The Universality of Modulation

A Technical Paper Prepared for SCTE/ISBE by

Tom Williams
Principal Architect
Cable Television Laboratories, Inc.
858 Coal Creek Circle
Louisville CO 80027-9750
303-661-3486
t.williams@cablelabs.com

Table of Contents

List of Figures

Introduction

This paper examines the common properties of modulation methods that use orthogonal basis functions, including single carrier (e.g., 64-QAM or 64-state quadrature amplitude modulation), multi-carrier (e.g., OFDM (orthogonal frequency division multiplex)), and spread spectrum. It also explains a time-frequency swapping technique that reveals OFDM and single carrier modulation are duals of one another, if the time axis and frequency axis are relabeled. This time-frequency swapping technique allows new modulation candidates to be created, such as duobinary (partial response signaling) OFDM, which has valuable properties for the cable plant.

Universality of Modulation

1. Modulation Techniques

Modulation techniques are used at carrier frequencies to send digital data over a distance, either by wires, wireless, or optically. Three popular modulation methods are single carrier, multi-carrier, and code division multiple access. All three have been used on cable networks at one time or another.

Orthogonality between signals is a property that allows one signal, comprising symbols, to be clearly received without interference from other signals' symbols.

$\sum x(n) \cdot y(n) = 0$ [eq. 1]

Two variables, x and y, are orthogonal over a range if:

Sum $x*y = 0$ over a range in $y \neq x$. [footnote 1]

Different modulation techniques use different techniques to achieve orthogonality.

In cable, single carrier modulation has been used extensively on downstream signal paths; examples include 64-QAM and 256-QAM. ATDMA (advanced time division multiple access), essentially a burst-mode single carrier transmission technique, has been used on upstream signal paths. Single carrier modulation comprises a time series of voltage impulses (symbols) that have been filtered to limit interference with other frequency bands. Figure 1 is a voltage vs. time diagram showing five sine(x)/x impulses with uniform time shifts. The illustrated different symbols are the same value, although shifted in time. The symbols can have any positive or negative values and have real-only values or be complex. The sampling instants are illustrated with five vertical lines. The waveforms do not interfere with each other because at each sampling instant, one symbol is at its peak while the others are passing through zero. Thus, orthogonality is maintained. For this system to work optimally, linear distortions, such as echoes, need to be removed prior to sampling, typically using an adaptive equalizer. Otherwise, the responses from the other symbols will not be at zero at the sampling instant and they will contribute distorting energy.

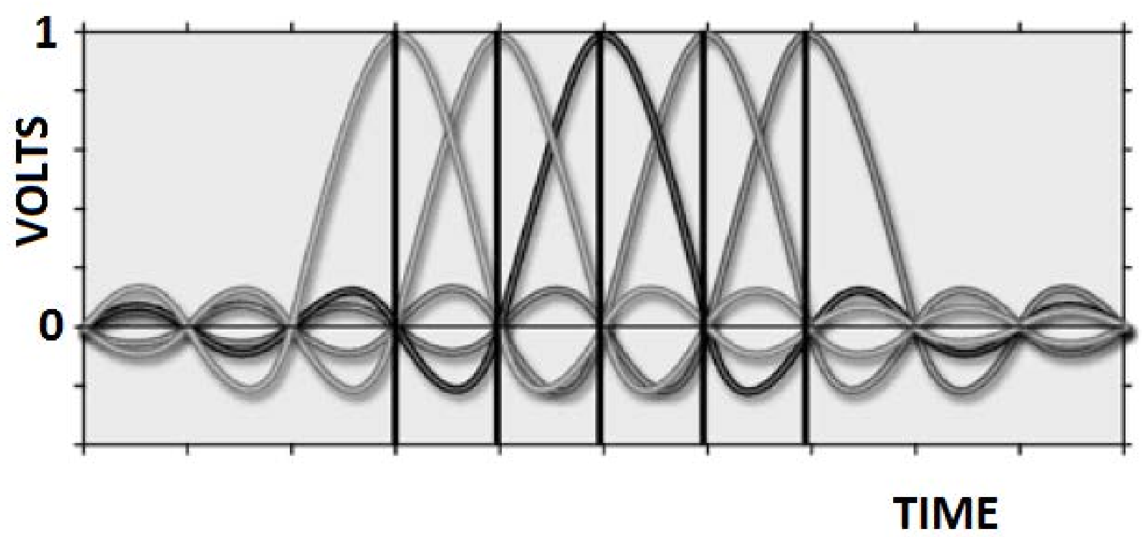

Figure 1 - Single Carrier Voltage Vs Time Diagram

Direct sequence spread spectrum (DSSS) has been used by the military to hide communications and radar signals by making them appear noise-like. In cable, a related technology called synchronous code division multiple access (S-CDMA) has been used for upstream noise immunity and for multiple access. [1] Multiple orthogonal codes are assigned to one or more users and simultaneous transmissions can occur on different codes without interference. This technique also requires equalized signals, or orthogonality between codes will be lost. Figure 2 is a timing diagram showing the creation of a pseudo-random sequence with a number of cascaded shift registers and an exclusive-or gate. The clocking rate is called the “chip” rate. A signal to be transmitted is generated by a data source inverting, or not inverting, the pseudo-random output of the shift register.

