Richard K. Martin and C. Richard Johnson Jr.

© ARTVILLE

Adaptive Equalization:

Transitioning from single-carrier to multicarrier systems

In the past, single-carrier communication was the modulation format of choice. Today, multicarrier modulation is being selected as the transmission scheme for the majority of new communications systems [1]. Examples include digital subscriber line (DSL) [2]; European digital video broadcast (DVB) and digital audio broadcast (DAB) [3], [4] [i.e., high-definition television (HDTV)]; wireless local area networks such as IEEE 802.11a [5], HIPERLAN/2 [6], and multimedia mobile access communication (MMAC) [7]; satellite digital audio radio services (SDARS) such as Sirius Satellite Radio and XM Radio [8]; and power line communications (PLC) [9].

One of the virtues of multicarrier systems is that they are resilient to multipath (or delay spread in the wire-line case), provided that the delay spread of the channel fits within a prespecified guard interval between blocks. However, the length of this guard interval, which is fixed, is much shorter than the block length. In general, the channel length is unknown, and, in some cases, it may exceed the length of the guard interval. This is known to be true in DSL, for example. In this case, an equalizer can be used to mitigate the problem. Whereas in a single-carrier system, the equalizer inverts the channel (in the absence of noise) to create an effective channel that is simply a delayed impulse, in multicarrier systems, the goal of the equalizer is to create an

 1053-5888/05/$20.00©2005IEEE

effective channel that may have multiple nonzero samples, so long as the length of the effective channel becomes shorter than the guard interval. This is referred to as channel shortening.

This article discusses the creation of adaptive algorithms for channel shortening, with particular attention to blind algorithms. The context is multicarrier modulation, though other applications of channel shortening are discussed. It is shown that the algorithms used for adaptive equalization are not easily applied to adaptive channel shortening. In a return to first principles, a property restoral design philosophy is put forth and several recent property-restoral-based approaches to adaptive channel shortening are reviewed. We conclude with a discussion of the limitations of the current approaches and a list of open problems in the area of adaptive channel shortening. Indeed, there is much more work to be done.

TODAY, MULTICARRIER MODULATION IS BEING SELECTED AS THE TRANSMISSION SCHEME FOR THE MAJORITY OF NEW COMMUNICATIONS SYSTEMS.

PROBLEM FORMULATION

Traditionally, equalization takes the form of channel inversion; thus, the effective channel impulse response is simply a delay. In this case, the problem is well defined. Given a noise-free transmission channel $h(k)$, the equalizer $w(k)$ is designed such that the effective channel $c(k)$ is simply an impulse with delay Δ

$$c(k) = h(k) \star w(k) = \delta(k - \Delta). \tag{1}$$

[Note that this notation is chosen to be consistent with the channel shortening literature rather than the traditional equalization literature. If conventional equalization notation is preferred, (1) would read $h(k) = c(k) \star f(k)$, as in [10], for example.] In the noisy case, a minimum mean-squared error (MMSE) design can be used to keep the noise gain small as well. If the residual interference is approximately Gaussian, then the MMSE design minimizes the bit error rate (BER).

One way to think of equalization is that, given the channel $h(k)$, we wish to shorten the channel to an impulse by convolving the channel with some filter $w(k)$. A more general problem statement is that we wish to shorten the effective channel to some window of predetermined length. This problem is not as well defined. Does the impulse response within the window matter? Does the shape of the residual channel taps outside the window matter? How does the BER of the system relate to the shape of the achieved impulse response? How do we design an algorithm to adapt a channel-shortening equalizer with these questions in mind? This tutorial article will address (though not fully resolve) these questions, with a contextual focus on multicarrier communication systems.

APPLICATIONS OF CHANNEL SHORTENING

In multicarrier communications, modulation is done on a block-by-block basis, with a guard interval between the blocks. A critical assumption underlying the successful operation of such a system is that the delay spread of the channel is no longer than this guard interval. If this assumption is violated, a channel-shortening equalizer (CSE) can be used to restore the validity of this assumption. This article focuses on the multicarrier context, but we now briefly review several other applications that can benefit from channel shortening.

Channel shortening was first applied to maximum likelihood sequence estimation (MLSE) in the 1970s. MLSE [11] is the optimal sequence estimator in the sense that it minimizes the error probability of a sequence. However, its complexity grows exponentially with the channel length. For many practical transmission schemes, this complexity is too high to be implemented [12], [13]. This complexity can be mitigated by employing a prefilter to shorten the transmission channel to a manageable length and then applying the MLSE to the output of the shortened effective channel. One approach is to design both the prefilter and the (shortened) target impulse response to minimize the mean-squared error (MSE) between the target and the convolution of the channel and prefilter [14], [15]. Other approaches use a decision feedback equalizer (DFE) to shorten the channel and then apply the MLSE [16], [17]. These cases essentially implement a standard DFE for a single-carrier system (so that decisions can be made) and then take the output of the feed-forward filter (before the subtraction of the feedback terms) and separately pass it through a Viterbi algorithm.

Channel shortening has also been proposed in conjunction with multi-user detection [18]. Consider a direct-sequence code division multiple access (DS-CDMA) system with L users, with a flat-fading channel for each user. The optimum multiuser detector is again the MLSE, yet complexity grows exponentially with the number of users. "Channel shortening" can be implemented to suppress $L - K$ of the scalar channels and retain the other K channels, effectively reducing the number of users from L to K. Then the MLSE can be used to recover the signals of the remaining K users [18]. In this context, "channel shortening" means reducing the number of flat-fading channels rather than reducing the number of taps of a single frequency-selective channel. However, the mathematics are quite similar.

BLIND ADAPTIVE METHODS

One straightforward way to design a channel shortener is to first identify the channel and then compute the "best" channel shortener (by some definition of "best") for the estimated channel. This may be done by periodically transmitting a predetermined sequence of symbols, called training symbols or pilot symbols [19], and then comparing the known channel input and output to estimate the channel. Such methods are called "trained," as opposed to "blind" techniques, which do not require knowledge of, or existence of, a training signal.

There are a variety of reasons for using blind rather than trained methods. Although trained methods are often adequate, they have several disadvantages. Since the training signal takes up time slots that could be used to send data, this configuration reduces the throughput of the system. Another disadvantage is that the training signal is not always known at the receiver, e.g., in a noncooperative (surveillance) environment. Finally, the faster the channel varies over time, the more often the training signal must be transmitted, further reducing the throughput. For these reasons, a blind receiver is often desired. Moreover, even when the training signal is available, a semiblind implementation, which combines the use of trained techniques during the training periods and blind techniques during periods of data transmission, can be used.

When the channel is modestly time varying, two receiver implementations are possible. The receiver can frequently update an estimate of the current channel and periodically recalculate the optimal channel shortener for the current channel; alternatively, it can directly update the channel shortener over time in an adaptive fashion. The former procedure can lead to higher performance but at a much higher computational cost; a direct adaptive approach can maintain near-optimal performance at a more manageable cost. This article emphasizes the conversion from blind, adaptive equalization algorithms to blind, adaptive channel-shortening algorithms, although trained, adaptive designs will be considered as well.

MODULATION FORMATS

Equalization requirements are different for traditional single-carrier systems, multicarrier systems, and single-carrier systems that make use of a cyclic prefix. In this section, we review single-carrier and multicarrier communication systems, with the goal of providing a context for the mathematical channel-shortening problem.

A typical single-carrier system modulates the input data by a complex exponential, transmits it through a passband channel, and then demodulates the signal, as depicted in Figure 1. If the channel is frequency selective, then the output is distorted in the frequency domain. In the time domain, this amounts to convolving the input with some unknown channel $h[n]$, which is possibly time varying as well. This problem can be addressed by using a blind, adaptive equalizer. Algorithms for adapting blind equalizers are discussed later in this article.

A popular heuristic way of thinking about multicarrier communications is that if we divide the frequency-selective channel into bins that are small enough, then the channel will be approximately flat fading in each bin. If we transmit a large number of independent, narrow-band single-carrier signals, then each one perceives the channel as flat fading. Then only a single-tap (scalar) equalizer is needed for each subchannel. It is more mathematically correct to think of sampling the frequency response of the channel, with the samples taken at the frequencies of the multiple carriers.

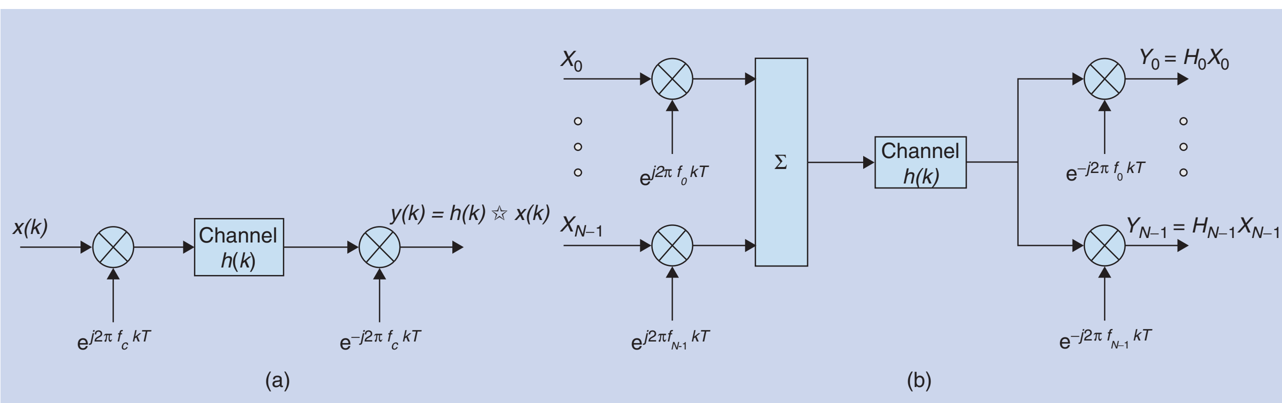
Figure 1(b) shows a multicarrier signal that consists of a sum of N narrow-band signals, each with its own carrier frequency f_i. Mathematically, the transmitted multicarrier signal is

$$x(k) = \sum_{i=0}^{N-1} X_i \, e^{j2\pi f_i kT}. \tag{2}$$

The carrier frequencies are linearly spaced, so that $f_i = f_c + i \cdot \Delta f$. The total bandwidth $N\,\Delta f$ is constrained to equal the sampling frequency $1/T$. Thus, the transmitted signal becomes

