# Adaptive Equalization

SHAHID U. H. QURESHI, SENIOR MEMBER, IEEE

*Invited Paper*

*Bandwidth-efficient data transmission over telephone and radio channels is made possible by the use of adaptive equalization to compensate for the time dispersion introduced by the channel. Spurred by practical applications, a steady research effort over the last two decades has produced a rich body of literature in adaptive equalization and the related more general fields of reception of digital signals, adaptive filtering, and system identification. This tutorial paper gives an overview of the current state of the art in adaptive equalization. In the first part of the paper, the problem of intersymbol interference (ISI) and the basic concept of transversal equalizers are introduced followed by a simplified description of some practical adaptive equalizer structures and their properties. Related applications of adaptive filters and implementation approaches are discussed. Linear and nonlinear receiver structures, their steady-state performance and sensitivity to timing phase are presented in some depth in the next part. It is shown that a fractionally spaced equalizer can serve as the optimum receive filter for any receiver. Decision-feedback equalization, decision-aided ISI cancellation, and adaptive filtering for maximum-likelihood sequence estimation are presented in a common framework. The next two parts of the paper are devoted to a discussion of the convergence and steady-state properties of least mean-square (LMS) adaptation algorithms, including digital precision considerations, and three classes of rapidly converging adaptive equalization algorithms: namely, orthogonalized LMS, periodic or cyclic, and recursive least squares algorithms. An attempt is made throughout the paper to describe important principles and results in a heuristic manner, without formal proofs, using simple mathematical notation where possible.*

## I. Introduction

The rapidly increasing need for computer communications has been met primarily by higher speed data transmission over the widespread network of voice-bandwidth channels developed for voice communications. A modulator–demodulator (modem) is required to carry digital signals over these analog passband (nominally 300- to 3000-Hz) channels by translating binary data to voice-frequency signals and back (Fig. 1). The thrust toward common carrier digital transmission facilities has also resulted in application of modem technology to line-of-sight terrestrial radio and satellite transmission, and recently to subscriber loops.

Analog channels deliver corrupted and transformed versions of their input waveforms. Corruption of the waveform —usually statistical—may be additive and/or multiplicative, because of possible background thermal noise, impulse noise, and fades. Transformations performed by the channel are frequency translation, nonlinear or harmonic distortion, and time dispersion.

In telephone lines, time dispersion results when the channel frequency response deviates from the ideal of constant amplitude and linear phase (constant delay). Equalization, which dates back to the use of loading coils to improve the characteristics of twisted-pair telephone cables for voice transmission, compensates for these nonideal characteristics by filtering.

A synchronous modem transmitter collects an integral number of bits of data at a time and encodes them into symbols for transmission at the signaling rate. In pulse amplitude modulation (PAM), each signal is a pulse whose amplitude level is determined by the symbol, e.g., amplitudes of $-3$, $-1$, 1, and 3 for quaternary transmission. In bandwidth-efficient digital communication systems, the effect of each symbol transmitted over a time-dispersive channel extends beyond the time interval used to represent that symbol. The distortion caused by the resulting overlap of received symbols is called intersymbol interference (ISI) [54]. This distortion is one of the major obstacles to reliable high-speed data transmission over low-background-noise channels of limited bandwidth. In its broad sense, the term "equalizer" applies to any signal processing device designed to deal with ISI.

It was recognized early in the quest for high-speed (4800-bit/s and higher rate) data transmission over telephone channels that rather precise compensation, or equalization, is required to reduce the intersymbol interference introduced by the channel. In addition, in most practical situations the channel characteristics are not known beforehand. For medium-speed (up to 2400-bit/s) modems, which effectively transmit 1 bit/Hz, it is usually adequate to design and use a compromise (or statistical) equalizer which compensates for the average of the range of expected channel amplitude and delay characteristics. However, the variation in the characteristics within a class of channels as in the lines found in the switched telephone network, is large enough so that automatic adaptive equalization is used nearly universally for speeds higher

Manuscript received July 16, 1984; revised March 6, 1985.

The author is with Transmission Products, Codex Corporation, Mansfield, MA 02048, USA.

0018-9219/85/0900-1349$01.00 ©1985 IEEE

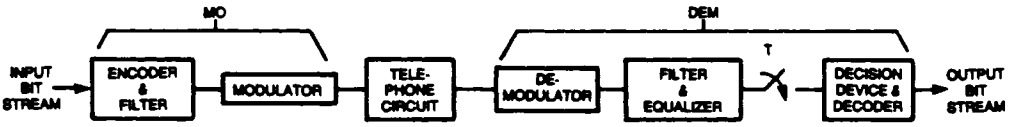


**Fig. 1.** Data transmission system.

than 2400 bits/s. Even 2400-bit/s modems now often incorporate this feature.

Voice-band telephone modems may be classified into one of three categories based on intended application: namely, for two-wire public switched telephone network (PSTN), four-wire point-to-point leased lines, and four-wire multipoint leased lines. PSTN modems can achieve 2400-bit/s two-wire full-duplex transmission by sending 4 bits/symbol and using frequency division to separate the signals in the two directions of transmission. Two-wire full-duplex modems using adaptive echo cancellation are now available for 2400- and 4800-bit/s transmission. Adaptive echo cancellation in conjunction with coded modulation will pave the way to 9600-bit/s full-duplex operation over two-wire PSTN circuits in the near future. At this time, commercially available leased-line modems operate at rates up to 16.8 kbits/s over conditioned point-to-point circuits, and up to 9.6 kbits/s over unconditioned multipoint circuits. An adaptive equalizer is an essential component of all these modems. (See [27] for a historical note on voice-band modem development.)

In radio and undersea channels, ISI is due to multipath propagation [87], [98], which may be viewed as transmission through a group of channels with differing relative amplitudes and delays. Adaptive equalizers are capable of correcting for ISI due to multipath in the same way as ISI from linear distortion in telephone channels. In radio-channel applications, an array of adaptive equalizers can also be used to perform diversity combining and cancel interference or jamming sources [6], [72]. One special requirement of radio-channel equalizers is that they be able to track the time-varying fading characteristics typically encountered. The convergence rate of the adaptation algorithm employed then becomes important during normal data transmission [87]. This is particularly true for 3-kHz-wide ionospheric high-frequency (HF), 3- to 30-MHz, radio channels which suffer from severe time dispersion and relatively rapid time variation and fading. Adaptive equalization has also been applied to slowly fading tropospheric scatter microwave digital radios, in the 4- to 11-GHz bands, at rates up to 200 Mbits/s [82].

In the last decade there has been considerable interest in techniques for full-duplex data transmission at rates up to 144 kbits/s over two-wire (nonloaded twisted-copper pair) subscriber loops [2], [22], [69], [106], [112]. Two competing schemes for achieving full-duplex transmission are time-compression multiplex or burst mode and adaptive echo cancellation. Some form of adaptive equalization is desirable, if not indispensible, for these baseband modems due to a number of factors: high transmission rates specially for the burst-mode scheme, attenuation distortion based on the desired range of subscriber loop lengths and gauges, and the presence of bridged taps, which cause additional time dispersion.

The first part of this paper, intended primarily for those not familiar with the field, is a simplified introduction to intersymbol interference and transversal equalizers, and an overview of some practical adaptive equalizer structures. In the concluding sections of the first part, we briefly mention other related applications of adaptive filters (such as echo cancellation, noise cancellation, and prediction) and discuss past and present implementation approaches.

Before presenting the introductory material, however, it seems appropriate to summarize the major areas of work in adaptive equalization, with reference to key papers and to sections of this article where these topics are discussed. (The interested reader should refer to Lucky [59] and Price [83] for a comprehensive survey of the literature and extensive bibliographies of work up to the early 1970s.) Unfortunately, use of some as yet undefined technical terms in the following paragraphs is unavoidable at this stage.

Nyquist's telegraph transmission theory [115] in 1928 laid the foundation for pulse transmission over band-limited analog channels. In 1960, Widrow and Hoff [109] presented a least mean-square (LMS) error adaptive filtering scheme which has been the workhorse adaptive equalization algorithm for the last decade and a half. However, research on adaptive equalization of PAM systems in the early 1960s centered on the basic theory and structure of zero-forcing transversal or tapped-delay-line equalizers with symbol interval tap spacing [55], [56]. In parallel, the theory and structure of linear receive and transmit filters [116], [117] were developed which minimize mean-square error for time-dispersive additive Gaussian noise channels [31]. By the late 1960s, LMS adaptive equalizers had been described and understood [33], [54], [84]. It was recognized that over highly dispersive channels even the best linear receiver falls considerably short of the matched filter performance bound, obtained by considering the reception of an isolated transmitted pulse [54]. Considerable research followed on the theory of optimum nonlinear receiver structures under various optimality criteria related to error probability [1], [59], [87]. This culminated in the development of the maximum-likelihood sequence estimator [24] using the Viterbi algorithm [25] and adaptive versions of such a receiver [17], [51], [61], [62], [88], [89], [103]. Another branch of research concentrated on a particularly simple suboptimum receiver structure known as the decision-feedback equalizer [3], [4], [12], [32], [71], [83], [93]. Linear feedback, or infinite impulse response (IIR), adaptive filters [47] have not been applied as adaptive equalizers due to lack of guaranteed stability, lack of a quadratic performance surface, and a minor performance gain over transversal equalizers [87]. As the advantages of double-sideband suppressed-carrier quadrature amplitude modulation (QAM) over single-sideband (SSB) and vestigial-sideband (VSB) modulation were recognized, previously known PAM equalizers were extended to complex-valued structures suitable for joint equalization of the in-phase and quadrature signals in a QAM receiver [15], [16], [49], [84], [119]. Transversal and decision-feedback equalizers

with forward-filter tap spacing that is less than the symbol interval were suggested in the late 1960s and early 1970s [6], [58], [71]. These fractionally spaced equalizers were first used in commercial telephone line modems [26], [53] and military tropospheric scatter radio systems [114] in the mid 1970s. Their theory and many performance advantages over conventional "symbol-spaced" equalizers have been the subject of several articles [39], [45], [91], [104]. The timing phase sensitivity of the mean-square error of symbol-spaced [64], fractionally spaced [45], [91], [104], and decision-feedback [94] equalizers has also been a research topic in the 1970s. Recently, interest in a nonlinear decision-aided receiver structure [85] now known as an ISI canceller has been revived by using a fractionally spaced equalizer as a matched filter [34], [80].

In the second part of the paper we develop the various receiver structures mentioned above and present their important steady-state properties. The first two sections are devoted to the definition of the baseband equivalent channel model and the development of an optimum receive filter which must precede further linear or nonlinear processing at the symbol rate. The next section on linear receivers shows that while the conventional, matched-filter plus symbol-spaced equalizer, and fractionally spaced forms of a linear receiver are equivalent when each is unrestricted (infinite in length), a finite-length fractionally spaced equalizer has significant advantages compared with a practical version of the conventional linear receiver. Nonlinear receivers are presented in the fourth section with a discussion of decision-feedback equalizers, decision-aided ISI cancellation, and adaptive versions of the maximum-likelihood sequence estimator. The final section of this part of the paper addresses timing phase sensitivity. A few important topics which have been excluded due to space limitations are: adaptive equalization of nonlinearities [5], [21], diversity-combining adaptive equalizer arrays to combat selective fades and interference in radio channels [6], [72], [114], and a particular passband equalizer structure [11], [77].

Until the early 1970s most of the equalization literature was devoted to equalizer structures and steady-state analysis [59], partly due to the difficulty of analyzing the transient performance of practical adaptive equalization algorithms. Since then some key papers [38], [65], [102], [111] have contributed to the understanding of the convergence of the LMS stochastic update algorithm for transversal equalizers, including the effect of channel characteristics on the rate of convergence. The third part of this paper is devoted to this subject and a discussion of digital precision considerations [7], [13], [36], [38]. The important topic of decision-directed convergence [60], [66] and self-recovering adaptive equalization algorithms [42], [95] has been omitted.

The demand for polled data communication systems using multipoint modems [26] which require fast setup at the central site receiver has led to the study of fast converging equalizers using a short preamble or training sequence. The fourth part of this paper summarizes three classes of fast-converging equalization algorithms. Some of the early work on this topic was directed toward orthogonalized LMS algorithms for partial response systems [8], [75], [88]. Periodic or cyclic sequences for equalizer training and methods for fast startup based on such sequences have been widely used in practice [43], [70], [76], [90], [91]. The third class of fast converging algorithms are self-orthogonalizing [37]. In 1973, Godard [41] described how the Kalman filtering algorithm can be used to estimate the LMS equalizer coefficient vector at each symbol interval. This was later recognized [20] to be a form of recursive least squares (RLS) estimation problem. Development of computationally efficient RLS algorithms has recently been a subject of intense research activity [46], [73], [78], [79], [87] leading to transversal [10], [11], [20] and lattice [52], [63], [74], [96], [97], [127] forms of the algorithm. Some of these algorithms have been applied to adaptive equalizers for HF radio modems [87], [126] which need to track a relatively rapidly time-varying channel. However, the extra complexity of these algorithms has so far prevented application to the startup problem of telephone line modems where periodic equalization [43] and other cost-effective techniques, e.g., [26], [125], are applicable.

Block least squares methods are widely used in speech coding [46], [122] to derive new adaptive filter parameters for each frame of the nonstationary speech waveform. Block implementations of adaptive filters have been suggested [130, and references therein] where the filter coefficients are updated once per block, and the output samples are computed a block at a time using transform-domain "high-speed convolution." Such implementations generally reduce the number of arithmetic operations at the expense of a more complex control structure, additional memory requirements, and a greater processing delay.

A brief view of the general direction of future work in adaptive equalization is given in the final part of the paper.

### *A. Intersymbol Interference*

Intersymbol interference arises in all pulse-modulation systems, including frequency-shift keying (FSK), phase-shift keying (PSK), and quadrature amplitude modulation (QAM) [54]. However, its effect can be most easily described for a baseband pulse-amplitude modulation (PAM) system. A model of such a PAM communication system is shown in Fig. 2. A generalized baseband equivalent model such as this can be derived for any linear modulation scheme. In this model, the "channel" includes the effects of the transmitter filter, the modulator, the transmission medium, and the demodulator.

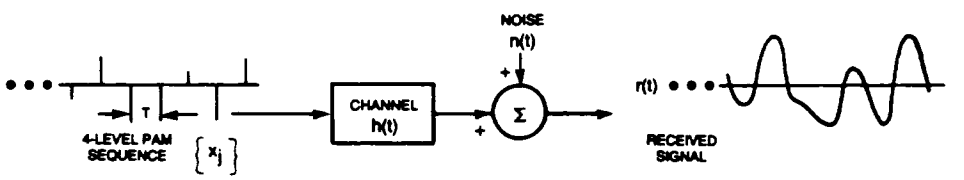


**Fig. 2.** Baseband PAM system model.

A symbol $x_m$, one of $L$ discrete amplitude levels, is transmitted at instant $mT$ through the channel, where $T$ seconds is the signaling interval. The channel impulse response $h(t)$ is shown in Fig. 3. The received signal $r(t)$ is

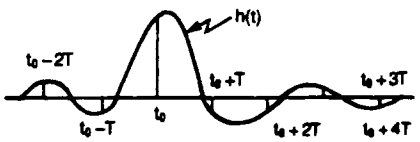


**Fig. 3.** Channel impulse response.

the superposition of the impulse responses of the channel to each transmitted symbol and additive white Gaussian noise $n(t)$

$$r(t) = \sum_j x_j h(t - jT) + n(t).$$

If we sample the received signal at instant $kT + t_0$, where $t_0$ accounts for the channel delay and sampler phase, we obtain

$$r(t_0 + kT) = x_k h(t_0) + \sum_{j \neq k} x_j h(t_0 + kT - jT) + n(t_0 + kT).$$

The first term on the right is the desired signal since it can be used to identify the transmitted amplitude level. The last term is the additive noise, while the middle sum is the interference from neighboring symbols. Each interference term is proportional to a sample of the channel impulse response $h(t_0 + iT)$ spaced a multiple $iT$ of symbol intervals $T$ away from $t_0$ as shown in Fig. 3. The ISI is zero if and only if $h(t_0 + iT) = 0$, $i \neq 0$; that is, if the channel impulse response has zero crossings at $T$-spaced intervals.

When the impulse response has such uniformly spaced zero crossings, it is said to satisfy Nyquist's first criterion. In frequency-domain terms, this condition is equivalent to

$$H'(f) = \sum_n H(f - n/T) = \text{constant for } |f| \leq 1/2T.$$

$H(f)$ is the channel frequency response and $H'(f)$ is the "folded" (aliased or overlapped) channel spectral response after symbol-rate sampling. The band $|f| \leq 1/2T$ is commonly referred to as the Nyquist or minimum bandwidth. When $H(f) = 0$ for $|f| > 1/T$ (the channel has no response beyond twice the Nyquist bandwidth), the folded response $H'(f)$ has the simple form

$$H'(f) = H(f) + H(f - 1/T), \qquad 0 \leq f \leq 1/T.$$

Fig. 4(a) and (d) shows the amplitude response of two linear-phase low-pass filters: one an ideal filter with Nyquist bandwidth and the other with odd (or vestigial) symmetry around $1/2T$ hertz. As illustrated in Fig. 4 (b) and (e), the folded frequency response of ech filter satisfies Nyquist's first criterion. One class of linear-phase filters, which is commonly referred to in the literature [54], [87], [113] and is widely used in practice [23], [118], is the raised-cosine family with cosine rolloff around $1/2T$ hertz.

In practice, the effect of ISI can be seen from a trace of the received signal on an oscilloscope with its time base synchronized to the symbol rate. Fig. 5 shows the outline of a trace (eye pattern) for a two-level or binary PAM system. If the channel satisfies the zero ISI condition, there are only two distinct levels at the sampling time $t_0$. The eye is then fully open and the peak distortion is zero. Peak distortion (Fig. 5) is the ISI that occurs when the data pattern is such that all intersymbol interference terms add to produce the maximum deviation from the desired signal at the sampling time.

The purpose of an equalizer, placed in the path of the received signal, is to reduce the ISI as much as possible to maximize the probability of correct decisions.

### *B. Linear Transversal Equalizers*

Among the many structures used for equalization the simplest is the transversal (tapped-delay-line or nonrecursive) equalizer shown in Fig. 6. In such an equalizer, the current and past values $r(t - nT)$ of the received signal are linearly weighted by equalizer coefficients (tap gains) $c_n$ and summed to produce the output. If the delays and tap-gain multipliers are analog, the continuous output of the equalizer $z(t)$ is sampled at the symbol rate and the samples go to the decision device. In the commonly used digital implementation, samples of the received signal at the symbol rate are stored in a digital shift register (or memory), and the equalizer output samples (sums of products) $z(t_0 + kT)$ or $z_k$ are computed digitally, once per symbol according to

$$z_k = \sum_{n=0}^{N-1} c_n r(t_0 + kT - nt)$$

where $N$ is the number of equalizer coefficients, and $t_0$ denotes sample timing.

The equalizer coefficients $c_n, n = 0, 1, \cdots, N-1$ may be chosen to force the samples of the combined channel and equalizer impulse response to zero at all but one of the $N$ $T$-spaced instants in the span of the equalizer. This is shown graphically in Fig. 7. Such an equalizer is called a zero-forcing (ZF) equalizer [55].

If we let the number of coefficients of a ZF equalizer increase without bound, we would obtain an infinite-length equalizer with zero ISI at its output. The frequency response $C(f)$ of such an equalizer is periodic, with a period equal to the symbol rate $1/T$ because of the $T$-second tap spacing. After sampling, the effect of the channel on the received signal is determined by the folded frequency response $H'(f)$. The combined response of the channel, in tandem with the equalizer, must satisfy the zero ISI condition or Nyquist's first criterion

$$C(f) H'(f) = 1, \qquad |f| \leq 1/2T.$$

From the above expression we see that an infinite-length zero-ISI equalizer is simply an inverse filter, which inverts the folded-frequency response of the channel. A finite-length ZF equalizer approximates this inverse. Such an inverse filter may excessively enhance noise at frequencies where the folded channel spectrum has high attenuation. This is undesirable, particularly for unconditioned telephone connections, which may have considerable attenuation distortion, and for radio channels, which may be subject to frequency-selective fades.

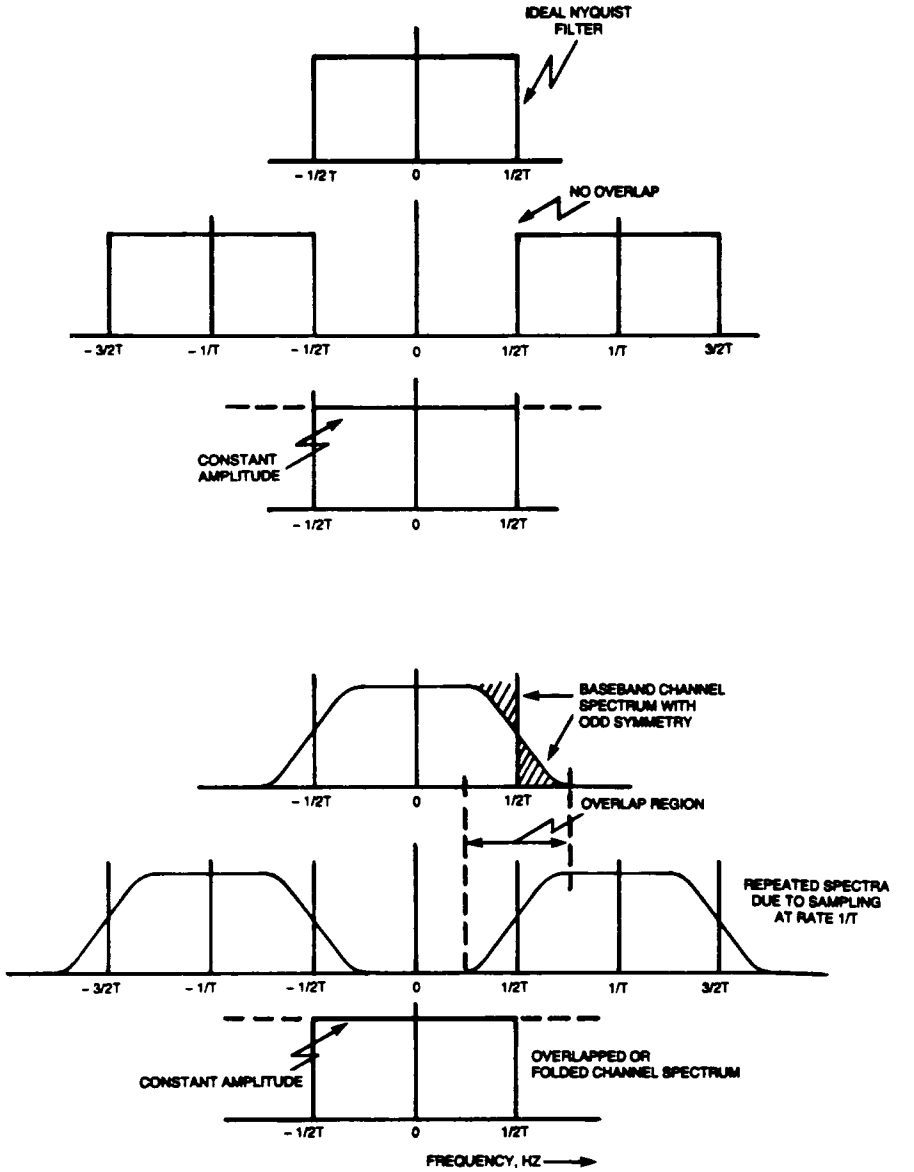


**Fig. 4.** Linear phase filters which satisfy Nyquist's first criterion.

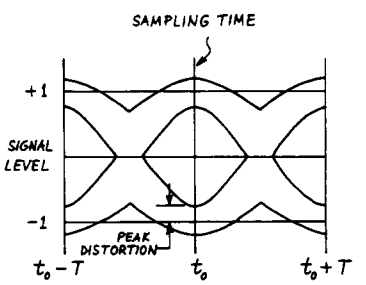


**Fig. 5.** Outline of a binary eye pattern.

Clearly, the ZF criterion neglects the effect of noise altogether. Also, a finite-length ZF equalizer is guaranteed to minimize the peak distortion or worst case ISI only if the peak distortion before equalization is less than 100 percent [55]; i.e., if a binary eye is initially open. However, at high speeds on bad channels this condition is often not met.