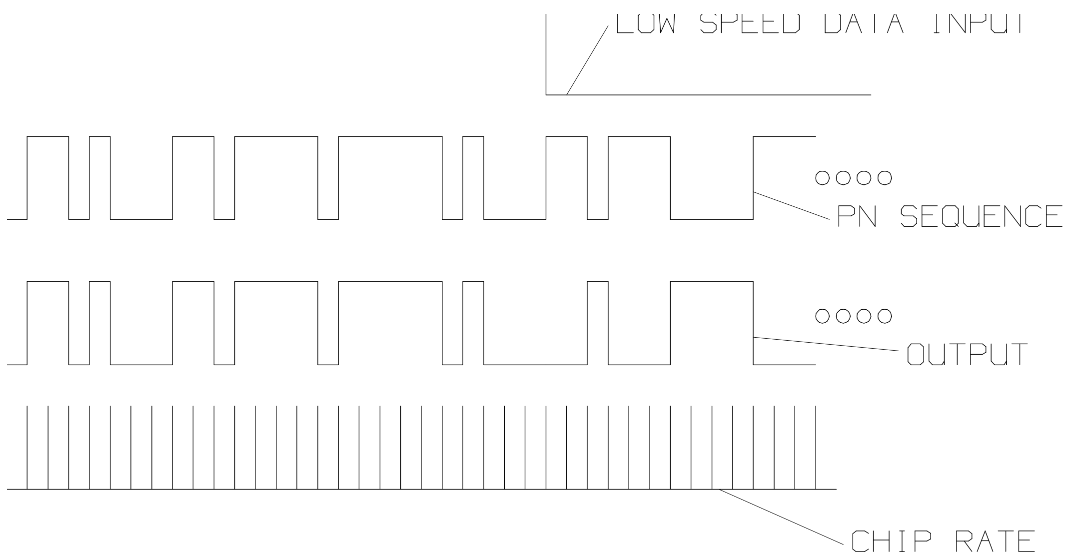

Figure 2 - Timing Diagram for a Spread Spectrum Signal

Figure Note: A low speed data source is exclusive-or'd with a high-speed PN (pseudo-noise) sequence to produce an output that appears noise-like.

Figure 3 is a block diagram for DSSS showing the transmission and reception of a spread-spectrum signal. At the receiver, a PN generator must be using both the same code as the transmitter, and be synchronized with it. With S-CDMA DOCSIS functionality, each circular time shift (excluding the initial chip which is not shifted) produces another basis function that is orthogonal to all the other shifts.

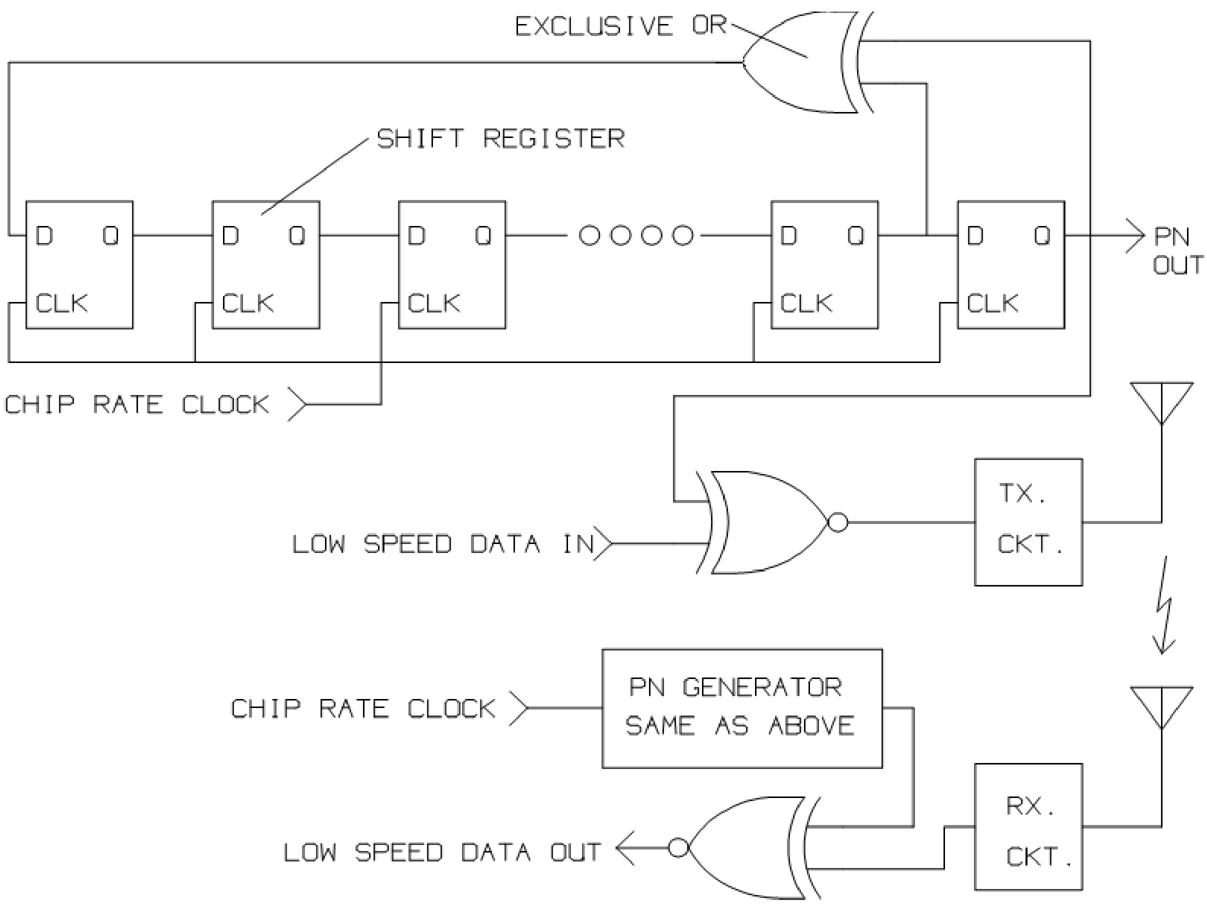

Figure 3 - Diagram for Direct Sequence Spread Spectrum

OFDM is used in DOCSIS® 3.1 technology as well as in numerous wireless standards. With OFDM many different subcarriers that are all harmonics of a fundamental are used to obtain orthogonality. Figure 4 shows an OFDM waveform with only four such subcarriers. Each of the harmonically-related subcarriers has a different magnitude and phase value. When all four subcarriers are combined (summed) for transmission, the result is a single composite signal. However, orthogonality allows the original subcarriers to be separated at the receiver, usually with a fast Fourier transform (FFT). A guard interval (GI), which is also known as a cyclic prefix (CP), is made by copying samples from the end of the signal and pasting it onto the beginning. This is done so that equalization can be performed with a circular convolution (or equivalent) and no interference is suffered from the previous OFDM block if there is an echo on the channel. This assumes that the echo is shorter than the GI.