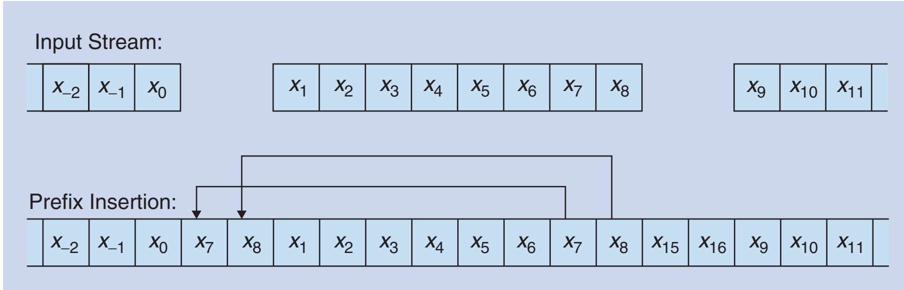
$$\begin{aligned} x(k) &= \sum_{i=0}^{N-1} X_i e^{j2\pi\, k\,(f_c + i\Delta f)T} \\ &= \underbrace{e^{j2\pi k f_c T}}_{\text{modulation}} \cdot \underbrace{\sum_{i=0}^{N-1} X_i e^{j\frac{2\pi}{N} ik}}_{\text{IFFT}}. \end{aligned} \tag{3}$$

Thus, multicarrier modulation can be performed by taking an N-point inverse fast Fourier transform (IFFT) of the input data and then modulating the result by a single carrier frequency. It is primarily this fact—that modulation and demodulation can be performed efficiently by the IFFT and fast Fourier transform (FFT)—that has made multicarrier modulation practical to implement.



[FIG1] (a) Single-carrier modulation. (b) Multicarrier modulation. If equalization or channel shortening is applied, the channel is immediately followed by the equalization or channel-shortening filter.

Circular convolution in the time domain amounts to pointwise multiplication in the frequency domain. The whole point of multicarrier communications is for the received narrowband signals to be obtained by point-wise multiplication of the input vector and the channel coefficients in the frequency domain. However, the convolution of the transmitted data block and the channel is actually a linear convolution. The typical solution is to trick the channel into thinking that the transmitted data is periodic, so that the convolution looks like a circular convolution. This is done by inserting a cyclic prefix before each transmitted block, as shown in Figure 2.



[FIG2] Insertion of the cyclic prefix, depicted for a block size of eight and a prefix length of two.

We can use a matrix interpretation to see how the cyclic prefix makes the linear convolution appear circular. The left side of Figure 3 shows a linear convolution as multiplication of a wide Toeplitz channel matrix of channel taps and a periodic input vector. The right side of Figure 3 shows the effective circular convolution as a multiplication of a square matrix and the nonperiodic input. Please see "Multicarrier and SCCP Parameter Selection Guidelines" for a description of how the FFT size and prefix length are chosen.

SINGLE-CARRIER CYCLIC PREFIX MODULATION

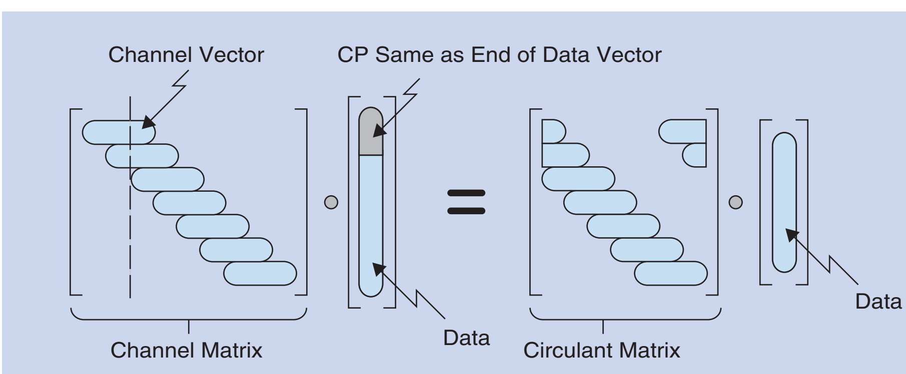
Single-carrier cyclic prefix (SCCP) modulation has been proposed as an alternative scheme that combines some of the advantages of single-carrier modulation with the equalization advantages of multicarrier modulation [20]–[22]. The idea is that the transmitter is single carrier, but transmission is conducted by blocks. Each block has a cyclic prefix added, as in multicarrier systems. The IFFT, frequency-domain scalar equalization, and FFT are all done at the receiver.

Figure 4 shows the relation between multicarrier and SCCP modulation. If a multicarrier system is operating correctly, the net system, in matrix form, is simply an $N \times N$ identity matrix. To convert this into an SCCP system, first note that the net effect of an FFT, an identity matrix, and an IFFT is still an identity matrix; this is shown in the bottom half of Figure 4, including the dashed boxes. Now note that the dashed initial FFT and IFFT cancel each other out. When this is done, we are left with a block-based transmitter that adds a cyclic prefix but has no FFT or IFFT. The receiver now has both the FFT and the IFFT, as well as the frequency-domain equalizer. Again, the cyclic prefix must be as long as the channel memory for equalization to be performed by a bank of scalars in the frequency domain. If the channel memory exceeds the CP length, we may once again use a channel shortener.

One of the advantages of SCCP modulation is that the transmitted samples still have a finite alphabet; since they are not Gaussian, they do not have the high peak-to-average power ratio of multicarrier signals. However, Louveaux et al. [23] have shown that for a fixed BER and adaptive bit loading, the achievable bit rate of a multicarrier system is always greater than or equal to the achievable bit rate of an SCCP system operating over the same channel and bandwidth.

ADAPTING SINGLE-CARRIER METHODS TO THE MULTICARRIER CASE

Adaptive equalizers (for single-carrier systems) have been studied extensively [24]. The most popular trained adaptive equalizer uses the least mean square (LMS) algorithm [25], [26] to minimize the mean-squared difference of the equalizer output and a training signal. The most popular blind adaptive equalizers are the decision-directed LMS algorithm and the constant modulus algorithm (CMA) [10], [27]. Decision-directed LMS makes use of the fact that the channel input is discrete or finite-alphabet, e.g., it may be ± 1. Hence, we can form a rough estimate (decision) of the input by simply quantizing the equalizer output to the nearest possible input value (e.g., by taking its sign when we expect a binary ± 1 source). This signal can then replace the training signal. The CMA conceptually makes



[FIG3] The cyclic prefix transforms a linear convolution of the channel and data into a circular convolution. Since the first few columns and the last few columns of the channel matrix get multiplied by the same data values (due to the redundancy of the prefix), then those columns can be combined by adding them together. This leads to the narrower channel matrix on the right, which is circulant.

MULTICARRIER AND SCCP PARAMETER SELECTION GUIDELINES

A detailed examination of Figure 3 and the underlying mathematics shows that for the channel to perceive the data as periodic, the redundant portion of the data (i.e., the prefix) must be as long as the memory L_h of the longest channel that is expected to be encountered. Thus,

1) *Prefix length (ν) must be greater than channel memory (L_h):* $\nu \geq L_h$

Although time variations of the channel are to be expected, the block convolution interpretation tacitly assumes that the channel is constant within each block. Thus, the block duration should be much less than the channel coherence time. We also want a small block size to keep the complexity low. This leads to rule 2:

2) *Block length must be much less than channel coherence time (in samples):* $N \ll \tau_{\text{coher}}/T$

The use of the cyclic prefix increases the symbol length to be the length of the data N plus the length of the cyclic prefix ν, reducing the throughput by a factor of $N/N+\nu$. Therefore,

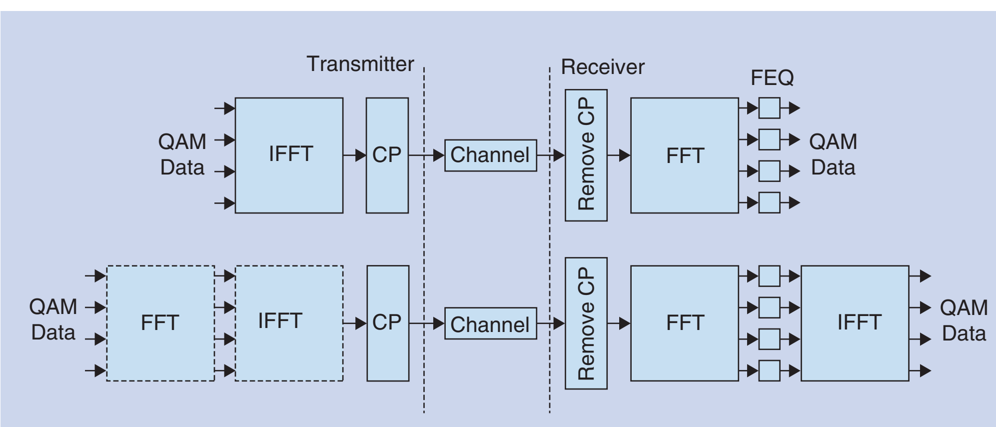
3) *Prefix length must be much less than the block length:* $\nu \ll N$.

Since L_h and τ_{coher} are beyond our control, these guidelines often cannot all be simultaneously satisfied. One resolution is to deliberately violate the first condition and then use a channel shortener to restore it. This is what is done in digital subscriber lines, for example.

use of the fact that the magnitude of the input is constant, even if the input itself is not. The CMA can also be used to equalize finite-alphabet, nonconstant modulus signals [10], in which case it is best viewed as a dispersion-minimizing algorithm.

Adapting blind algorithms from the equalization problem to the channel-shortening problem can be difficult. One issue is that, even if the input to the channel is finite-alphabet, the output of a short channel need not be. This creates problems for both multicarrier and SCCP systems. Moreover, the channel input in the multicarrier case is the IFFT of a finite-alphabet vector, so it is no longer finite-alphabet; in fact, it will have a nearly Gaussian distribution. This is because each IFFT output is a weighted sum of the N IFFT inputs (i.e., the actual data), and for large N, the central limit theorem indicates that the distribution approaches a Gaussian. This has dire consequences for the CMA, which does not converge at all when the input is Gaussian.