The least mean-square (LMS) equalizer [54] is more robust. Here the equalizer coefficients are chosen to minimize the mean-square error—the sum of squares of all the ISI terms plus the noise power at the output of the equalizer. Therefore, the LMS equalizer maximizes the signal-to-distortion ratio at its output within the constraints of the equalizer time span and the delay through the equalizer.

The delay introduced by the equalizer depends on the position of the main or reference tap of the equalizer. Typically, the tap gain corresponding to the main tap has the largest magnitude.

If the values of the channel impulse response at the sampling instants are known, the $N$ coefficients of the ZF and the LMS equalizers can be obtained by solving a set of $N$ linear simultaneous equations for each case.

Most current high-speed voice-band telephone-line modems use LMS equalizers because they are more robust in the presence of noise and large amounts of ISI, and superior to the ZF equalizers in their convergence properties. The same is generally true of radio-channel modems [72], [82], [114] except in one case [23] where the ZF equalizer was selected due to its implementation simplicity.

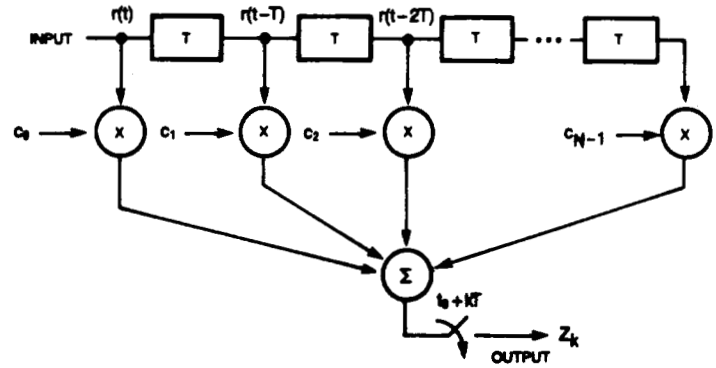


**Fig. 6.** Linear transversal equalizer.

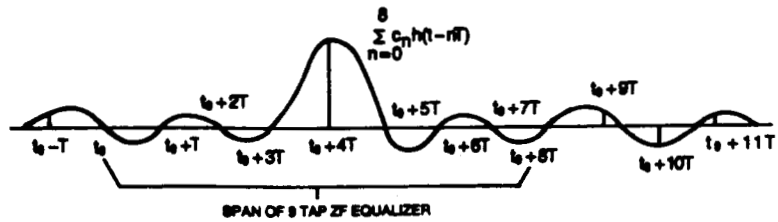


**Fig. 7.** Combined impulse response of a channel and zero-forcing equalizer in tandem.

### C. Automatic Synthesis

Before regular data transmission begins, automatic synthesis of the ZF or LMS equalizers for unknown channels, which involves the iterative solution of one of the above-mentioned sets of simultaneous equations, may be carried out during a training period. (In certain applications, such as microwave digital radio systems, and remote site receivers in a multipoint telephone modem network [43], the adaptive equalizers are required to bootstrap in a decision-directed mode (see Section I-E) without the help of a training sequence from the transmitter.)

During the training period, a known signal is transmitted and a synchronized version of this signal is generated in the receiver to acquire information about the channel characteristics. The training signal may consist of periodic isolated pulses or a continuous sequence with a broad, uniform spectrum such as the widely used maximum-length shift-register or pseudo-noise (PN) sequence [9], [54], [76], [118], [119]. The latter has the advantage of much greater average power, and hence a larger received signal-to-noise ratio (SNR) for the same peak transmitted power. The training sequence must be at least as long as the length of the equalizer so that the transmitted signal spectrum is adequately dense in the channel bandwidth to be equalized.

Given a synchronized version of the known training signal, a sequence of error signals $e_k = z_k - x_k$ can be computed at the equalizer output (Fig. 8), and used to adjust the equalizer coefficients to reduce the sum of the squared errors. The most popular equalizer adjustment method involves updates to each tap gain during each symbol interval. Iterative solution of the coefficients of the equalizer is possible because the mean-square error (MSE) is a quadratic function of the coefficients. The MSE may be envisioned as an $N$-dimensional paraboloid (punch bowl) with a bottom or minimum. The adjustment to each tap gain is in a direction opposite to an estimate of the gradient of the MSE with respect to that tap gain. The idea is to move the set of equalizer coefficients closer to the unique optimum set corresponding to the minimum MSE. This symbol-by-symbol procedure developed by Widrow and Hoff [109] is commonly referred to as the continual or stochastic update method because, instead of the true gradient of the mean-square error,

$$\partial E\left[e_k^2\right]/\partial c_n(k)$$

a noisy but unbiased estimate

$$\partial e_k^2/\partial c_n(k) = 2e_k r(t_0 + kT - nT)$$

is used. Thus the tap gains are updated according to

$$c_n(k+1) = c_n(k) - \Delta e_k r(t_0 + kT - nT), \qquad n = 0,1,\cdots,N-1$$

where $c_n(k)$ is the $n$th tap gain at time $k$, $e_k$ is the error signal, and $\Delta$ is a positive adaptation constant or step size.

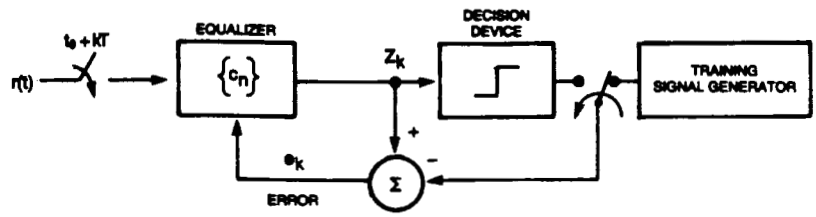


**Fig. 8.** Automatic adaptive equalizer.

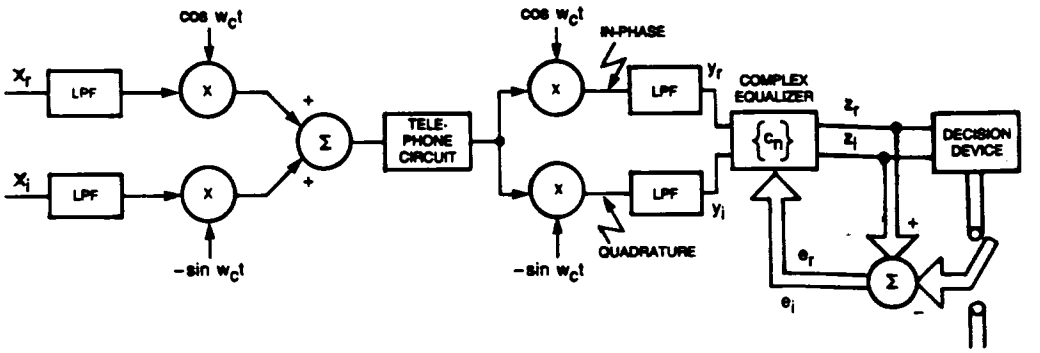


**Fig. 9.** QAM system with baseband complex adaptive equalizer.

### *D. Equalizer Convergence*

The exact convergence behavior of the stochastic update method is hard to analyze (see Section III-B). However, for a small step size and a large number of iterations, the behavior is similar to the steepest descent algorithm, which uses the actual gradient rather than a noisy estimate.

Here we list some general convergence properties: a) fastest convergence (or shortest settling time) is obtained when the (folded)-power spectrum of the symbol-rate sampled equalizer input is flat, and when the step size $\Delta$ is chosen to be the inverse of the product of the received signal power and the number of equalizer coefficients; b) the larger the variation in the above-mentioned folded-power spectrum, the smaller the step size must be, and therefore the slower the rate of convergence; c) for systems where sampling causes aliasing (channel foldover or spectral overlap), the convergence rate is affected by the channel-delay characteristics and the sampler phase, because they affect the aliasing. This will be explained more fully later.

### *E. Adaptive Equalization*

After the initial training period (if there is one), the coefficients of an adaptive equalizer may be continually adjusted in a decision-directed manner. In this mode, the error signal $e_k = z_k - \hat{x}_k$ is derived from the final (not necessarily correct) receiver estimate $\{\hat{x}_k\}$ of the transmitted sequence $\{x_k\}$. In normal operation, the receiver decisions are correct with high probability, so that the error estimates are correct often enough to allow the adaptive equalizer to maintain precise equalization. Moreover, a decision-directed adaptive equalizer can track slow variations in the channel characteristics or linear perturbations in the receiver front end, such as slow jitter in the sampler phase.

The larger the step size, the faster the equalizer tracking capability. However, a compromise must be made between fast tracking and the excess mean-square error of the equalizer. The excess MSE is that part of the error power in excess of the minimum attainable MSE (with tap gains frozen at their optimum settings). This excess MSE, caused by tap gains wandering around the optimum settings, is directly proportional to the number of equalizer coefficients, the step size, and the channel noise power. The step size that provides the fastest convergence results in an MSE which is, on the average, 3 dB worse than the minimum achievable MSE. In practice, the value of the step size is selected for fast convergence during the training period and then reduced for fine tuning during the steady-state operation (or data mode).

### *F. Equalizers for QAM Systems*

So far we have only discussed equalizers for a baseband PAM system. Modern high-speed voice-band modems almost universally use phase-shift keying (PSK) for lower speeds, e.g., 2400 to 4800 bits/s, and combined phase and amplitude modulation or, equivalently, quadrature amplitude modulation (QAM) [54], for higher speeds, e.g., 4800 to 9600 or even 16 800 bits/s. At the high rates, where noise and other channel distortions become significant, modems using coded forms of QAM such as trellis-coded modulation [27], [105], [107] are being introduced to obtain improved performance. QAM is as efficient in bits per second per hertz as vestigial or single-sideband amplitude modulation, yet enables a coherent carrier to be derived and phase jitter to be tracked using easily implemented decision-directed carrier recovery techniques [49]. A timing waveform with negligible timing jitter can also be easily recovered from QAM signals. This property is not shared by vestigial sideband amplitude-modulation (AM) systems [128].

Fig. 9 shows a generic QAM system, which may also be used to implement PSK or combined amplitude and phase modulation. Two double-sideband suppressed-carrier AM signals are superimposed on each other at the transmitter and separated at the receiver, using quadrature or orthogonal carriers for modulation and demodulation. It is convenient to represent the in-phase and quadrature channel low-pass filter output signals in Fig. 9 by $y_r(t)$ and $y_i(t)$, as the real and imaginary parts of a complex valued signal $y(t)$. (Note that the signals are real, but it will be convenient to use complex notation.)

The baseband equalizer [84], with complex coefficients $c_n$, operates on samples of this complex signal $y(t)$ and produces complex equalized samples $z(k) = z_r(k) + jz_i(k)$, as shown in Fig. 10. This figure illustrates more concretely the concept of a complex equalizer as a set of four real transversal filters (with cross-coupling) for two inputs and two outputs. While the real coefficients $c_{rn}$, $n = 0, \cdots, N-1$, help to combat the intersymbol interference in the in-phase and quadrature channels, the imaginary coefficients $c_{in}$, $n = 0, \cdots, N-1$, counteract the cross-interference between the two channels. The latter may be caused by

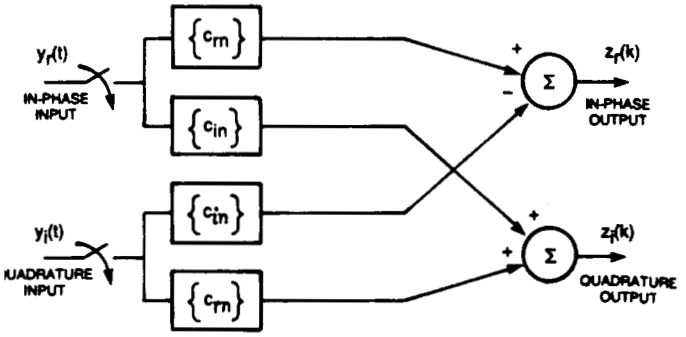


**Fig. 10.** Complex transversal equalizer for QAM modems.

asymmetry in the channel characteristics around the carrier frequency.

The coefficients are adjusted to minimize the mean of the squared magnitude of the complex error signal, $e(k) = e_r(k) + je_i(k)$, where $e_r$ and $e_i$ are the differences between $z_r$ and $z_i$, and their desired values. The update method is similar to the one used for the PAM equalizer except that all variables are complex-valued

$$c_n(k+1) = c_n(k) - \Delta e_k y^*(t_0 + kT - nT), \qquad n = 0,1,\cdots,N-1$$

where $y^*$ is the complex conjugate of $y$. Again, the use of complex notation allows the writing of this single concise equation, rather than two separate equations involving four real multiplications, which is what really has to be implemented.

The complex equalizer can also be used at passband [6], [16] to equalize the received signal before demodulation, as shown in Fig. 11. Here the received signal is split into its in-phase and quadrature components by a pair of so-called phase-splitting filters, with identical amplitude responses and phase responses that differ by 90°. The complex passband signal at the output of these filters is sampled at the symbol rate and applied to the equalizer delay line in the same way as at baseband. The complex output of the equalizer is demodulated, via multiplication by a complex exponential as shown in Fig. 11, before decisions are made and the complex error computed. Further, the error signal is remodulated before it is used in the equalizer adjustment algorithm. The main advantage of implementing the equalizer in the passband is that the delay between the demodulator and the phase error computation circuit is reduced to the delay through the decision device. Fast phase jitter can be tracked more effectively because the delay through the equalizer is eliminated from the phase correction loop. The same advantage can be attained with a baseband equalizer by putting a jitter-tracking loop after the equalizer.

### G. Decision-Feedback Equalizers

We have discussed placements and adjustment methods for the equalizer, but the basic equalizer structure has remained a linear and nonrecursive filter. A simple nonlinear equalizer [3], [4], [32], [71], [93], which is particularly useful for channels with severe amplitude distortion, uses decision feedback to cancel the interference from symbols which have already been detected. Fig. 12 shows such a decision-feedback equalizer (DFE). The equalized signal is the sum of the outputs of the forward and feedback parts of the equalizer. The forward part is like the linear transversal equalizer discussed earlier. Decisions made on the equalized signal are fed back via a second transversal filter. The basic idea is that if the value of the symbols already detected are known (past decisions are assumed to be correct), then the ISI contributed by these symbols can be canceled exactly, by subtracting past symbol values with appropriate weighting from the equalizer output. The weights are samples of the tail of the system impulse response including the channel and the forward part of the equalizer.

The forward and feedback coefficients may be adjusted simultaneously to minimize the MSE. The update equation

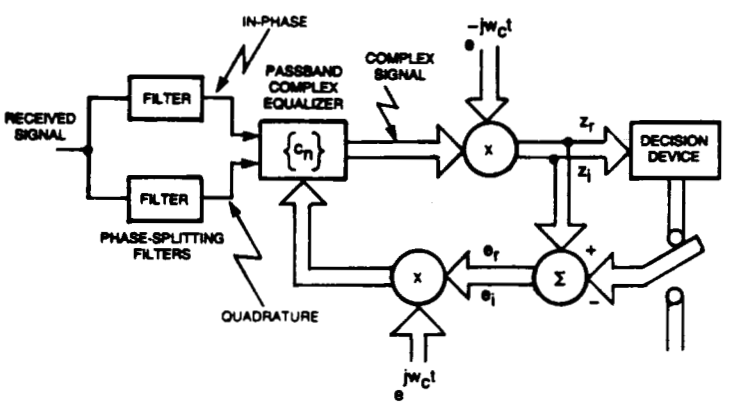


**Fig. 11.** Passband complex adaptive equalizer for QAM system.

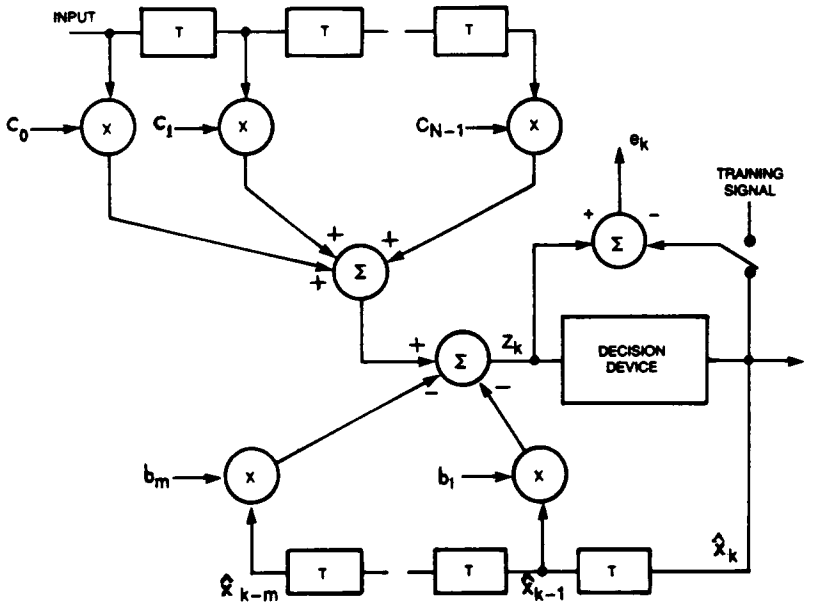


**Fig. 12.** Decision-feedback equalizer.

for the forward coefficients is the same as for the linear equalizer. The feedback coefficients are adjusted according to

$$b_m(k+1) = b_m(k) + \Delta e_k \hat{x}_{k-m}, \qquad m = 1, \cdots, M$$

where $\hat{x}_k$ is the $k$th symbol decision, $b_m(k)$ is the $m$th feedback coefficient at time $k$, and there are $M$ feedback coefficients in all. The optimum LMS settings of $b_m, m = 1, \cdots, M$, are those that reduce the ISI to zero, within the span of the feedback part, in a manner similar to a ZF equalizer. Note that since the output of the feedback section of the DFE is a weighted sum of noise-free past decisions, the feedback coefficients play no part in determining the noise power at the equalizer output.

Given the same number of overall coefficients, does a DFE achieve less MSE than a linear equalizer? There is no definite answer to this question. The performance of each type of equalizer is influenced by the particular channel characteristics and sampler phase, as well as the actual number of coefficients and the position of the reference or main tap of the equalizer. However, the DFE can compensate for amplitude distortion without as much noise enhancement as a linear equalizer. The DFE performance is also less sensitive to the sampler phase [94].

An intuitive explanation for these advantages is as follows: The coefficients of a linear transversal equalizer are selected to force the combined channel and equalizer impulse response to approximate a unit pulse. In a DFE, the ability of the feedback section to cancel the ISI, because of a number of past symbols, allows more freedom in the choice of the coefficients of the forward section. The combined impulse response of the channel and the forward section may have nonzero samples following the main pulse. That is, the forward section of a DFE need not approximate the inverse of the channel characteristics, and so avoids excessive noise enhancement and sensitivity to sampler phase.

When a particular incorrect decision is fed back, the DFE output reflects this error during the next few symbols as the incorrect decision traverses the feedback delay line. Thus there is a greater likelihood of more incorrect decisions following the first one, i.e., error propagation. Fortunately, the error propagation in a DFE is not catastrophic. On typical channels, errors occur in short bursts that degrade performance only slightly.

### *H. Fractionally Spaced Equalizers*

A fractionally spaced transversal equalizer [6], [39], [45], [58], [71], [91], [104] is shown in Fig. 13. The delay-line taps

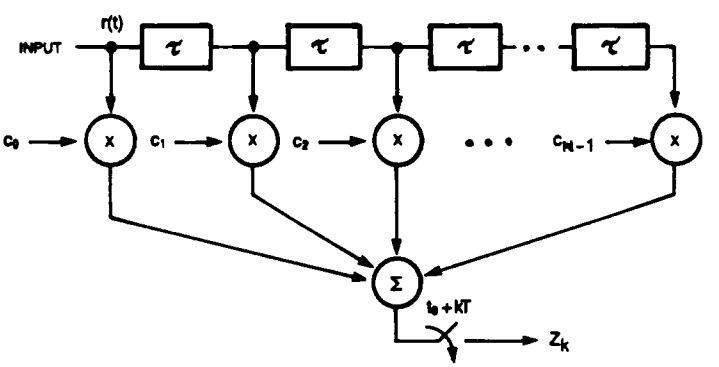


**Fig. 13.** Fractionally spaced equalizer.

of such an equalizer are spaced at an interval $\tau$ which is less than, or a fraction of, the symbol interval $T$. The tap spacing $\tau$ is typically selected such that the bandwidth occupied by the signal at the equalizer input is $|f| < 1/2\tau$, i.e., $\tau$-spaced sampling satisfies the sampling theorem. In an analog implementation, there is no other restriction on $\tau$, and the output of the equalizer can be sampled at the symbol rate. In a digital implementation $\tau$ must be $KT/M$, where $K$ and $M$ are integers and $M > K$. (In practice, it is convenient to choose $\tau = T/M$, where $M$ is a small integer, e.g., 2.) The received signal is sampled and shifted into the equalizer delay line at a rate $M/T$ and one output is produced each symbol interval (for every $M$ input samples). In general, the equalizer output is given by

$$z_k = \sum_{n=0}^{N-1} c_n r(t_0 + kT - nKT/M).$$

The coefficients of a $KT/M$ equalizer may be updated once per symbol based on the error computed for that symbol, according to

$$c_n(k+1) = c_n(k) - \Delta e_k r(t_0 + kT - nKT/M), \quad n = 0, 1, \cdots, N-1.$$

It is well known (see Section II) that the optimum receive filter in a linear modulation system is the cascade of a filter matched to the actual channel, with a transversal $T$-spaced equalizer [14], [24], [31]. The fractionally spaced equalizer, by virtue of its sampling rate, can synthesize the best combination of the characteristics of an adaptive matched filter and a $T$-spaced equalizer, within the constraints of its length and delay. A $T$-spaced equalizer, with symbol-rate sampling at its input, cannot perform matched filtering. An FSE can effectively compensate for more severe delay distortion and deal with amplitude distortion with less noise enhancement than a $T$ equalizer.

Consider a channel whose amplitude and envelope delay characteristics around one band edge $f_c - 1/2T$ hertz differ markedly from the characteristics around the other band edge $f_c + 1/2T$ hertz in a QAM system with a carrier frequency of $f_c$ hertz. Then the symbol-rate sampled or folded channel-frequency response is likely to have a rapid transition in the area of spectral overlap. It is difficult for a typical $T$ equalizer, with its limited degrees of freedom (number of taps), to manipulate such a folded channel into one with a flat frequency response. An FSE, on the other hand, can independently adjust the signal spectrum (in amplitude and phase) at the two band-edge regions before symbol-rate sampling (and spectral overlap) at the equalizer output, resulting in significantly improved performance.