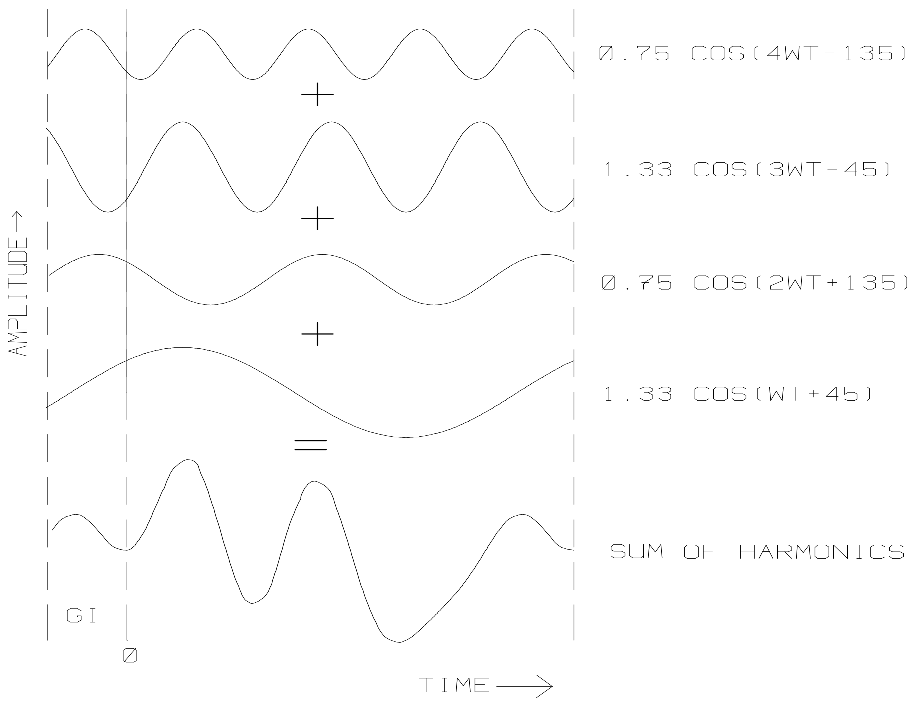

Figure 4 - An OFDM Signal in the Time Domain Containing Four Subcarriers

Note that in Figure 4, an OFDM signal in the time domain contains only four subcarriers, first, second, third, and fourth harmonics.

Figure 5 show the same OFDM signal of Figure 4, but in the frequency domain.

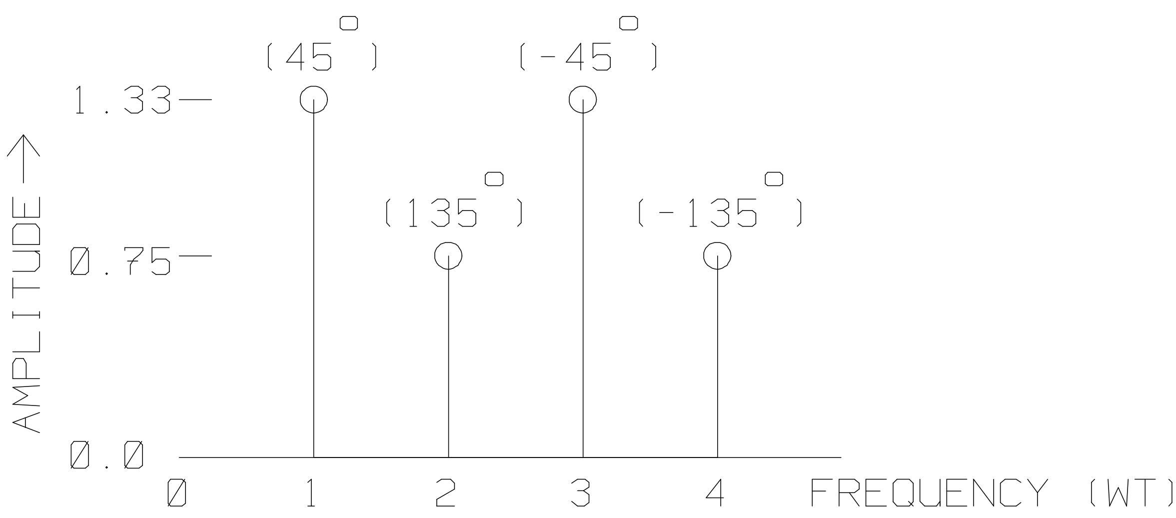

Figure 5 - The OFDM signal of Figure 4 Viewed In The Frequency Domain

Figure 6 shows a spectral plot of an OFDM signal that is affected by a deep frequency-selective fade. This fading occurs frequently in wireless channels where a sum of echo components cancels the signal, or at least at some subcarrier frequencies. This is an environment ideal for OFDM where the faded subcarriers lost in the noise can be recovered using forward error correction (FEC).