One algorithm that has been successfully adapted to the channel-shortening case is the trained adaptive MMSE algorithm. It is a generalization of the LMS algorithm, and it is the focus of the next section. Please see "The Folklore of Equalizer Design" for modification of adaptive equalization rules of thumb to the multicarrier case.



[FIG4] Comparison of multicarrier modulation (top) and single-carrier cyclic prefix modulation (bottom).

TRAINED ADAPTIVE MMSE CHANNEL SHORTENING

The LMS algorithm [25] compares the output of the equalizer to a delayed training sequence and forms an error signal from the difference. The equalizer coefficients are adapted such that the expectation of the square of the error is minimized. This is depicted in Figure 5(a). When the goal is to make the effective channel (i.e., the channel-equalizer combination) into something other than a delay, the model must change.

The trained, adaptive MMSE channel shortener [15] accounts for the fact that the target impulse response may be something other than simply a delay. Specifically, the target is a channel that has been shortened to some particular finite length, though there is still a bulk delay Δ associated with this target impulse response. However, we are free to choose not only the channel shortener, but the coefficients of the target impulse response as well, so as to minimize the same mean square error as the LMS algorithm. This is depicted in Figure 5(b). A gradient descent of the MSE can be used to adapt the filters, as with the LMS algorithm. However, there are two differences. First, in the simplest gradient calculation, one filter is held constant while the other adapts, and then vice-versa, with one filter using a much larger step size than the other. This allows one filter to "follow" the other, so that the two updates do not conflict [15]; however, this configuration also leads to very slow convergence times [28]. Second, more critically, a constraint must be applied since setting both filters to all zeros leads to a perfect MSE of zero. Falconer and Magee proposed to maintain a unit-norm target impulse response by renormalizing at each

THE FOLKLORE OF EQUALIZER DESIGN

The past three decades of research on adaptive equalizers have provided a collection of "rules of thumb" that are useful starting points for selecting the parameters of the equalizer before adaptation commences. Analogous guidelines can be followed for adaptive channel shorteners.

Rule	Single Carrier	Multicarrier
Initialization	Single spike at equalizer center	Single spike (equalizer and/or target response)
Equalizer length	3–5 times channel length	3–5 times difference of channel and prefix lengths
Effect of zero on unit circle	Noninvertible channel	Several noninvertible subchannels
Desired delay	Channel peak + half equalizer length	Location of window of channel with largest energy + half shortener length

iteration [15]. The unit-norm constraint could also be implemented by adding a vector CMA term to the cost function rather than by renormalizing [29]. Alternatively, one filter tap could be constrained to unity [14].

The adaptive MMSE channel shortener of [15] was designed for channel shortening in conjunction with MLSE for a single-carrier system. Hence, it assumes that time-domain training is available. However, when training is available for a multicarrier system, it is typically provided on selected tones in the frequency domain [19]; hence, the required time-domain training will not be available. For example, if every fourth frequency-domain input is known at the receiver, the IFFT of this partially known vector is not known. This can be mitigated somewhat by resorting to decision direction, but at the cost of adding a large delay since decisions cannot be made until an entire block has been received and passed through the FFT.

PROPERTY RESTORAL

Even though decision-directed and constant-modulus-based designs cannot be used to create blind, adaptive channel shorteners, we can make use of the underlying philosophy used to develop those algorithms. This philosophy is the concept of "property restoral" [30, Chapter 6]. The idea is to look for and restore properties of the transmitted sequence that ought to be present in the equalized received sequence. In the single-carrier case, the transmitted sequence can be constant modulus or can have a finite-alphabet; hence, algorithms have been designed to restore those properties. In the multicarrier case, several properties are available for creating blind, adaptive channel shorteners.

1) A cyclic prefix is present, so each symbol has redundancy in its data [31], [32].

2) The channel is desired to be shorter than the cyclic prefix length. If it is, then the autocorrelation of the output data (assuming an uncorrelated source) should be short as well [33], [34].

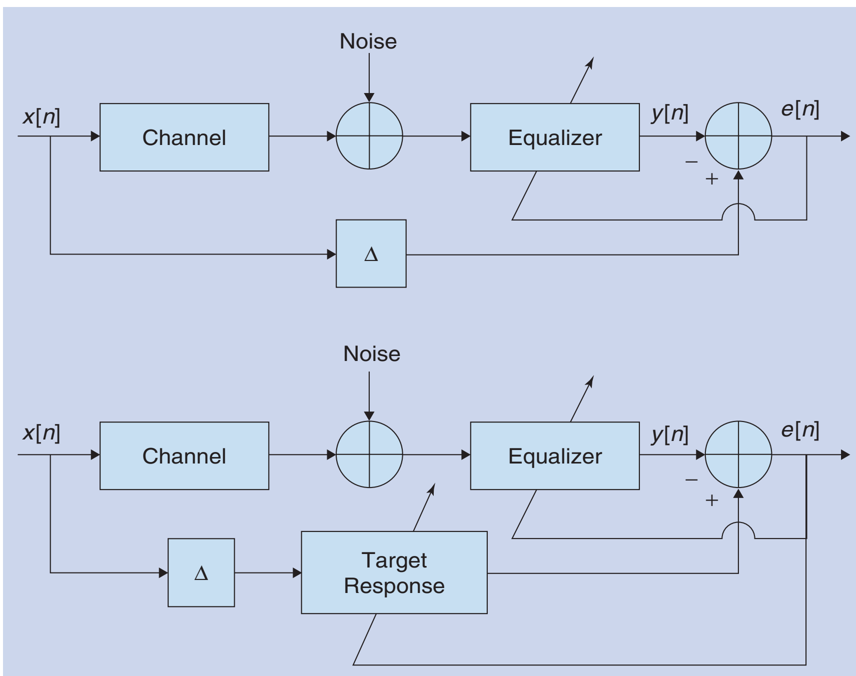
3) Often zeros are transmitted on the band edges, which is like frequency-domain training [35], [36].

4) The frequency-domain (not time-domain) data is finite alphabet [37].

We now discuss algorithms that attempt to restore each of these properties.

CYCLIC-PREFIX RESTORATION

In multicarrier or SCCP modulation, the transmitted sequence has redundancy due to the cyclic prefix. This redundancy has often been exploited for carrier frequency offset (CFO) estimation, under the assumption that the channel is shorter than the cyclic prefix [38] or that the channel is not time-dispersive at all [39]. This anticipated redundancy can also be exploited in the



[FIG5] Comparison of (a) the LMS trained adaptive equalizer and (b) the MMSE trained adaptive channel shortener, where Δ is the chosen transmission delay.

property restoral sense to create a blind, adaptive channel shortener. The multicarrier equalization by restoration of redundancy (MERRY) algorithm adapts the channel shortener with the goal of restoring this redundancy [31].

Figure 6 illustrates the concept of redundancy restoration. Consider a block-based transmission scheme (multicarrier or SCCP) with a data block size of eight samples and a prefix length of two samples. At the transmitter, the redundancy of the cyclic prefix can be represented by $x_1 = x_9$ and $x_2 = x_{10}$. At the receiver, samples y_2 and y_{10} would still be equal in the absence of a channel. However, these samples consist of the convolution of the channel and the input sequence. If the channel is no longer than the cyclic prefix, then the convolution for y_{10} only uses the x data in the end of the symbol, and the convolution for y_2 only uses the redundant data in the prefix, making the two y values equal. However, if the channel is longer than the prefix (as shown in Figure 6), then the excess channel taps create terms that will be different in the two convolution sums. These undesirable terms are shown in brackets. If we adapt the channel shortener to make y_2 and y_{10} equal in the mean square sense, then these undesirable terms will go away. The only way to do this is to remove all of the channel taps except those within the length of the prefix.

ONE OF THE VIRTUES OF MULTICARRIER SYSTEMS IS THAT THEY ARE RESILIENT TO MULTIPATH (OR DELAY SPREAD IN THE WIRE LINE CASE), PROVIDED THAT THE DELAY SPREAD OF THE CHANNEL FITS WITHIN A PRESPECIFIED GUARD INTERVAL BETWEEN BLOCKS.

Formally, the MERRY algorithm attempts to minimize the expectation of the square of a "cyclic difference," which is the difference between two y values separated by the data block length. The two values that we choose will depend on the symbol synchronization Δ (our estimate of the boundaries of the data block). That is, we wish to perform a gradient descent of

$$J_{\text{merry}} = \mathrm{E}\left[\, | y_{\nu+\Delta} - y_{\nu+\Delta+N} |^2 \right]. \tag{4}$$

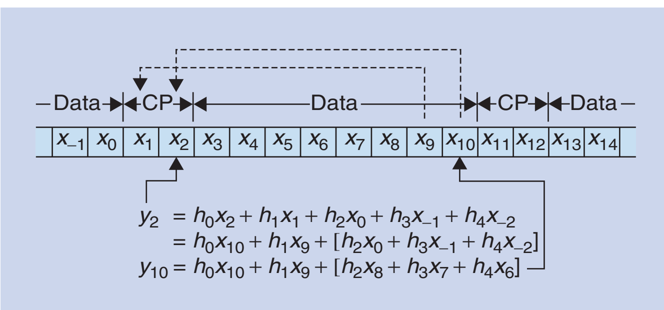
This leads to a very simple LMS-like form for the algorithm:

$$\begin{aligned} &\text{Given } \Delta, \text{ for symbol } k = 0, 1, 2, \ldots, \\ \widetilde{\mathrm{r}}(k) &= \mathrm{r}(Mk + \nu + \Delta) - \mathrm{r}(Mk + \nu + N + \Delta), \\ e(k) &= \mathrm{w}^T(k)\,\widetilde{\mathrm{r}}(k), \\ \widehat{\mathrm{w}}(k+1) &= \mathrm{w}(k) - \mu e(k)\,\widetilde{\mathrm{r}}^*(k), \\ \mathrm{w}(k+1) &= \frac{\widehat{\mathrm{w}}(k+1)}{\|\widehat{\mathrm{w}}(k+1)\|}. \end{aligned} \tag{5}$$