A related property of an FSE is the insensitivity of its performance to the choice of sampler phase. This distinction between the conventional $T$-spaced and fractionally spaced equalizers can be heuristically explained as follows: First, symbol-rate sampling at the input to a $T$ equalizer causes spectral overlap or aliasing, as explained in connection with Fig. 4. When the phases of the overlapping components match they add constructively, and when the phases are 180° apart they add destructively, which results in the cancellation or reduction of amplitude as shown in Fig. 14. Variation in the sampler phase or timing instant corresponds to a variable delay in the signal path; a linear phase component with variable slope is added to the signal spectrum. Thus changes in the sampler phase strongly influence the effects of aliasing; i.e., they influence the amplitude and delay characteristics in the spectral overlap region of the sampled equalizer input. The minimum MSE achieved by the $T$ equalizer is, therefore, a function of the sampler phase. In particular, when the sample phase causes cancellation of the band-edge ($|f| = 1/2T$ hertz) components, the equalizer cannot manipulate the null into a flat spectrum at all, or at least without significant noise enhancement (if the null is a depression rather than a total null).

In contrast, there is no spectral overlap at the input to an

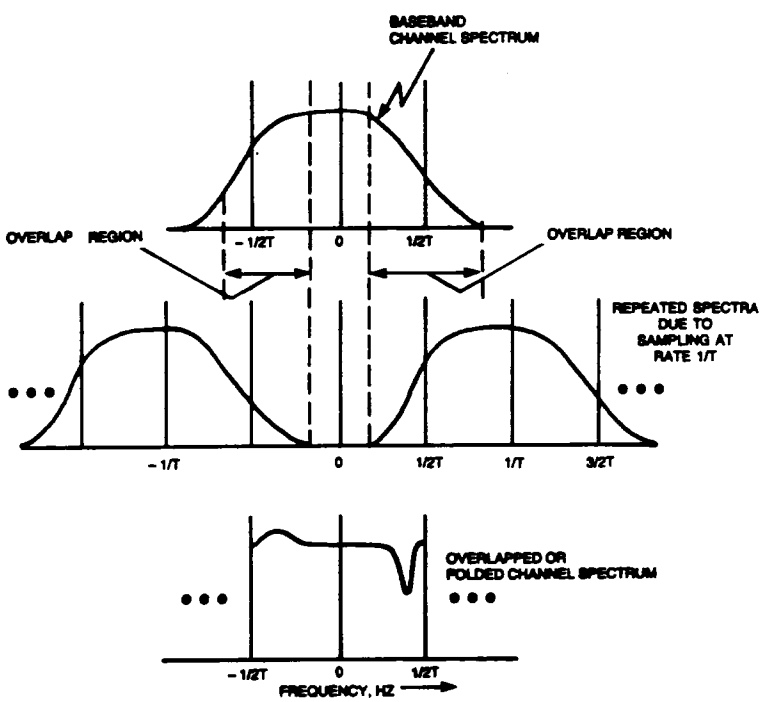


**Fig. 14.** Spectral overlap at the input to a $T$ equalizer.

FSE. Thus the sensitivity of the minimum MSE, achieved with an FSE with respect to the sampler phase, is typically far smaller than with a $T$ equalizer.

Comparison of numerical performance results of $T$ and $T/2$ equalizers for QAM systems operating over representative voice-grade telephone circuits [91] has shown the following properties: a) a $T/2$ equalizer with the same number of coefficients (half the time span) performs almost as well or better than a $T$ equalizer; b) a pre-equalizer receive shaping filter is not required with a $T/2$ equalizer; c) for channels with severe band-edge delay distortion, the $T$ equalizer performs noticeably worse than a $T/2$ equalizer regardless of the choice of sampler phase.

### *I. Other Applications*

While the primary emphasis of this paper is on adaptive equalization for data transmission, a number of the topics covered are relevant to other applications of adaptive filters. In this section, generic forms of adaptive filtering applications are introduced to help in establishing the connection between the material presented in later sections and the application of interest. But first let us briefly mention applications of automatic or adaptive equalization in areas other than data transmission over radio or telephone channels.

One such application is generalized automatic channel equalization [57], where the entire bandwidth of the channel is to be equalized without regard to the modulation scheme or transmission rate to be used on the channel. The tap spacing and input sample rate are selected to satisfy the sampling theorem, and the equalizer output is produced at the same rate. During the training mode, a known signal is transmitted, which covers the bandwidth to be equalized. The difference between the equalizer output and a synchronized reference training signal is the error signal. The tap gains are adjusted to minimize the MSE in a manner similar to that used for an automatic equalizer for synchronous data transmission.

Experimental use of fixed transversal equalizers has also been made in digital magnetic recording systems [99]. The recording method employed in such a case must be linearized by using an ac bias. Having linearized the "channel," equalization can be employed to combat intersymbol interference at increased recording densities, using a higher symbol rate or multilevel coding.

Fig. 15 shows a general form of an adaptive filter with input signals $x$ and $y$, output $z$, and error $e$. The parameters of an LMS adaptive filter are updated to minimize the mean-square value of the error $e$. In the following paragraphs, we point out how the adaptive filter is used in different applications by listing how $x$, $y$, $z$, and $e$ are interpreted for each case.

*Equalization:*

- $y$ Received signal (filtered version of transmitted data signal) plus noise uncorrelated with the data signal.
- $x$ Detected data signal.
- $z$ Equalized signal used to detect received data.
- $e$ Residual intersymbol interference plus noise.

*Echo Cancellation:* Echo cancellation is a form of a general system identification problem, where the system to be identified is the echo path linear system. The coefficients of a transversal echo canceler converge in the mean to the echo path impulse response samples.

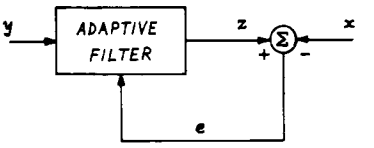


**Fig. 15.** General form of an adaptive filter.

*i) Voice:*

- $y$ Far-end voice signal plus uncorrelated noise.
- $x$ Echo of far-end voice plus near-end voice plus noise.
- $z$ Estimated echo of far-end voice.
- $e$ Near-end voice plus residual echo plus noise.

Adaptation is typically carried out in the absence of the near-end voice signal. When double talk is detected (both near- and far-end signals present), update of the echo canceler coefficients is inhibited [69].

*ii) Data:*

- $y$ Transmitted data signal.
- $x$ Echo of transmitted data signal plus received signal plus noise.
- $z$ Estimated echo of transmitted data signal.
- $e$ Received signal plus residual echo plus noise.

Filter adaptation is typically required to be continued in the presence of a large interfering received signal which is uncorrelated with the transmitted data [69], [108]. A method proposed in [22] involves locally generating a delayed replica of the received signal and subtracting it from $e$ before using the residual for echo canceler update.

*Noise Cancellation:*

- $y$ Noise source correlated with noise in $x$.
- $x$ Desired signal plus noise.
- $z$ Estimate of noise in $x$.
- $e$ Desired signal plus residual noise.

One example is that of canceling noise from the pilot's speech signal in the cockpit of an aircraft [110]. In this case, $y$ may be the pickup from a microphone in the pilot's helmet, and $x$ is the ambient noise picked up by another microphone placed in the cockpit. See [110] for a number of other interesting applications of noise and periodic interference cancellation, e.g., to electrocardiography.

*Prediction:*

- $y$ Delayed version of original signal.
- $x$ Original signal.
- $z$ Predicted signal.
- $e$ Prediction error or residual.

A well-known example is linear predictive coding (LPC) of speech where the end result is the set of estimated LPC coefficients [46]. Due to the nonstationary nature of the speech signal, LPC coefficients are typically obtained separately for each new frame (10 to 25 ms) of the speech signal.

In adaptive differential pulse code modulation (ADPCM) of speech, the purpose of adaptive prediction is to generate a residual signal with less variance so that it can be quantized and represented by fewer bits for transmission [46]. In this case:

- $y$ Reconstructed speech signal = quantized residual plus past prediction.
- $x$ Original speech signal.
- $z$ Prediction
- $e$ Residual to be quantized for transmission.

Note that the reconstructed speech signal is used for $y$ instead of a delayed version of the original speech signal, and the predictor coefficients are updated using the quantized residual instead of $e$. Both the reconstructed speech signal and the quantized residual signal are also available at the ADPCM decoder so that the predictor coefficients at the decoder can be adapted in a manner identical to that used at the ADPCM encoder.

*Adaptive Arrays:* A further generalization of the adaptive filter of Fig. 15 is shown in Fig. 16 where a number of input signals are processed through an array of adaptive

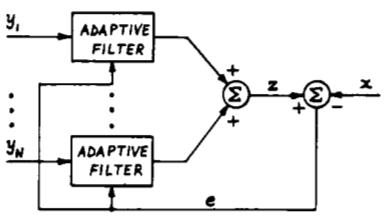


**Fig. 16.** Adaptive filter array.

filters whose outputs are summed together. Such adaptive arrays are useful in diversity combining [6], [114] and in dealing with jamming or spatially distributed interference [72], [110].

### *J. Implementation Approaches*

One may divide the methods of implementing adaptive equalizers into the following general categories: analog, hardwired digital, and programmable digital.

Analog adaptive equalizers, with inductor–capacitor ($LC$) tapped-delay lines and switched ladder attenuators as tap gains, were among the first implementations for voice-band modems. The switched attenuators later gave way to field-effect transistors (FETs) as the variable gain elements. Analog equalizers were soon replaced by digitally implemented equalizers for reduced size and increased accuracy. Recently, however, there is renewed interest in large-scale integrated (LSI) analog implementations of adaptive filters based on switched capacitor technology [2], [112]. Here the equalizer input is sampled but not quantized. The sampled analog values are stored and transferred as charge packets. In one implementation [2], the adaptation circuitry is digital. The variable tap gains are typically stored in digital memory locations and the multiplications between the analog sample values and the digital tap gains take place in analog fashion via multiplying digital-to-analog converters. In another case [112], a five-tap adaptive transversal filter has been fabricated on a single integrated circuit (IC) using an all-analog implementation approach combining switched capacitor and charge-coupled device (CCD) technologies. The IC consists of a five-tap CCD delay line, a convolver, five correlators (one for each tap gain), an offset error canceler, and an error signal generator. Integrators and four-quadrant analog multipliers are implemented in switched capacitor technology. The IC can be configured for use as an echo canceler, a linear equalizer, or a decision-feedback equalizer at sampling rates up to 250 kHz. Analog or mixed analog–digital implementations have significant potential in applications, such as digital radio and digital subscriber loop transmission, where symbol rates are high enough to make purely digital implementations difficult.

The most widespread technology of the last decade for voice-band modem adaptive equalizer implementation may be classified as hardwired digital technology. In such implementations, the equalizer input is made available in sampled and quantized form suitable for storage in digital shift registers. The variable tap gains are also stored in shift registers and the formation and accumulation of products takes place in logic circuits connected to perform digital arithmetic. This class of implementations is characterized by the fact that the circuitry is hardwired for the sole purpose of performing the adaptive equalization function with a predetermined structure. Examples include the early units based on metal–oxide–semiconductor (MOS) shift registers and transistor–transistor logic (TTL) circuits. Later implementations [53], [100] were based on MOS LSI circuits with dramatic savings in space, power dissipation, and cost.

A hardwired digital adaptive filtering approach described in [106] for a 144-kbit/s digital subscriber loop modem is based on a random-access memory (RAM) table lookup structure [124]. Both an echo canceler (EC) and a decision-feedback equalizer (DFE), where inputs are the transmit and receive binary data sequences, respectively, can be implemented in this way. An output signal value is maintained and updated in the RAM for each of the $2^N$ possible states of an $N$-tap transversal filter with a binary input sequence. Such a structure is not restricted to be linear and, therefore, can adapt to compensate for nonlinearities [123]. A two-chip realization of a digital subscriber loop modem based on a joint EC-DFE RAM structure is described in [121].

The most recent trend in implementing voice-band modem adaptive equalizers is toward programmable digital signal processors [43], [81], [101], [120]. Here, the equalization function is performed in a series of steps or instructions in a microprocessor or a digital computation structure specially configured to efficiently perform the type of digital arithmetic (e.g., multiply and accumulate) required in digital signal processing. The same hardware can then be time-shared to perform functions such as filtering, modulation, and demodulation in a modem. Perhaps the greatest advantage of programmable digital technology is its flexibility, which permits sophisticated equalizer structures and training procedures to be implemented with ease.

For microwave digital radio systems, adaptive equalizers have been implemented both in the passband at the intermediate frequency (IF) stage and at baseband [98]. Passband equalizers are analog by necessity, e.g., an amplitude slope equalizer [23] at 70-MHz IF, and a dynamic resonance equalizer using p-i-n and varactor diodes [82] at 140-MHz IF. Three-tap $T/2$ transversal equalizers have been imple-

mented in the passband (at 70-MHz IF) using quartz surface acoustic wave filters, analog correlators, and tap multipliers for a 4-PSK 12.6-Mbit/s digital radio [114]. Baseband transversal equalizers have been implemented using a combination of analog and digital or all-digital circuitry. A five-tap zero-forcing equalizer using lumped delay elements, hybrid integrated circuits for variable-gain and buffer amplifiers, and emitter-coupled logic for tap control is described in [23] for a 16-QAM 90-Mbit/s digital radio. A five-tap LMS transversal equalizer [82], and all digital DFEs have also been reported [98], [114].

## II. Receiver Structures and their Steady-State Properties

### A. Baseband Equivalent Model

To set a common framework for discussing various receiver configurations, we develop a baseband equivalent model of a passband data transmission system. We start from a generic QAM system (Fig. 9) since it can be used to implement any linear modulation scheme.

The passband transmitted signal can be compactly written as

$$s(t) = \mathrm{Re}\left[\sum_n x_n a(t - nT)\exp(j2\Pi f_c t)\right]$$

where $\{x_n\}$ is the complex sequence of data symbols with in-phase (real) and quadrature (imaginary) components, $x_r$ and $x_i$, respectively, such that $x_n = x_{rn} + jx_{in}$, $a(t)$ is the transmit pulse shape, and $f_c$ is the carrier frequency. We shall assume that the baseband transmit spectrum $A(f)$ is band-limited to $|f| \leqslant (1+\alpha)/2T$ hertz where the rolloff factor $\alpha$, between 0 and 1, determines the excess bandwidth over the minimum $|f| \leqslant 1/2T$. (Note that greater than 100-percent excess bandwidth is sometimes used in radio and baseband subscriber loop transmission systems.)

The received signal is

$$r(t) = s(t) * h_p(t) + n_p(t)$$

where $h_p(t)$ is the passband channel impulse response, $n_p(t)$ is "passband" Gaussian noise, and the operator $*$ represents convolution. The in-phase and quadrature outputs of the receive low-pass filters, $y_r(t)$ and $y_i(t)$, may be represented in complex notation as $y(t) = y_r(t) + jy_i(t)$

$$y(t) = g(t) * [r(t)\exp(-j2\Pi f_c t)]$$

where $g(t)$ is the impulse response of the receive filter. Assume that the receive filter completely rejects the double-frequency signal components produced by demodulation and centered around $2f_c$. Then the baseband received signal may be written as

$$y(t) = \sum_n x_n h(t - nT) + n(t) \qquad (1)$$

where

$$h(t) = g(t) * h_b(t) * a(t) \qquad (2)$$

is the complex-valued impulse response of the baseband equivalent model (Fig. 17). The real-valued passband channel impulse response $h_p(t)$ and the complex-valued baseband channel impulse response, $h_b(t)$, are related according to

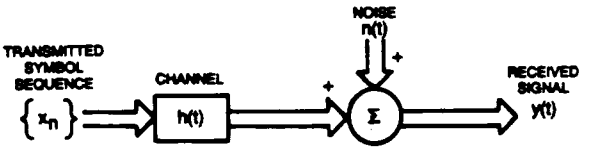


**Fig. 17.** General complex-valued baseband equivalent channel model.

$$h_p(t) = \mathrm{Re}[h_b(t)\exp(j2\Pi f_c t)].$$

The corresponding frequency-domain relationship is

$$H_p(f) = H_b[(f + f_c) + H_b^*(f - f_c)]/2.$$

Thus $h_b(t)$ is the impulse response of the positive frequency part of the channel spectral response translated to baseband.

The noise waveform $n(t)$ in (1) is also complex-valued, i.e.,

$$n(t) = g(t) * [n_p(t)\exp(-j2\Pi f_c t)].$$

The receive low-pass filters (in Fig. 9) typically perform two functions: rejection of the "double-frequency" signal components, and noise suppression. The latter function is accomplished by further shaping the baseband signal spectrum. In the baseband equivalent model, only the first of these functions of the receive filters is performed by $g(t)$ and absorbed in $h(t)$ given in (2). For simplicity, the noise $n(t)$ in the baseband equivalent model is assumed to be white with jointly Gaussian real and imaginary components.

### B. Optimum Receive Filter

Given the received signal $y(t)$, what is the best receive filter? This question has been posed and answered in different ways by numerous authors. Here, we follow Forney's development [24]. He showed that the sequence of $T$-spaced samples, obtained at the correct timing phase, at the output of a matched filter is a set of sufficient statistics for estimation of the transmitted sequence $\{x_n\}$. Thus such a receive filter is sufficient regardless of the (linear or nonlinear) signal processing which follows the symbol-rate sampler.

For the baseband equivalent model derived in the previous section, the receive filter must have an impulse response $h^*(-t)$, where the superscript $*$ denotes complex conjugate. The frequency response of this matched filter is $H^*(f)$, where $H(f)$ is the frequency response of the channel model $h(t)$.

If the data sequence $\{x_n\}$ is uncorrelated with unit power, i.e.,

$$E[x_n x_m^*] = \delta_{nm}$$

then the signal spectrum at the matched filter output is $|H(f)|^2$, and the noise power spectrum is $|H(f)|^2$. After $T$-spaced sampling, the aliased or "folded" signal spectrum is

$$S_{hh}(f) = \sum_n |H(f - n/T)|^2, \qquad 0 \leqslant f \leqslant 1/T$$

and the noise power spectrum is $N_0 S_{hh}(f)$.

If the transmission medium is ideal, then the baseband signal spectrum $H(f)$ at the matched filter input is determined solely by the transmit signal-shaping filters. From the discussion in Section I-A it is clear that if ISI is to be avoid-

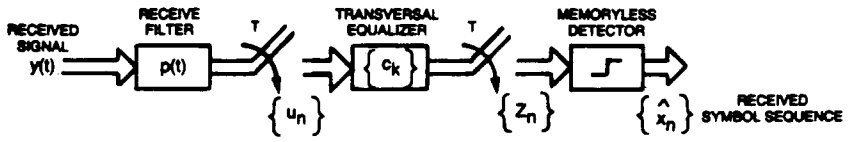


**Fig. 18.** Conventional linear receiver.

ed, the composite of the transmit filter and receive matched filter response must satisfy the Nyquist criterion, i.e.,

$$S_{hh}(f) = R_0, \quad 0 \leqslant f \leqslant 1/T$$

where, in general,

$$R_0 = T\int_0^{1/T} S_{hh}(f)\, df.$$

Therefore, the overall Nyquist amplitude response $|H(f)|^2$ must be equally divided between the transmit and receive filters. For instance, each filter may have an amplitude response which is a square root of a raised-cosine characteristic [54], [113]. For such an ideal additive white Gaussian noise (AWGN) channel, the matched filter, symbol-rate sampler, and a memoryless detector comprise the optimum receiver.

Let us now consider the more interesting case of an AWGN channel with linear distortion. For such a channel, the simple linear receiver described above is no longer adequate. The symbol-rate sampled sequence, though still providing a set of sufficient statistics, now contains intersymbol interference in addition to noise. The current received symbol is distorted by a linear combination of past and future transmitted symbols. Therefore, a memoryless symbol-by-symbol detector is not optimum for estimating the transmitted sequence. Nonlinear receivers which attempt to minimize some measure of error probability are the subject of Section II-D, where the emphasis is on techniques which combine nonlinear processing with adaptive filters.

It is instructive to first study linear receivers, which attempt to maximize signal-to-noise ratio (SNR) (minimize
less detection.

### C. Linear Receivers

We begin by reviewing the conventional linear receiver comprising a matched filter, a symbol-rate sampler, an infinite-length symbol-spaced equalizer, and a memoryless detector. In the following section we show that the matched-filter, sampler, symbol-spaced equalizer combination is a special case of a more general infinite-length fractionally spaced transversal filter/equalizer. In fact, this general filter may be used as the receive filter for any receiver structure without loss of optimality. Sections II-C3 and II-C4 present a contrast between practical forms of the conventional and fractionally spaced receiver structures.

*1) Matched Filter and Infinite-Length Symbol-Spaced Equalizer:* If further processing of the symbol-rate sampled sequence at the output of a matched receive filter is restricted to be linear, this linear processor takes the general form of a $T$-spaced infinite-length transversal or nonrecursive equalizer followed by a memoryless detector [31], [54]. Let the periodic frequency response of the transversal equalizer be $C(f)$. The equalized signal spectrum is $S_{hh}(f)C(f), 0 \leqslant f \leqslant 1/T$. The optimum $C(f)$ is one which minimizes the MSE at its output. The MSE is given by

$$\epsilon = T\int_0^{1/T} |1 - S_{hh}(f)\,C(f)|^2 + N_0 S_{hh}(f)|C(f)|^2\, df \quad (3)$$

where the first term is the residual ISI power, and the second term is the output-noise power. Differentiating the integrand with respect to $C(f)$ and equating the result to zero, one obtains the minimum MSE (MMSE) equalizer frequency response

$$C(f) = 1/[N_0 + S_{hh}(f)]. \quad (4)$$

Thus the best (MMSE) matched-filtered equalized signal spectrum is given by

$$S_{hh}(f)/[N_0 + S_{hh}(f)]. \quad (5)$$

Substituting (4) in (3), one obtains the following expression for the minimum MSE achievable by a linear receiver:

$$\epsilon_{\min}\,(\text{linear}) = T\int_0^{1/T} N_0/[N_0 + S_{hh}(f)]\, df. \quad (6)$$

The receiver structure (Fig. 18) comprising a matched filter, a symbol-rate sampler, an infinite-length $T$-spaced equalizer, and a memoryless detector is referred to as the conventional linear receiver.

If the equalizer response $C(f)$ is designed to satisfy the zero-ISI constraint then $C(f) = 1/S_{hh}(f)$.

The overall frequency response of the optimum zero-forcing (ZF) linear receiver is given by $H^*(f)/S_{hh}(f)$, which forces the ISI at the receiver output to zero. The MSE achieved by such a receiver is

$$\epsilon_{ZF}(\text{linear}) = T\int_0^{1/T} N_0/S_{hh}(f)\, df. \quad (7)$$

$\epsilon_{ZF}(\text{linear})$ is always greater than or equal to $\epsilon_{\min}(\text{linear})$ because no consideration is given to the output-noise power in designing the ZF equalizer. At high signal-to-noise ratios the two equalizers are nearly equivalent.

*2) Infinite-Length Fractionally Spaced Transversal Filter:* In this section, we derive an alternative form of an optimum linear receiver.