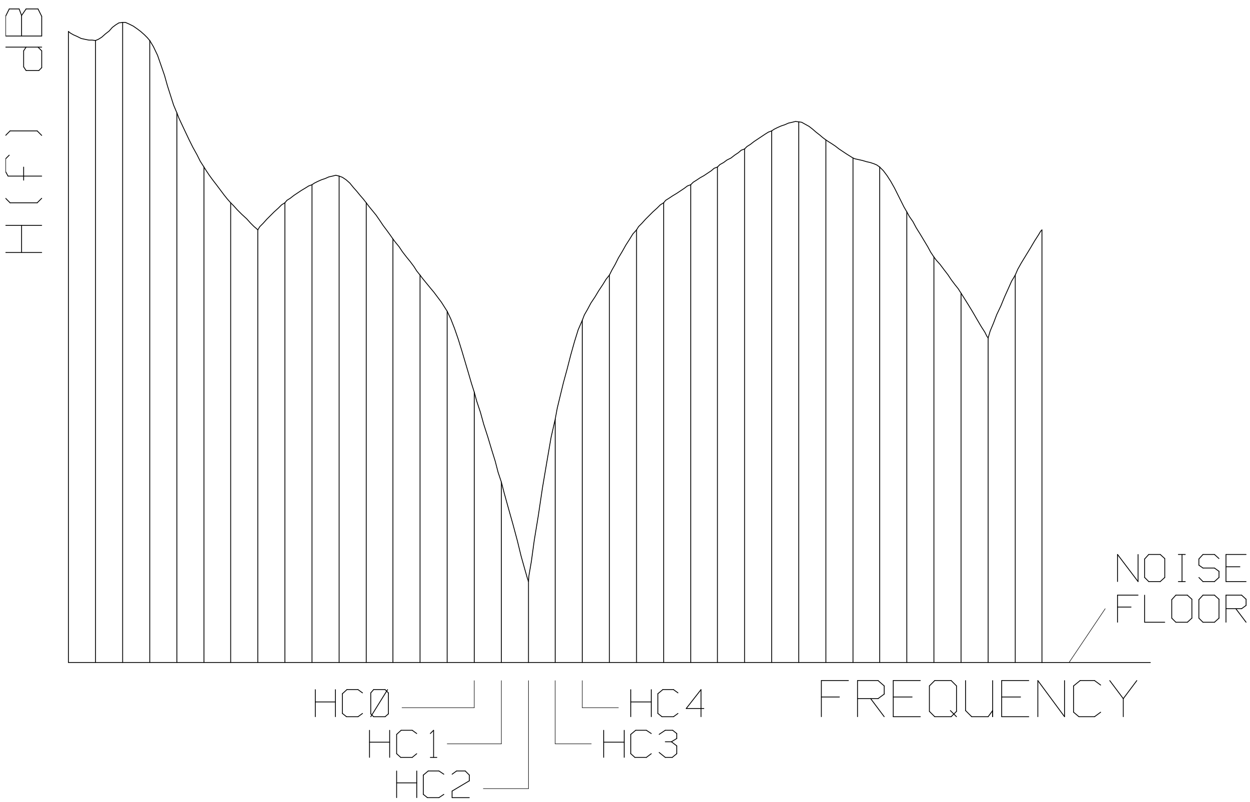

Figure 6 - A received OFDM Signal in the Frequency Domain

Figure 7 is a diagram of time vs. frequency showing common impairments. This diagram helps with the understanding of the effects of different impairments on different modulation types.

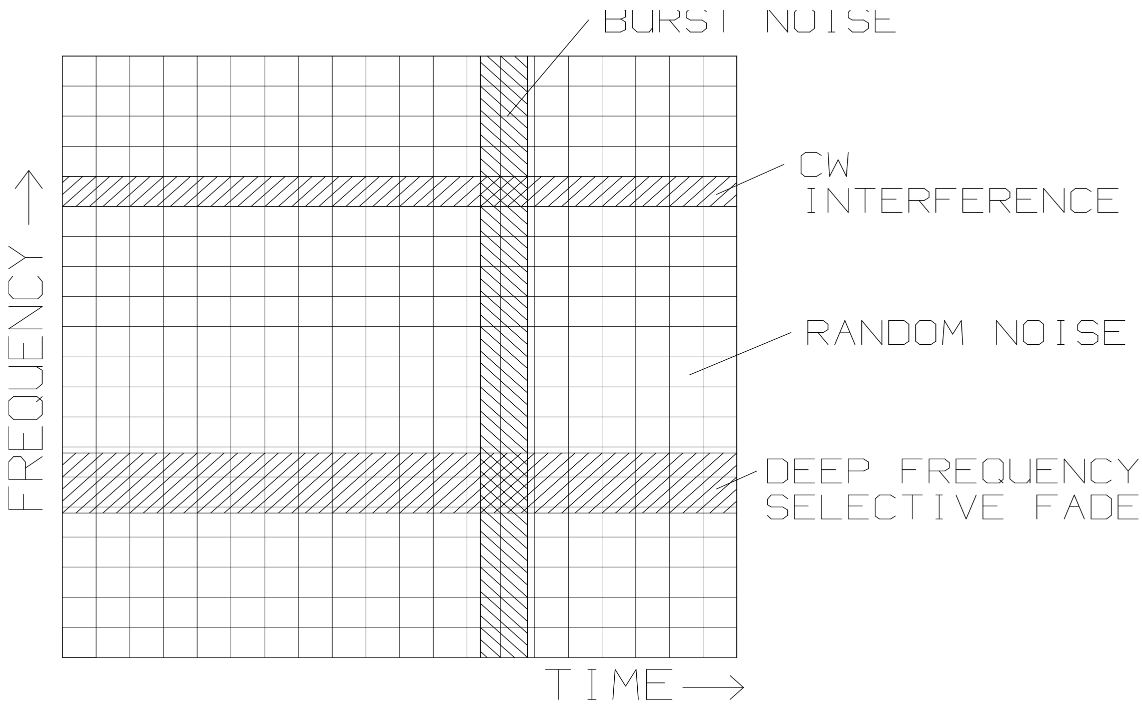

Figure 7 - Time and Frequency Relationships Between Common Impairments

Random thermal, or Gaussian, noise is present at all frequencies and all times, so there is no modulation technique that has a relative advantage in AWGN (additive white Gaussian noise). The Shannon Hartley Theorem states the maximum data capacity in a channel with AWGN [6]. If the noise is non-white (i.e., its spectrum is non-flat), the maximum capacity can be determined from the “water-pour” method of transmit power distribution. In cable systems, having a non-flat signal-to-noise ratio occurs due to cable loss varying with frequency, and nonlinear distortion products, which are random noise-like if the distortion was created by digital carriers.

