Multicarrier Modulation for Data Transmission: An Idea Whose Time Has Come

John A. C. Bingham

HE PRINCIPLE OF TRANSMITTING DATA BY dividing it into several interleaved bit streams, and using these to modulate several carriers, was used more than 30 years ago in the Collins Kineplex system [1], and has been of continuing, albeit peripheral, interest ever since. Now, however, interest is increasing because modems based on the principle are being used—or being considered for use—for transmission of data and facsimile on the following:

- General Switched Telephone Network (GSTN)
- 60–108 kHz Frequency-Division Multiplexed (FDM) group-band
- Cellular radio

In addition, high-speed data is being considered for transmission on the High-rate Digital Subscriber Line (HDSL).

The technique has been called by many names—orthogonally multiplexed Quadrature Amplitude Modulation (QAM) [2], orthogonal FDM [3], and dynamically assigned multiple QAM [4]—but we will refer to it by a generic name: Multicarrier Modulation (MCM). A more general form of the technique, which uses more complex signals as carriers [5], has been developed recently as vector coding [6] and structured channel signalling [7] [8]. Unless otherwise stated, the discussion here will concentrate on the special MCM form.

The reasons for the interest in MCM depend upon the transmission medium, and have also changed over the years as signal processing techniques (mainly digital) have improved, but the two most important ones are first, that an MCM signal can be processed in a receiver without the enhancement (by as much as 8 dB in some media) of noise or interference that is caused by linear equalization of a single-carrier signal, and second, that the long symbol time used in MCM produces a much greater immunity to impulse noise and fast fades.

The first seven sections of this article will discuss the following: the general technique of parallel transmission on many carriers; the performance that can be achieved on an undistorted channel; algorithms for achieving that performance; dealing with channel impairments; improving the performance through coding; and methods of implementation. The last two sections discuss duplex operation of MCM and the possible use of this on the GSTN.

Multiplexing

MCM is a form of FDM; the basic principle is shown in Figure 1. Input data at Mf_s b/s are grouped into blocks of M bits at a block ("symbol") rate of f_s. The M bits are used, m_n bits[1] for the carrier at $f_{c,n}$ to modulate N_c carriers, which are spaced Δf apart across any usable frequency band; that is,

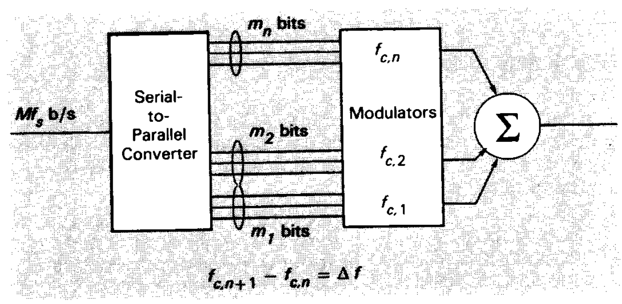


Fig. 1. Basic multicarrier transmitter.

$$f_{c,n} = n\,\Delta f \text{ for } n = n_1 \text{ to } n_2 \tag{1}$$

and

$$M = \sum_{n=n_1}^{n_2} m_n$$

where

$$N_c = n_2 - n_1 + 1$$

The modulated carriers are summed for transmission, and must be separated in the receiver before demodulation. Three methods have been used for this separation:

- First, the earliest MCM modems borrowed from conventional FDM technology, and used filters to completely separate the bands. The transmitted power spectra for just three sub-bands of a multicarrier system are shown in Figure 2a.

[1]Each of the ms typically = 2 to 8.

0163-6804/90/0005-0005 $01.00 © 1990 IEEE

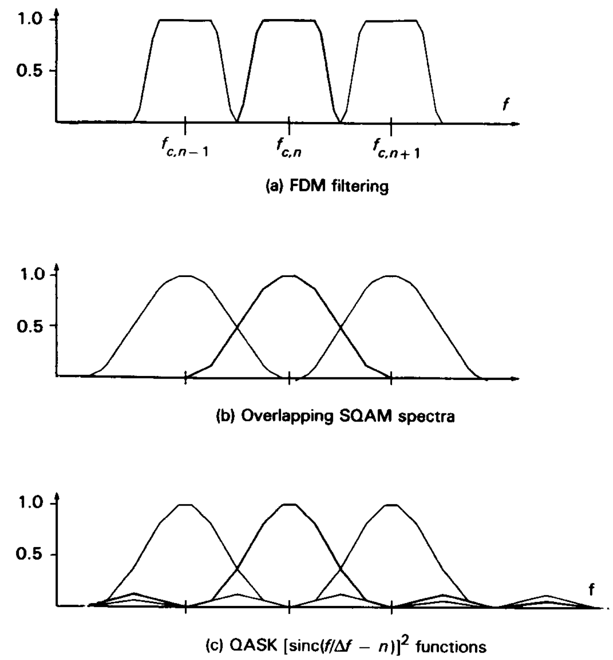


Fig. 2. MCM transmit power spectra.

Because of the difficulty of implementing very sharp filters, each of the signals must use a bandwidth, $(1 + \alpha)f_s$, which is greater than the Nyquist minimum, f_s; the efficiency of band usage is $f_s/\Delta f = 1/(1 + \alpha)$.