Let us start with the conventional receiver structure comprising a matched filter, symbol-rate sampler, $T$-spaced transversal equalizer, and a memoryless detector. As a first step, linearity permits us to interchange the order of the $T$-spaced transversal equalizer and the symbol-rate sampler. Next, the composite response $H^*(f)C(f)$ of the matched filter in cascade with a $T$-spaced transversal equalizer may be realized by a single continuous-time filter. Let us assume that the received signal spectrum $H(f) = 0$ outside the range $|f| \leqslant 1/2\tau, \tau \leqslant T$, and the flat noise spectrum is also limited to the same band by an ideal antialiasing filter with cutoff frequency $1/2\tau$. The composite matched-filter equalizer can then be realized by an infinite-length continuous-time transversal filter with taps spaced at $\tau$-second

intervals and frequency response $H^*(f)C(f), |f| \leq 1/2\tau$, which is periodic with period $1/\tau$ hertz. The continuous output of this composite transversal filter may be sampled at the symbol rate without any further restriction on the tap-spacing $\tau$. If the frequency response $C(f)$ is selected according to (4), the symbol-rate output of the composite filter is identical to the corresponding output (5) in the conventional MMSE linear receiver.

To implement this composite matched-filter equalizer as a fractionally spaced digital nonrecursive filter, the effective tap-spacing must be restricted to $KT/M$, where $K$ and $M$ are relatively prime integers, $K < M$, and the fraction $KT/M \leq \tau$. The desired frequency response $H^*(f)C(f)$ of this fractionally spaced digital transversal filter is periodic with period $M/KT$ hertz, and limited to the band $|f| \leq 1/2\tau < M/2KT$ since $H(f)$ is limited to the same bandwidth. An ideal antialiasing filter with a cutoff frequency $M/2T$ hertz is assumed before the rate $M/T$ sampler.

The operation of the digital filter may be visualized as follows. Each symbol interval, $M$ input samples are shifted into digital shift register memory, every $K$th sample in the shift register is multiplied by a successive filter coefficient, and the products summed to produce the single output required per symbol interval.

Note that the $M/T$ rate input of the fractionally spaced digital filter has the signal spectrum

$$\begin{array}{ll} H(f), & |f| \leq M/2KT \\ 0, & M/2KT < |f| \leq M/2T. \end{array}$$

The frequency response of the digital filter is $H^*(f)C(f), |f| \leq M/2KT$. Thus if the output of this filter was produced at the rate $M/T$, the output signal spectrum would be

$$\begin{array}{ll} H(f)H^*(f)C(f), & |f| \leq 1/2KT \\ 0, & M/2KT < |f| \leq M/2T. \end{array}$$

When the filter output is produced at the symbol rate, it has the desired aliased signal spectrum

$$\sum_n |H(f - n/T)|^2 C(f - n/T), \qquad 0 \leq f \leq 1/T.$$

Noting that $C(f)$ is periodic with period $1/T$, this output-signal spectrum is recognized as $S_{hh}(f)C(f)$, which is the same as for the conventional MMSE linear receiver (5), provided $C(f)$ is selected according to (4).

This simple development proves the important point that an infinite-length fractionally spaced digital transversal filter is at once capable of performing the functions of the matched filter and the $T$-spaced transversal equalizer of the conventional linear receiver.

Let us further show that the symbol-rate sampled outputs of a fractionally spaced digital filter form a set of sufficient statistics for estimation of the transmitted sequence under the following conditions. The digital filter with tap spacing $KT/M$ has the frequency response $H^*(f)C(f), |f| \leq M/2KT$, where the received signal spectrum $H(f)$ is zero outside the band $|f| \leq M/2KT$, $C(f)$ is periodic with period $1/T$, and $C(f)$ is information lossless. A sufficient condition for $C(f)$ to be information lossless is that $C(f)$ is invertible, i.e., $C(f) \neq 0$, $0 \leq f \leq 1/T$. However, it is only necessary that $C(f)/S_{hh}(f)$ is invertible, i.e., $C(f)/S_{hh}(f) \neq 0$, $0 \leq f \leq 1/T$. In words, this condition implies that $C(f)$ may not introduce any nulls or transmission zeros in the Nyquist band, except at a frequency where the signal (and the noise power spectrum at the matched-filter output) may already have a null.

The above result shows that with an appropriately designed $C(f)$, the symbol-rate outputs of a fractionally spaced filter, with frequency response $H^*(f)C(f)$, may be used without loss of optimality, for any linear or nonlinear receiver, regardless of the criterion of optimality.

In a linear receiver, where a memoryless detector operates on the symbol-rate outputs of the fractionally spaced filter, the function $C(f)$ may be designed to minimize the MSE at the detector input. The optimum $C(f)$ is then obtained using the same procedure as outlined in the previous section for the conventional linear receiver. Thus the MMSE $KT/M$-spaced filter frequency response is

$$H^*(f)/[N_0 + S_{hh}(f)], \qquad |f| \leq M/2KT. \tag{8}$$

The MMSE achieved by this filter is, of course, the same as $\epsilon_{\min}$ (linear) derived earlier (6) for the conventional receiver structure.

If the criterion of optimality used in designing $C(f)$ is to force the intersymbol interference at the detector input to zero, one obtains the overall filter response

$$H^*(f)/S_{hh}(f), \qquad |f| \leq M/2KT. \tag{9}$$

The MSE achieved in this case is the same as $\epsilon_{ZF}$(linear) derived earlier (7).

*3) Fixed-Filter and Finite-Length Symbol-Spaced Equalizer:* The conventional MMSE linear receiver is impractical for two reasons. First, constraints of finite length and computational complexity must be imposed on the matched filter as well as the $T$-spaced transversal equalizer. Secondly, in most applications, it is impractical to design, beforehand, a filter which is reasonably matched to the variety of received signal spectra resulting from transmissions over different channels or a time-varying channel. Thus the most commonly used receiver structure comprises a fixed filter, symbol rate sampler, and finite-length $T$-spaced adaptive equalizer (Fig. 18). The fixed-filter response is either matched to the transmitted signal shape or is designed as a compromise equalizer which attempts to equalize the average of the class of line characteristics expected for the application. For the present discussion, let us assume that the fixed filter has an impulse response $p(t)$ and a frequency response $P(f)$. Then, the $T$-spaced sampled output of this filter may be written as

$$u_k = \sum_n x_n q(kT - nT) + v_k \tag{10}$$

where

$$q(t) = p(t) * h(t)$$

and

$$v_k = \int n(t)p(kT - t)\,dt.$$

Denoting the $N$ equalizer coefficients at time $kT$ by the column vector $\boldsymbol{c}_k$, and the samples stored in the equalizer delay line by the vector $\boldsymbol{u}_k$, the equalizer output is given by

$$z_k = \boldsymbol{c}_k^T \boldsymbol{u}_k$$

where the superscript $T$ denotes transpose. Minimizing the MSE $E[|z_k - x_k|^2]$ leads to the set of optimum equalizer coefficients

$$\boldsymbol{c}_{\text{opt}} = \boldsymbol{A}^{-1}\boldsymbol{\alpha} \tag{11}$$

where $\boldsymbol{A}$ is an $N \times N$ Hermitian covariance matrix $E[\boldsymbol{u}_k^* \boldsymbol{u}_k^T]$, and $\boldsymbol{\alpha}$ is an $N$-element cross-correlation vector $E[\boldsymbol{u}_k^* x_k]$.

Using the assumption that the data sequence $\{x_k\}$ is uncorrelated with unit power, it can be shown that the elements of the matrix $\boldsymbol{A}$ and vector $\boldsymbol{\alpha}$ are given by

$$a_{i,j} = \sum_k q^*(kT)\, q(kT + iT - jT) + N_0 \int p^*(t)\, p(t + iT - jT)\, dt \tag{12}$$

and

$$\alpha_i = q^*(-iT). \tag{13}$$

The MMSE achieved by this conventional suboptimum linear receiver is given by

$$\epsilon_{\min}(\text{con}) = 1 - \boldsymbol{\alpha}^{*T}\boldsymbol{A}^{-1}\boldsymbol{\alpha}. \tag{14}$$

Alternatively, the $N$ equalizer coefficients may be chosen to force the samples of the combined channel and equalizer impulse response to zero at all but one of the $N$ $T$-spaced instants in the span of the equalizer. The zero-forcing equalizer coefficient vector is given by

$$\boldsymbol{c}_{ZF} = \boldsymbol{Q}^{-1}\boldsymbol{\delta} \tag{15}$$

where $\boldsymbol{Q}$ is an $N \times N$ matrix with elements

$$q_{i,j} = q(iT - jT) \tag{16}$$

and $\boldsymbol{\delta}$ is a vector with only one nonzero element, that element being unity. The MSE achieved by the zero-forcing suboptimal receiver is given by

$$\epsilon_{ZF}(\text{con}) = \epsilon_{\min}(\text{con}) + (\boldsymbol{c}_{ZF} - \boldsymbol{c}_{\text{opt}})^{*T}\boldsymbol{A}(\boldsymbol{c}_{ZF} - \boldsymbol{c}_{\text{opt}})$$

where the quadratic form is the excess MSE over the LMS solution.

It is instructive to derive expressions for the equalizer and its performance as the number of coefficients is allowed to grow without bound. Since the $N \times N$ matrices $\boldsymbol{A}$ and $\boldsymbol{Q}$ are Toeplitz, their eigenvalues can be obtained by the discrete Fourier transform (DFT) of any row or column as $N \to \infty$ [44]. Thus by taking the DFT of (11) for $\boldsymbol{c}_{\text{opt}}$ we obtain the frequency spectrum of the infinite-length $T$-spaced LMS equalizer

$$C_{\text{opt}}(f) = \frac{Q_{\text{eq}}^*(f)}{|Q_{\text{eq}}(f)|^2 + N_0 S_{pp}(f)} \tag{17}$$

where

$$Q_{\text{eq}}(f) = \sum_n Q(f - n/T) \tag{18}$$

is the aliased spectrum of $q(t)$, and

$$S_{pp}(f) = \sum_n |P(f - n/T)|^2 \tag{19}$$

is the aliased power spectrum of $p(t)$. The minimum achievable MSE is given by

$$\epsilon_{\min}(\text{con}) = T\int_0^{1/T} 1 - \frac{|Q_{\text{eq}}(f)|^2}{N_0 S_{pp}(f) + |Q_{\text{eq}}(f)|^2}\, df = T\int_0^{1/T} \frac{N_0}{N_0 + |Q_{\text{eq}}(f)|^2 / S_{pp}(f)}\, df. \tag{20}$$

The corresponding expressions for the zero-forcing equalizer are

$$C_{ZF}(f) = 1/Q_{\text{eq}}(f) \tag{21}$$

and

$$\epsilon_{ZF}(\text{con}) = T\int_0^{1/T} N_0 S_{pp}(f)/|Q_{\text{eq}}(f)|^2\, df. \tag{22}$$

When $P(f)$ is a matched filter, i.e., $P(f) = H^*(f)$, the above expressions reduce to those given in Section II-B1 because $S_{pp}(f) = Q_{\text{eq}}(f) = S_{hh}(f)$.

The smallest possible MSE (zero-ISI matched filter bound) is achieved when in (20) we have

$$|Q_{\text{eq}}(f)|^2 / S_{pp}(f) = S_{hh}(f) = R_0, \qquad 0 \leq f \leq 1/T.$$

This occurs when the channel amplitude characteristic is ideal and perfect equalization is achieved by the matched filter. The greater the deviation of $|Q_{\text{eq}}(f)|^2/S_{pp}(f)$ from its average $R_0$, the greater $\epsilon_{\min}(\text{con})$. The aliased power spectrum $S_{pp}(f)$, as defined in (19), is independent of the phase characteristics of $P(f)$ or the sampler phase. The value of the squared absolute value $|Q_{\text{eq}}(f)|^2$ of the aliased spectrum $Q_{\text{eq}}(f)$, on the other hand, is critically dependent on the sampler phase in the rolloff region due to aliasing. Thus the minimum MSE achieved by the conventional receiver is dependent on the sampler phase even when the number of $T$-spaced equalizer coefficients is unlimited ($N \to \infty$).

When $N$ is finite, the value of $\epsilon_{\min}(\text{con})$ depends on the channel impulse response $h(t)$, the noise power spectral density $N_0$, the choice of the fixed receive filter $p(t)$, and the number $N$ of $T$-spaced equalizer coefficients. Therefore, it is difficult to say much about the performance or degree of suboptimality of this receiver structure without resorting to numerical computations for particular examples. As noted above, one general characteristic of this receiver structure is the sensitivity of its performance to the choice of sampler phase. This point is discussed further in Section II-E.

*4) Finite-Length Fractionally Spaced Transversal Equalizer:* This suboptimum linear receiver structure is simply a practical form of the infinite-length structure discussed in Section II-C2. We shall restrict our attention to the digitally implemented fractionally spaced equalizer (FSE) with tap spacing $KT/M$ (see Fig. 19). The input to the FSE is the received signal sampled at rate $M/T$

$$y(kT/M) = \sum_n x_n h(kT/M - nT) + n(kT/M). \tag{23}$$

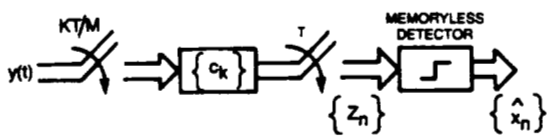


**Fig. 19.** Linear receiver based on a fractionally spaced transversal equalizer.

Each symbol interval, the FSE produces an output according to

$$z(kT) = \sum_{n=0}^{N-1} c_n y(kT - nKT/M). \quad (24)$$

Denoting the $N$ equalizer coefficients at time $kT$ by the vector $\boldsymbol{c}_k$ and the $N$ most recently received samples (spaced $KT/M$ seconds apart) by the vector $\boldsymbol{y}_k$, the equalizer output may be written as

$$z_k = \boldsymbol{c}_k^T \boldsymbol{y}_k.$$

Minimizing the MSE $E[|z_k - x_k|^2]$ leads to the set of optimum equalizer coefficients

$$\boldsymbol{c}_{\text{opt}} = \boldsymbol{A}^{-1}\boldsymbol{\alpha} \quad (25)$$

where $\boldsymbol{A}$ is an $N \times N$ covariance matrix $E[\boldsymbol{y}_k^* \boldsymbol{y}_k^T]$, and $\boldsymbol{\alpha}$ is an $N$-element cross-correlation vector $E[\boldsymbol{y}_k^* x_k]$.

Using the assumption that the data sequence $\{x_k\}$ is uncorrelated with unit power, it can be shown that the elements of the matrix $\boldsymbol{A}$ and vector $\boldsymbol{\alpha}$ are given by

$$a_{i,j} = \sum_k h^*(kT - iKT/M)\, h(kT - jKT/M) + N_0 \delta_{ij} \quad (26)$$

$$\alpha_i = h^*(-iKT/M). \quad (27)$$

The MMSE achieved by the FSE is given by

$$\epsilon_{\min}(\text{FSE}) = 1 - \boldsymbol{\alpha}^{*T} \boldsymbol{A}^{-1} \boldsymbol{\alpha}. \quad (28)$$

On the surface, the FSE development is quite similar to that of the conventional $T$-spaced LMS equalizer given in the previous section. There are, however, significant differences. First, unlike the $T$ equalizer, the FSE does not require a fixed receive shaping filter $p(t)$. Secondly, note that while the FSE input covariance matrix $\boldsymbol{A}$ is Hermitian, it is not Toeplitz. In fact, each diagonal periodically takes one of $M$ different values. Due to the non-Toeplitz cyclostationary nature of the $\boldsymbol{A}$ matrix, it is no longer possible to obtain the eigenvalues of $\boldsymbol{A}$ by simply taking the DFT of one of its rows even as $N \to \infty$. However, it is possible to decompose the set of infinite equations

$$\boldsymbol{A}\boldsymbol{c} = \boldsymbol{\alpha}$$

into $M$ subsets each with $M$ Toeplitz submatrices. Using this procedure, it can be shown [39] that as $N \to \infty$, a fraction $(M - K)/M$ of the eigenvalue are equal to $N_0$, and the remaining eigenvalues are of the form $(M/K)\, S_{hh}(f) + N_0$, where

$$S_{hh}(iM/NKT) = \sum_n |H(iM/NKT - n/T)|^2,$$
$$i = 0, 1, 2, \cdots, (NK/M) - 1. \quad (29)$$

The frequency response of the optimum FSE approaches (8) as $N \to \infty$, and its MSE approaches $\epsilon_{\min}$ (linear) given in (6).

As the noise becomes vanishingly small, an infinitely long FSE has a set of zero eigenvalues. This implies that there are an infinite number of solutions which produce the same minimum MSE. The nonunique nature of the infinite FSE is evident from the fact that when both signal and noise vanish in the frequency range $1/2T < |f| \leq M/2T$, the infinite FSE spectrum $C(f)$ can take any value in this frequency range without affecting the output signal or MSE.

Gitlin and Weinstein [39] show that for transmission systems with less than 100-percent excess bandwidth, the matrix $\boldsymbol{A}$ is nonsingular for a finite-length FSE even as the noise becomes vanishingly small. Therefore, there exists a unique set of optimum equalizer coefficients $\boldsymbol{c}_{\text{opt}}$ given by (25).

Deviation of the coefficient vector $\boldsymbol{c}_k$ from the optimum results in the following excess MSE over $\epsilon_{\min}$(FSE) given in (28):

$$(\boldsymbol{c}_k - \boldsymbol{c}_{\text{opt}})^{*T} \boldsymbol{A} (\boldsymbol{c}_k - \boldsymbol{c}_{\text{opt}}).$$

This quadratic form may be diagonalized to obtain $\boldsymbol{d}_k^{*T} \boldsymbol{\Lambda} \boldsymbol{d}_k$, where the diagonal matrix $\boldsymbol{\Lambda}$ has the eigenvalues of $\boldsymbol{A}$ along its main diagonal, and $\boldsymbol{d}_k$ is the transformed coefficient deviation vector according to

$$\boldsymbol{d}_k = \boldsymbol{V}(\boldsymbol{c}_k - \boldsymbol{c}_{\text{opt}}) \quad (30)$$

where the columns of the diagonalizing matrix $\boldsymbol{V}$ are the eigenvectors of $\boldsymbol{A}$. From the analysis of the infinite FSE, one would expect that when the number of FSE coefficients is "large," a significant fraction $(M - K)/M$ of the eigenvalues of $\boldsymbol{A}$ are relatively small. If the $i$th eigenvalue is very small, the $i$th element of the deviation vector $\boldsymbol{d}_k$ will not contribute significantly to the excess MSE. It is, therefore, possible for coefficient deviations to exist along the eigenvectors corresponding to the small eigenvalues of $\boldsymbol{A}$ without significant impact on the MSE. Thus many coefficient vectors may produce essentially the same MSE.

The most significant difference in the behavior of the conventional suboptimum receiver and the FSE is a direct consequence of the higher sampling rate at the input to the FSE. Since no aliasing takes place at the FSE input, it can independently manipulate the spectrum in the two rolloff regions to minimize the output MSE after symbol-rate sampling. Thus unlike the $T$-spaced equalizer, it is possible for the FSE to compensate for timing phase as well as asymmetry in the channel amplitude or delay characteristics without noise enhancement. This is discussed further in Section II-E.

### *D. Nonlinear Receivers*

The MMSE linear receiver is optimum with respect to the ultimate criterion of minimum probability of symbol error only when the channel does not introduce any amplitude distortion, i.e., $S_{hh}(f) = R_0$, $0 \leq f \leq 1/T$. The linear receive filter then achieves the matched filter (mf) bound for MSE

$$\epsilon_{\min}(\text{mf}) = T \int_0^{1/T} N_0/[N_0 + R_0]\, df$$

and a memoryless threshold detector is sufficient to minimize the probability of error. When amplitude distortion is present in the channel, a linear receive filter, e.g., FSE, can reduce ISI and output MSE by providing the output signal spectrum

$$S_{hh}(f)/[N_0 + S_{hh}(f)].$$

The corresponding output error power spectrum is $N_0/[N_0 + S_{hh}(f)]$. Thus noise power is enhanced at those frequencies where $S_{hh}(f) < R_0$. A memoryless detector oper-

ating on the output of this received filter no longer minimizes symbol error probability.

Recognizing this fact, several authors have investigated optimum or approximately optimum nonlinear receiver structures subject to a variety of criteria [59]. Most of these receivers use one form or another of the maximum *a posteriori* probability rule to maximize either the probability of detecting each symbol correctly [1] or of detecting the entire transmitted sequence correctly. The classical maximum-likelihood receiver [54] consists of $m^k$ matched filters, where $k$ is the length of the transmitted sequence whose symbols are drawn from a discrete alphabet of size $m$.

The complexity of the classical receiver, which grows exponentially with the message length, can be avoided by using the Viterbi algorithm. This recursive algorithm which was originally invented to decode convolutional codes was recognized to be a maximum-likelihood sequence estimator (MLSE) of the state sequence of a finite-state Markov process observed in memoryless noise [25]. Forney [24] showed that if the receive filter is a whitened matched filter, its symbol rate outputs at the correct sampling times form a set of sufficient statistics for estimation of the information sequence. Thus the transmission system between the data source and the Viterbi algorithm (VA) can be considered as a discrete channel, as shown in Fig. 20. The state and hence the input sequence of the discrete channel can be estimated by the VA which observes the channel output corrupted by additive white Gaussian noise [24]. The computational complexity of the MLSE is proportional to $m^{L-1}$, the number of discrete channel states, where $L$ is the number of terms in the discrete channel pulse response.

The MLSE maximizes the mean time between error (MTBE) events, a reasonable criterion for practical automatic repeat request (ARQ) data communications systems where efficiency is measured by throughput (i.e., the number of blocks of data correctly received versus the total number of blocks transmitted).

The symbol error probability of the MLSE [24] is estimated by an expression of the form

$$\Pr(e) \simeq KQ[d_{\min}/2\sigma].$$

The minimum Euclidean distance between any two valid neighboring sequences is $d^2_{\min}$, $\sigma^2$ is the mean square white Gaussian noise power at the input to the VA, and $d^2_{\min}/\sigma^2$ is the effective SNR.

$$Q(x) = (1/2\pi)\int_x^\infty \exp(-y^2/2)\,dy$$

is the Gaussian probability of error function, and the error coefficient $K$ may be interpreted as the average number of ways in which minimum distance symbol errors can occur.

The lower bound on the probability of error for binary transmission over an ideal AWGN channel is given by

$$Q\left[(2\epsilon_{\min}(\mathrm{mf}))^{-1/2}\right].$$

The MLSE approaches this lower bound at high SNR for all channels except those with extremely severe ISI. For instance, for a class IV partial response system with a discrete channel model of the form $1 - D^2$, the symbol error rate achieved by the MLSE is about 4 times the lower bound given above. In decibels this difference is small (about 0.5 dB for a symbol error rate of $10^{-4}$) and goes to zero as the effective SNR goes to infinity.

For unknown and/or slowly time-varying channels, the MLSE can be made adaptive by ensuring that both the whitened matched filter and the channel model used by the VA adapt to the channel response. Magee and Proakis [62] proposed an adaptive version of the VA which uses an adaptive identification algorithm to provide an estimate of the discrete channel pulse response. Structures with adaptive whitened matched filters (WMF) were proposed in [61] and [88].

Forney [24] derived the WMF as the cascade connection of a matched filter $H^*(f)$, a symbol-rate sampler, and a $T$-spaced transversal whitening filter whose pulse response is the anticausal factor of the inverse filter $1/S_{hh}(f)$. The noise at the output of the WMF is white and the ISI is causal. Price [83] showed that the WMF is also the optimum forward filter in a zero-forcing decision-feedback equalizer, where the feedback transversal filter can exactly cancel the causal ISI provided all past decisions are correct. (Earlier, Monsen [71] had derived the optimum forward filter for an LMS DFE.)