Burst noise occurs locally in time, but often has a wide spectrum. Single carrier, with FEC, can be effective against burst noise for correcting corrupted TD (time domain) symbols. However, an OFDM receiver will perform a FFT on the sequence and spread the burst noise contamination to all FD (frequency domain) symbols.

A continuous wave (CW) interferer can be continuous in time but localized in frequency. OFDM with FEC can repair the localized damage to a limited number of subcarriers. However, with single carrier modulation the CW interferer will affect all symbols, turning a constellation point into a donut shape.

Likewise, a deep frequency-selective fade can be overcome by OFDM with FEC, as mentioned previously.

There are other important considerations for selecting a modulation technique for a particular radio frequency (RF) signal path, such as tolerance to frequency offsets and tolerance to phase noise (both of which increase the cost of local oscillators), and peak-to-average power ratio, which makes transmitters consume more power, decreasing battery life. Furthermore, receiver designers have a number of design tricks, or “secret sauces” to mitigate the effects of impairments, such as noise cancelers.

Let's make a signal to transmit from symbols:

$$E = \lfloor e_1, e_2, e_3, e_4 ......e_j \rfloor \quad \text{[eq. 2]}$$

Where E is the signal to be transmitted, and e_n are the individual component symbols.

We can formulate a modulation matrix, C, with rows and columns, where the rows are (hopefully) orthogonal to each other. *Each of the three modulation techniques described previously is nothing more than a different set of row functions, also known as orthogonal basis functions.* For single carrier, the modulation matrix is simply an identity matrix, with a single diagonal row of 1s and 0s elsewhere. For DSSS the rows may be from a Walsh matrix, with a single circular shift between rows. For OFDM, the rows may be complex exponentials (sine and cosine waves), where the first row is the first harmonic, and the second row is the second harmonic, etc. This is illustrated in Figure 8.

$$C = \begin{bmatrix} c(1,1) & c(1,2) & \bullet & \bullet & \bullet & c(1,k) \\ c(2,1) & c(2,2) & \bullet & \bullet & \bullet & c(2,k) \\ c(3,1) & \bullet & \bullet & \bullet & \bullet & c(3,k) \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ c(j,1) & c(j,2) & \bullet & \bullet & \bullet & c(j,k) \end{bmatrix} \quad \text{[eq. 3]}$$

The principle of orthogonality between rows is restated as:

$$\sum_{n=1}^{n=k} c(x,n) \cdot c(y,n) = 0 \quad \text{[eq. 4]}$$

For all rows where x ≠ y.

So, a signal for transmission, F, is made by simply multiplying the input sequence by the multiplication matrix, C:

$$F = E \cdot C = \lfloor f_1, f_2, f_3, f_4 ......f_j \rfloor \quad \text{[eq. 5]}$$

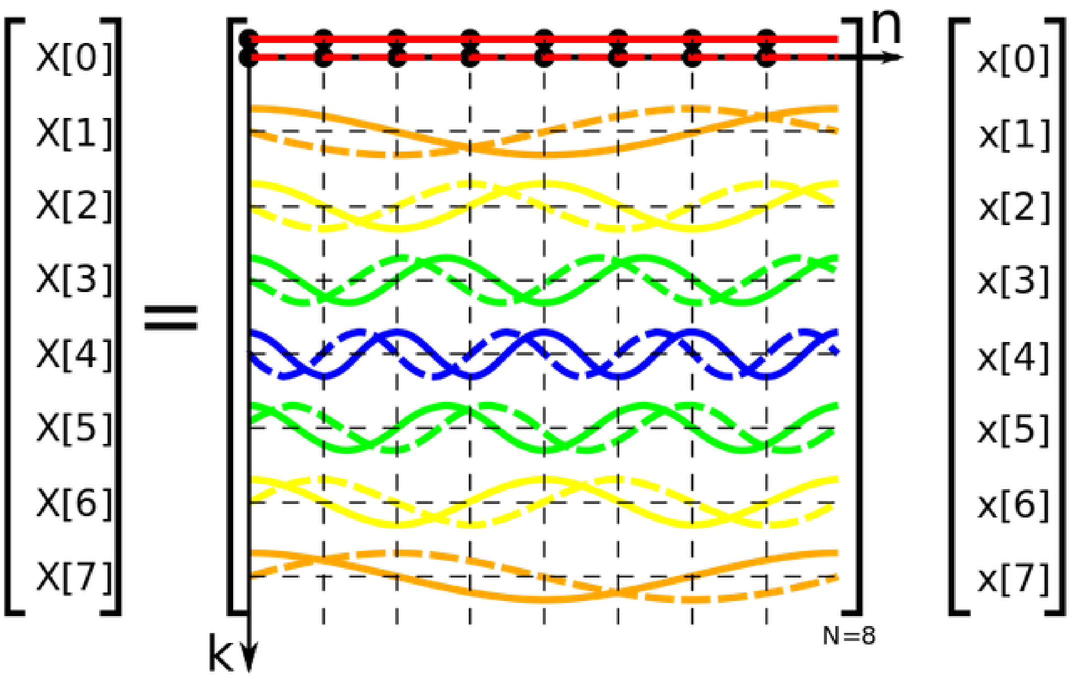

Figure 8 – Eight-Row Matrix as Sines And Cosines, Making OFDM Modulation

In Figure 8, cosine waves are solid and sine waves are dashed. X[1] forms an upper sideband and X[7] forms a matching lower sideband.

2. Rotation of Time and Frequency Axes

Delving deeper into the time and frequency relationships between symbols, single carrier modulation and multicarrier OFDM may be viewed as fundamentally a same modulation technique, apart from a 90-degree rotation in the time-frequency plot.

There is a duality between time and frequency that can be observed in discrete Fourier transform (DFT) and inverse discrete Fourier transform (IDFT) equations:

$$f[k] = \frac{1}{N}\sum_{n=0}^{N-1} F[n]e^{+j\frac{2\pi}{N}nk}$$

[eq. 6]

$$F[n] = \sum_{k=0}^{N-1} f[k]e^{-j\frac{2\pi}{N}nk}$$

[eq. 7]

The differences between equations are only scale factor and a negative sign in front of the complex exponential. The equations are almost the same. If you were shown a set of transform pairs, you could not identify which plot was time and which was frequency.