- Second [9–13] the efficiency of band usage was increased to almost 100% by using Staggered Quadrature Amplitude Modulation (SQAM); the individual transmit spectra of the modulated carriers still use an excess bandwidth of α, but they overlap at the -3 dB frequencies (as shown in Figure 2b), and the composite spectrum is flat. If $\alpha \leq 1$, each sub-band overlaps only its immediate neighbors, and orthogonality of the sub-bands—with resultant separability in the receiver—is achieved by staggering the data (that is, offsetting it by half a symbol period) on alternate in-phase and quadrature sub-channels. The amount of filtering required is less than for complete separation, but it is still considerable, and the total number of carriers must be small (typically less than 20).
- Third [2] [4] [14–16], the carriers are "keyed" by the data, using Quadrature Amplitude Shift Keying (QASK). The individual spectra are now sinc functions, as shown in Figure 2c; they are not bandlimited but, as we shall see, the signals can still be separated in the receiver; the frequency-division is achieved, not by bandpass filtering, but by baseband processing. The big advantage of this approach is that both transmitter and receiver can be implemented using efficient Fast Fourier Transform (FFT) techniques.

Maximum Achievable Bit Rate: Seeking the Shannongri-la of Data Transmission

The performance of a data transmission system is usually analyzed and measured in terms of the probability of error at a given bit rate and Signal-to-Noise Ratio (SNR). It is, however, more useful for our purpose—and, indeed, more appropriate for modern data communication systems that use any combination of compression, error correction, and flow control—to consider the attainable bit rate at a given error rate and SNR. For single-carrier signals that are equalized with either a Linear Equalizer (LE) or a Decision-Feedback Equalizer (DFE) this can be done by inverting the well-known error rate formulas (e.g., those for LEs [17] [18] and DFEs [3]).

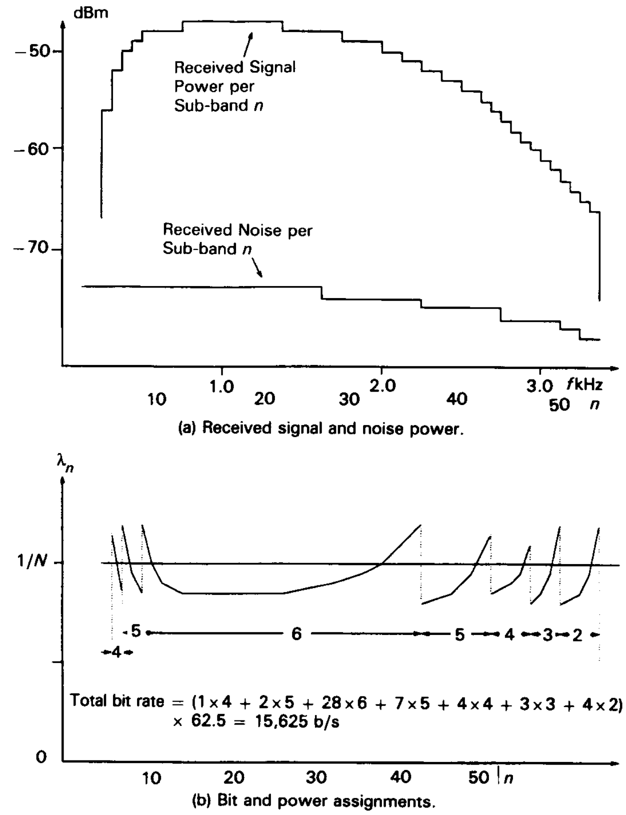


Fig. 3. Adaptive loading for a badly distorted GSTN channel.

The variables for a multicarrier signal are the number of bits per symbol, m_n, and the proportion, γ_n, of the total transmitted power, P, that are allotted to each sub-band. The aggregate bit rate is approximately maximized if these variables are chosen so that the bit error rates in all the sub-bands are equal. This has not been proved rigorously, but it is intuitively reasonable; the dependence of error rates on the m_n and γ_n is such that if the error rates are unbalanced, the rate in one band will increase much more than it will decrease in another band.

In order to calculate the attainable bit rate for a channel with transfer function $H(f)$ and noise power spectrum at the input to the receiver $U(f)$,[2] we can approximate $H(f)$ and $U(f)$ by segments H_n and U_n centered about carrier frequencies $f_{c,n}$, as defined in Equation (1). This is illustrated in Figure 3a for a badly distorted and noisy voiceband channel with $f = 62.5$ Hz;[3] the signal power received in each sub-band is calculated assuming that the total transmit power of -9 dBm is distributed equally across the sub-bands (i.e., if all the γ_n were equal); the total noise power in the 0.3 to 3.4 kHz band is -57 dBm.

The probability of bit error, $\mathcal{P}$, in the symbol-by-symbol detection (i.e., without the benefit of any coding across symbols)

[2] The possible non-whiteness of the "noise" is important for HDSL, where the principal impairment is strongly correlated Near-End Cross-Talk (NEXT).

[3] This is one of the carrier separations used in Telebit's "Trailblazer" modem; the reason for such a choice (62.5 = 8,000/128) will become clear later.

of the QAM signal in sub-band n—assuming no interference from the signals in the other bands—is

$$\mathcal{P}_n = K\beta Q\left\{\left|\frac{3}{Ln^2-1}\ \frac{\gamma_n P|H_n|^2}{U_n}\right|^{1/2}\right\} \qquad (2)$$

where

$$L_n^{\ 2} = 2^{m}n,$$

$$\beta = 4(1 - 1/L)/m_n,$$

and K is an error-rate multiplier, which is a little less than 6 if, as is most usual, differential phase modulation and a 3-tap scrambler are used. Q is defined, as usual, by

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} \exp(-y^2/2)\,dy, \qquad (3)$$

$\mathcal{P}$ is the total transmitted power, and γ_n is the proportion of that total allotted to sub-band n.