At this point it is helpful to discuss simpler forms of nonlinear receivers, i.e., decision feedback equalization and

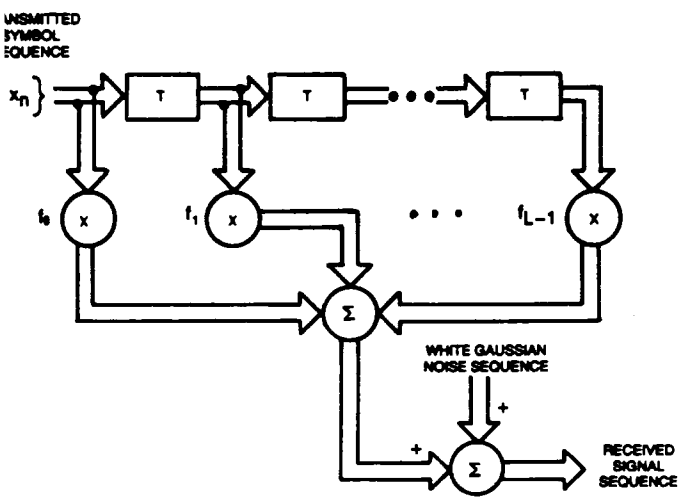


**Fig. 20.** Discrete channel model.

general decision-aided ISI cancellation, before returning to the topic of adaptive receiver filtering for maximum-likelihood sequence estimation.

*1) Decision Feedback Equalizers:* The MLSE unravels ISI by deferring decisions and weighing as many preliminary decision sequences as the number of states in the discrete channel model. Thus in most cases, the MLSE makes use of all the energy in the discrete channel impulse response to maximize the effective SNR. By contrast, a DFE makes memoryless decisions and cancels all trailing ISI terms. Even when the WMF is used as the receive filter for both the MLSE and the DFE, the latter suffers from a reduced effective SNR, and error propagation, due to its inability to defer decisions.

*a) Infinite length:* A decision-feedback equalizer (DFE) takes advantage of the symbols which have already been detected (correctly with high probability) to cancel the intersymbol interference due to these symbols without noise enhancement. An infinite-length DFE receiver takes the general form (Fig. 21) of a forward linear receive filter, symbol-rate sampler, canceler, and memoryless detector. The symbol-rate output of the detector is then used by the feedback filter to generate future outputs for cancellation.

As pointed out by Belfiore and Park [4], an equivalent structure to the DFE receiver of Fig. 21 is the structure shown in Fig. 22. The latter may be motivated from the point of view that given the MMSE forward filter, e.g., an infinite-length FSE, we know that the sequence of symbol-rate samples at the output of this filter form a set of sufficient statistics for estimating the transmitted sequence. Then what simple form of nonlinear processing could further reduce the MSE? To this end, let us examine the power spectra of the two components of distortion: noise and ISI. Noise at the output of the MMSE forward filter (see Section II-C1) has the power spectrum

$$N_0 S_{hh}(f)/[N_0 + S_{hh}(f)]^2, \qquad 0 \leq f \leq 1/T$$

and the residual ISI has the power spectrum

$$|1 - S_{hh}(f)/[N_0 + S_{hh}(f)]|^2 = N_0^2/[N_0 + S_{hh}(f)]^2, \qquad 0 \leq f \leq 1/T.$$

Since noise and ISI are independent, the power spectrum of the total distortion or error sequence is given by the sum of the noise and ISI power spectra, i.e.,

$$|\hat{E}(f)|^2 = N_0/[N_0 + S_{hh}(f)], \qquad 0 \leq f \leq 1/T. \quad (31)$$

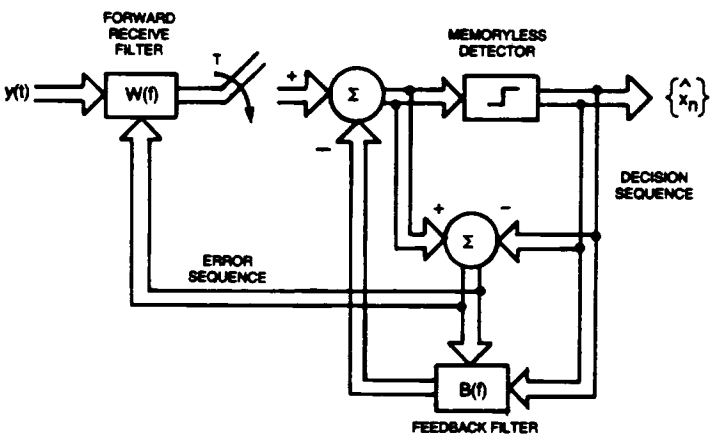


**Fig. 21.** Conventional decision-feedback receiver.

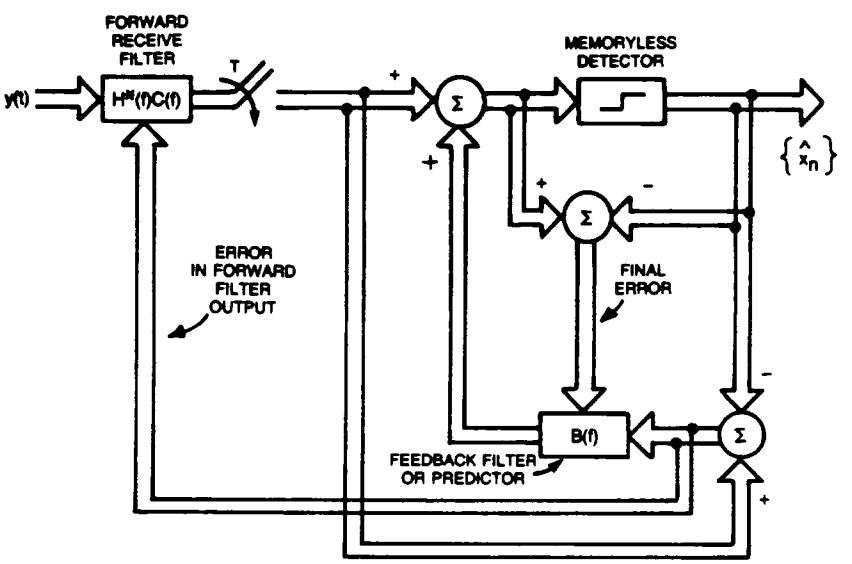


**Fig. 22.** Predictor form of decision-feedback receiver.

The error sequence is white if and only if $S_{hh}(f)$ is a constant; e.g., $S_{hh}(f) = R_0$, $|f| \leq 1/2T$, when the channel has no amplitude distortion. In this case, further reduction in MSE is not possible. However, for channels with amplitude distortion, the power of the error sequence at the output of the forward filter can be reduced further by linear prediction [122], provided past samples of the error sequence are available. An estimate of these past error samples can be obtained by decision feedback via a memoryless detector, as shown in Fig. 22. To complete the picture, it remains to derive the optimum predictor spectrum as the number of predictor coefficients grows without bound.

The error sequence at the predictor output has the spectrum

$$E(f) = \hat{E}(f) + \hat{E}(f)B(f)$$

where $B(f)$ is the desired spectrum of the infinitely long predictor, i.e.,

$$B(f) = \sum_{n=1}^{\infty} b_n e^{-j2\pi fnT}.$$

The optimum $B(f)$ is one which minimizes the final MSE

$$\epsilon(\text{DFE}) = T\int_0^{1/T} |1 + B(f)|^2 |\hat{E}(f)|^2 \, df.$$

The solution is available from the theory of one-step predictors and Toeplitz quadratic forms [44]. There exists a factorization of the inverse power spectrum of the error sequence, such that

$$\Gamma(f)\Gamma^*(f) = 1/|\hat{E}(f)|^2$$

where

$$\Gamma(f) = \sum_{n=0}^{\infty} \gamma_n \exp(-j2\pi fnT)$$

i.e., $\{\gamma_n\}$ is causal. Then the optimum $B(f)$ is given by

$$B(f) = \Gamma(f)/\gamma_0 - 1, \qquad 0 \leq f \leq 1/T$$

where the normalizing factor $\gamma_0$ is the average value of $\Gamma(f)$ in the range $|f| \leq 1/2T$. The minimum achievable MSE is

$$\epsilon_{\min}(\text{DFE}) = 1/|\gamma_0|^2. \tag{32}$$

Given the error sequence with power spectrum $|\hat{E}(f)|^2$, the optimum predictor produces a white error sequence with power spectral density $1/|\gamma_0|^2$. Note that since both error sequences, before and after prediction, contain noise and ISI components, neither sequence is Gaussian. The MMSE given in (32) can be expressed directly in terms of the folded power spectrum $S_{hh}(f)$. Note that

$$T\int_0^{1/T} \ln|\hat{E}(f)|^2 \, df = T\int_0^{1/T} \ln|E(f)|^2 \, df - T\int_0^{1/T} \ln|1 + B(f)|^2 \, df.$$

The first integral on the right-hand side is $\ln(1/|\lambda_0|^2)$ since $|E(f)|^2 = 1/|\lambda_0|^2$. The second integral is zero because $1 + B(f)$ has all its zeros inside the unit circle [122]. Thus using (32), we have

$$\epsilon_{\min}(\text{DFE}) = \exp\left[T\int_0^{1/T} \ln|\hat{E}(f)|^2 \, df\right] = \exp\left[-T\int_0^{1/T} \ln\left(1 + S_{hh}(f)/N_0\right) df\right]. \tag{33}$$

This expression is identical to the MMSE for an infinite-length conventionl DFE [93], which proves the equivalence of the predictor and conventional DFE structures.

We can write the following equivalence relationship by comparing the two DFE structures shown in Figs. 21 and 22:

$$W(f) = H^*(f)C(f)[1 + B(f)]. \tag{34}$$

After some inspection, it becomes evident that the infinite-length forward filter $W(f)$ in the conventional DFE structure is the cascade of a matched filter and the anticausal factor of the optimum $C(f)$ given in (4) [71]. So long as the length of the forward filter in each of the two DFE structures is unconstrained, the two structures remain equivalent even when the feedback (or prediction) filter is reduced to a finite length. Note, however, that while the forward filter $W(f)$ in the conventional structure depends on the number of feedback coefficients, the forward filter in the predictor structure is independent of the predictor coefficients.

An alternative to the optimum mean-square DFE receiver is the zero-forcing formulation [4], [83]. The forward filter is again of the form (34). However, $C(f)$ is designed to satisfy the zero-ISI constraint, i.e., $C(f) = 1/S_{hh}(f)$. Thus the forward filter $W(f)$ is the cascade of a matched filter and the anticausal factor of the inverse filter $1/S_{hh}(f)$. The noise at the output of $W(f)$ is white and the ISI is causal (as in the case of Forney's whitened matched filter). The causal ISI is completely canceled by an infinite-length decision-feedback filter. The final error sequence consists solely of noise which is white and Gaussian with power

$$\epsilon_{ZF}(\text{DFE}) = \exp\left[-T\int_0^{1/T} \ln\left(S_{hh}(f)/N_0\right) df\right]. \tag{35}$$

As expected, $\epsilon_{\min}(\text{mf}) \leq \epsilon_{\min}(\text{DFE}) \leq \epsilon_{ZF}(\text{DFE})$.

*b) Finite length:* Neglecting zero-forcing equalizers, there are four possible structures for finite-length DFE receivers, based on the conventional or FSE forward filters, and conventional or predictor forms of feedback filters. Let us denote these as

- Type 1: Conventional forward filter + conventional feedback filter.
- Type 2: Conventional forward filter + predictor feedback filter.
- Type 3: FSE forward filter + conventional feedback filter.
- Type 4: FSE forward filter + predictor feedback filter.

*Type 1:* To the forward equalizer structure described in Section II-C3, we add $N_b$ feedback coefficients. Denoting the latter at time $kT$ by the column vector $\boldsymbol{b}_k$, the samples stored in the forward equalizer delay line by the vector $\boldsymbol{u}_k$, and the past $N_b$ decisions by the vector $\boldsymbol{x}_k$, the equalizer output is given by

$$z_k = \boldsymbol{c}_k^T \boldsymbol{u}_k - \boldsymbol{b}_k^T \boldsymbol{x}_k.$$

Minimizing the MSE with respect to the feedback coefficients leads to

$$\boldsymbol{b}_k = \boldsymbol{Q}\boldsymbol{c}_k \tag{36}$$

where $\boldsymbol{Q}$ is an $N_b \times N$ matrix with elements given by (16). Using (36) and proceeding as in Section II-C1, it can be shown that the set of optimum forward coefficients is given by

$$\boldsymbol{c}_{\text{opt}} = \hat{A}^{-1}\boldsymbol{\alpha}. \tag{37}$$

The $N \times N$ matrix $\hat{\boldsymbol{A}}$ has elements $\hat{a}_{i,j}$ similar to $a_{i,j}$ given

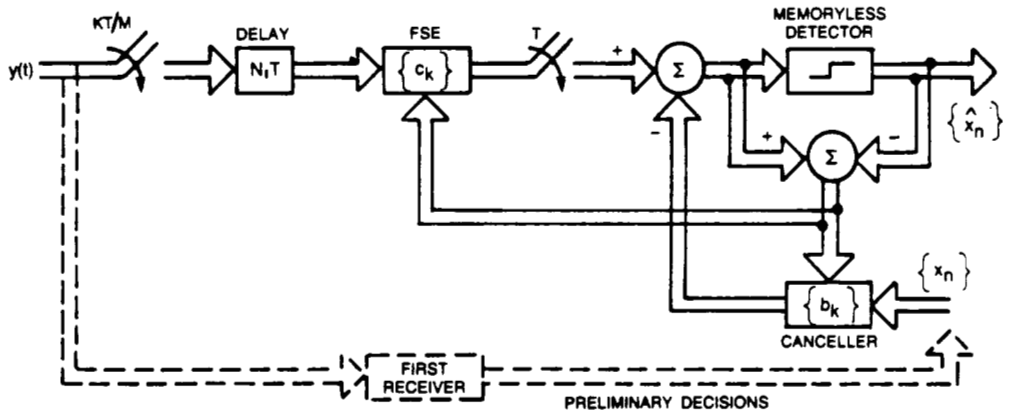


**Fig. 23.** Decision-aided ISI cancellation.

by (12), except that the summation over $k$ now excludes the set $1 \leqslant k \leqslant N_b$ (which is in the span of the feedback filter). Finally

$$\boldsymbol{b}_{\text{opt}} = \boldsymbol{Q}\boldsymbol{c}_{\text{opt}}.$$

*Type 2:* In this case, the forward equalizer coefficients can be obtained from (11) independently of the predictor coefficients. Let $\hat{\boldsymbol{e}}_k$ be an $N_b$-element vector at time $kT$ consisting of the $N_b$ most recent error signals before prediction at instants $k, k-1, \cdots, k-N_b-1$, with the forward equalizer coefficients at their optimal values. Then, the set of optimum predictor coefficients $\boldsymbol{b}_k$ is the solution of the normal equations

$$E\left[\hat{\boldsymbol{e}}^*_{k-1}\hat{\boldsymbol{e}}^T_{k-1}\right]\boldsymbol{b}_k = E\left[\hat{\boldsymbol{e}}^*_k\hat{e}_k\right]. \tag{38}$$

*Type 3:* This case is similar to Type 1, except that the forward filter structure described in Section II-C4 is used. The set of optimum forward coefficients can be obtained using

$$\boldsymbol{c}_{\text{opt}} = \hat{\boldsymbol{A}}^{-1}\boldsymbol{\alpha}$$

where the matrix $\hat{\boldsymbol{A}}$ has elements $a_{i,j}$ similar to $\hat{a}_{i,j}$ given by (26), except that the summation over $k$ now excludes the set $1 \leqslant K \leqslant N_b$. The optimum feedback coefficients are given by

$$\boldsymbol{b}_{\text{opt}} = \boldsymbol{H}\boldsymbol{c}_{\text{opt}}$$

where $\boldsymbol{H}$ is an $N_b \times N$ matrix with elements $h_{i,j} = h(iT - jKT/M)$.

*Type 4:* This type of DFE is similar to Type 2. The optimum forward coefficients given by (25) still apply and the predictor coefficients can be obtained by solving a set of normal equations similar to that given in (38).

For a direct comparison of the performance of the four types of DFE structures, resort must be made to numerical solution for particular channel characteristics, number of equalizer coefficients, etc. One general comment that can be made is that for an equal number of forward and feedback coefficients, the DFE structures of Type 1 and 3 will always achieve an output MSE at least as low or lower than the MSE achieved by structures of Type 2 and 4, respectively. This is true because unlike Type 1 and 3 structures, independent solution of the forward equalizer and the feedback predictor coefficients in Type 2 and 4 DFE structures does not in general guarantee joint minimization of the final MSE.

One important factor in the practical performance of DFE structures of all types is the effect of error propagation on the final error probability. Assuming correct decisions, the improvement in output signal-to-MSE ratio provided by a DFE reduces the probability of occurrence of the first error. However, once an error occurs it tends to propagate due to incorrect decision feedback. Bounds on the error multiplication factor have been developed by Duttweiler *et al.* [12] and by Belfiore and Park [4].

*2) Decision-Aided ISI Cancellation:* The concept of decision feedback of past data symbols to cancel intersymbol interference can theoretically be extended to include future data symbols. If all past and future data symbols were assumed to be known at the receiver, then given a perfect model of the ISI process, all ISI could be canceled exactly without any noise enhancement. Such a hypothetical receiver could, therefore, achieve the zero-ISI matched-filter bound on performance. In practice, the concept of decision-aided ISI cancellation can be implemented by using tentative decisions and some finite delay at the receiver. In the absence of tentative decision errors, the ISI due to these finite number of future data symbols (as well as past data symbols) can be canceled exactly before final receiver decisions are made. Such a receiver structure with a two-step decision process was proposed by Proakis [85]. Recently, Gersho and Lim [34] observed that the MSE performance of this receiver structure could be improved considerably by an adaptive matched filter in the path of the received signal prior to ISI cancellation. In fact, the zero-ISI matched-filter bound on MSE could be achieved in the limit assuming correct tentative decisions. A general theoretical treatment is available in [80].

Consider the block diagram of Fig. 23 where correct data symbols are assumed to be known to the canceler. Let the forward filter have $N$ coefficients fractionally spaced at $KT/M$-second intervals. Then the final output of the structure is given by

$$z_k = \boldsymbol{c}_k^T\boldsymbol{y}_k - \boldsymbol{b}_k^T\boldsymbol{x}_k$$

where $\boldsymbol{c}_k$ and $\boldsymbol{y}_k$ are the forward filter coefficient and input vectors, respectively, with $N$ elements as defined earlier in Section II-C2, $\boldsymbol{b}_k$ is the vector of the canceler coefficients spaced $T$ seconds apart, and $\boldsymbol{x}_k$ is the vector of data symbols. Each of these vectors is of length $N_1 + N_2$. The data vector $\boldsymbol{x}_k$ has elements $x_{k+N_1}, \cdots x_{k+1}, x_{k-1}, \cdots, x_{k-N_2}$. Note that the current data symbol $x_k$ is omitted. The canceler has $N_1$ noncausal and $N_2$ causal coefficients

$b_{-N_1}, \cdots, b_{-1}, b_1, \cdots, b_{N_2}$. To minimize the MSE $E[|z_k - x_k|^2]$, we can proceed as in Section II-D1b for the DFE. First, setting the derivative of the MSE with respect to $\boldsymbol{b}_k$ to zero we obtain

$$\boldsymbol{b}_k = \boldsymbol{H}\boldsymbol{c}_k$$

where $\boldsymbol{H}$ is an $(N_1 + N_2) \times N$ matrix with elements

$$h_{i,j} = h(iT - jKT/M), \qquad \begin{matrix} i = -N_1, \cdots, 1, 1, \cdots, N_2 \\ j = 0, 1, \cdots, N-1. \end{matrix}$$

Using this result, it can be shown that the optimum set of forward filter coefficients is given by

$$\boldsymbol{c}_{\text{opt}} = \hat{\boldsymbol{A}}^{-1}\boldsymbol{\alpha}$$

where the matrix $\hat{\boldsymbol{A}}$ has elements $\hat{a}_{i,j}$ similar to $a_{i,j}$ given by (26), except that the summation over $k$ now excludes the set $k = -N_1, \cdots, 1, 1, \cdots, N_2$ (which is in the span of the canceler). The vector $\boldsymbol{\alpha}$ is the same as in (27). Finally, the optimum canceler coefficient vector $\boldsymbol{b}_{\text{opt}} = \boldsymbol{H}\boldsymbol{c}_{\text{opt}}$ consists of $T$-spaced samples of the impulse response of the channel and the forward filter in cascade, omitting the reference sample.

The role of the forward filter in this structure may be better understood from the following point of view. Given an optimum (LMS) desired impulse response (DIR) for the canceler (with the reference sample forced to unity), the forward filter equalizes the channel response to this desired impulse response with least MSE. In fact both the DIR, modeled by the canceler coefficients, and the forward filter coefficients are being jointly optimized to minimize the final MSE. The same basic idea was used by Falconer and Magee [17] to create a truncated DIR channel for further processing by the Viterbi algorithm. In [17], a unit energy constraint was imposed on the DIR while in the ISI canceler structure the reference sample of the DIR is constrained to be unity. Another difference between the two structures is the use of a forward filter with fractional tap spacing rather than a predetermined "matched" filter followed by a $T$-spaced equalizer. The use of a fractionally spaced forward filter for creating a truncated DIR was proposed in [88] noting its ability to perform combined adaptive matched filtering and equalization.

As we allow the lengths of the forward filter and the canceler to grow without bound, the forward filter evolves into a matched filter with frequency response [34], [80]

$$C(f) = \frac{H^*(f)}{N_0 + R_0}.$$

The canceler models the $T$-spaced impulse response of the channel and forward filter in cascade, all except the reference sample. The frequency response of the canceler may be written as

$$B(f) = \frac{S_{hh}(f)}{N_0 + R_0} - T\int_0^{1/T} \frac{S_{hh}(f)}{N_0 + R_0}\, df = \frac{S_{hh}(f) - R_0}{N_0 + R_0}.$$

Note that since all the quantities in the above expression for the canceler frequency response are real, the canceler coefficients must be Hermitian symmetric.

The output noise power spectrum is given by

$$N_0|C(f)|^2 = N_0 S_{hh}(f)/[N_0 + R_0]^2$$

and the ISI power spectrum may be written as

$$|1 - S_{hh}(f)/[N_0 + R_0]^2 + B(f)|^2 = N_0^2/[N_0 + R_0]^2.$$

The output MSE, obtained by integrating the sum of the noise and ISI power spectra, is equal to the zero-ISI matched filter bound, i.e.,

$$\epsilon_{\min}(\text{mf}) = N_0/[N_0 + R_0].$$

The critical question regarding decision aided ISI cancellation is the effect of tentative decision errors on the final error probability of the receiver. Published results answering this question are not yet available. However, reduced MSE has also been reported [5] when a version of the decision-aided receiver structure is used to cancel nonlinear ISI.

*3) Adaptive Filters for MLSE:* Adaptive receive filtering prior to Viterbi detection is of interest from two points of view. First, for unknown and/or slowly time-varying channels the receive filter must be adaptive in order to obtain the ultimate performance gain from maximum-likelihood sequence estimation. Secondly, the complexity of the MLSE becomes prohibitive for practical channels with a larger number of ISI terms. Therefore, in a practical receiver, an adaptive receive filter may be used to limit the time spread of the channel as well as to track slow time variation in the channel characteristics [17], [89].