Figure 9 is a time-frequency plot. With a single carrier signal, such as a pulse amplitude modulation (PAM) signal, each symbol is very short in time, wide in bandwidth, and a next symbol occurs sequentially in time. An OFDM subcarrier is narrow in frequency and long in duration. Many OFDM subcarriers operate simultaneously in time. In Figure 9, there are 32 time symbols or 32 frequency symbols. By rotating a PAM transmission 90 degrees, you have an OFDM transmission, and vice-versa.

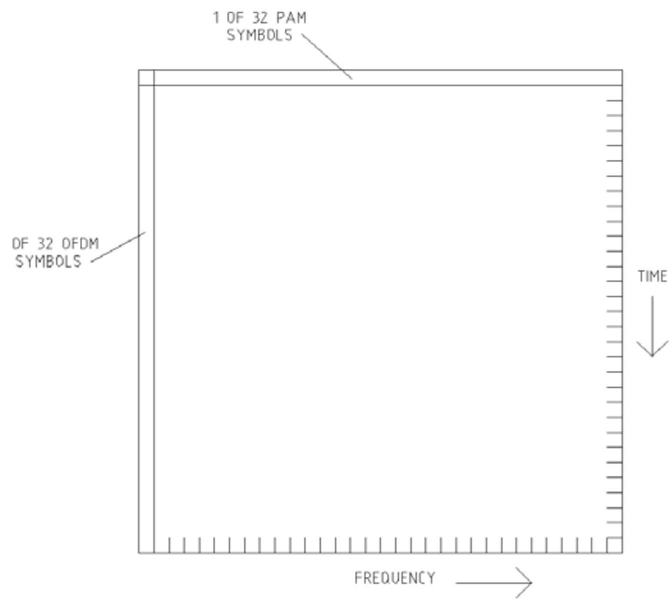

Figure 9 - 32x32 Block Time-Frequency Plot

Likewise, by rotating many TDMA single carrier sequential transmissions from several transmitters you have several OFDMA simultaneous transmissions from several transmitters. [2]

While OFDM (without TD tapering) has out-of-band energy splatter, this characteristic is analagous to PAM signals having a sine(x)/x response in time if the channel rolloff factor (alpha) is small or zero. So OFDM, using TD tapering, does the analagous operation as PAM using a rolloff factor, alpha. [3]

So a symbol could be observed to be 32x1 or 1x32. Note that dispersion, or other linear distortion, occurs along the time axis, but not the frequency axis. Dispersion along the frequency axis would be indicative of non-linear distortion.

Performing a 90 degree rotation on a sequence is nothing more than performing a FFT on a sequence, and a -90 degree rotation can be done with an inverse fast Fourier transform (IFFT) on the sequence.

This rotational view of modulation techniques reveals some interesting cases if studied. For example, conventional duobinary (partial response) transmission is well-known. [4] If it is rotated 90 degrees on a time-frequency axis, you can obtain a new transmission method: frequency domain (FD) duobinary, or duobinary OFDM. This is simulated in Figure 10, which is a screen shot from a digital oscilloscope with an internal FFT. The TD plot (on top) has an envelope shaped like a half-cosine, and the FD plot (created by the oscilloscope) is flat, just a time-frequency dual to conventional duobinary.