We would like to solve Equation (2) for m_n, but this cannot be done explicitly because m_n occurs in three places on the righthand side. Kalet [19] developed upper and lower bounds for the symbol error rate by considering the limits of $4(1 - 1/L)$, but it is adequate for our purpose[4] to consider only an average value of β. For a practical range of m_n from 2 to 8 β varies from 1 to 15/32, so an average value of 4 for the combined error-rate multiplier, βK, will suffice. Then, as shown in [19], Equation (2) can be inverted, and the total number of bits that can be transmitted in one symbol with error probability $\mathcal{P}$ using N_c sub-bands can be written:

$$M = \sum_{n=n_1}^{n_2} \log_2\left|1 + \frac{3}{[Q^{-1}(\mathcal{P}/4)]^2}\ \frac{\gamma_n P|H_n|^2}{U_n}\right| \qquad (4)$$

where

$$\sum_{n=n_1}^{n_2} \gamma_n = 1$$

Ideally, the optimum power distribution, γ_n, should be calculated by a "water-pouring" procedure that is similar to that of Gallager [20], but for high SNRs (corresponding to most acceptable error rates), the optimum γ_n are approximately equal. The most efficient use is made of the channel if the symbol rate, f_s, is made equal to the carrier separation, Δf, and both are made very small. Then the summation in Equation (4) can be approximated by an integration, and the maximum bit rate

$$B_{tot} = \underset{\Delta f \to 0}{Lt}\ [Mf_s] \qquad (5)$$

$$= \int_{f_1}^{f_u} \log_2\left[1 + \frac{3}{[Q^{-1}(\mathcal{P}/4)]^2}\ \frac{P|H(f)|^2}{WU(f)}\right] df$$

where the frequency range, f_l to f_u, is that for which the integrand is > 2 (i.e., the range over which QAM transmission is possible), and $W\ (= f_u - f_l)$ is the measure of that range.

As pointed out by Kalet and Zervos [3], Equation (5) is very similar to the bit rate for a Single-Carrier QAM (SCQAM) signal equalized by a DFE, which was originally shown by Price [18]. In fact, the only difference is in the frequency range of the integration; for the single-carrier signal with DFE it should be extended to that for which the integrand is greater than zero, but in practice the extra contribution to the integral is usually insignificant.

It should be noted that Equation (5) assumes that the number of bits per carrier is continuously variable but, in practice, each m_n must be integer.[5] It was shown in [17] that the effects of this quantizing can be mitigated by adjusting the γ_n to re-equalize the error rates in all the sub-bands, and it has been found from numerous simulations that the total bit rate achieved in this way is only slightly less than that given by Equation (5).

Thus, the aggregate bit rate for MCM is approximately equal to that for SCQAM/DFE; for channels with attenuation distortion or non-white noise this may be considerably greater than for SCQAM with a linear equalizer.

Adaptive Loading

It was shown that if the ratio $|H(f)|^2/U(f)$ varies significantly across the band and a fixed loading is used [21], the error rate in the too-heavily-loaded sub-bands may be very high, and the overall error rate may be greater than for a single-carrier signal [17]! The m_n must be varied in order to keep all the sub-band error rates, $\mathcal{P}_n$, equal; the following procedure for calculating the γ_n and integer m_n was described [16].

Given a set of signal-to-"noise"[6] ratios, measured in the receiver when the far transmitter is transmitting at the maximum permitted level in all sub-bands, calculate the terms, $\Delta P_{m,n}$ of an "incremental power" matrix, where $\Delta P_{m,n} = P_{m,n} - P_{m-1,n}$ ($P_{m,n}$ = the transmit power needed in sub-band n to transfer m bits per symbol at some predefined error rate), and clearly, $P_{0,n} = 0$.

Then assign bits one at a time to carriers, each time choosing the carrier that requires the least incremental power. This can be described algorithmically:

- Search row 1 for the smallest $\Delta P_{i,n}$
- Assign one more bit to sub-band n
- Increment M and P_{tot}; that is,
  $M' = M + 1$ and $P_{tot}' = P_{tot} + \Delta P_{i,n}$

[4]Equation (1) is exact only for square constellations (i.e., m_n even) anyway. For $m_n = 5$ and 7, the "cross" constellations are slightly more efficient, and P is slightly lower; for $m_n = 3$ all constellations are less efficient, and P is significantly higher.

[5]Coding schemes to allow non-integer m_n have been discussed for use on the DSL, but it is not clear how much they would increase the capacity.

[6]The equivalent noise should be the power sum of Gaussian noise, NEXT, and inter-symbol and inter-channel interferences.

- Move all terms of column n up one place; that is, $\Delta P_{i,n}' = \Delta P_{i+1,n}$
- Repeat search

For the preferred mode of operation for multicarrier—at the highest rate achievable with a predefined error rate—the assignment should be stopped when P_{tot} just exceeds P, the available power. If, however, transmission at a given bit rate (a synchronous "bit pump") is insisted upon, then the process should be stopped at the appropriate value of M. P_{tot} may then be less or more than permitted (that is, the specified error rate was pessimistic or optimistic, respectively); all allotted powers must be scaled to adjust P_{tot} to the correct value.

