.

H. Signal Spectra

We compute the spectrum (frequency domain waveforms) for AM and FM signals using the conventional techniques of e.g. [4] with sine waves for the message signal $m(t)$. The spectrum illustrates the concept of sidebands, and we show how AM leads to one pair of sidebands, whereas FM leads to many sidebands with amplitudes given by Bessel functions.

The AM sidebands with $m(t) = cos(2\pi f_m t)$ arise from the multiplication of two cosine waves, yielding sum and difference frequencies

$$s(t) A_c cos 2\pi f_c t + A_c cos(2\pi f_m t) cos(2\pi f_c t) \tag{19}$$

$$= A_c cos 2\pi f_c t + A_c cos[2\pi(f_c + f_m)t] cos[2\pi(f_c - f_m)t] \tag{20}$$

However, even though the mathematics is clear, it is often difficult for students to understand how modulating a carrier in amplitude can result in sidebands. A very convincing demonstration of sidebands is to generate $s(t)$ in Matlab using $f_c = 440$ Hz and slowly ramping f_m from 0 to 110 Hz over about 20 seconds. The signal $s(t)$ can be played through the sound card and speakers using the

Matlab *soundsc* function. The effect for f_m from 0 to 10 Hz is a slow tremolo. Near 20-30 Hz the sound becomes rough and with unclear pitch. Above 30 Hz, we start to hear 3 distinct tones. At 110 Hz, we have 3 tones at 330, 440 and 550 Hz, yielding a pleasant major chord with notes E, A, C#. This may be repeated using $f_c = 220$ Hz with similar results.

For FM, the same demonstration allows multiple sidebands to be heard. The relative amplitude of the sidebands can be varied with the modulation index, which is most apparent with $f_c = 440$ Hz and $f_m = 110$ Hz. A convincing demonstration of the complexity of FM spectra is to generate a bell sound using combined amplitude/frequency modulation with time varying amplitude and modulation index[5]

$$s(t) = a(t)cos[2\pi f_c t + \beta(t) sin 2\pi f_m)t \tag{21}$$

with $f_m = f_c$, $a(t) = exp(-t/\tau_1), \beta(t) = b_0 exp(-t/\tau_2)$, both decaying exponentially. The student can experiment with the values of $f_c, f_m, b_0, \tau_1, \tau_2$. Other functions for $a(t), \beta(t)$ can yield interesting sounds.

I. Discussion

Having completed this outline of the basic structure of the modulation of a carrier wave, and convinced the student that sidebands are real, we continue with the traditional course material in [4].

We can show how the traditional analog AM and FM receivers (envelope detector, frequency discriminator) are equivalent to the general I-Q demodulator along with the the appropriate demapping block. We introduce the idea of the superheterodyne receiver, and how fixed IF filters are used to select the desired signal and filter out undesired signals on adjacent channels. These filters are located in the IF amplifier chain before the general I-Q demodulator.

Up to this point, we have deliberately omitted some fine points such as pulse shapes longer than the symbol time T, and the expression for a digital message

$$m(t) = \sum_k c_k g(t - kT) \tag{22}$$

Complex signals are mentioned, but not emphasized, since real signals are easier to visualize the first time. We have mentioned the matched filter, and justified it intuitively, but have not derived it. We have not mentioned MSK or OQPSK. Spread spectrum may be briefly mentioned. System performance calculations in terms of bit error rate are deliberately omitted, and left for a later course. We have outlined the basic structure of the physical layer to give context to the traditional material, and further details can be explored in later courses.

The course makes extensive use of MATLAB so that students can plot the waveforms in time and frequency domain and thus visualize what is going on. The entire structure of the general I-Q transmitter, channel and receiver are implemented in MATLAB [3] and the source code made available to the students to experiment with and modify.

IV. Summary

This top-down approach to the first course in telecommunications creates motivation, context and structure for studying the traditional modulations and receivers. The link budget provides examples of practical communication systems, and provides intuition about the power and bandwidth needed to build particular systems to cover specified distances, and how for a fixed power, there is a tradeoff between distance and bandwidth (data rate). The general modulator structure unifies all modulation types AM, FM, PSK etc in a common framework. The general demodulator structure can demodulate any signal type. This framework will be useful to students, even if this is the only communications course they take.

Appendix

Webster's defines "communications" as "the exchange of thoughts, messages or information, as by speech, signals, writing or behaviour". as well as "the technology employed in transmitting messages". Since "tele" means distance, telecommunications, according to Webster's, is "the science and technology of communications at a distance by electronic transmission of impulses, as by telegraph, cable, telephone, radio or television" (or cellphones or computer networks).

Thus telecommunications is about how to send a message between locations separated by some distance. The message (thoughts or information) may be in the form of speech, music, still image, moving video, text or data. At each location, there will be people who will see or hear the message, or devices which will read it electronically.

In human communications (without the 'tele'), messages are expressed (transmitted) in a form which can be detected (received) by human senses (e.g. sight and hearing). Telecommunications provides a means for these human senses to work at a distance[2].

References

[1] http://www.ece.uvic.ca/ peter/35000/350link.html
[2] http://www.ece.uvic.ca/ peter/35000/hbook350.txt
[3] http://www.ece.uvic.ca/ peter/35000/amfmdem3.txt
[4] S. Haykin, "Communication Systems", 4th Edition, Wiley, 2000.
[5] J. Chowning, "The Synthesis of Complex Audio Spectra by Means of Frequency Modulation", *J. Audio Eng. Soc.* 21, no. 7, pp. 526-534, 1973.