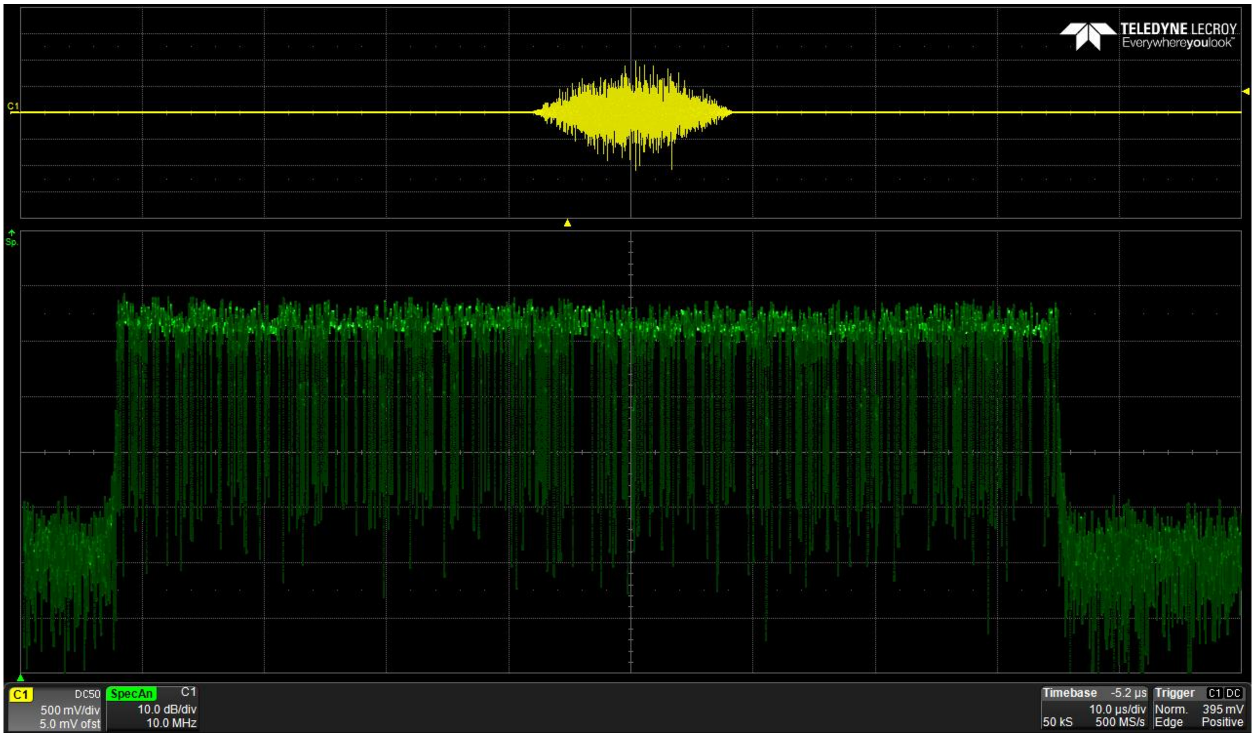

Figure 10 - FD Duobinary Block Transmission

If all symbols have a same magnitude, it has a desirable property of abrupt drop of energy out-of-band, naturally reducing interference with neighboring channels. Also, the burst's energy rises and falls gradually in time, causing less interblock interference if an uncorrected short echo is on the channel. On the negative side, the noise performance is poorer, due to being fundamentally duobinary, but that deficiency is ameloriated somewhat by getting more symbols per second within a given bandwidth. It also has a non-flat power vs. time which makes it less desirable for use with power-limited transmitters, such as cell phones. See Appendix A – Detail on Duobinary Modulation.

The desirable characteristic of a gentle rise and fall in transmit power level is a result of duobinary modulation summing each subcarrier with the one next to it, provided it has a same magnitude. This desirable characterisitc would be reduced if subcarriers with random or non-equal magnitudes were summed.

One possible application for duobinary OFDM is for very narrow bandwidth OFDM transmissions with a small number of subcarriers, such as ham radio, where the spectral splatter could cause adjacent channel interference. Another use is signaling with a small number of bits in a narrow bandwidth, such as "acks" or acknowledgements.

Conclusion

For three common modulation methods, using orthogonal basis functions can be considered to be the same process as matrix multiplication, where the rows are orthogonal to each other. Likewise, by relabeling the time and frequency axes, single carrier modulation and OFDM multicarrier modulation are comparable.

Abbreviations

ATDMA	advanced time division multiple access
AWGN	additive white Gaussian noise
BPSK	binary phase shift keying
CDMA	code division multiple access
CP	cyclic prefix (see also GI)
CW	continuous wave
dB	decibel
DFT	discrete Fourier transform
DOCSIS	Data-Over-Cable Service Interface Specifications
DSSS	direct sequence spread spectrum
FD	frequency domain
FEC	forward error correction
FFT	fast Fourier transform
GI	guard interval (see also CP)
I	in-phase
IDFT	inverse discrete Fourier transform
IFFT	inverse fast Fourier transform
PAM	pulse amplitude modulation
PN	pseudo-noise
PRS	partial response signaling
Q	quadrature
QAM	quadrature amplitude modulation
OFDM	orthogonal frequency division multiplex
OFDMA	orthogonal frequency division multiple access
PRS	partial response signaling
QPSK	quadrature phase shift keying
RF	radio frequency
SC	single carrier
S-CDMA	synchronous code division multiple access
SCTE	Society of Cable Telecommunications Engineers
TD	time domain

Bibliography & References

[1] DOCSIS 2.0 Radio Frequency Interface Specification, CM-SP-RFIv2.0-C02-090422, page 71

[2] US patent 5,815,488

[3] Digital Telephony, 3rd Edition, John Wiley and Sons, by John Bellamy, Appendix C.

[4] Digital Telephony, 3rd Edition, John Wiley and Sons, by John Bellamy, Appendix C, pp. 185-188, and pp. 310-311.

[5] Digital Telephony, 3rd Edition, John Wiley and Sons, by John Bellamy, Appendix C, pp. 587-592

[6] Communications in the Presence of Noise, by Claude Shannon, reprinted Proceedings of the IEEE, VOL 26, NO. 2, 1998

Appendix A – Detail on Duobinary Modulation

A basic modulation technique, such a BPSK (binary phase shift keying) can be created by connecting a periodic series of positive or negative impulses to a lowpass filter having a sine(x)/x impulse response, as illustrated in the top left side of Figure 11 (and in Figure 1). This produces a raised cosine frequency response on the modulated signal on the top right side of Figure 11. The abruptness of the FD roll-off is a factor commonly called “alpha,” and depends on damping applied to the sine(x)/x waveform. Duobinary modulation is simply a different impulse response. The impulse response lasts over two symbol periods, not one, as illustrated on the lower left side of Figure 11. In the frequency domain the response is cosine, not raised cosine, and is illustrated in the lower right side of Figure 11.

Passing a two-level complex (I and Q) signal through a duobinary filter produces a 9-PRS (partial response signaling) signal as illustrated in Figure 12. To compare the power of the two signals, the quadrature phase shift keying (QPSK) signal has four equally probable states. The 9-PRS signal has a single state in the middle with a probability of 0.25, four high power corner states with a combined probability of 0.25, and four intermediate power levels for a combined probability of 0.5.

If the voltage difference between A and B is assumed to be 1.0, the power of the 9-PRS constellation is .25*0 + .5*1.0 + .25*1.414^2 = 1 watt. If the voltage difference on the QPSK constellation is set to be 0.707 between points C and D, the QPSK power is also 1 watt. So, a noise vector required to make a slicing error on the QPSK signal is .707 volt, and 0.5 volt on the 9-PRS signal, a difference of 3dB.

Looking at the spectrum in Figure 10, observe that three power levels are visible, the peak subcarrier level, an intermediate subcarrier level, and a zero-power subcarrier level, where the energy drops to the origin.

For conventional TD duobinary, the occupied bandwidth of a QPSK signal, relative to 9-PRS, is greater by the channel roll-off factor, commonly called alpha. [6] For DOCSIS single carrier modulations, the value is around 5% more for the downstream and 25% more for the upstream. For FD duobinary, if the number of subcarriers is the same, with equal subcarrier spacing, the occupied bandwidth will be the same. The bandwidth advantage for FD duobinary is a more abrupt drop of energy out of band, allowing closer carrier spacing.

Other candidate modulations for duobinary OFDM are discussed in ref. [4].