The resulting power distribution for the channel of Figure 2a is shown in Figure 2b. The discontinuities occur because of the integer constraint on the number of bits; if Δf is small, then the SNR can change only slightly from one sub-band to the next, so that if, for example, the SNR is decreasing, and $m_n = m_{n-1} - 1$, the nth carrier will require approximately 3 dB less power than the $(n-1)$th carrier for the same error rate. The algorithm is clearly not water-pouring in the classical sense, but since it puts every increment of transmit power where it will be most effective, it appears to be optimum for multicarrier transmission using QAM constellations and symbol-by-symbol detection.

Feedback from Receiver to Transmitter

Adaptive loading requires that the receiver measure the sub-band SNRs, calculate the best power and bit assignments, and send this information back to the transmitter. This may seem like a big increase in complexity, but it should be noted that all single-carrier systems that make best use of a channel also require some feedback. This can be used in three different ways:

- Many present fixed-symbol-rate systems use a "fall-back" procedure that requires the feedback of error-rate information.
- Better use of a channel might be made by calculating and feeding back an optimum symbol rate, and then using some form of Maximum Likelihood Sequence Estimation in the receiver.
- Another approach is to combine trellis coding with an adaptive symbol-rate and a DFE. A conventional DFE cannot be used, however, because of error propagation, and the function of the feedback part of the DFE must be implemented in the transmitter using a generalization of Tomlinson precoding; this requires the feedback of much the same detailed channel characteristics as are needed for MCM.

Adaptive Loading When NEXT is the Dominant Impairment

For high-speed transmission on the subscriber loop, NEXT is usually more harmful than noise. If this NEXT is mainly from other MCM transmitters, a unilateral decision to change the spectral distribution of one transmitted signal would change the conditions under which the other transmitters make their decisions; clearly some coordinated strategy for assigning all the sub-band powers is needed. Work is being done on this but it is too early to predict the results.

Modulation and Demodulation

Modulation is performed on M bits (a symbol or block) of data at a time—preferably using an Inverse FFT (IFFT)—and samples of the transmit signal are generated at a sampling rate, f_{samp}. For greatest efficiency f_{samp} should be equal to Δf multiplied by an integer power of two. If $f_{samp} = 2N_{tot}\,\Delta f$, then N_{tot} carriers are available for modulation, but the channel will usually be such that only N_c carriers can be used. If these are at frequencies $n_1\Delta f$ to $n_2\Delta f$, as defined in Equation (1), modulation of a total of M bits, m_n at a time, is most easily accomplished by calculating N_c complex numbers (each selected from a constellation with 2^{m_n} points), augmenting them with $n_1 - 1$ zeros in front and $N_{tot} - n_2$ zeros behind, and performing an N_{tot}-point IFFT.

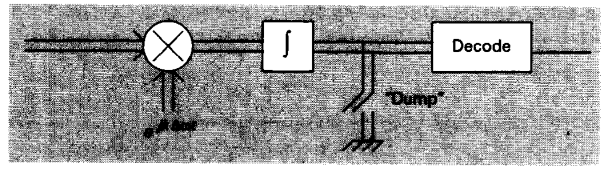


Fig. 4. Integrate and dump detection for QASK.

Modulation via an IFFT is equivalent to multicarrier QASK in which the fundamental baseband pulse shape is a rectangle, g(t). That is,

$$g_n(t) = 1/T \text{ for } 0 \le t < T, \text{ and } = 0 \text{ otherwise.} \qquad (6)$$

In the receiver the signal is demodulated by assembling N_{tot} samples into a block, and performing a real-to-complex FFT. This is equivalent to demodulating each sub-band separately, and then doing an integrate-and-dump on each product, as shown in Figure 4. If the received baseband pulse in sub-band n is defined as $g_n'(t)$, then the output from the demodulator resulting from an input to another sub-band $(n-k)$ is $g_n'(t)$ multiplied by a cosine or sine wave of the difference frequency $k\Delta f$; that is,

$$h_{n,n-k}(i) = \int_{iT}^{(i+1)T} g_n'(t)\,.\exp(jk2\pi\Delta f t)\;dt. \qquad (7)$$

If the channel is non-distorting, so that $g_n(t) = g_n'(t) = 1/T$, then these integrals over a time $1/\Delta f$ are zero for all non-zero k. That is,

$$h_{n,n-k}(i) = 1 \text{ for } i = k = 0\text{, and } = 0 \text{ otherwise,} \qquad (8)$$

and orthogonality between the sub-bands is maintained.

Correcting for the Effects of Channel Impairments

Linear Distortion

The primary effect of attenuation and/or delay distortion in the channel is that each subcarrier is received with a different amplitude and/or phase, so that the channel can be grossly characterized by a single complex number for each sub-band. These are learned from a training signal of unmodulated carriers (a "comb"), and inverted to generate the complex coefficients of a set of one-tap equalizers. All subsequent received samples are then multiplied by these inverses.

A secondary effect is that $g_n'(t)$ is not rectangular, and also overlaps into the preceding and following symbol periods. Moreover, even with an undistorted—but necessarily band-limited—channel, the sub-bands near the ends of the band are asymmetrical, and distort their g_ns. Thus, there is both Inter-Channel Interference (ICI) $(h_{n,n-k}(0) \neq 0)$, and Inter-Symbol Interference (ISI) $(h_{n,n}(\pm 1) \neq 0)$, and even the combination of the two $(h_{n,n-k}(\pm 1) \neq 0)$; orthogonality of the sub-bands is lost.

It can be seen that the impulse response of each sub-band depends only on the channel, and that the transient at the beginning and end of each $g_n'(t)$ is independent of the separation of the carriers (that is, of the symbol period, T). One way of dealing with distortion would be to increase T enough that distortion becomes insignificant, but in general this is not possible.[7] Four other ways have been described; these are discussed below.