By a development similar to that given in Section II-C2, it can be shown that a fractionally spaced transversal filter can model the characteristics of the WMF proposed by Forney [24]. However, the constraint on this filter to produce zero anticausal ISI makes it difficult to derive an algorithm for updating the filter coefficients in an adaptive receiver.

The general problem of adaptive receive filtering for MLSE may be approached as follows. We know from Section II-C2 that an LMS FSE produces the composite response of a matched filter and an LMS $T$-spaced equalizer, when the MSE is defined with respect to a unit pulse DIR. In general, an LMS FSE frequency response can always be viewed as the composite of a matched filter and an LMS $T$-spaced equalizer response

$$H^*(f)\,G(f)/[N_0 + S_{hh}(f)] \tag{39}$$

where $G(f)$ is the DIR frequency response with respect to which the MSE is minimized. After symbol-rate sampling, the signal spectrum at the output of the FSE is given by $S_{hh}(f)G(f)/[N_0 + S_{hh}(f)]$, $|f| \leq 1/2T$, and the output noise power spectrum may be written as

$$N_0 S_{hh}(f)|G(f)|^2/[N_0 + S_{hh}(f)]^2, \qquad |f| \leq 1/2T.$$

The selection of the DIR is therefore the crux of the problem. Fredericsson [30] has shown that from an effective MSE point of view, best performance is obtained when the DIR is selected such that its power spectrum is

$$|G(f)|^2 = [N_0 + S_{hh}(f)]/(R_0 + N_0), \qquad |f| \leq 1/2T. \tag{40}$$

With this optimum DIR and the receive filter selected according to (39), the residual ISI power spectrum at the input to the Viterbi algorithm is given by

$$|G(f) - S_{hh}(f)\,G(f)/[N_0 + S_{hh}(f)]|^2 = N_0/[N_0 + S_{hh}(f)](R_0 + N_0).$$

Summing the residual ISI and noise power spectra we

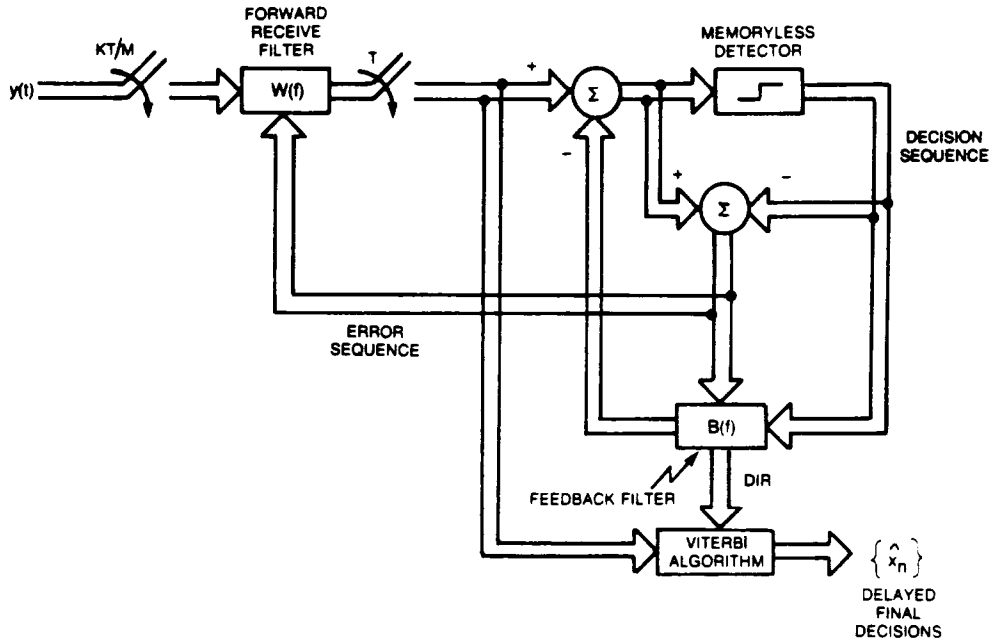


**Fig. 24.** Adaptive MLSE receiver with causal desired impulse response (DIR) developed via decision-feedback equalization.

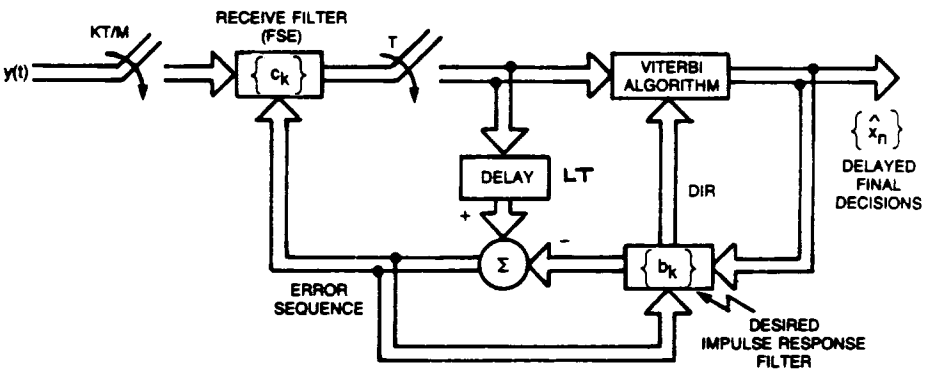


**Fig. 25.** General form of adaptive MLSE receiver with finite-length desired impulse response.

obtain the power spectrum of the combined error sequence at the input to the Viterbi algorithm. The error sequence is found to be white with mean-squared value $N_0/(R_0 + N_0)$, which is equal to the matched-filter bound. Note that the Gaussian noise component of the error sequence is approximately white at a moderately high SNR. However, the total error sequence while white is not Gaussian due to residual ISI.

Let us select the DIR such that $G(f)$ is the causal factor of the power spectrum given in (40). Then the resulting FSE (39) may be recognized as the optimum forward filter of an infinite-length conventional LMS DFE. At moderately high SNR this FSE approaches the WMF. An MLSE receiver using such a receive filter is shown in Fig. 24. This receiver can easily be made adaptive by updating both the FSE and the feedback filter coefficients to jointly minimize the mean-squared value of the error sequence. The feedback filter coefficients also provide all but the first DIR coefficient for use by the Viterbi algorithm. The first DIR coefficient is assumed to be unity.

Fredericsson [30] points out the difficulty of obtaining a general explicit solution for the optimum truncated DIR of a specified finite length. However, when the DIR is limited to two or three terms [17], [30], it invariably takes the form of one or the other familiar class of partial response systems [48] with a null at one or both band edges. As mentioned in Section II-D, the Viterbi algorithm is able to recover most of the nearly 3-dB loss which otherwise results from use of bandwidth-efficient partial-response systems.

Several methods of jointly optimizing the fractionally spaced receive filter and the truncated DIR are available which minimize the MSE at the input to the Viterbi algorithm (VA). These methods differ in the form of constraint [17], [68] on the DIR which is necessary in this optimization process to exclude the selection of the null DIR corresponding to no transmission through the channel. The general form of such a receiver is shown in Fig. 25.

One such constraint is to restrict the DIR to be causal and to restrict the first coefficient of the DIR to be unity. In this case, the delay $LT$ in Fig. 25 is equal to the delay through the VA and the first coefficient of $\{b_k\}$ is constrained to be unity.

If the causality constraint is removed (as in [34] for decision-aided ISI cancellation) but the reference (or center) coefficient of the DIR is constrained to unity, another form of the receiver shown in Fig. 25 is obtained [103]. In this

case the delay $LT$ is equal to the delay through the VA plus $N_1T$ where the DIR has $2N_1 + 1$ coefficients with its center coefficient constrained to be unity. As before, this structure can easily be made adaptive by updating the FSE and DIR coefficients in a direction opposite to the gradient of the squared error with respect to each coefficient.

The least restrictive constraint on the DIR is the unit energy constraint proposed by Falconer and Magee [17]. This leads to yet another form of the receiver structure shown in Fig. 25. However, the adaptation algorithm for updating the desired impulse response coefficients $\{b_k\}$ is considerably more complicated (see [17]). Note that the fixed predetermined WMF and $T$-spaced prefilter combination of [17] has been replaced in Fig. 25 by a general fractionally spaced adaptive filter.

A common characteristic of the above mentioned truncated DIR suboptimum MLSE receiver structures is that the sample sequence at the input to the VA contains residual ISI (with respect to the DIR), and that the noise sequence is not white. Bounds on the performance of the VA in the presence of correlated noise and residual ISI are developed in [89]. In practice, so long as the DIR power spectrum more or less matches the nulls or high-attenuation regions of the channel folded power spectrum, a reasonable length FSE can manipulate the channel response to the truncated DIR without significant noise enhancement or residual ISI. However, the degree to which the noise sequence at the VA input is uncorrelated depends on the constraints imposed on the DIR. As mentioned earlier, as the causal DIR length is allowed to grow, the forward filter in the structure of Fig. 25 approaches the WMF at moderately high SNR, resulting in uncorrelated noise at the VA input. On the other hand, if the DIR is noncausal, the receive filter approaches a matched filter as the DIR length approaches the length of the original channel impulse response, resulting in the noise to be colored according to the folded power spectrum of the received signal. For such receiver structures, the modified VA [103], which takes the noise correlation into account, is more appropriate.

### E. Timing Phase Sensitivity

As noted in earlier sections, conventional suboptimum receiver structures based on $T$-spaced equalizers suffer from extreme sensitivity to sampler timing phase. The inherent insensitivity of the performance of fractionally spaced equalizers to timing phase was heuristically explained in Section I-H. Here we present an overview of the influence of timing phase on the performance of various receiver structures.

Let us reconsider the conventional linear receiver of Section II-C1 in the presence of a timing phase offset $t_0$. Assume that $H(f) = 0$, $|f| > 1/T$, i.e., the channel has at most 100-percent excess bandwidth. The sampled noise sequence at the output of the matched filter is given by

$$n_k = n(kT + t_0) = \int n(t)\, h^*(t - kT - t_0)\, dt$$

where $h^*(-t)$ is the impulse response of the matched filter $H^*(f)$. The power spectrum, $N_0S_{hh}(f)$, of the noise sequence $\{n_k\}$ is independent of the timing phase $t_0$.

The sampled signal spectrum, however, is a function of $t_0$ according to

$$S_{hh}(f, t_0) = \exp(-j2\pi f t_0)\big[\,|H(f)|^2 + |H(f - 1/T)|^2 \cdot \exp(j2\pi t_0/T)\big], \qquad 0 \leq f \leq 1/T.$$

Note that when $t_0 = 0$, the alias $|H(f - 1/T)|^2$ adds constructively to $|H(f)|^2$, while for $t_0 = T/2$, destructive aliasing takes place in the foldover region around $1/2T$ hertz. In particular, if the channel power spectrum is the same at the two band edges, a null is created in the sampled signal spectrum at $f = 1/2T$ hertz when $t_0 = T/2$.

The MSE corresponding to (3) may be written as a function of timing phase, i.e.,

$$\epsilon(t_0) = T\int_0^{1/T} |1 - S_{hh}(f, t_0)\, C(f)|^2 + N_0 S_{hh}(f)|C(f)|^2\, df$$

where $C(f)$ is the periodic frequency response of the $T$-spaced equalizer which follows the sampler. Proceeding to optimize $C(f)$, as in Section II-B1, the minimum MSE may be derived as a function of $t_0$

$$\epsilon_{\min}(t_0) = T\int_0^{1/T} N_0/\big[N_0 + S_{hh}^2(f, t_0)/S_{hh}(f)\big]\, df. \tag{41}$$

Note that if $S_{hh}^2(f, t_0)/S_{hh}(f)$ is small in the foldover region, due to poor choice of $t_0$, the intergrand becomes relatively large in that range of frequencies. This leads to a larger $\epsilon_{min}$. Clearly, timing phase is unimportant when there is no excess bandwidth and therefore no aliasing.

The above development shows that for systems with excess bandwidth, the performance of the conventional MMSE linear receiver with a fixed matched filter is sensitive to choice of timing phase due to the inability of the $T$-spaced equalizer to invert a "null" in the sampled signal spectrum without excessive noise enhancement. This sensitivity can be avoided if the matched filter spectrum $H^*(f)$ is adjusted by a linear phase factor, $\exp(j2\pi f t_0)$, to explicitly compensate for the timing offset $t_0$. An infinite-length fractionally spaced equalizer obtains its insensitivity to timing phase in this way.

For instance, given a timing offset $t_0$, an infinite length $T/2$ equalizer synthesizes the frequency response

$$W(f) = H^*(f) \exp(j2\pi f t_0)/\big[N_0 + S_{hh}(f)\big], \qquad |f| \leq 1/T.$$

The signal spectrum at the equalizer input is the DFT of the $T/2$-sampled channel impulse response $\{h(kT/2 + t_0)\}$, i.e., $H(f, t_0) = H(f)\exp(-j2\pi f t_0), |f| \leq 1/T$. Aliasing, due to symbol-rate sampling, takes place at the equalizer output after $W(f)$ has explicitly compensated for the timing offset. This results in the equalized signal spectrum

$$H(f, t_0)W(f) + H(f - 1/T, t_0)W(f - 1/T) = S_{hh}(f)/\big[N_0 + S_{hh}(f)\big], \qquad 0 \leq f \leq 1/T.$$

Thus the residual ISI, noise, and MSE at the $T/2$ equalizer output are all independent of the timing phase.

In a practical conventional receiver structure, the matched filter is typically replaced by a filter matched to the transmitted pulse or the received pulse over the average channel. Mazo [64] has shown that under some assumptions, the best timing phase for an infinite-length $T$-spaced equalizer is one which maximizes the energy in the band edge, $f = 1/2T$ hertz, component in the symbol-rate sampled signal at the equalizer input. However, on channels with severe delay distortion and moderately large excess band-

width, e.g., 50-percent, even the timing phase which maximizes band-edge energy can lead to near nulls elsewhere in the folded spectrum. On such a channel, a $T$-spaced equalizer performs poorly regardless of the choice of timing phase. Conditions for amplitude depressions to occur in the folded spectrum are given in [91]. For instance, asymmetrical delay distortion in the two band-edge regions can cause the phase responses of the two aliasing components to differ by $\pi$ radians a few hundred hertz away from the band edge. Thus while the aliases add at the band edge, they subtract where they are opposite in phase. This cancellation may be accentuated by amplitude distortion in the channel causing the amplitudes of the canceling aliases to be nearly equal.

By contrast, a finite-length FSE maintains its ability to compensate for a timing offset in such a way as to equalize with the minimum of noise enhancement.

The results discussed above can be extended to decision-feedback receiver structures. Consider a conventional DFE consisting of a matched filter, a symbol-rate sampler, and infinite-length $T$-spaced forward and feedback filters. The MSE of such a receiver can be derived as a function of the timing phase (using the method given in [94])

$$\epsilon_{\min}(t_0) = \exp\left\{-T\int_0^{1/T} \ln\left[1 + S_{hh}^2(f, t_0)/N_0 S_{hh}(f)\right] df\right\}. \tag{42}$$

As $t_0$ is varied, the greatest deviation in $\epsilon_{\min}(t_0)$ occurs for a channel such that

$$|H(f)|^2 = \begin{cases} 1, & |f| \leqslant (1-\alpha)/2T \\ 0.5, & (1-\alpha)/2T \leqslant |f| \leqslant (1+\alpha)/2T \end{cases}$$

where $\alpha$ is the rolloff factor. For this channel

$$S_{hh}(f, t_0)|_{t_0=0} = S_{hh}(f) = 1, \qquad 0 \leqslant f \leqslant 1/T$$

and

$$S_{hh}(f, t_0)|_{t_0=T/2} = \begin{cases} 1, & |f| \leqslant (1-\alpha)/2T \\ 0, & (1-\alpha)/2T \leqslant f \leqslant (1+\alpha)/2T. \end{cases}$$

Using this result in (42) we obtain for the DFE

$$\epsilon_{\min}(t_0)|_{t_0=0} = 1/(1 + 1/N_0)$$

and

$$\epsilon_{\min}(t_0)|_{t_0=T/2} = 1/(1 + 1/N_0)^{(1-\alpha)}.$$

For $N_0 = 0.01$ (20-dB SNR) and $\alpha = 0.1$, the MSE of the decision-feedback equalizer degrades by 2 dB when $t_0$ is varied from the best choice $t_0 = 0$ to the worse choice $t_0 = T/2$. The corresponding result for the conventional linear receiver is obtained by using (41)

$$\epsilon_{\min}(t_0)|_{t_0=0} = N_0/(1 + N_0)$$

and

$$\epsilon_{\min}(t_0)|_{t_0=T/2} = \alpha + (1-\alpha)N_0/(1 + N_0).$$

For $N_0 = 0.01$ and $\alpha = 0.1$, the MSE of the linear receiver degrades by 10.4 dB as $t_0$ is varied from 0 to $T/2$.

This example illustrates the ability of the conventional DFE to compensate for the spectral null created by poor choice of timing phase with much less noise enhancement than a conventional linear receiver. However, this relative insensitivity may not translate into actual performance insensitivity from an error probability point of view due to error propagation in the DFE. In order to cancel the ISI due to a spectral null created by a bad timing phase, the feedback filter must develop a relatively long impulse response with large magnitude coefficients. Such a DFE is likely to suffer from severe error propagation.

The performance of a DFE with a fractionally spaced forward filter is, of course, insensitive to timing phase by virtue of explicit compensation of the timing offset by the forward filter before symbol-rate sampling. The same comment applies to a fractionally spaced receive filter used in conjunction with a maximum-likelihood sequence estimator.

## III. Least Mean-Square Adaptation

In this section we expand upon the topics briefly introduced in Sections I-C through I-E, that is, LMS adaptation algorithms, their convergence properties, and excess MSE. The effect of finite precision in digital implementations is also discussed. The results of this section are applicable to other forms of adaptive filters, e.g., an echo canceler, with appropriate reinterpretation of terms.

The deterministic gradient algorithm, which is of little practical interest, is presented first to set the stage for a discussion of the LMS or stochastic gradient algorithm.

### A. Deterministic Gradient Algorithm

When the equalizer input covariance matrix $\boldsymbol{A}$ and the cross-correlation vector $\boldsymbol{\alpha}$ (see Sections II-C3 and II-C4) are known, one can write the MSE as a function of $\boldsymbol{A}$, $\boldsymbol{\alpha}$, and the equalizer coefficient vector $\boldsymbol{c}_k$ according to

$$\epsilon_k = \boldsymbol{c}_k^{*T}\boldsymbol{A}\boldsymbol{c}_k - 2\,\mathrm{Re}\left[\boldsymbol{c}_k^{*T}\boldsymbol{\alpha}\right] + 1.$$

Taking the gradient of the MSE with respect to $\boldsymbol{c}_k$ gives

$$\partial\epsilon_k/\partial\boldsymbol{c}_k = 2(\boldsymbol{A}\boldsymbol{c}_k - \boldsymbol{\alpha}). \tag{43}$$

Thus a deterministic (or exact) gradient algorithm for adjusting $\boldsymbol{c}_k$ to minimize $\epsilon_k$ can be written as

$$\boldsymbol{c}_{k+1} = (\boldsymbol{I} - \Delta\boldsymbol{A})\boldsymbol{c}_k + \Delta\boldsymbol{\alpha} \tag{44}$$

where $\Delta$ is the step-size parameter. (This update procedure is also known as the steepest descent algorithm.) Using the fact that $\boldsymbol{\alpha} = \boldsymbol{A}\boldsymbol{c}_{\mathrm{opt}}$, from (11) or (25), and subtracting $\boldsymbol{c}_{\mathrm{opt}}$ from both sides of (44), we obtain

$$\boldsymbol{c}_{k+1} - \boldsymbol{c}_{\mathrm{opt}} = (\boldsymbol{I} - \Delta\boldsymbol{A})(\boldsymbol{c}_k - \boldsymbol{c}_{\mathrm{opt}}). \tag{45}$$

In order to analyze the stability and convergence of the deterministic gradient algorithm, we use coordinate transformation to diagonalize the set of equations (45) so that

$$\boldsymbol{d}_{k+1} = (\boldsymbol{I} - \Delta\boldsymbol{\Lambda})\boldsymbol{d}_k \tag{46}$$

where the transformed coefficient deviation vector is defined in (30).

Since $\boldsymbol{\Lambda}$ is a diagonal matrix, it is clear from (46) that the $i$th element of $\boldsymbol{d}_k$ decays geometrically according to

$$d_{ik} = (1 - \Delta\lambda_i)^k d_{i0}, \qquad i = 0, 1, \cdots, N-1 \tag{47}$$

where $\lambda_i$ is the $i$th eigenvalue of $\boldsymbol{A}$, and $d_{i0}$ is the initial value of the $i$th transformed tap-gain deviation. Recall from

Section II-C2 that given $\boldsymbol{d}_k$, the MSE at step $k$ may be written as the sum of $\epsilon_{\min}$ and the excess MSE

$$\epsilon_k = \epsilon_{\min} + \boldsymbol{d}_k^{*T}\boldsymbol{\Lambda}\boldsymbol{d}_k = \epsilon_{\min} + \sum_{i=0}^{N-1} \lambda_i(1 - \Delta\lambda_i)^{2k}|d_{i0}|^2. \quad (48)$$

Given finite initial deviations, the deterministic algorithm is stable and the MSE converges to $\epsilon_{\min}$ provided

$$0 < \Delta < 2/\lambda_{\max} \quad (49)$$

where $\lambda_{\max}$ is the maximum eigenvalue of $\boldsymbol{A}$. If all the eigenvalues of $\boldsymbol{A}$ are equal to $\lambda$ and $\Delta$ is selected to be $1/\lambda$, the excess MSE will be reduced to zero in one adjustment step of the deterministic gradient algorithm (44). For a $T$-spaced equalizer, this condition corresponds to a flat folded power spectrum at the equalizer input.

When the eigenvalues of $\boldsymbol{A}$ have a large spread, i.e., the ratio $p = \lambda_{\max}/\lambda_{\min}$ is large, no single value of the step size $\Delta$ leads to fast convergence of all the tap-gain deviation components. When $\Delta = 2/(\lambda_{\max} + \lambda_{\min})$, the two extreme tap-gain deviation components converge at the same rate according to $[(p-1)/(p+1)]^k$ [33]. All other components converge at a faster rate with a time constant at most as large as $(p+1)/2$ iterations. The impact of the large eigenvalue spread on the convergence of the excess MSE is somewhat less severe because tap-gain deviations corresponding to the small eigenvalues contribute less to the excess MSE (see (48)).

Two possibilities for speeding up the convergence of this deterministic algorithm can be devised. The first of these involves the use of a variable step size $\Delta_k$ in (44). This leads to the relationships

$$d_{ik} = d_{i0}\prod_{n=0}^{k-1}(1 - \Delta_n\lambda_i), \qquad i = 0,1,\cdots,N-1$$

corresponding to (47). Observe that, if the eigenvalues are known beforehand, complete convergence can be obtained in $N$ steps by using a variable step size provided the $N$ values of the step size are selected such that $\Delta_n = 1/\lambda_n$, $n = 0,1,\cdots,N-1$. Thus each step reduces one of the tap-gain deviation components to zero.

The second method of obtaining fast convergence is to replace the scalar $\Delta$ in (44) by the precomputed inverse matrix $\boldsymbol{A}^{-1}$. This is tantamount to direct noniterative solution, since regardless of initial conditions or the eigenvalue spread of $\boldsymbol{A}$, or the equalizer size $N$, optimum solution is obtained in one step.