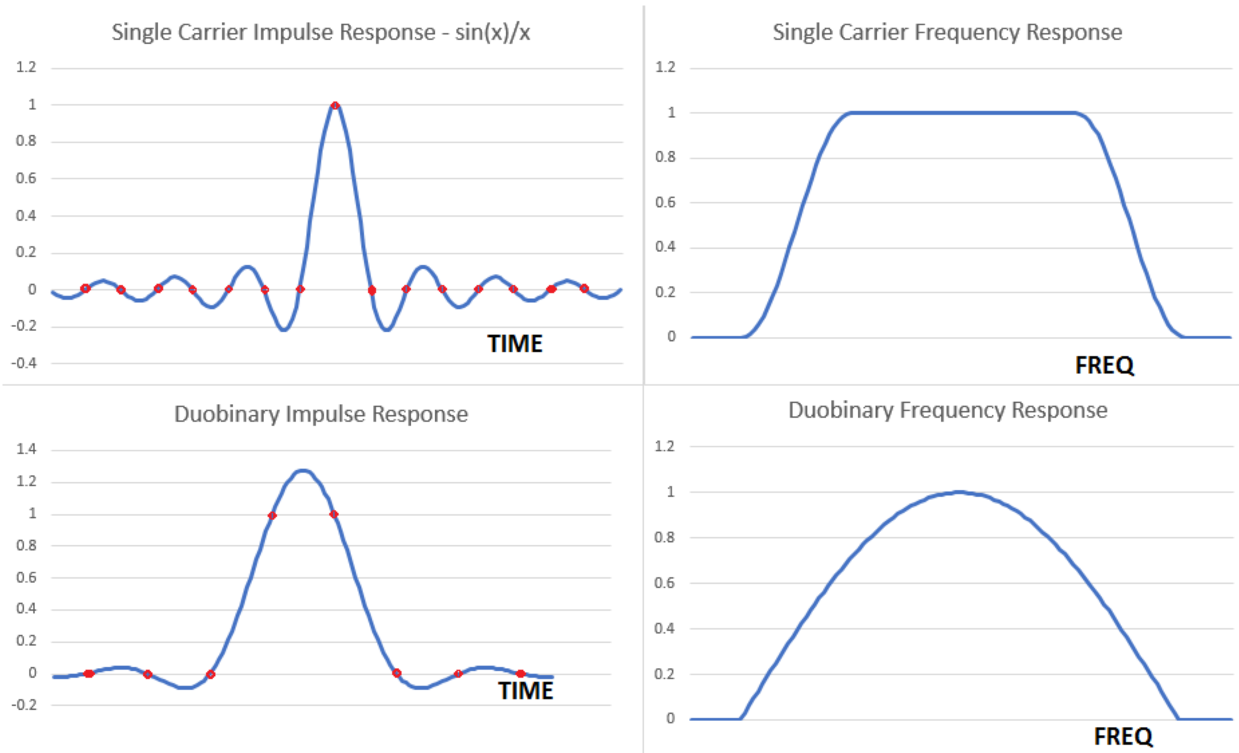

Figure 11 - Impulse and Spectral Responses Comparison of Single Carrier (QPSK) and Duobinary (9-PRS)

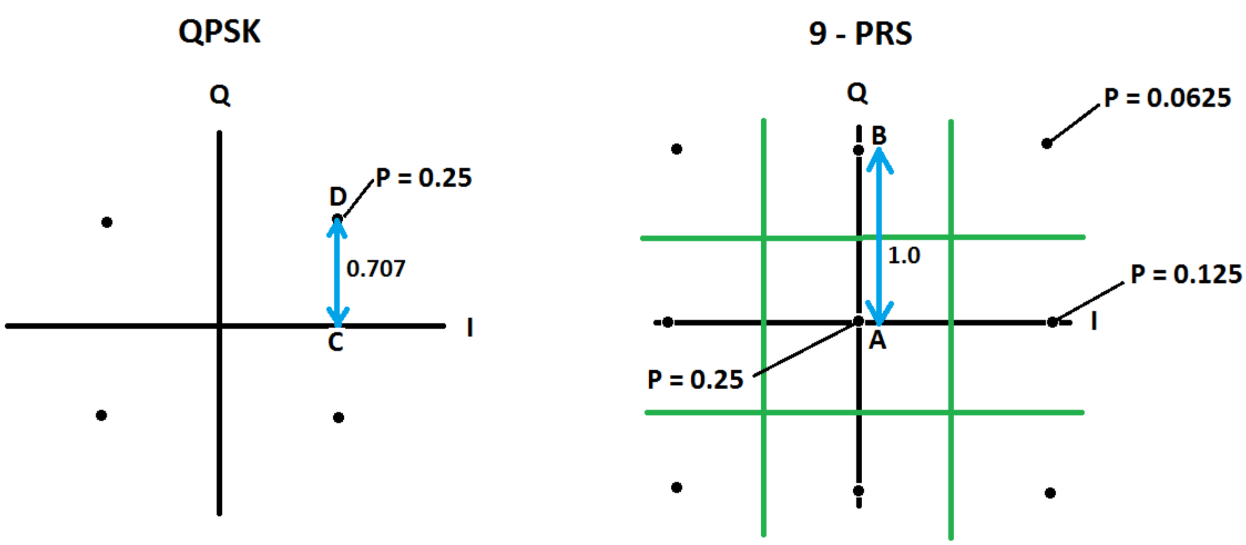

Figure 12 - Relative Power Calculations for Error Thresholds

Figure Note: Both signals have the same RF power. On duobinary, not all states are equally probable.

Footnotes:

1. For complex numbers, this would be Sum x*conj(y)

2. It is also possible to use non-orthogonal signals for communications. For example, a non-orthogonal spread-spectrum signal can work in the presence of other signals by taking advantage of spreading gain. In this case, interference is non-zero, but hopefully tolerable.

3. It is also useful to view equalization as another matrix multiply.

4. If the sine-shaped TD duobinary signal illustrated as the top trace in Figure 9 had a cyclic prefix inserted, the waveform would have a fish outline. That is, the shape would gain a tail.