Guard-Period

The transients in the $g_n'(t)$ can be avoided [1] [14] [22] by postponing the integration in Equation (5) for a time T_g, and increasing the total symbol time to $T_s = T + T_g$, while still, of course, retaining $T = 1/\Delta f$. One commercial modem for the GSTN [4] uses $T = 128$ ms and $T_g = 7$ ms. This limits the MSE from ISI and ICI on even the worst lines to less than 1%, but it does reduce the total bit rate by 5.2%.

Passband Channel Equalization

The reduction in bit rate caused by the use of a guard-period can be avoided by linearly equalizing the received signal. Because of the reduction of MSE achieved by integrating over a long symbol period, the equalizer can be much less complex than that for SCQAM; furthermore, it may be acceptable in some media to adapt it only during training, and freeze it during data reception.

(It should be noted that although the signal is being linearly equalized, this approach does not incur the large noise-enhancement penalty of single-carrier modulation. The loading is calculated from, and the performance determined by, the sub-band SNRs, which are reduced only slightly by the amplitude equalization across each sub-band; the equalization across the full band acts mainly like a delay equalizer plus many separate Automatic Gain Controls, or AGCs.)

The conclusion that can be drawn from [23] is that for such a simple equalizer, a Tapped Delay Line (TDL) structure using time-domain convolution is the most efficient. The training signal for this should be an unmodulated subset of the carriers, and the taps could be calculated either iteratively, by a conventional Least Mean Square (LMS) algorithm that takes advantage of the cyclic nature of the signal, or by performing an FFT of the signal to calculate the channel characteristics, inverting these, and performing an IFFT to calculate the taps.[8]

The optimum lengths of the data symbol and the TDL are a subject for further investigation. Clearly, as the length of the symbol is reduced, the effects of ISI and ICI become relatively more important, and the complexity of the equalizer must be increased. The limit of this would be reached when the equalizer had $2N_c$ parameters, and, since it would then equalize the channel response to all N_c carriers, it could also take over the role of the one-tap complex baseband equalizers.

Baseband Equalization

The ICI terms defined by setting $i = 0$ in Equation (6) form an $N_c \times N_c$ matrix, with the terms off the main diagonal decreasing only very slowly (approximately as $1/k$). This would require an extremely complicated equalizer, and baseband equalization is not used for QASK signals. It can be used, however, for SQAM signals [13], because each sub-band is filtered so as to limit interference to the two adjacent bands; the ICI matrix then has terms only on the main and two adjacent diagonals.

[7] The DSP memory, the processing requirements (proportional to T and $\log_2 T$, respectively), and the delay through the modem all become prohibitive.

[8] This is typical of the judicious mixture of frequency- and time-domain processing that is used in MCM. See [23] for a discussion of the trade-offs, and for more references on frequency-domain processing.

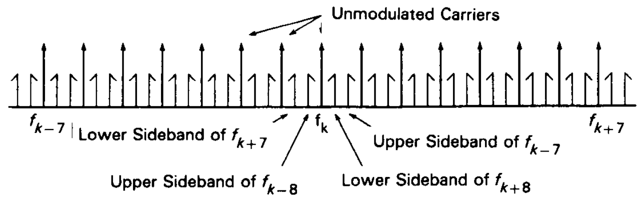


Fig. 5. Multicarrier spectrum with sidebands resulting from 60 Hz phase jitter.

Vector Coding, Structured Channel Signaling

Holsinger [5] showed that orthogonality of the sub-band signals through a distorted channel can be achieved by using, as "carriers," the eigenvectors of the auto-correlation matrix. This approach is presently attracting considerable interest [6–8], but it is too soon to know whether it can compete in computational efficiency with passband equalization.

Combination of Different Methods

The above methods are not mutually exclusive, and it is likely that some combination will provide the best compromise between amount of computation and total bit rate; passband equalization with a very short guard-space ($T_g/T \simeq 1$ to 2%) seems to be a very promising combination.

Phase Jitter

Phase jitter affects MCM and SCQAM quite differently. If a composite signal of unmodulated carriers is subjected to phase jitter of frequency f_j and amplitude less than about 10°, then each carrier at $n\Delta f$ will generate just two significant sidebands at $n\Delta f + f_j$. The carriers and their sidebands are shown in Figure 5 for the case where $f_j/\Delta f = 7.68$[9].

Both detection methods in the receiver—an FFT or demodulation followed by an integrate and dump—result in equivalent filter shapings of sinc functions centered at the carrier frequencies.

It can be seen, therefore, that the sidebands of at least two other carriers[10] contribute to the distortion seen by any given carrier. Since the data modulated onto these other carriers is uncorrelated with that on the carrier under consideration, the jitter is seen as random distortion about each point in the constellation, as shown in Figure 6a. That is, the jitter power (the total power in all the sidebands) is spread evenly over all carriers and over all data patterns on those carriers, and it can be added to the noise on a power basis.

In contrast, a single-carrier constellation is rotated by the jitter, as shown in Figure 6b; the outer points are clearly more susceptible, and the overall effect upon the error rate with added noise will be greater than for MCM.

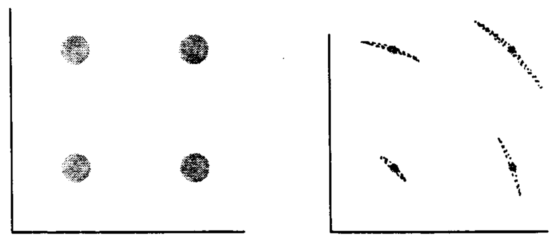
Tracking Phase Jitter

Although the effects of phase jitter are less for MCM than they are for SCQAM, they should not be ignored; identifiable, discrete components of jitter should be tracked. Identification is easier in a multicarrier receiver because much of the signal processing involves FFTs, but tracking is harder because of the long symbol period.