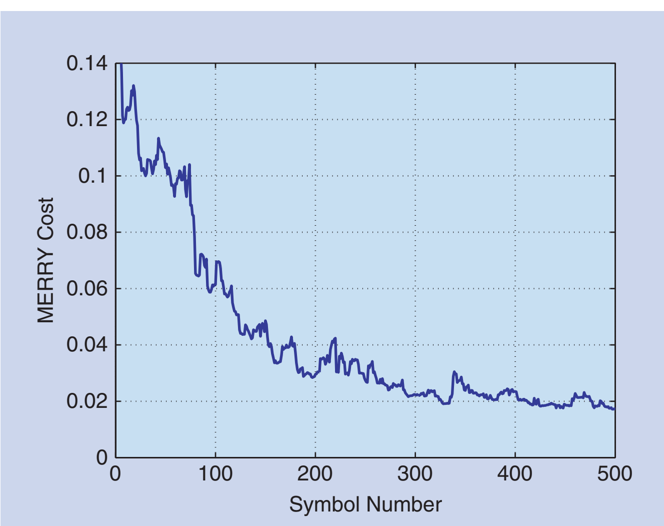
In other words, we form a difference vector of received data samples (each pair separated by N samples), then use that as the "regressor" vector for an LMS-like algorithm. The only other difference between MERRY and LMS is that MERRY must be constrained to avoid the all-zero solution [since there is no "desired" signal to compare to in the error $e(k)$], which can be done by constraining a single tap or the norm of the filter to unity. This is enforced by the renormalization in the last line of (5). The cyclic difference can only be measured once per symbol, so we can only update the adaptive algorithm once per block. This is one of the weaknesses of the MERRY algorithm. Its strength is that it has the same cost surface as the trained, adaptive MMSE channel shortener, with two caveats.

The first caveat is that the MMSE algorithm maintains a unit-norm constraint on the target response. The easiest way to implement the MERRY algorithm is to renormalize the channel shortener after each iteration, as discussed above. This leads to a unit-norm *channel shortener*, rather than a unit-norm *target response*. This can be corrected by adding a Lagrangian term to the cost function [40].

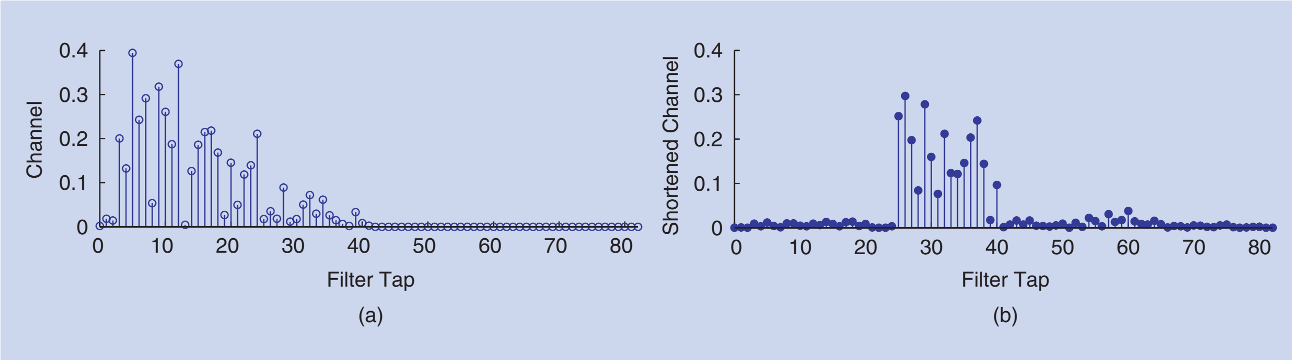
The second caveat is that MERRY shortens the channel to the length of the cyclic prefix. Technically, it is acceptable to only shorten the channel *memory* to the length of the prefix,



[FIG6] Loss of the prefix-induced redundancy due to a long channel.



[FIG7] MERRY cost versus time.



[FIG8] (a) Channel and (b) shortened channel.

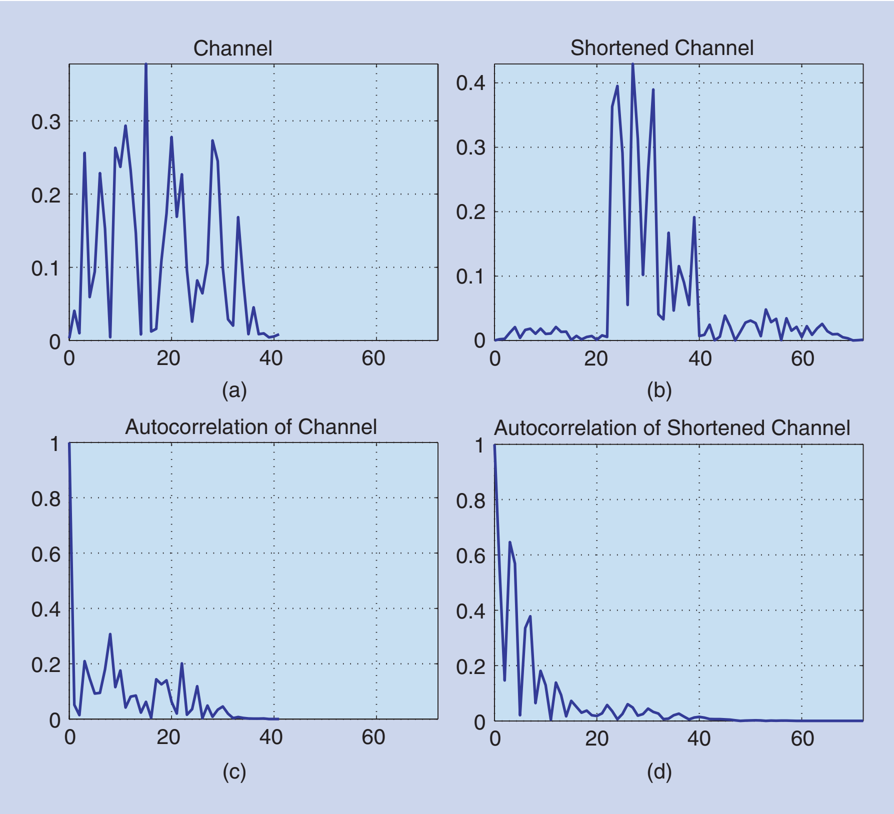
i.e., to shorten the channel length to the prefix length plus one. Thus, we are shortening the channel to be one tap shorter than we really intend to. However, for typical prefix lengths, such as 32 in assymmetric digital subscriber line (ADSL) or hundreds of taps for broadcast television or radio, this difference is negligible.

Figures 7 and 8 give an example of the operation of the MERRY algorithm. Figure 7 shows the value of the MERRY cost as the channel shortener adapts. The channel and channel shortener are each modeled as length-42 finite impulse response (FIR) filters. Convergence is on the order of 500 symbols in this case. The unshortened channel is shown in Figure 8(a), and the final shortened effective channel is shown in Figure 8(b). Observe that the energy has been compressed to lie primarily in 16 consecutive samples, with a delay of 25 samples.

AUTOCORRELATION SHORTENING

Regardless of the modulation format, if the input is uncorrelated and the channel is in fact short, then the autocorrelation of the received data will also be short. We can view a short autocorrelation as a property that is degraded by a long channel. We can then form a blind, adaptive channel shortener that attempts to restore this property. Figure 9 illustrates this principle. The two plots in Figure 9(a) show a long channel and its autocorrelation function. Then a channel shortener is applied, and the resulting short channel and its autocorrelation are shown in Figure 9(b). The short channel has most of its energy in 17 taps; hence, its autocorrelation has most of its energy in lags zero through 16. The first question to be answered is: Since a short channel has a short autocorrelation, is the converse always true? The answer is that it is not always true [33], but it appears to be true often enough to be a useful tool for developing an adaptive channel shortener.

Several algorithms have been proposed to restore a short autocorrelation. The first of these, the sum-squared autocorrelation minimization (SAM) algorithm [33], performs a gradient descent of the excess autocorrelation. Specifically, the cost function to be minimized is the sum of the squares of autocorrelations of all lags greater than the desired channel memory. Again, as with MERRY, the all-zero solution minimizes this cost function, and this degenerate case must be avoided. One simple approach is to renormalize the filter after each iteration or to hold one of the filter taps equal to unity.



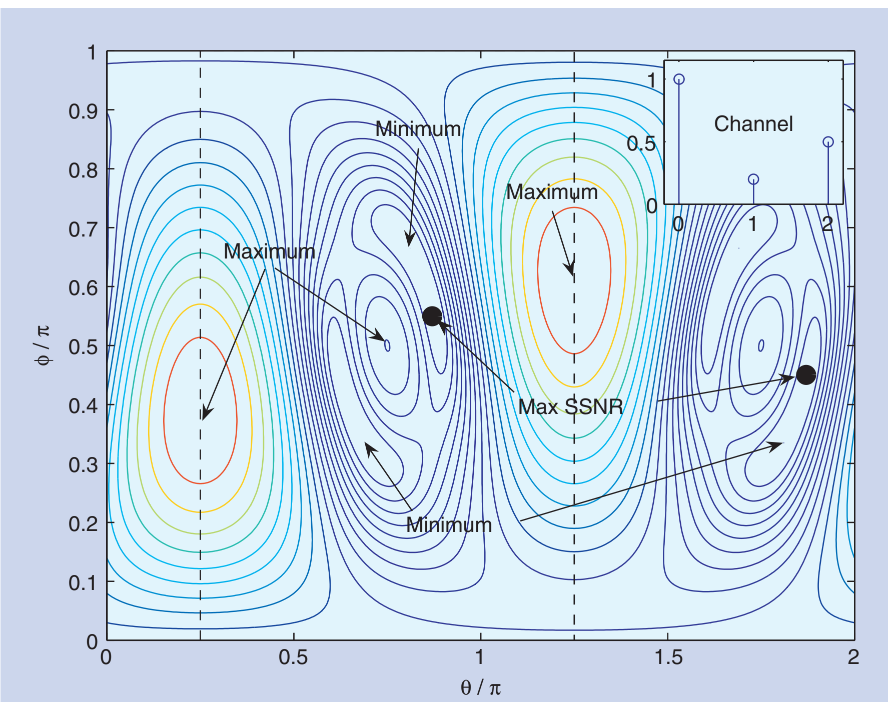
[FIG9] Shortening a channel shortens its autocorrelation. Is the converse true? Not always. In this example, the desired channel memory and maximum autocorrelation lag are both 16.