### B. LMS Gradient Algorithm

In practice, the channel characteristics are not known beforehand. Therefore, the gradient of the MSE cannot be determined exactly and must be estimated from the noisy received signal. The LMS gradient algorithm [109] is obtained from the deterministic gradient algorithm (44) by replacing the gradient

$$2(\boldsymbol{A}\boldsymbol{c}_k - \boldsymbol{\alpha}) = 2E\left[\boldsymbol{y}_k^*\left(\boldsymbol{y}_k^T\boldsymbol{c}_k - x_k\right)\right]$$

by its unbiased but noisy estimate $\boldsymbol{y}_k^*(\boldsymbol{y}_k^T\boldsymbol{c}_k - x_k)$. The equalizer coefficients are adjusted once in every symbol interval according to

$$\boldsymbol{c}_{k+1} = \boldsymbol{c}_k - \Delta\boldsymbol{y}_k^*\left(\boldsymbol{y}_k^T\boldsymbol{c}_k - x_k\right) = \boldsymbol{c}_k - \Delta\boldsymbol{y}_k^*e_k \quad (50)$$

where $\boldsymbol{y}_k$ is the equalizer input vector, $x_k$ is the received data symbol, and $e_k$ is the error in the equalizer output.

Subtracting $\boldsymbol{c}_{\text{opt}}$ from both sides of (50) allows us to write

$$(\boldsymbol{c}_{k+1} - \boldsymbol{c}_{\text{opt}}) = (\boldsymbol{I} - \Delta\boldsymbol{y}_k^*\boldsymbol{y}_k^T)(\boldsymbol{c}_k - \boldsymbol{c}_{\text{opt}}) - \Delta\boldsymbol{y}_k^*e_{k\,\text{opt}}$$

where $e_{k\,\text{opt}}$ is the instantaneous error if the optimum coefficients were used. The transformed coefficient deviation vector $\boldsymbol{d}_k$ is now a random quantity. Neglecting the dependence of $\boldsymbol{c}_k$ on $\boldsymbol{y}_k$, we see that the mean of $\boldsymbol{d}_k$ follows the recursive relationship (46), i.e.,

$$E[\boldsymbol{d}_{k+1}] = (\boldsymbol{I} - \Delta\boldsymbol{\Lambda})E[\boldsymbol{d}_k]. \quad (51)$$

The ensemble average of the MSE evolves according to

$$\epsilon_k = \epsilon_{\min} + E\left[\boldsymbol{d}_k^{*T}\boldsymbol{\Lambda}\boldsymbol{d}_k\right]$$

where the second term on the right is the average excess MSE

$$\epsilon_{\Delta k} = \sum_{i=0}^{N-1}\lambda_i E\left[|d_{ik}|^2\right].$$

This quantity is difficult to evaluate exactly in terms of the channel and equalizer parameters. Using the assumption that the equalizer input vectors $\boldsymbol{y}_k$ are statistically independent [38], [65], [84], [102], the following approximate recursive relationship can be derived:

$$E\left[|\boldsymbol{d}_{k+1}|^2\right] = \boldsymbol{M}E\left[|\boldsymbol{d}_k|^2\right] + \Delta^2\epsilon_{\min}\boldsymbol{\lambda} \quad (52)$$

where $\boldsymbol{\lambda}$ is the vector of eigenvalues of $\boldsymbol{A}$ and the $N \times N$ matrix $\boldsymbol{M}$ has elements

$$M_{ij} = (1 - 2\Delta\lambda_i)\delta_{ij} + \Delta^2\lambda_i\lambda_j.$$

Let $\rho$ be the ratio of the maximum to the effective average eigenvalue of $\boldsymbol{A}$.

Three important results can be derived [38] using this line of analysis [102] and eigenvalue bounds [38]:

1) The LMS algorithm is stable if the step size $\Delta$ is in the range

$$0 < \Delta < 2/(N\rho\bar{\lambda}) \quad (53)$$

where $\bar{\lambda}$ is the average eigenvalue of $\boldsymbol{A}$ and is equal to the average signal power at the equalizer input, defined according to $E[\boldsymbol{y}_k^T\boldsymbol{y}_k^*]/N$. When $\Delta$ satisfies (53), all eigenvalues of the matrix $\boldsymbol{M}$ in (52) are less than one in magnitude permitting mean-square coefficient deviations in (52) to converge.

2) The excess MSE follows the recursive relationship

$$\epsilon_{\Delta k+1} = \left[1 - 2\Delta\bar{\lambda} + \Delta^2N\rho(\bar{\lambda})^2\right]\epsilon_{\Delta k} + \Delta^2\epsilon_{\min}N\rho(\bar{\lambda})^2. \quad (54)$$

If $\Delta$ is selected to minimize the excess MSE at each iteration, we would obtain

$$\Delta_k = \epsilon_{\Delta k}/(\epsilon_{\min} + \epsilon_{\Delta k})(N\rho\bar{\lambda}).$$

Initially $\epsilon_{\Delta k} \gg \epsilon_{\min}$, so that fastest convergence is obtained with an initial step size

$$\Delta_0 = 1/(N\rho\bar{\lambda}). \quad (55)$$

Note that $\Delta_0$ is half as large as the maximum permissible step size for stable operation.

3) The steady-state excess MSE can be determined from (54) to be

$$\epsilon_\Delta = \Delta\epsilon_{\min}N\rho\bar{\lambda}/(2 - \Delta N\rho\bar{\lambda}). \quad (56)$$

A value of $\Delta = \Delta_0$ results in $\epsilon_\Delta = \epsilon_{\min}$, that is, the final MSE

is 3 dB greater than the minimum achievable MSE. In order to reduce $\epsilon_\Delta$ to $\gamma\epsilon_{min}$, where $\gamma$ is a small fraction, $\Delta$ must be reduced to

$$\Delta = 2\gamma/[(1+\gamma)N\rho\bar{\lambda}] = 2\Delta_0\gamma/(1+\gamma). \quad (57)$$

For instance, a reduction of $\Delta$ to 0.1 $\Delta_0$ results in a steady-state MSE which is about 0.2 dB greater than the minimum achievable MSE.

Note that the impact of a distorted channel, with an eigenvalue ratio $\rho > 1$, on the excess MSE, its rate of convergence, and the choice of $\Delta$ is the same as if the number of equalizer coefficients $N$ was increased to $N\rho$. For a $T$-spaced equalizer, some of the eigenvalues (and hence $\rho$) depend on the timing phase and the channel envelope delay characteristics in the band-edge regions.

The above results apply equally to $T$-spaced and fractionally spaced equalizers, except that the recursion for the excess MSE (54) for a $KT/M$-spaced equalizer is given by

$$\epsilon_{\Delta K+1} = \left[1 - 2\Delta(M/K)\bar{\lambda} + \Delta^2 N(M/K)\rho(\bar{\lambda})^2\right] + \Delta^2\epsilon_{min}N(M/K)\rho(\bar{\lambda})^2.$$

As mentioned earlier, only about $K/M$ of the eigenvalues of the correlation matrix $\boldsymbol{A}$ are significant for a $KT/M$ FSE. For the same average signal power at the equalizer input, equal to the average eigenvalue $\bar{\lambda}$, the significant eigenvalues are generally $M/K$ times larger for an $N$-coefficient $KT/M$ FSE compared with the eigenvalues for an $N$-coefficient $T$-spaced equalizer. Thus the eigenvalue ratio $\rho$ for an FSE should be computed only over the significant eigenvalues of $\boldsymbol{A}$. Note that $\boldsymbol{A}$ and $\rho$ for an FSE are independent of timing phase and channel envelope delay characteristics, and $\rho = 1$ when the unequalized amplitude shape is square-root of Nyquist.

For well-behaved channels ($\rho$ approximately 1), $\epsilon_{\Delta k}$ for an $N$-coefficient $KT/M$ FSE, using the best initial step size given by (55), initially converges faster as $(1 - M/KN)^k$ compared with $(1 - 1/N)^k$ for an $N$-coefficient $T$ equalizer using the same best initial step size. Conversely, an $MN/K$-coefficient FSE, using a step size $K/M$ times as large, generally exhibits the same behavior with respect to the convergence and steady-state value of excess MSE as an $N$-coefficient $T$-spaced equalizer.

As a rule of thumb, the symbol-by-symbol LMS gradient algorithm with the best initial $\Delta$ leads to a reduction of about 20 dB in MSE for well-behaved channels ($\rho$ approximately 1) in about five times the time span $T_{eq}$ of the equalizer. At this time it is desirable to reduce $\Delta$ by a factor of 2 for the next $5T_{eq}$ seconds to permit finer tuning of the equalizer coefficient. Further reduction in excess MSE can be obtained by reducing $\Delta$ to its steady-state value according to (57). (For distorted channels, an effective time span of $\rho T_{eq}$ should be substituted in the above discussion.)

### C. Digital Precision Considerations

The above discussion may suggest that it is desirable to continue to reduce $\Delta$ in order to reduce the excess MSE to zero in the steady state. However, this is not advisable in a practical limited precision digital implementation of the adaptive equalizer. Observe [36] that as $\Delta$ is reduced, the coefficient correction terms in (50), on the average, become smaller than half the least significant bit of the coefficient; adaptation stalls and the MSE levels off. If $\Delta$ is reduced further, the MSE increases if the channel characteristics change at all or if some adjustments made at peak errors are large enough to perturb the equalizer coefficients.

The MSE which can be attained by a digital equalizer of a certain precision can be approximated as follows [38]. Let the equalizer coefficients be represented by a uniformly quantized number of $B$ bits (including sign) in the range $(-1,1)$. Then the real and imaginary parts of the equalizer coefficients will continue to adapt, on the average, so long as

$$\Delta(\epsilon\bar{\lambda}/2)^{1/2} \geq 2^{-B}. \quad (58)$$

It is desirable to select a compromise value of $\Delta$ such that the total MSE $\epsilon = \epsilon_{min} + \epsilon_\Delta$, with $\epsilon_\Delta$ predicted by infinite-precision (analog) analysis (56), is equal to the lower limit on $\epsilon$ determined by digital precision (58). The required precision can be estimated by substituting $\Delta$ from (57) into (58)

$$2^B \geq \left[N\rho(1+\gamma)/\sqrt{2}\gamma\right](\bar{\lambda}/\epsilon)^{1/2} \quad (59)$$

where $(\bar{\lambda}/\epsilon)$ can be recognized as the desired equalizer input-power-to-output-MSE ratio.

As an example, consider a 32-tap $T$-spaced equalizer ($N = 32$) for a well-behaved channel ($\rho = 1$). Select $\gamma = 0.25$ (corresponding to a 1-dB increase in output MSE over $\epsilon_{min}$) and a desired equalizer output-signal-power-to-MSE ratio of 24 dB (adequate for 9.6-kbit/s transmission). Let the input-to-output power scale factor for this equalizer be 2, so that $(\bar{\lambda}/\epsilon)^{1/2} = 10^{27/20}$. Solving (59) for $B$, we find that 12-bit precision is required.

The required coefficient precision increases by 1 bit for each doubling of the number of coefficients, and for each 6-dB reduction in desired output MSE. However, for a given $\epsilon_{min}$, each 6-dB reduction in excess MSE $\epsilon_\Delta$ requires a 2-bit increase in the required coefficient precision. This becomes the limiting factor in some adaptive filters, e.g., an echo canceler, which must track slow variations in system parameters in the presence of a large uncorrelated interfering signal [108].

Note that the precision requirement imposed by the LMS gradient adaptation algorithm is significantly more stringent than a precision estimate based on quantization noise due to roundoff in computing the sum of products for the equalizer output. If each product is rounded individually to $B$ bits and then summed, the variance of this roundoff noise is $N2^{-2B}/3$. Assuming an equalizer output signal power of 1/6, the signal-to-output roundoff noise ratio is $2^{2B}/2N$, which is 54 dB for $B = 12$ and $N = 32$: 30 dB greater than the desired 24-dB signal-to-MSE ratio in our example.

Roundoff in the coefficient update process, which has been analyzed in [7], is another source of quantization noise in adaptive filters. This roundoff causes deviation of the coefficients from the values they take when infinite precision arithmetic is used. The MSE $\epsilon_r$ contributed by coefficient roundoff [7] is approximated by $\epsilon_r = N2^{-2B}/6\Delta$. Again using an output signal power value of 1/6, the signal-to-coefficient roundoff noise ratio is $\Delta 2^{2B}/N$. For our example, we obtain $\Delta = 0.0375$ from (59) using $\gamma = 0.25$, $N = 32$, $\rho = 1$, and $\bar{\lambda} = 1/3$. The signal-to-coefficient roundoff noise ratio is 43 dB for $B = 12$, which is 19 dB greater than the desired 24-dB signal-to-MSE ratio, suggesting that the effect of coefficient roundoff is insignificant.

Moreover, each time we reduce $\Delta$ by a factor of 2, the coefficient precision $B$ must be increased by 1 bit in order to prevent adaptation from stalling according to (58). This reduction in $\Delta$ and the corresponding increase in $B$ further reduces the MSE due to coefficient roundoff by 3 dB. Using the expression for $\epsilon_r$ given above, and $\Delta$ from (57), one can rewrite (58) as

$$\epsilon/\epsilon_r \geq 6\rho(1+\gamma)/\gamma.$$

Since $\rho \geq 1$ and $\gamma$ is typically a small fraction, the MSE due to coefficient roundoff, $\epsilon_r$, is always small compared to $\epsilon$ provided coefficient precision is sufficient to allow adaptation to continue for small desired values of $\gamma$ and $\Delta$. Thus coefficient roundoff noise can be neglected in the process of estimating LMS adaptive filter precision requirements.

Simple update schemes using nonlinear multipliers to produce the coefficient correction terms have been devised to address the precision problem and reduce implementation complexity at some penalty in performance [13]. Two such schemes are:

1) Sign-bit multiplication update

$$c_{k+1} = c_k - \Delta \operatorname{sgn}(y_k^*) \operatorname{sgn}(e_k). \tag{60}$$

2) Power-of-two multiplication update

$$c_{k+1} = c_k - \Delta f(y_k^*) f(e_k) \tag{61}$$

where

$$f(x) = \operatorname{sgn}(x) 2^{[\log_2|x|]}$$

and $[\cdot]$ is used here to denote the greatest integer less than or equal to the argument.

In (60) and (61), the $\operatorname{sgn}(\cdot)$ and $f(\cdot)$ functions are applied separately to each element of a vector and to the real and imaginary parts of a complex quantity. Analysis and simulation results show [13] that while sign-bit multiplication update (60) suffers a significant degradation in convergence time compared to the true LMS gradient update (50), the loss in performance is small when the power-of-two multiplication update (61) is used.

Whenever some of the eigenvalues of the input matrix $\boldsymbol{A}$ for an adaptive filter become vanishingly small, the output MSE is relatively insensitive to coefficient deviations corresponding to these eigenvalues. In some practical situations this may lead to numerical instability: when input signals are not adequately dense across the entire bandwidth of the adaptive filter (e.g., a signal consisting of a few tones at the input to a long adaptive equalizer or voice echo canceler), or an FSE which naturally has a set of relatively small eigenvalues, as mentioned in Section II-C4. In these cases, finite precision errors tend to accumulate along the eigenvectors (or frequency components) corresponding to small eigenvalues without significantly affecting the output MSE. If a natural converging force constraining these components (e.g., background noise) is weak, and the digital precision errors have a bias, these fluctuations may eventually take one or more filter coefficients outside the allowed numerical range, e.g., $(-1,1)$. This overflow can be catastrophic unless saturation arithmetic is used in the coefficient update process to prevent it.

Another solution is to use a so-called tap-leakage algorithm [40]

$$c_{k+1} = (1 - \Delta\mu) c_K - \Delta y_k^* e_k$$

where a decay factor $(1 - \Delta\mu)$ has been introduced into the usual stochastic gradient algorithm (50). The tap-leakage algorithm seeks to minimize a modified mean-square cost function

$$\mu c_k^* c_k^T + E\left[|e_k|^2\right]$$

which is a sum of the MSE and an appropriate fraction $\mu$ of the squared length of the $N$-dimensional coefficient vector. The decay tends to force the coefficient magnitudes and, by Parseval's theorem, the equalizer power spectral response toward zero. For an FSE, a value of $\mu$ on the order of a small eigenvalue is suggested [40].

The tap-leakage algorithm is similar to an adaptive predictor update algorithm for an ADPCM system [129] where the predictor coefficients are forced to decay toward predetermined compromise values during silence intervals. This prevents the effect of channel errors on the decoder adaptive predictor coefficients from persisting.

## IV. Fast Converging Equalizers

The design of update algorithms to speed up the convergence of adaptive filters has been a topic of intense study for more than a decade. Rapid convergence is important for adaptive equalizers designed for use with channels, such as troposcatter and HF radio, whose characteristics are subject to time variations [87]. In voice-band telephone applications, reduction of the initial setup time of the equalizer is important in polling multipoint networks [26] where the central site receiver must adapt to receive typically short bursts of data from a number of transmitters over different channels.

In this section we present an overview of three classes of techniques devised to speed up equalizer convergence.

### A. Orthogonalized LMS Algorithms

Recall from Section III-A that for the deterministic gradient algorithm, no single value of the step size $\Delta$ leads to fast convergence of all the coefficient deviation components when the eigenvalues of the equalizer input covariance matrix $\boldsymbol{A}$ have a large spread. Using the independence assumption, the same is true regarding the convergence of the mean of the coefficient deviations for the LMS gradient algorithm (see (51) in Section III-B). The excess MSE is a sum of the mean-square value of each coefficient deviation weighted by the corresponding eigenvalues of $\boldsymbol{A}$. Slow decay of some of these mean-square deviations, therefore, slows down the convergence of the excess MSE. Substituting the best initial $\Delta$ from (55) in (54), we obtain the recursion

$$\epsilon_{\Delta k+1} = (1 - 1/N\rho)\epsilon_{\Delta k} + \epsilon_{\min}/N\rho. \tag{62}$$

Observe that the initial decay of $\epsilon_{\Delta k}$ is geometric with a time constant of approximately $N\rho$ symbol intervals. Thus for the same length equalizer, a severely distorted channel ($\rho = 2$) will cause the rate of convergence of the LMS gradient algorithm to be slower by a factor of 2 compared to that for a good channel. The inadequacy of the LMS gradient algorithm for fast start-up receivers becomes obvious if we consider a 9.6-kbit/s, 2400-Bd modem with an equalizer spanning 32 symbol intervals. For a severely distorted channel, more than 320 equalizer adjustments over a 133-ms interval would be required before data transmission could begin.

For partial-response systems [48], [50], where a controlled amount of intersymbol interference is introduced to obtain a desired spectral shape, the equalizer convergence problem is fundamental. It can be shown that $\rho = 2$ for an ideal cosine-shaped spectrum at the equalizer input for a Class IV SSB partial-response [8] or a Class I QAM partial-response [90] system. Noting this slow convergence, Chang [8] suggested the use of a prefixed weighting matrix to transform the input signals to the equalizer tap gains to be approximately orthonormal. All eigenvalues of the transformed equalizer input covariance matrix are then approximately equal resulting in faster equalizer convergence.

Another orthogonalized LMS update algorithm [75], [88] is obtained by observing that the decay of the mean of all the transformed coefficient deviation components could be speeded up by using a diagonal matrix $\mathrm{diag}(\Delta_i)$ instead of the scaler $\Delta$ in (51), such that each element $\Delta_i$ of this matrix is the inverse of the corresponding element $\lambda_i$ of $\Lambda$. Transforming $\mathrm{diag}(\Delta_i)$ back to the original coordinate system, we obtain the orthogonalized LMS update algorithm

$$\boldsymbol{c}_{k+1} = \boldsymbol{c}_k - \boldsymbol{P}\boldsymbol{y}_k^* e_k. \tag{63}$$

As in Chang's scheme [8], the best value of the weighting matrix $\boldsymbol{P}$ is given by

$$\boldsymbol{P} = \boldsymbol{V}^{*T}\mathrm{diag}(\Delta_i)\boldsymbol{V} = \boldsymbol{V}^{*T}\boldsymbol{\Lambda}^{-1}\boldsymbol{V} = \boldsymbol{A}^{-1}$$

where the columns of the diagonalizing matrix $\boldsymbol{V}$ are eigenvectors of $\boldsymbol{A}$ (see (30)). In practice, $\boldsymbol{A}$ is not known beforehand, therefore, $\boldsymbol{P}$ can only approximate $\boldsymbol{A}^{-1}$. For instance, in partial-response systems where the dominant spectral shape is known beforehand, we can use $\boldsymbol{P} = \boldsymbol{S}^{-1}$, where $\boldsymbol{S}$ is the covariance matrix of the partial-response shaping filter.

A practical advantage of this algorithm (63) over Chang's structure is that the weighting matrix is in the path of the tap-gain corrections rather than the received signal. The computation required, therefore, need not be carried out to as much accuracy.

As we shall see in Section IV-C, the fastest converging algorithms are obtained when $\boldsymbol{P}$ is continually adjusted to do the best job of orthogonalizing the tap-gain corrections.

### *B. Periodic or Cyclic Equalization*

As mentioned in Section I-C, one of the most widely used [118], [119] methods of training adaptive equalizers in high-speed voice-band modems is based on *PN* training sequences with periods significantly greater than the time span of the equalizer. Here we discuss the techniques [70], [76], [90] which can be used to speed up equalizer convergence in the special case when the period of the training sequence is selected to be equal to the time span of the equalizer.

*1) Periodic or Averaged Update:* Consider a training sequence $\{x_k\}$ with period $NT$ for a $T$-spaced equalizer with $N$ coefficients. Let the equalizer coefficients be adjusted periodically, every $N$-symbol intervals, according to the following LMS algorithm with averaging:

$$c_n(k+1) = c_n(k) - \Delta \sum_{j=0}^{N-1} e_{kN+j} y^*(kNT + jT - nT),$$
$$n = 0,1,\cdots,N-1.$$

Let us denote the error sequence during the $k$th period by the vector $\boldsymbol{e}_k$ whose $j$th element is given by

$$e_{kN+j} = \sum_{n=0}^{N-1} c_n(k)\, y(kNT + jT - nT) - x_{kN+j},$$
$$j = 0,1,\cdots,N-1.$$

The update algorithm can be written in matrix notation as

$$\boldsymbol{c}_{k+1} = \boldsymbol{c}_k - \Delta \boldsymbol{Y}_k^* \boldsymbol{e}_k \tag{64}$$

where the $n,j$th element of the $N \times N$ matrix $\boldsymbol{Y}_k$ is given by $y(kNT + jT - nT)$. Substituting $\boldsymbol{e}_k = \boldsymbol{Y}_k^T \boldsymbol{c}_k - \boldsymbol{x}_k$, we obtain

$$\boldsymbol{c}_{k+1} = \boldsymbol{c}_k - \Delta\left(\boldsymbol{Y}_k^* \boldsymbol{Y}_k \boldsymbol{c}_k - \boldsymbol{Y}_k^* \boldsymbol{x}_k\right).$$

Neglecting noise and using the fact that the periodic training sequence can be designed to be white [53], i.e.,

$$\frac{1}{N}\sum_{n=0}^{N-1} x_n x_k^* = \delta_{nk}$$

we have

$$\boldsymbol{Y}_k^* \boldsymbol{Y}_k = N\boldsymbol{A}_p \quad \text{and} \quad \boldsymbol{Y}_k^* \boldsymbol{x}_k = N\boldsymbol{\alpha}_p.$$

The elements of the matrix $\boldsymbol{A}_p$ and vector $\boldsymbol{\alpha}_p$ are given by

$$a_{i,j} = \sum_{n=0}^{N-1} h^*(nT - iT)\, h(nT - jT)$$

and

$$\alpha_i = h^*(-iT) \tag{65}$$

where $h(nT)$, $n = 0,1,\cdots,N-1$, are the $T$-spaced samples of the periodic channel response to a sequence of periodic impulses spaced $NT$ seconds apart. Thus the periodic update algorithm may be written as

$$\boldsymbol{c}_{k+1} = \left(\boldsymbol{I} - \Delta N \boldsymbol{A}_p\right)\boldsymbol{c}_k + \Delta N \boldsymbol{\alpha}_p. \tag{66}$$

Note the similarity of (66) to the deterministic gradient algorithm (44). However, in this case the matrix $\boldsymbol{A}_p$ is not only Toeplitz but also circulant, i.e., each row of $\boldsymbol{A}_p$ is a circular shift of another row. The elements $a_{i,j}$ of the $i$th row are the coefficients $R_{i-j}$ of the periodic autocorrelation function of the channel. The final solution to the difference equation (66) is the "optimum" set of periodic equalizer coefficients: $\boldsymbol{c}_{p\,\mathrm{opt}} = \boldsymbol{A}_p^{-1}\boldsymbol{\alpha}_p$. Substituting in (66), we obtain

$$\boldsymbol{c}_{k+1} - \boldsymbol{c}_{p\,\mathrm{opt}} = \left(\boldsymbol{I} - \Delta N \boldsymbol{A}_p\right)\left(\boldsymbol{c}_k - \boldsymbol{c}_{p\,\mathrm{opt}}\right). \tag{67}$$

For a noiseless ideal Nyquist channel $R_{i-j} = \delta_{ij}$, i.e., $\boldsymbol{A}_p = \boldsymbol{I}$. Therefore, a single averaged adjustment with $\Delta = 1/N$ results in perfect periodic equalization.