One method [24] processes one complete symbol to calculate the remanent phase error (the difference between the input

[9] $f = 7.8125$ Hz is the preferred carrier separation in the Trailblazer, and $f_j = 60$ Hz, the most common jitter frequency in the U.S.

[10] The number of contributing carriers reduces to two in the special case of $f_j/\Delta f$ being an integer.



(a) Multicarrier. (b) Single carrier.

Fig. 6. Effects of phase jitter on one quadrant of a 16-point constellation.

phase and the locally generated tracking phase), passes the error through narrow-band feedback filters as described in [17], and uses the outputs to update a phase predictor which generates the tracking phase for the next symbol. It has been found that discrete jitter components can be tracked almost perfectly.

Non-Linear Distortion

A multicarrier signal is the sum of many independent modulated sinewaves, and its sampled amplitude has an almost Gaussian distribution. Therefore, its peak-to-average ratio is much higher than that of SCQAM, and it is more susceptible to non-linear distortion. The most severe component of this is usually a negative cubic term ("saturation"), and it appears that if this can be quantified it can be, at least partially, corrected in the receiver by operating on the samples with a complementary nonlinearity.

Impulse Noise

Because a multicarrier signal is integrated over a long symbol period, the effects of impulse noise are much less than for SCQAM; indeed, this was one of the original motivations for MCM [25]. Tests reported to the Consultative Committee for International Telephone and Telegraph (CCITT) [26] showed that the threshold level for noise to cause errors can be as much as 11 dB higher for MCM.

Single-Frequency Interference

There is an interesting time/frequency duality involved here. An SCQAM signal is sensitive to impulses in the time domain in the same way that an MCM signal might be sensitive to impulses in the frequency domain (single-tone interference). The advantage of MCM lies in the fact that the sources of these interferers are discrete,[11] and their frequencies are usually stable (in contrast to the time of occurrence of impulses in the time domain); they can be recognized during training and avoided (that is, nearby carriers are not used) by the adaptive loading algorithm.

Fades

Mobile radio channels often suffer wideband fades, in which the SNR across the whole frequency decreases alarmingly for a short time. A single-carrier system might have a very low error rate between these fades, but would suffer from a very high one during a fade; the overall error rate might still be intolerable.

On the other hand, in a multicarrier system both the signal and the noise are integrated over the whole symbol period; the average SNR and resultant error rate are usually still tolerable.

[11] A tone at 2,600 Hz, which is used in some single-freqency signalling systems, is the most notorious interferer in the US.

Trellis Code Modulation

The advantages of TCM—about 3.5 dB of coding gain with present-generation codes and perhaps up to 5 dB with future codes—are now widely recognized. Early applications of trellis coding to MCM [25] [27] used encoding in the conventional way; that is, from symbol to symbol. Only a few carriers were used, and the delay through the Viterbi decoder was just tolerable because the symbols were fairly short. However, when MCM was first introduced to the mainstream of modem technology, it was clear that the proposed symbol period of 138 ms would be so long as to make MCM and conventional trellis coding incompatible.

The justification for trellis coding of SCQAM in general and decoding by the Viterbi algorithm in particular is that the noise is white (or almost so); that is, samples of it are almost uncorrelated from symbol to symbol. The time/frequency duality of single-/multicarrier can be exploited here by recognizing that samples of the noise, averaged over one symbol, are also uncorrelated from one frequency sub-band to the next, and that therefore trellis coding can be applied in the same way [28].

Following the terminology of [29], let the m_n bits for input to sub-band n be designated $x_n^1, x_n^2, \ldots x_n^m$. Then x_n^1 and x_n^2 should be input to the encoder to generate the output set z_n^0, z_n^1, z_n^2, which together with the uncoded bits $x_n^3, \ldots x_n^m$ are used to define a point in the appropriate constellation. The state of the encoder after encoding sub-band n is then used as the initial state for encoding sub-band $(n + 1)$.

As a result of the adaptive loading, the number of bits, m_n, and therefore the size of the nth constellation will probably vary with n, but this does not matter. The three encoded bits define one of eight sets of points, each containing $2^{(m_n - 3)}$ points, and the Viterbi decoding determines these three bits and, hence, the set; identification of a point within the defined set can then be done one sub-band at a time, even though the size of the set may vary from one sub-band to the next.

Any of the codes that have been developed for single-carrier could be used for MCM, but since a decoder will have to deal with constellations of varying sizes, it would be preferable to use codes and signal mappings that allow constellations to grow smoothly, such as were described in [30].

Block Processing of a Convolutional Code

It is highly desirable that all of the data in one symbol (block) be decoded in the same symbol period and from only the signals received within that block. This would not be possible, however, if both conventional encoding and decoding were used, because, first, a conventional encoder uses its state after encoding the last sub-band as the initial state for encoding the first sub-band of the next symbol, and second, a conventional Viterbi decoder makes a decision about a symbol only after receiving K_d more symbols, where K_d, the "look-back" distance or decoding delay, is typically between five and eight times the constraint length, l, of the code—about twenty for the common eight-state codes. Consequently, the last K_d sub-bands could not be decoded until the next symbol had been received and demodulated.