Several variants of SAM have recently been proposed. The sum-absolute autocorrelation minimization (SAAM) [41] algorithm replaces the squares of the autocorrelation with their absolute values, leading to better performance in the presence of impulse noise. Impulse noise is a common problem in digital subscriber lines. The single lag autocorrelation minimization (SLAM) [42] algorithm greatly reduces the complexity of SAM by considering only the single autocorrelation term whose lag is just barely greater than the desired channel memory. The intuition for this is that the autocorrelation tends to decrease as the lag increases; hence, the first lag outside of the prefix will tend to dominate the remaining lag terms. Miyajima and Ding [43] proposed an approach similar to SAM, based on oversampling and subspace methods. Loosely speaking, they search for a filter in the null space of an autocorrelation matrix based on the same lag as SLAM, subject to the requirement that the filter must not be in the null space of the autocorrelation matrix based on zero lag (which keeps the filter away from the all-zero solution).

IN MULTICARRIER COMMUNICATIONS, MODULATION IS DONE ON A BLOCK-BY-BLOCK BASIS, WITH A GUARD INTERVAL BETWEEN THE BLOCKS.

The autocorrelation of a filter does not change if its poles or zeros are flipped over the unit circle. Since the SAM, SAAM, and SLAM algorithm costs depend on the autocorrelation of the effective channel, they will not change if we invert any of the zeros of the channel shortener. For a filter of order L, there are up to 2^L combinations of inverting or noninverting zeros. The unfortunate consequence of this is that, for any given point on the cost surface, there are up to 2^L points of identical cost elsewhere; in particular, each minimum will be repeated 2^L times. Some of the global minima of the SAM cost may be close to the global optimum of a more traditional cost surface, whereas other global minima of the SAM cost may not be in desirable locations.

Figure 10 illustrates this effect for a unit-norm filter of length three (i.e., second order). The plot shows contours of the SAM cost surface, and the maximum shortening signal-to-noise ratio (SNR) solution of [44] is shown as a large dot. The channel is shown in the picture-in-picture in the upper right of Figure 10, and for this example, the desired channel length is two taps rather than three. There are four global minima of the SAM cost function, and only two of them are near the maximum shortening SNR solution (and the negative of that solution, hence the second dot). The dashed line represents a plane of symmetry for the SAM cost function, which will also be a plane of symmetry for the SAAM and SLAM cost functions. These extra minima require care in initialization or some additional procedure for re-initialization upon capture of a SAM minimum with poor shortening SNR performance.



[FIG10] **Contours of the SAM cost function. The channel shortener is length three and has a unit norm; hence, it is parameterized by two angles in spherical coordinates. The channel (shown in the picture-in-picture) is length three, and we wish to shorten it to a length of two. The large dots are the maximum shortening SNR solution [44], which SAM attempts to approximate.**

NULL-TONE RESTORATION

Another common property of multicarrier signals is the presence of null tones in the transmitted data. For example, in IEEE 802.11a, 12 of the 64 tones are null tones, with six null tones located at each edge of the frequency band. This provides a buffer to limit adjacent channel interference. It has also been suggested in [35] that this can be viewed as oversampling the transmitted signal (before transmission, rather than at the receiver) since, of the 64 inputs, 52 are data and 12 are zeros. A blind, adaptive channel-shortening algorithm can be derived with the goal of restoring the values of these tones to zero at the output of the receiver's FFT [35], [36]. This results in a carrier nulling algorithm (CNA). A block diagram of this algorithm is given in Figure 11.

Formally, the CNA cost function is the average power of the outputs on the tones that should be null.

$$J_{\text{cna}} = \sum_{\text{null tones } i} \mathrm{E}[|z_i|]^2, \qquad (6)$$

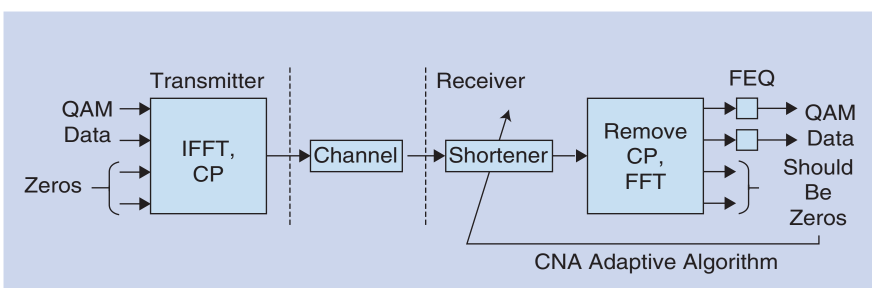
where z_i is the ith output of the demodulating FFT in Figure 11. The CNA algorithm is a constrained gradient descent of this cost function. This leads to a very simple LMS-like structure, although the computational complexity is somewhat higher due to several matrix-vector products.

$$\text{Given } \Delta, \text{ for symbol } k = 0, 1, 2, \ldots,$$
$$\hat{\mathrm{w}}(k+1) = \mathrm{w}(k) - \mu\, \mathrm{R}_k^H\, \mathrm{F}_T^H\, \underbrace{\mathrm{F}_T \mathrm{R}_k\, \mathrm{w}(k)}_{\mathrm{z}^{\text{null}}(k)},$$
$$w(k+1) = \frac{\hat{\mathrm{w}}(k+1)}{\|\hat{\mathrm{w}}(k+1)\|}, \qquad (7)$$

where $\mathbf{R}_k$ is a Toeplitz matrix of received data (to implement the filter/data convolution) and F_T^H is a truncated IFFT matrix containing the columns of an IFFT matrix corresponding to the positions of the null tones. The vector $\mathrm{z}^{\text{null}}(k)$ contains the outputs z_i of the tones that are supposed to be zero. The delay Δ implicitly appears in the data matrix $\mathbf{R}_k$, since the indices of the data values depend on an estimate of the symbol boundaries.

The CNA algorithm has much in common with the MERRY algorithm. Like MERRY, CNA can only update once per symbol. This is because the cost function is measured at the output of the FFT, once per block. Also, as with the MERRY and SAM algorithms, a constraint is required for CNA to avoid the all-zero solution. De Courville et al. [35] chose to implement a unit-norm constraint on the channel shortener via periodic renormalization. Romano and Barbarossa [36] did not specify their choice of constraint. Assuming that the unit-norm constraint is used, the CNA algorithm solves for the eigenvector corresponding to the minimum eigenvalue of the autocorrelation matrix of the outputs on the null tones [35], whereas MERRY seeks the eigenvector corresponding to the minimum eigenvalue of the autocorrelation matrix of a difference of two vectors of received samples [31].

Analysis of the CNA algorithm is surprisingly difficult. De Courville et al. [35] show that the zero-forcing equalizer (not a more generic channel shortener) minimizes the CNA cost function. Hence, CNA should be used in multicarrier systems that do not employ a cyclic prefix. Romano and Barbarossa [36] state that in the single-input, multiple-output case (e.g., for oversampled received data or multiple receive antennas), if the locations of the null inputs are cycled through all possible tones (i.e., frequency hopping at the transmitter), then CNA converges to perfectly shorten the channel. However, in practice, the null tones are located at the band edges and are not frequency-hopped.



[FIG11] Block diagram of the carrier nulling algorithm (CNA). For simplicity, the null tones are shown as grouped rather than split with half at each band edge, though the latter is the case in practice.

FREQUENCY-DOMAIN, FINITE-ALPHABET METHODS

The time-domain data in a multicarrier system is not finite-alphabet, but the frequency domain data at the output of the demodulating FFT is finite alphabet. This means that a decision-directed or constant modulus cost function can be proposed in the frequency domain. However, now there are N tones, so the cost must be summed over the N outputs. For example, we might have the frequency-domain decision-directed and constant modulus cost functions [37]

$$J_{dd} = \sum_{\text{tones } i} \beta_i\, \mathrm{E}\left[(Q\{z_i\} - z_i)^2\right], \qquad (8)$$

$$J_{cm} = \sum_{\text{tones } i} \beta_i\, \mathrm{E}\left[(z_i^2 - \gamma_i)^2\right], \qquad (9)$$

where each β_i is a weight, $Q\{\cdot\}$ finds the nearest constellation point in a finite alphabet, and γ_i is the desired average squared modulus of the final output z_i on tone i. The choice of nonuniform βs can be used to provide unequal error protection across the tones. The CNA algorithm can be thought of as using a special case of (8); it is a decision-directed algorithm in which we compare the null tones to a finite alphabet that is simply the set $\{0\}$, so $Q\{z_i\} = 0$ always. Aside from CNA, adaptive channel shorteners that make use of these types of frequency-domain cost functions have not been studied in the literature.

If the channel is short, the output will be quadrature amplitude modulation (QAM) data on each nonnull tone, but the modulus of the points will not be correct until after the bank of one-tap frequency-domain equalizers (FEQs). Thus, the frequency-domain cost must be measured at the output of the FEQ. This means that the channel shortener and FEQ, which are connected in series, will both be adapting based on the N outputs of the FEQ. Typically, adaptive devices are analyzed under the assumption that each device operates independently, and this sort of adaptation of a series of elements is not well understood [45].

Lin et al. [46] have proposed a trained, nonadaptive design that operates in the frequency domain. Their method maximizes the energy at the output of the pilot tones divided by the energy of the null tones. In principle, this idea could be used to create a trained, adaptive algorithm that restores both the pilots and the null tones, like a combination of CNA and frequency-domain LMS.

ALTERNATIVE STRUCTURES

The mathematical problem we wish to solve is how to design blind, adaptive channel-shortening algorithms. However, since we are focusing on the context of multicarrier communication systems, for completeness we now present alternative multicarrier equalizer structures that do not involve channel shortening. The common theme of these alternative structures is that they all operate on a tone-by-tone basis, with one filter or linear combiner per tone.