To achieve full block decoding the look-back distance in the decoder must be curtailed towards the end of the block. This can be done in two ways:

- The encoder can be modified by constraining l bits at the end of the symbol in order to force the 2^l state encoder into a known final state. Then all $(M - l)$ unconstrained bits can be decoded with no reduction of coding gain. This is easier to do with a feedforward encoder, but it would seem to be feasible even with a non-linear feedback encoder such as is described in CCITT Recommendation V.32.
- The Viterbi decoder can be modified to decode the last K_d sub-bands by tracing back the path from the smallest final

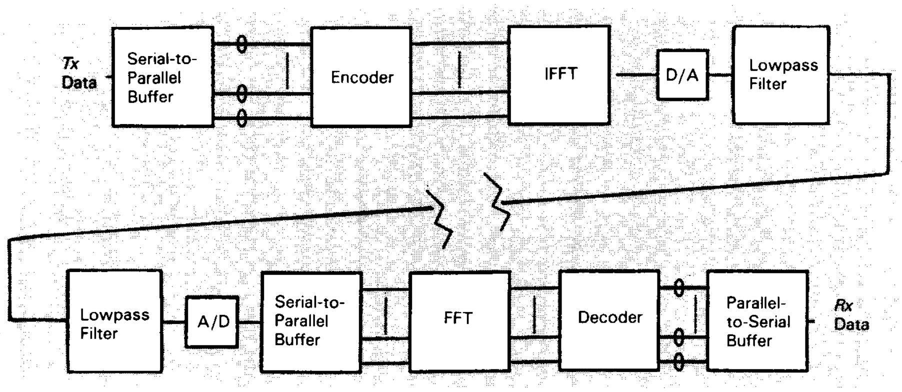


Fig. 7. Basic multicarrier "mo-dem."

metric, and decoding all of the remaining bits from the nodes on that path. This means that the coding gains for the last few carriers decrease more or less linearly from the maximum to about 0 dB for the last carrier. This effect can be anticipated in the original loading of these carriers, and will probably reduce the overall bit rate by about four bits per symbol.

Implementation

The first Echo Cancelers (ECs) to be developed were signal-driven; a generic one is shown in Figure 8a. The transmit signal is input to both the four-wire-to-two-wire connector ("hybrid") and an echo emulator (usually a TDL), which slowly learns the characteristics of the echo path, calculates samples of the estimated echo (usually by time-domain convolution), and subtracts them from the corrupted receive signal. It was soon recognized, however, that an EC can be greatly simplified if its input is the transmit data instead of the modulated and filtered

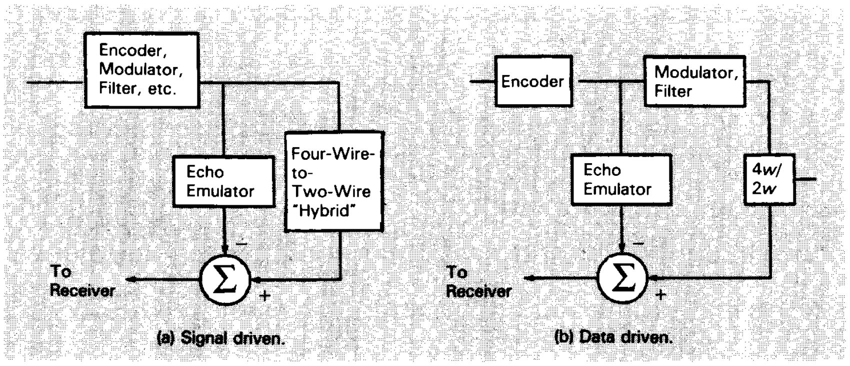


Fig. 8. Echo cancelers.

Multicarrier modulation solves both these problems by making the most advanced techniques usable at any speed—thus achieving the highest possible speed on any line—and by selecting that speed during the modem training, without any external control.

Duplex Operation

Simultaneous transmission and reception may be desirable for any one of many reasons, but the speed and error rate required in the reverse channel will vary greatly with the application. The use of MCM would allow the reverse channel to be placed in the optimum freqency band and to use the minimum transmitted power; this could relax the requirements for echo canceling—or, alternatively, extend the dynamic range of the modem—by as much as 18 dB.

Acknowledgments

I am very grateful to Adam Lender for the original encouragement to write this article and for many helpful suggestions, to John Cioffi for many enlightening discussions, and to an anonymous reviewer for the opportunity to see things through a reader's eyes.

References

[1] M. L. Doelz, E. T. Heald, and D. L. Martin, "Binary Data Transmission Techniques for Linear Systems," *Proc. IRE*, vol. 45, pp. 656–661, May 1957.

[2] B. Hirosaki, "An Orthogonally Multiplexed QAM System Using the Discrete Fourier Transform," *IEEE Trans. Commun.*, vol. COM-29, pp. 982–989, July 1981.

[3] I. Kalet and N. A. Zervos, "Optimized Decision Feedback Equalization versus Optimized Orthogonal Frequency Division Multiplexing for High-Speed Data Transmission Over the Local Cable Network," *IEEE Int'l. Conf. Commun. Rec.*, pp. 1,080–1,085, Sept. 1989.

[4] J. Fegreus, "Prestissimo," *Digital Rev.*, pp. 82–87, Apr. 1986.

[5] J. L. Holsinger, "Digital Communication Over Fixed Time-Continuous Channels with Memory—With Special Application to Telephone Channels," *Lincoln Lab. Tech. Rep. No. 366*, MIT, Cambridge, MA, Oct. 1964

[6] S. Kasturia, J. Aslanis, and J. M. Cioffi, "Vector Coding for Partial-Response Channels," *IEEE Trans. Info. Theory*, to be published.