The most popular alternative equalization structure for a multicarrier system is called a "per-tone equalizer" (PTEQ) [47], which performs equalization tone-by-tone in the frequency domain. For each tone, a linear combination of the outputs of a sliding FFT of the received data is formed to remove the inter-carrier interference from the desired tone. An equivalent (and computationally cheaper) solution is to linearly combine the desired tone and a vector of differences of the received samples [47]. During data transmission, each of the N linear combiners forms an output once per block, whereas a channel shortener forms the output of a single linear combiner N times per block; thus, the computational complexity of processing the data is the same for the two filters. However, a PTEQ has N times as many coefficients to initialize compared to a channel shortener. Hence, initialization complexity is much higher for a PTEQ than for a channel shortener.

It is a curious coincidence that the difference vector used by the PTEQ (without the ith FFT output) is the same difference vector $\tilde{\mathbf{r}}$ used in the MERRY algorithm, and the ith FFT output is used in adapting the FEQ that necessarily follows the MERRY algorithm. Thus, the regressor vector (the vector used in an LMS-like gradient descent update rule) for an adaptive PTEQ linear combiner for tone i consists of the concatenation of the regressor vectors for the MERRY linear combiner and an adaptive FEQ for tone i. This implies a similarity of behavior of MERRY and a PTEQ adapting via LMS.

Another alternative structure is a bank of filters in the time domain, or a time-domain equalizer filter bank (TEQ-FB) [48]. If the filters in both structures are complex, then the PTEQ and TEQ-FB are mathematically identical. The TEQ-FB performs each linear combination in the time domain, and again interference is removed tone-by-tone. However, now each of the N filters must produce N outputs per block, so the TEQ-FB has a computational complexity during data transmission that is about N times larger than a PTEQ or channel shortener.

A third alternative equalization structure that operates tone-by-tone is frequency-domain equalization for discrete multitone, dubbed FEQ-DMT by its creators, Trautmann and Fliege [49]. Discrete multitone is the name given to FFT-based wire-line multicarrier transmissions, but this approach applies to wireless transmissions as well. Like the CNA channel shortener, the FEQ-DMT exploits the fact that many multicarrier systems transmit zeros on a subset of the tones. Rather than forcing the corresponding outputs to zero (as CNA does), FEQ-DMT uses the fact that the intercarrier interference is the only signal present on these tones; hence, the outputs on the supposed null tones can be weighted and added or subtracted from the data tones to remove the interference from the data tones. This leads to a bank of linear combiners, much like the PTEQ. The difference is that in the PTEQ, the signals being combined for tone i were the ith FFT output and a collection of differences of received samples, whereas the FEQ-DMT combines the ith FFT output with the output signals on the null tones. Trautmann and Fliege [49] have shown that if ν null tones are used, they provide the same amount of protection from channel distortion as the use of a cyclic prefix of length ν. That is, if a cyclic prefix is used and no null tones are present, *or* if no prefix is used but the null tones are present and the FEQ-DMT receiver is used, then the data can be recovered perfectly in the absence of noise. Their proof can be extended to show that the use of both a cyclic prefix and null tones (with an FEQ-DMT receiver) provides protection against multipath with twice as much delay spread.

[CHANNEL SHORTENING WAS FIRST APPLIED TO MAXIMUM LIKELIHOOD SEQUENCE ESTIMATION IN THE 1970S.]

All of these alternative structures have very good performance in time-invariant environments. However, even aside from issues of computational complexity, these structures have disadvantages. They all have many times the number of parameters as a single channel shortener. If we wish to adapt all of these parameters based on a finite amount of data, the adaptation speed will be slow. A common rule of thumb is that convergence of an adaptive algorithm requires about 100 times as many data points as there are parameters (filter coefficients) to be estimated. Thus, if we can use the data in an optimal manner, the convergence and tracking speed of a single channel shortener should be much faster than for a bank of filters. For example, the SAM channel shortener converges within several data blocks [33], whereas a recursive least squares per tone equalizer takes about 200 blocks to converge [50], presumably because there are about 250 times as many parameters to update in the latter case.

CONSEQUENCES FOR THE PERFORMANCE SURFACE

All of the adaptive channel-shortening algorithms discussed above optimize some heuristic cost function in the hopes that this will lead to a high bit rate (in a point-to-point link with fixed maximum BER) or low BER (in a broadcast system with fixed transmitted bit rate). However, it is difficult to establish a direct relation between the heuristic performance surfaces and the bit rate or BER surfaces.

Many channel-shortening designs have been proposed by nominating a heuristic performance metric and then finding the filter that optimizes that particular metric. Examples include the maximum shortening SNR design [44], the minimum mean squared error design [15], the minimum delay spread design [51], and the minimum interblock interference design [52]. However, none of these designs is optimal in terms of the performance metric of interest for wireless channels, which is the BER. Figure 12, taken from [40], shows a plot of the (uncoded) BER versus the SNR for these channel shorteners. Note that for some algorithms, the use of a channel shortener actually degrades the BER.

The difficulty is that the BER is a highly nonlinear and multimodal function of the filter taps; hence, it is difficult to optimize. For QPSK data on each tone with a high SNR, the BER can be approximated by [53, pp. 225–226]

$$P_e = (2/N_u) \sum_{i \in \text{data tones}} Q\left(\sqrt{\text{SNR}_i}\right), \quad (10)$$

where N_u is the number of data-carrying tones. The effective SNR at the output of each data tone, SNR_i, is a generalized Rayleigh quotient with respect to the channel shortener,

$$\text{SNR}_i = \frac{\mathbf{w}^H \mathbf{B}_i \mathbf{w}}{\mathbf{w}^H \mathbf{A}_i \mathbf{w}}. \quad (11)$$

The matrices $\mathbf{A}_i$ and $\mathbf{B}_i$ are Hermitian positive semidefinite. They depend on the channel impulse response, the correlation of the noise and interference, the residual (unshortened) portions of the effective channel, and the ith row of the DFT matrix. These matrices have been analyzed with fewer and fewer approximations in recent literature; see [54] for a relatively simple approach that only makes a few approximations.

The form of (10) shows just how difficult it is to optimize the BER. For a single-carrier system, the summation is absent, so the BER can be minimized by maximizing the SNR at the output of the receiver. This, in turn, can be accomplished by minimizing the MSE. In contrast, since there are many outputs in a multicarrier system, the summation must be included in (10), creating a challenging optimization problem.

In wire-line, point-to-point communication systems such as DSL, typically the BER is constrained and the data rate is increased until the BER reaches the prespecified upper limit. In this case, the SNR on each tone still has the form of (11), but the performance measure becomes the capacity (i.e., the number of bits that can be transmitted per block without exceeding the BER upper limit)

$$B = \sum_{i \in \text{data tones}} \log_2\left(1 + \frac{\text{SNR}_i}{\Gamma}\right), \quad (12)$$

where the SNR gap Γ is determined in part by the desired BER threshold [2]. This problem has been well studied in the DSL literature (see, e.g., [55]), although the solutions are computationally intensive and optimality is not guaranteed.

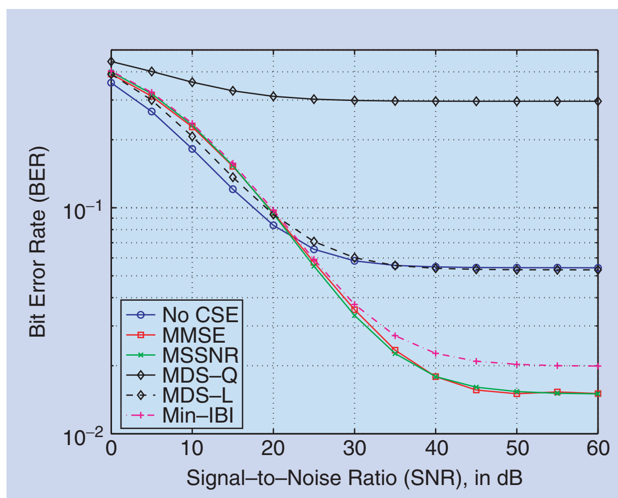
In terms of adaptive algorithm design, it is important to realize that the cost functions used to create adaptive algorithms, such as mean-squared error [15] or MERRY cost [31], do not necessarily have a direct relation to the BER (10) or bit rate (12). That is, these cost functions are essentially heuristics that can be easily adaptively optimized, but optimizing these proxy cost functions will not necessarily optimize the BER or bit rate. More sophisticated adaptive algorithms are needed to address this incongruity.

EQUALIZATION REQUIREMENTS ARE DIFFERENT FOR TRADITIONAL SINGLE-CARRIER SYSTEMS, MULTICARRIER SYSTEMS, AND SINGLE-CARRIER SYSTEMS THAT MAKE USE OF A CYCLIC PREFIX.

DISRUPTING FACTORS

The adaptive channel-shortening algorithms discussed above make various assumptions that are often not true, and when the assumptions are invalid, the algorithms behave differently than expected. In particular, we have assumed that 1) the transmitted signal is white, i.e., its autocorrelation function is a delta function, and 2) the transmitted null tones are in fact zero, or at least any signal on these tones is uncorrelated with the transmitted data signal. In fact, these two assumptions are mutually exclusive. We now discuss the ways in which these assumptions can be violated, as well as the effect such a violation has on the behavior of the various adaptive algorithms.

Generally, the data bits on the used carriers are uncorrelated with each other if the source coding and interleaving have been



[FIG12] BER versus SNR for several nonadaptive channel-shortening designs, reproduced from [40]. "No CSE" is the case when there is no channel-shortening equalizer. Otherwise, each design is optimal under a different design criterion, yet none is optimal in terms of BER. The designs considered are the MMSE [14], [15], MSSNR [44], minimum delay spread design [51] with quadratic or linear weighting (MDS-Q, MDS-L), and minimum inter-block interference (Min-IBI) [52].

done correctly. However, if null tones are present, then the output of the transmitter IFFT will not be uncorrelated. To see this, assume that the frequency domain data vector is $\mathbf{X}^T = [\mathbf{B}^T, \mathbf{0}^T]$, where $\mathbf{B}$ is a collection of QAM data and $\mathbf{0}$ is a vector of zeros. Assume that $\mathbf{B}$ is white and examine the autocorrelation of its inverse Fourier transform $\mathbf{x}$:

$$E[\mathbf{B}^*\mathbf{B}^T] = \mathbf{I}, \tag{13}$$

$$E[\mathbf{X}^*\mathbf{X}^T] = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & 0 \end{bmatrix}, \tag{14}$$

$$\begin{aligned} E[\mathbf{x}^*\mathbf{x}^T] &= E_N\left[(\mathcal{F}_N^*\mathbf{X})^*(\mathcal{F}_N^*\mathbf{X})^T\right] \\ &= \mathcal{F}_N E[\mathbf{X}^*\mathbf{X}^T]\mathcal{F}_N^H \neq \mathbf{I}. \end{aligned} \tag{15}$$

Because the autocorrelation matrix of $\mathbf{X}$ has a block of zeros, it is not identity, and cancellation of the FFT and IFFT matrices cannot occur. The more null tones there are, the more the IFFT output is correlated with itself.

Another similar way in which the uncorrelated input assumption can be broken is if an orthogonal frequency division multiple access (OFDMA) protocol is used. In the uplink of an OFDMA, the tones are partitioned into disjoint sets and allocated to the various users [56]. If there are four users, for example, each user transmits on a quarter of the tones and assumes that the remaining tones are null tones. The tones for a given user may have a continuous allocation, they may have an equally spaced allocation, or they may have a random allocation [56]. Assuming that one or more of the user's uplink channels is longer than the guard interval, channel shortening will need to be performed at the base station. In this case, all of the channels must be simultaneously shortened, either by a single channel shortener or by multiple channel shorteners. The crux of the problem is that each channel has a different user, and since each user uses only a subset of the channels, each user's transmitted signal is highly colored. It is not equivalent to model this as a single transmitter that uses all of the tones because each user has a different channel.

Both the MERRY and SAM algorithms assume an uncorrelated transmitted signal. Simulations have shown that for MERRY, if the transmitted signal is correlated, there will be residual energy in the channel outside of the desired window. However, it is difficult to prove this mathematically. For SAM, if the input is correlated up to some time separation τ_1 and the channel has memory τ_2, the received data is correlated up to a time separation of $\tau_1 + \tau_2$. Thus, if we shorten the correlation of the received data to length ν, we have to shorten the autocorrelation of the channel (and hence its memory) to $\nu - \tau_1$ rather than ν. In other words, we wind up making the channel significantly shorter than the cyclic prefix. By allocating effort to reducing some of the channel taps inside the window, we must spend less effort reducing the taps outside of the window, and the residual interference outside the desired window becomes higher than the case in which all of our assumptions hold.

The null tone assumption used by CNA is often violated as well. The null tones are almost always present, but they are usually not truly null. One of the predominant disadvantages of a multicarrier system is that the transmitted data has a high peak-to-average-power ratio (PAPR). This is because the time-domain data is a linear combination of N independent QAM symbols, and the central limit theorem says that for large N, the resulting data has a nearly Gaussian probability distribution. One common approach to reducing the PAPR is to transmit small values on the null tones in such a way that the PAPR is reduced [57]. Thus, these small correction terms are necessarily signal dependent. In this case, not only are the null tones not null, but the values on these tones are correlated with the signal. The presence of these correction terms raises the noise floor, and their correlation with the signal distorts the cost function.

[ONE OF THE ADVANTAGES OF SCCP MODULATION IS THAT THE TRANSMITTED SAMPLES STILL HAVE A FINITE ALPHABET; SINCE THEY ARE NOT GAUSSIAN, THEY DO NOT HAVE THE HIGH PEAK-TO-AVERAGE POWER RATIO OF MULTICARRIER SIGNALS.]

OPEN PROBLEMS

Research on adaptive channel shortening is truly just beginning. The few algorithms that have been proposed so far are not fully understood, and presumably there are many more algorithms awaiting discovery. We conclude with a description of issues that are not yet completely resolved or that have yet to be addressed in the literature.

Virtually all channel-shortening (adaptive and nonadaptive) designs in the literature take the form of a constrained optimization problem [58]. In the nonadaptive case, the ever-present constraint leads to algorithms that maximize a generalized Rayleigh quotient (or a product of many of them). In the adaptive case, this usually leads to algorithms that descend some cost surface subject to a unit-norm or unit-tap constraint on the channel shortener. Virtually all authors have chosen to implement the unit-norm constraint, and they have implemented it by renormalizing the filter after each update [15], [28], [31], [33], [35], [42]. One possible reason for this choice is that Al-Dhahir and Cioffi [14] have shown that the MMSE channel shortener with a unit-tap constraint always has a higher MSE than one with a unit-norm constraint, with the implication that a unit-norm constraint is preferable. However, the MSE is not directly related to the BER in a multicarrier system, so the implications of this result are debatable. In any case, the behavior of adaptive channel shorteners that renormalize after each update is not yet well understood. The study of the behavior of unit-norm constrained adaptive channel shorteners is thus an open problem. Along the same lines, other channel shorteners could be proposed that use different constraints or that enforce the constraint by some means other than renormalization.

The number of algorithms that perform adaptive channel shortening is currently very small compared to the number of adaptive equalization algorithms. Algorithm construction and analysis thus forms yet another open problem. Since most of the current algorithms only update once per block, they tend to have very slow convergence rates. For a wireless environment in which the channel changes significantly every ten or 20 blocks, such algorithms have difficulty tracking the changing environment. Fast versions of existing algorithms are of interest, as are algorithms that can update multiple times per block.

The performance of adaptive channel shorteners is heavily dependent on the choice of synchronization delay. Choosing this delay is equivalent to choosing the desired boundaries of each data block at the output of the channel shortener. Even within the range of delays that lead to relatively good performance, the performance is not a smooth function of the delay. One possible contribution to be made is the proposal and study of a heuristic method for choosing the delay in a blind fashion. Another contribution would be to determine analytically how the cost function for each algorithm depends on the delay, or at least to give bounds on the performance as a function of delay. One of the virtues of per-tone structures is that they are less sensitive to the choice of delay, but this problem is a significant issue for the design of a channel shortener.

Finally, MIMO and multiuser extensions are becoming increasingly important. If there are multiple transmit and receive antennas, the channel shortener(s) at the receiver must simultaneously shorten many channels. If this is in the context of a multicarrier code-division multiple access (MC-CDMA) system, then shortening the channels for the various users must be done before the signals can be despread and the different users' signals can be separated.

CONCLUDING REMARKS

The design of blind, adaptive channel shorteners requires fundamentally different approaches than the design of adaptive equalizers. Although the property restoral concept is still a good starting point, the properties that are expected to be present in a system requiring channel shortening are typically different from the properties expected to be present in a system requiring equalization. Several simple approaches have been proposed in the literature, but they are limited by various assumptions and they optimize a heuristic cost function rather than the bit rate or BER. Algorithms that address these issues would be of great utility. Additionally, more powerful algorithms that converge more rapidly would be of interest, especially since it is difficult to apply standard gradient descent algorithm acceleration techniques to the constrained gradient descent algorithms that have been proposed for adaptive channel shortening.

ACKNOWLEDGMENTS

The authors wish to thank Jaiganesh Balakrishnan of Texas Instruments, Bangalore, India; Jonathon Chambers of Cardiff University, United Kingdom; Brian Evans of the University of Texas, Austin, USA; Bill Sethares of the University of Wisconsin—Madison, USA; Koen Vanbleu of Broadcom UK Ltd., Belgium; and Geert Ysebaert of Alcatel Telecom, Belgium, for fruitful discussions of adaptive channel shortening.

AUTHORS

Richard K. Martin obtained dual B.S. degrees in physics and electrical engineering from the University of Maryland, College Park, in 1999 (summa cum laude), and an M.S. degree and a Ph.D. in electrical engineering from Cornell University in 2001 and 2004, respectively. Since August 2004, he has been an assistant professor at the Air Force Institute of Technology (AFIT) in Dayton, Ohio, where he is the signal processing curriculum chair and the graduate electrical engineering program chair. The AFIT Student Association has voted him "Instructor of the Quarter" for the Electrical and Computer Engineering Department on two occasions. His research interests include multicarrier equalization, blind, adaptive algorithms, reduced complexity equalizer design, and sparse adaptive filters. He has authored nine journal papers, 17 conference papers, and the book *Theory and Design of Adaptive Filters Answer Book*. He has three patents pending.

C. Richard Johnson, Jr. received a Ph.D. in electrical engineering, along with the first Ph.D. minor in art history, from Stanford University in 1977. He joined the faculty of Cornell University in 1981, where he is a Stephen H. Weiss Presidential Fellow and a professor of electrical and computer engineering. His primary research interest has been blind, adaptive equalization for mitigating intersymbol interference in communication systems. He has received several teaching awards, including selection by Eta Kappa Nu as the C. Holmes MacDonald Outstanding Teacher among young professors of electrical engineering in the United States in 1983, and selection in 2004 as a Stephen H. Weiss Presidential Fellow. He is a Fellow of the IEEE.

REFERENCES

[1] Z. Wang and G.B. Giannakis, "Wireless multicarrier communications," *IEEE Signal Processing Mag.*, vol. 17, pp. 29–48, May 2000.

[2] T. Starr, J.M. Cioffi, and P.J. Silvermann, *Understanding Digital Subscriber Line Technology*. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[3] *Radio Broadcasting System, Digital Audio Broadcasting (DAB) to Mobile, Portible, and Fixed Receivers*, ETS 300 401, 1995–1997.

[4] *Digital Video Broadcasting (DVB); Framing Structure, Channel Coding and Modulation for Digital Terrestrial Television*, ETSI EN 300 744 V1.4.1, 2001.

[5] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE 802.11a, 1999.

[6] *Broadband Radio Access Networks (BRAN); High Performance Radio Local Area Networks (HIPERLAN) Type 2; System Overview,* ETR101 683 114, 1999.

[7] R.D.J. van Nee, G.A. Awater, M. Morikura, H. Takanashi, M.A. Webster, and K.W. Halford, "New high-rate wireless LAN standards," *IEEE Commun. Mag.*, vol. 37, no. 12, pp. 82–88, Dec. 1999.

[8] D.H. Layer, "Digital radio takes to the road," *IEEE Spectr.*, vol. 38, pp. 40–46, July 2001.

[9] S. Galli, A. Scaglione, and K. Dostert, "Broadband is power: Internet access through the power line network," *IEEE Commun. Mag.*, vol. 41, no. 5, pp. 82–83, May 2003.