[7] J. W. Lechleider, "Structured Channel Signaling," TIA Tech. Subcommittee T1E1.4, Contribution No. 89-067, Mar. 1989.

[8] J. W. Lechleider, "The Optimum Combination of Block Codes and Receivers for Arbitrary Channels," *IEEE Trans. Commun.*, to be published May 1990.

[9] G. A. Franco and G. Lachs, "An Orthogonal Coding Technique for Communications," *IRE Int'l. Conv. Rec.*, vol. 9, pt. 8, pp. 126–133, 1961.

[10] R. W. Chang, "Synthesis of Band-Limited Orthogonal Signals for Multichannel Data Transmission," *Bell Syst. Tech. J.*, vol. 45, pp. 1,775–1,796, Dec. 1966.

[11] B. R. Saltzberg, "Performance of an Efficient Parallel Data Transmission System," *IEEE Trans. Commun. Tech.*, vol. COM-15, pp. 805–811, Dec. 1967.

[12] M. S. Zimmerman and A. L. Kirsch, "The AN/GSC-10 Variable Rate Data Modem for HF Radio," *IEEE Trans. Commun. Tech.*, vol. COM-15, pp. 197–204, Apr. 1967.

[13] B. Hirosaki, "An Analysis of Automatic Equalizers for Orthogonally Multiplexed QAM Systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 73–83, Jan. 1980.

[14] S. B. Weinstein and P. M. Ebert, "Data Transmission by Freqency-Division Multiplexing Using the Discrete Fourier Transform," *IEEE Trans. Commun. Tech.*, vol. COM-19, pp. 628–634, Oct. 1971.

[15] P. Baran, "Packetized Ensemble Modem," U. S. Patent No. 4,438,511, Mar. 1984.

[16] D. Hughes-Hartogs, "Ensemble Modem Structure for Imperfect Transmission Media," U.S. Patents Nos. 4,679,227 (July 1987), 4,731,816 (Mar. 1988), and 4,833,706 (May 1989).

[17] J. A. C. Bingham, *The Theory and Practice of Modem Design*, New York: John Wiley and Sons, May 1988.

[18] R. Price, "Non-Linearly Feedback Equalized PAM vs. Capacity for Noisy Noisy Filter Channels," *IEEE Int'l. Conf. Commun.*, June 1970.

[19] I. Kalet, "The Multitone Channel," *IEEE Trans. Commun.*, vol. 37, pp. 119–124, Feb. 1989.

[20] R. G. Gallager, *Information Theory and Reliable Communication*, New York: John Wiley and Sons, 1968.

[21] B. Hirosaki *et al.*, "A 19.2 kb/s Voice-Band Data Modem Based on Orthogonally Multiplexed QAM Techniques," *IEEE Int'l. Conf. Commun. Rec.*, pp. 661–665, Aug. 1985.

[22] W. E. Keasler and D. L. Bitzer, "High Speed Modem Suitable for Operating with a Switched Network," U. S. Patent No. 4,206,320, June 1980.

[23] N. Jablon, "Complexity of Frequency Domain Adaptive Filtering for Data Modems," *Asilomar Conf. on Circuits and Syst.*, 1989.

[24] J. A. C. Bingham, "Method and Apparatus for Correcting for Frequency Offset, Phase Jitter, and Timing Offset in Multicarrier Modems," U. S. Patent Application filed Mar. 1990.

[25] B. Hirosaki, S. Hasegawa, and A. Sabato, "Advanced Group-Band Modem Using Orthogonally Multiplexed QAM Technique," *IEEE Trans. Commun.*, vol. COM-34, pp. 587–592, June 1986.

[26] Telebit Corporation, "Comparative Performance Results for Asymmetrical Duplex, V.32 (extended), and Multicarrier Modems," CCITT SG XVII, Contribution D56, Sept. 1989.

[27] A. Ruiz and J. Cioffi, "A Frequency Domain Approach to Combined Spectral Shaping and Coding," *IEEE Int'l. Conf. Commun. Rec.*, pp. 1,711–1,715, June 1987.

[28] D. Decker *et al.*, "Multi-Channel Trellis Encoder/Decoder," U. S. Patent Application filed Aug. 1988.

[29] IBM Europe, "Trellis-Coded Modulation Schemes with 8-State Systematic Encoder and 90° Symmetry for Use in Data Modems Transmitting 3–7 Bits per Modulation Interval," CCITT SG XVII, Contribution D180, Oct. 1983.

[30] G. Ungerboeck, "Trellis-Coded Modulation with Redundant Signal Sets," *IEEE Commun. Mag.*, vol. 25, pp. 5–21, Feb. 1987.

Biography

John A. C. Bingham received a B.Sc. degree from Imperial College, London in 1956 and an M.S.E.E. from Stanford University in 1961.

From 1959 to 1963 and 1966 to 1970, he was with Lenkurt Electric, San Carlos, CA, working on computational problems of filter design and data transmission. From 1972 to 1985, he was the Manager of the Advanced Development Department at Racal-Vadic, Milpitas, CA, where, in 1973, he invented the VA 3400, the first full-duplex 1,200 b/s modem. In 1985, he was a Visiting Scholar at ETH, Zurich, Switzerland. He is now a Senior Scientist at Telebit Corporation, Sunnyvale, CA, working on all types of data transmission.

He is the author of one book and about twenty papers, and holds five patents.