[10] C.R. Johnson, Jr., P. Schniter, T.J. Endres, J.D. Behm, D.R. Brown, and R.A. Casas, "Blind equalization using the constant modulus criterion: A review," *Proc. IEEE*, vol. 86, pp. 1927–1950, Oct. 1998.

[11] G.D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. 18, pp. 363–378, May 1972.

[12] F. Magee, "A comparison of compromise viterbi algorithm and standard equalization techniques over band-limited channels," *IEEE Trans. Commun.*, vol. 23, pp. 361–367, Mar. 1975.

[13] W. Younis and N. Al-Dhahir, "Joint prefiltering and MLSE equalization of space-time coded transmissions over frequency-selective channels," *IEEE Trans. Veh. Technol.*, vol. 51, pp. 144–154, Jan. 2002.

[14] N. Al-Dhahir and J.M. Cioffi, "Efficiently computed reduced-parameter input-aided MMSE equalizers for ML detection: A unified approach," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 903–915, May 1996.

[15] D.D. Falconer and F.R. Magee, "Adaptive channel memory truncation for maximum likelihood sequence estimation," *Bell Sys. Tech. J.*, vol. 9, pp. 1541–1562, Nov. 1973.

[16] D. Messerschmitt, "Design of a finite impulse response for the Viterbi algorithm and decision-feedback equalizer," in *Proc. Int. Conf. Comm.*, Minneapolis, MN, June 1974, pp. 37D.1–37D.5.

[17] W. Lee and F. Hill, "A maximum-likelihood sequence estimator with decision-feedback equalization," *IEEE Trans. Commun.*, vol. 25, no. 9, pp. 971–979, Sept. 1977.

[18] I. Medvedev and V. Tarokh, "A channel-shortening multiuser detector for DS-CDMA systems," in *Proc. 53rd Vehicular Technology Conf.*, Rhodes, Greece, May 2001, vol. 3, pp. 1834–1838.

[19] L. Tong, B.M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions," *IEEE Signal Processing Mag.*, vol. 21, no. 6, pp. 12–25, Nov. 2004.

[20] H. Sari, G. Karam, and I. Jeanclaude, "Frequency-domain equalization of mobile radio and terrestrial broadcast channels," in *Proc. IEEE Global Comm. Conf. (GLOBECOM)* 94, San Fransisco, CA, Nov. 1994, pp. 1–5.

[21] H. Sari, G. Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *IEEE Commun. Mag.*, vol. 33, no. 2, pp. 100–109, Feb. 1995.

[22] D.D. Falconer, S.L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Commun. Mag.*, vol. 40, pp. 58–66, Apr. 2002.

[23] J. Louveaux, L. Vandendorpe, and T. Sartenaer, "Cyclic prefixed single carrier and multicarrier transmission: Bit rate comparison," *IEEE Commun. Lett.*, vol. 7, no. 4, pp. 180–182, Apr. 2003.

[24] S. Qureshi, "Adaptive equalization," *Proc. IEEE*, vol. 73, no. 9, pp. 1349–1387, Sept. 1985.

[25] B. Widrow, J. McCool, M.G. Larimore, and C.R. Johnson, Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, no. 8, pp. 1151–1162, Aug. 1976.

[26] B. Widrow, "Thinking about thinking: The discovery of the LMS algorithm," *IEEE Signal Processing Mag.*, vol. 22, no. 1, pp. 100–102, 106, Jan. 2005.

[27] D.N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867–1875, Nov. 1980.

[28] J.S. Chow, J.M. Cioffi, and J.A.C. Bingham, "Equalizer training algorithms for multicarrier modulation systems," in *Proc. IEEE Int. Conf. Comm.*, Geneva, Switzerland, May 1993, pp. 761–76.

[29] D.L. Jones, "Property-restoral algorithms for blind equalization of OFDM," in *Proc. IEEE Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2003, pp. 619–622.

[30] J.R. Treichler, C.R. Johnson, Jr., and M.G. Larimore, *Theory and Design of Adaptive Filters*. New York: Prentice-Hall, 2001.

[31] R.K. Martin, J. Balakrishnan, W.A. Sethares, and C.R. Johnson, Jr., "A blind, adaptive TEQ for multicarrier systems," *IEEE Signal Processing Lett.*, vol. 9, no. 11, pp. 341–343, Nov. 2002.

[32] R.K. Martin, J. Balakrishnan, W.A. Sethares, W. Chung, and C.R. Johnson, Jr., "Methods and system for equalizing data in multicarrier receiver systems," patent pending, filed on March 17, 2003, application #20030210743.

[33] J. Balakrishnan, R.K. Martin, and C.R. Johnson, Jr., "Blind, adaptive channel shortening by sum-squared auto-correlation minimization (SAM)," *IEEE Trans. Signal Processing*, vol. 51, no. 12, pp. 3086–3093, Dec. 2003.

[34] J. Balakrishnan, R.K. Martin, and C.R. Johnson, Jr., "Methods and system for equalizing data," patent pending, filed on March 17, 2003, application #20030210742.

[35] M. de Courville, P. Duhamel, P. Madec, and J. Palicot, "Blind equalization of OFDM systems based on the minimization of a quadratic criterion," in *Proc. IEEE Int. Conf. Comm.*, Dallas, TX, June 1996, pp. 1318–1321.

[36] F. Romano and S. Barbarossa, "Non-data aided adaptive channel shortening for efficient multi-carrier systems," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Hong Kong SAR, China, Apr. 2003, vol. 4, pp. 233–236.

[37] R.K. Martin and C.R. Johnson, Jr., "Blind, adaptive per tone equalization for multicarrier receivers," in *Proc. Conf. Information Sciences and Systems*, Princeton, NJ, Mar. 2002.

[38] U. Tureli, H. Liu, and M.D. Zoltowski, "OFDM blind carrier estimation: ESPRIT," *IEEE Trans. Commun.*, vol. 48, pp. 1459–1461, Sept. 2000.

[39] J.-J. van de Beek, M. Sandell, and P.O. Borjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Signal Processing*, vol. 45, no. 7, pp. 1800–1805, July 1997.

[40] R.K. Martin, J.M. Walsh, and C.R. Johnson, Jr., "Low complexity MIMO blind, adaptive channel shortening," *IEEE Trans. Signal Processing*, vol. 53, no. 4, pp. 1324–1334, Apr. 2005.

[41] R. Nawaz and J.A. Chambers, "Robust blind adaptive channel shortening in impulsive noise environments," in *Proc. 12th Eur. Signal Processing Conf.*, Vienna, Austria, Sept. 2004.

[42] R. Nawaz and J.A. Chambers, "A novel single lag auto-correlation minimization (SLAM) algorithm for blind adative channel shortening," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, Philadelphia, PA, Mar. 2005, vol. 3, pp. 885–888.

[43] T. Miyajima and Z. Ding, "Second-order statistical approaches to channel shortening in multicarrier systems," *IEEE Trans. Signal Processing*, vol. 52, no. 11, pp. 3253–3264, Nov. 2004.

[44] P.J.W. Melsa, R.C. Younce, and C.E. Rohrs, "Impulse response shortening for discrete multitone transceivers," *IEEE Trans. Comm.*, vol. 44, pp. 1662–1672, Dec. 1996.

[45] J.M. Walsh and C.R. Johnson, Jr., "Series feedforward interconnected adaptive devices," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Canada, May 2004, vol. 2, pp. 445–448.

[46] Y. Lin, L. Liang, and S. Phoong, "A semi-blind eigen approach to time-domain equalizer design for VDSL systems," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, Mar. 2005, vol. 3, pp. 341–344.

[47] K. Van Acker, G. Leus, M. Moonen, O. van de Wiel, and T. Pollet, "Per tone equalization for DMT-based systems," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 109–119, Jan. 2001.

[48] M. Milosevic, L.F.C. Pessoa, B.L. Evans, and R. Baldick, "DMT bit rate maximization with optimal time domain equalizer filter bank architecture," in *Proc. IEEE Asilomar Conf. Signals, Systems, and Comp.*, Pacific Grove, CA, Nov. 2002, vol. 1, pp. 377–382.

[49] S. Trautmann and N.J. Fliege, "Perfect equalization for DMT systems without guard interval," *IEEE J. Select. Areas Commun.*, vol. 20, no. 5, pp. 987–996, June 2002.

[50] G. Ysebaert, K. Vanbleu, G. Cuypers, M. Moonen, and T. Pollet, "Combined RLS-LMS initialization for per tone equalizers in DMT-receivers," *IEEE Trans. Signal Processing*, vol. 51, no. 7, pp. 1916–1927, July 2003.

[51] R. Schur and J. Speidel, "An efficient equalization method to minimize delay spread in OFDM/DMT systems," in *Proc. IEEE Int. Conf. Commun.*, Helsinki, Finland, June 2001, vol. 5, pp. 1481–1485.

[52] S. Celebi, "Interblock interference (IBI) minimizing time-domain equalizer (TEQ) for OFDM," *IEEE Signal Processing Lett.*, vol. 10, no. 8, pp. 232–234, Aug. 2003.

[53] G.L. Stuber, *Principles of Mobile Communication*. Boston: Kluwer, 1996.

[54] G. Arslan, B.L. Evans, and S. Kiaei, "Equalization for discrete multitone receivers to maximize bit rate," *IEEE Trans. Signal Processing*, vol. 49, no. 12, pp. 3123–3135, Dec. 2001.

[55] K. Vanbleu, G. Ysebaert, G. Cuypers, M. Moonen, and K. Van Acker, "Bitrate maximizing time-domain equalizer design for DMT-based systems," *IEEE Trans. Commun.*, vol. 52, no. 6, pp. 871–876, June 2004.

[56] H. Wang and B. Chen, "Asymptotic distributions and peak power analysis for uplink OFDMA signals," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, May 2004, vol. 4, pp. 17–21.

[57] A. Gatherer and M. Polley, "Controlling clipping probability in DMT transmission," in *Proc. Asilomar Conf.*, Pacific Grove, California, Nov. 1997, pp. 578–584.

[58] R.K. Martin, K. Vanbleu, M. Ding, G. Ysebaert, M. Milosevic, B.L. Evans, M. Moonen, and C.R. Johnson, Jr., "Unification and evaluation of equalization structures and design algorithms for discrete multitone modulation systems," IEEE Trans. Signal Processing, vol. 53, no. 10, Oct. 2005.