In general, the transformed coefficient deviation vector after the $k$th periodic update is given by

$$\boldsymbol{d}_k = \left(\boldsymbol{I} - \Delta N \boldsymbol{\Lambda}_p\right)^k \boldsymbol{d}_0$$

where $\boldsymbol{\Lambda}_p$ is a diagonal matrix with eigenvalues $\lambda_{ip}$, $i = 0,1,\cdots,N-1$, equal to the coefficients of the discrete Fourier transform (DFT) of the channel periodic autocorrelation function $R_n$, $n = 0,1,\cdots,N-1$. Thus

$$\lambda_{pi} = \sum_{n=0}^{N-1} R_n \exp(-j2\pi ni/N)$$
$$= \left|\sum_k H(i/NT - k/T)\right|^2, \qquad i = 0,1,\cdots,N-1. \tag{68}$$

Using these results, it can be shown that in the absence of

noise the perfect periodic equalizer has a frequency response equal to the inverse of the folded channel spectrum at $N$ uniformly spaced discrete frequencies

$$C(n/NT) = 1 \Big/ \sum_k H(n/NT - k/T), \quad n = 0,1,\cdots,N-1.$$

The excess MSE after the $k$th update is given by

$$\epsilon_{p\Delta k} = \boldsymbol{d}_k^{*T} \boldsymbol{\Lambda} \boldsymbol{d}_k = \sum_{i=0}^{N-1} \lambda_{pi} (1 - \Delta N \lambda_{pi})^{2k} |d_{i0}|^2.$$

If the initial coefficients are zero, then $|d_{i0}|^2 = 1/\lambda_{pi}$, and

$$\epsilon_{p\Delta k} = \sum_{i=0}^{N-1} (1 - \Delta N \lambda_{pi})^{2k}. \tag{69}$$

Each component of this sum converges provided $0 < \Delta < 2/(N\lambda_{p\max})$. Fastest convergence is obtained when $\Delta = 2/[N(\lambda_{p\max} + \lambda_{p\min})]$.

*2) Stochastic Update:* So far we have examined the convergence properties of the periodic update or LMS steepest descent algorithm with averaging. It is more common, and as we shall see, more beneficial to use the continual or stochastic update method, where all coefficients are adjusted in each symbol interval according to (50). Proceeding as in Section III-B and noting that in the absence of noise $e_{k\,\mathrm{opt}} = 0$ for a periodic input, we obtain

$$(\boldsymbol{c}_{k+1} - \boldsymbol{c}_{p\mathrm{opt}}) = (\boldsymbol{I} - \Delta \boldsymbol{y}_k^* \boldsymbol{y}_k^T)(\boldsymbol{c}_k - \boldsymbol{c}_{p\mathrm{opt}}).$$

Let us define an $N \times N$ matrix

$$\boldsymbol{B}_k \triangleq (\boldsymbol{I} - \Delta \boldsymbol{y}_k^* \boldsymbol{y}_k^T).$$

Note that since $\boldsymbol{y}_k$ is periodic with period $N$, $\boldsymbol{B}_k$ is a circulant matrix and

$$\boldsymbol{B}_{k+1} = \boldsymbol{U}^{-1} \boldsymbol{B}_k \boldsymbol{U}$$

where the $N \times N$ cyclic shift matrix $\boldsymbol{U}$ is of the form

$$\boldsymbol{U} = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 1 \\ 1 & 0 & 0 & \cdot & \cdot & 0 \end{bmatrix}.$$

Note that $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{U}^N = \boldsymbol{I}$. Consider the first $N$ updates from time zero to $N-1$. Then

$$\begin{aligned}(\boldsymbol{c}_N - \boldsymbol{c}_{p\mathrm{opt}}) &= \boldsymbol{B}_{N-1}\boldsymbol{B}_{N-2}\cdots\boldsymbol{B}_0(\boldsymbol{c}_0 - \boldsymbol{c}_{p\mathrm{opt}}) \\ &= (\boldsymbol{U}^{-N+1}\boldsymbol{B}_0\boldsymbol{U}^{N-1})(\boldsymbol{U}^{-N+2}\boldsymbol{B}_0\boldsymbol{U}^{N-2})\cdots\boldsymbol{B}_0(\boldsymbol{c}_0 - \boldsymbol{c}_{p\mathrm{opt}}) \\ &= (\boldsymbol{U}\boldsymbol{B}_0)^N(\boldsymbol{c}_0 - \boldsymbol{c}_{p\mathrm{opt}}).\end{aligned}$$

In general

$$(\boldsymbol{c}_{kN} - \boldsymbol{c}_{p\mathrm{opt}}) = (\boldsymbol{U}\boldsymbol{B}_0)^{kN}(\boldsymbol{c}_0 - \boldsymbol{c}_{p\mathrm{opt}}). \tag{70}$$

The convergence of the coefficient deviations, therefore, depends on the eigenvalues of $\boldsymbol{U}\boldsymbol{B}_0$. Using the fact that $\boldsymbol{y}_0^*\boldsymbol{y}_0^T$ is singular with rank 1, after some manipulation, the characteristic equation $\det(\lambda\boldsymbol{I} - \boldsymbol{U}\boldsymbol{B}_0) = 0$ can be reduced to the form

$$\lambda^N + \Delta N \sum_{n=0}^{N-1} \lambda^n R_n - 1 = 0. \tag{71}$$

Here

$$R_n = (\boldsymbol{y}_k^T \boldsymbol{Y}_{k+n}^*)/N, \qquad n = 0,1,\cdots,N-1$$

are the coefficients of the periodic autocorrelation function of the equalizer input (defined earlier in terms of the periodic impulse response of the channel). When $R_n = 0$ for $n \geq 1$, corresponding to an ideal Nyquist channel, all roots of (71) are equal to the $N$th roots of $(1 - \Delta N R_0)$. Thus perfect equalization is obtained after $N$ updates with $\Delta = 1/(NR_0)$.

Let $\lambda_{\max}$ be the maximum magnitude root of (71). Then the coefficient deviations after every $N$ stochastic updates are reduced in magnitude according to (70) provided $|\lambda_{\max}^N| < 1$. Moreover, from the maximum modulus principle of holomorphic functions [92] we have the condition that

$$|\lambda_{\max}^N| \leq \max_{0 \leq i \leq N-1} \left| 1 - \Delta N \sum_{n=0}^{N-1} R_n \exp(-j2\pi ni/N) \right|$$

or

$$|\lambda_{\max}^N| \leq \max_{0 \leq i \leq N-1} |1 - \Delta N \lambda_{pi}| \tag{72}$$

where $\lambda_{pi}$ are the eigenvalues of $\boldsymbol{A}_p$ in the periodic update method. Since (72) holds with equality only when $R_n = 0$ for $n \geq 1$, we reach the important conclusion for periodic training that for the ideal Nyquist channel the stochastic and averaged update algorithms converge equally fast, but for all other channels the stochastic update algorithm results in faster convergence. This behavior has also been observed for equalizer convergence in the presence of random data [71].

In the presence of noise, an exact expression for the excess MSE for the stochastic update algorithm is difficult to derive for periodic training sequences. However, assuming zero initial coefficients the following expression is a good approximation to results obtained in practice for moderately high SNR:

$$\epsilon_{\Delta kN} = \sum_{i=0}^{N-1} \left[ (1 - \Delta N \lambda_i)^{2k} + \Delta^2 N \lambda_i^2 \epsilon_{p\min} \right] \tag{73}$$

where $\lambda_i$ are the roots of (71) (the effect of noise can be included by defining $R_n = E[\boldsymbol{y}_k^T \boldsymbol{y}_{k+N}^*]/N$). For an ideal channel, $\lambda_i = R_0 = 1$ for all $i$, and the excess MSE for periodic training converges to $\epsilon_{p\min}$ after $N$ adjustments with $\Delta = 1/N$, where $\epsilon_{p\min}$ is the minimum achievable MSE for periodic training.

The significant difference in the convergence behavior of the stochastic gradient algorithm for random data and periodic training is now apparent by comparing (54) and (73). The well-controlled correlation properties of periodic training sequences tend to reduce the average settling time of the equalizer by about a factor of two compared to the settling time in the presence of random data.

An important question regarding periodic equalization is that once the coefficients have been optimized for a periodic training sequence, how close to optimum is that set of coefficients for random data. The answer depends primarily on the selected period of the training sequence (and hence the equalizer span) relative to the length of the channel impulse response. When the period is long enough to contain a sufficiently large percentage (say 95 percent) of the energy of the channel impulse response, the edge

effects in the periodic channel response and equalizer coefficients are small. In frequency-domain terms, the discrete tones of the periodic training sequence are adequately dense to obtain representative samples of the channel spectrum. Under these conditions, the excess MSE due to the periodicity of the training sequence is small compared to the excess MSE due to the large value of $\Delta$ which must be selected for fast initial convergence. After rapid initial convergence has been obtained in this manner, it may be desirable to make finer adjustments to the equalizer using a pseudo-random training sequence with a longer period, or begin decision-directed adaptation using randomized customer data.

*3) Application to Fractionally Spaced Equalizers:* The averaged and stochastic update methods of periodic training are also applicable to fractionally spaced equalizers [91]. The period of the training sequence is still equal to the time span of the equalizer. Thus a sequence with period $NT$ can be used to train an equalizer with $NM/K$ coefficients spaced $KT/M$ seconds apart.

The equalizer coefficients may be adjusted periodically, every $N$ symbol intervals, according to the averaged update algorithm

$$c_n(k+1) = c_n(k) - \Delta \sum_{j=0}^{N-1} e_{kN+j} y^*(kNT - jT - nKT/M), \qquad n = 0,1,\cdots,NM/K - 1.$$

In the absence of noise, the elements of the matrix $\boldsymbol{A}_p$ and vector $\boldsymbol{\alpha}_p$ (given in (65) for a $T$-spaced equalizer) are given below for a $KT/M$-spaced equalizer

$$a_{i,j} = \sum_{n=0}^{N-1} h^*(nT - iKT/M)\, h(nT - jKT/M)$$

and

$$\alpha_i = h^*(-iKT/M).$$

The matrix $\boldsymbol{A}_p$ is no longer Toeplitz or circulant and, therefore, its eigenvalues cannot be obtained by DFT techniques (68). However, it can be shown that the $N$ significant eigenvalues are samples of the channel folded power spectrum, that is,

$$\lambda_{pi} = (M/K) \sum_k |H(i/NT - k/T)|^2, \qquad i = 0,1,\cdots,N-1 \tag{74}$$

and the remaining $N(M-K)/K$ eigenvalues are zero. The frequency response of the perfect periodic equalizer at $N$ uniformly spaced discrete frequencies is given by

$$C(n/NT) = H^*(n/NT) \Big/ \sum_k |H(n/NT - k/T)|^2, \qquad n = 0,1,\cdots,NM/K - 1.$$

Starting with zero initial coefficients, the excess MSE after the $k$th averaged update is related to the eigenvalues (74) according to

$$\epsilon_{p\Delta k} = \sum_{i=0}^{N-1} (1 - \Delta N \lambda_{pi})^{2k}.$$

If the channel folded power spectrum is flat, i.e., $\lambda_{pi} = M/K$, $i = 0,1,\cdots,N-1$, then a single averaged update with $\Delta = K/MN$ results in a matched filter which also removes intersymbol interference. In contrast with $T$-spaced equalizers, the ideal amplitude shape of the unequalized system for fast convergence of a fractionally spaced equalizer is a square root of Nyquist rather than Nyquist. Moreover, the convergence of an FSE is not affected by sampler timing phase or channel delay distortion since the channel power spectrum and hence $\lambda_{pi}$ are independent of phase-related parameters.

The stochastic update algorithm (50) for periodic training of fractionally spaced equalizers can be analyzed along the lines of Section IV-B2. For instance, for a $T/2$-spaced equalizer, neglecting noise, the coefficient deviation recursion (70) still applies. However, all matrices are now $2N \times 2N$ and the cyclic shift matrix $\boldsymbol{U}$ now produces a double shift. The eigenvalues of $\boldsymbol{UB}_0$ are the roots of the characteristic equation $\det(\lambda \boldsymbol{I} - \boldsymbol{UB}_0) = 0$, or

$$(\lambda^N - 1)\left(\lambda^N + \Delta N \sum_{n=0}^{N-1} \lambda^n R_n - 1\right) = 0 \tag{75}$$

where

$$R_n = (1/N) \sum_{k=0}^{2N-1} y(kT/2)\, y^*(kT/2 + nT), \qquad n = 0,1,\cdots,N-1$$

are the $T$-spaced samples of the channel periodic autocorrelation. Note that half the roots of (75) are the $N$th roots of unity. Coefficient deviation components corresponding to these roots do not converge. Appropriate selection of initial coefficients, e.g., $\boldsymbol{c}_0 = 0$, ensures that these components are zero. The remaining roots of (75) are dependent on the properties of $R_n$, whose DFT may be recognized to be equal to the samples of the channel folded power spectrum (74). Therefore, like the averaged update algorithm, the convergence of the stochastic update algorithm is also independent of phase-related parameters. When the folded power spectrum is flat, $R_n = R_0\, \delta_{n0}$, perfect equalization can be obtained in $N$ adjustments with $\Delta = 1/NR_0$. When the channel power spectrum is not Nyquist ($R_n \neq 0, n \geq 1$), the maximum modulus principle (72) applies ensuring that the stochastic update algorithm will result in faster convergence than that obtained by periodic or averaged update. Comments with regard to the selection of an adequately long period of the training sequence given in Section IV-B2 still apply.

*4) Accelerated Processing:* One technique for reducing the effective settling time of an equalizer involves performing equalizer coefficient update iterations as often as permissible by the computational speed limitations of the implementation. For instance, for a periodic equalizer, one period of the received sequence of samples may be stored in the equalizer delay line and iterative updates using the averaged or stochastic update algorithm may be made at a rate faster than the usual [76]. This update rate may be selected to be independent of the modem symbol rate since the sequence of samples already stored in the equalizer delay line may be circularly shifted as often as required to produce new output samples. Based on each such output new coefficient correction terms can be computed using a circular shift of the locally stored periodic training sequence. Thus after the equalizer delay line has been filled with a set of received samples, the best periodic

equalizer for that particular set of received samples can be determined almost instantly by accelerated processing given unlimited computational speed. If this set of received samples is representative of the channel response to the periodic training sequence then the equalizer obtained by such accelerated processing is a good approximation to the optimum periodic equalizer. However, all sources of aperiodicity, e.g., initial transients and noise, in these received samples degrade performance, since new received samples are not used in this method to reduce the effect of noise by averaging. Modified versions of the accelerated processing method are possible which reprocess some previously processed samples and then accept a new input sample as it becomes available, thus updating the equalizer coefficients several times per symbol interval.

*5) Orthogonalized Periodic Equalizer:* As discussed in Section IV-A, the convergence of the LMS gradient algorithm can be improved by inserting an orthogonalizing matrix $\boldsymbol{P}$ in the path of the coefficient corrections (63). The desired value of $\boldsymbol{P}$ is the inverse of the equalizer input correlation matrix $\boldsymbol{A}$. For a periodic equalizer the average update algorithm (64) can be modified in a similar manner to

$$\boldsymbol{c}_{k+1} = \boldsymbol{c}_k - \boldsymbol{P}\boldsymbol{Y}_k^* e_k. \tag{76}$$

However, in this case the orthogonalizing matrix $\boldsymbol{P}$ can be replaced by a single inverse filter in the path of the periodic equalizer input sequence before it is used for coefficient adjustment [90]. This simplification is a consequence of the fact that, as discussed in Section IV-B1, the input correlation matrix $\boldsymbol{A}_p$ for a $T$-spaced periodic equalizer is Toeplitz and circulant. The inverse $\boldsymbol{P}$ of the circulant matrix $N\boldsymbol{A}_p$ is also circulant. For periodic input, the transformation performed by $\boldsymbol{P}$ is equivalent to a nonrecursive filtering operation (or periodic convolution) with coefficients equal to the elements of the first row of $\boldsymbol{P}$.

The fast settling periodic equalizer structure with a single inverse filter is also applicable when the equalizer coefficients are updated symbol-by-symbol. The modified stochastic update algorithm is given by

$$c_n(k+1) = c_n(k) - \Delta e_k \sum_{j=0}^{N-1} p_j y^*(kT - jT - nT), \qquad n = 0,1,\cdots,N-1.$$

A block diagram of the equalizer structure and performance curves showing significant improvement in the settling time for partial response QAM systems are given in [90].

*6) Discrete Fourier Transform Techniques:* Throughout the discussion on periodic equalization, we have taken advantage of the circulant property of the equalizer input to use the discrete Fourier transform (DFT) for analysis. As pointed out in Section IV-B1, in the absence of noise the perfect periodic equalizer has a frequency response equal to the inverse of the folded channel spectrum at $N$ uniformly spaced frequencies. Therefore, the equalizer coefficients can be directly obtained by transmitting a periodic training sequence and using the following steps at the receiver [43], [70], [90]:

1) Compute the DFT of one period of the equalizer input

$$Y_i = \sum_{k=0}^{N-1} y_k \exp(-j2\pi ik/N), \qquad i = 0,1,\cdots,N-1.$$

2) Compute the desired equalizer spectrum according to

$$C_i = X_i Y_i^* / |Y_i|^2, \qquad i = 0,1,\cdots,N-1$$

where $X_i$ is the precomputed DFT of the training sequence.

3) Compute the inverse DFT of the equalizer spectrum to obtain the periodic equalizer coefficients

$$c_n = (1/N) \sum_{i=0}^{N-1} C_i \exp(j2\pi ni/N), \qquad n = 0,1,\cdots,N-1.$$

A number of modifications may be made to improve performance of this direct computation method in the presence of noise and other distortions, such as frequency translation, which adversely affect periodicity. For instance, when the equalizer input is not strictly periodic with period $NT$ due to channel-induced frequency translation, its DFT at any frequency suffers from interference from adjacent components. The effect of this interference can be minimized by windowing a longer sequence of input samples before taking the DFT. The window function should be selected such that its Fourier transform has reduced sidelobe energy while preserving the property of zero response at $1/NT$-hertz intervals. A $2NT$ second triangular window has both these properties [43]. A second minor modification can be made to step 2 by adding a constant estimate of the expected flat noise power spectral components to the denominator.

### C. Recursive Least Squares (RLS) Algorithms

The orthogonalized LMS algorithms of Section IV-A can provide rapid convergence when the overall received signal spectral shape is known beforehand; for example, in partial-response systems. In certain voice-band modem applications special training sequences can be used to design fast equalizer startup algorithms, such as those in the last section. However, in general, a self-orthogonalizing method, such as one of the RLS algorithms described in this section, is required for rapidly tracking adaptive equalizers (or filters) when neither the reference signal nor the input (received) signal (or channel) characteristics can be controlled.

As discussed in earlier sections, the rate of convergence of the output MSE of an LMS gradient adaptive equalizer is adversely affected by the eigenvalue spread of the input covariance matrix. This slow convergence is due to the fundamental limitation of a single adjustable step size parameter $\Delta$ in the LMS gradient algorithm. If the input covariance matrix is known *a priori* then an orthogonalized LMS gradient algorithm can be derived, as in Section IV-A, where the scalar $\Delta$ is replaced by a matrix $\boldsymbol{P}$. Most rapid convergence is obtained when $\boldsymbol{P}$ is the inverse of the equalizer input covariance matrix $\boldsymbol{A}$, thus rendering the adjustments to the equalizer coefficients independent of one another.

In [41], Godard applied the Kalman filter algorithm to the estimation of the LMS equalizer coefficient vector under some assumptions on the equalizer output error and input statistics. The resulting algorithm has since been recognized to be the fastest known equalizer adaptation algorithm. It is an ideal self-orthogonalizing algorithm [37] in that the received equalizer input signals are used to build up the inverse of the input covariance matrix which is applied to the coefficient adjustment process. A disadvantage of the

Kalman algorithm is that it requires on the order of $N^2$ operations per iteration for an equalizer with $N$ coefficients.

Falconer and Ljung [20] showed that the Kalman equalizer adaptation algorithm can be derived as a solution to the exact least squares problem without any statistical assumptions. An advantage of this approach is that the "shifting property" previously used for fast recursive least squares identification algorithms [73] can be applied to the equalizer adaptation algorithm. This resulted in the so-called fast Kalman algorithm [20] which requires on the order of $N$ operations per iteration.

A third class of recursive least squares algorithms known as adaptive lattice algorithms [127] were first described for adaptive identification in [74] and for adaptive equalization in [63], [96], and [97]. Like the fast Kalman algorithm, adaptive lattice algorithms are recursive in time, requiring of the order of $N$ operations per iteration. However, unlike the Kalman algorithms, adaptive lattice algorithms are order-recursive. That is, the number of equalizer coefficients (and the corresponding lattice filter sections) can be increased to $N + 1$ without affecting the already computed parameters of the $N$th-order equalizer. Low sensitivity of the lattice coefficients to numerical perturbations is a further advantage.

In the remainder of this section, we shall briefly review the least square criterion and its variants, introduce the "shifting property" and the structure of the fast Kalman and adaptive lattice algorithms, and summarize some important results, complexity estimates, and stability considerations.

*1) The Least Squares Criterion:* The performance index for recursive least squares (RLS) algorithms is expressed in terms of a time average instead of a statistical or ensemble average as in LMS algorithms. The RLS equalizer adaptation algorithm is required to generate the $N$-coefficient vector $\boldsymbol{c}_n$ at time $n$ which minimizes the sum of all squared errors as if $\boldsymbol{c}_n$ were used over all the past received signals, i.e., $\boldsymbol{c}_n$ minimizes

$$\sum_{k=0}^{n} |x_k - \boldsymbol{y}_k^T \boldsymbol{c}_n|^2. \tag{77}$$

This leads to the so-called prewindowed RLS algorithm, where the input samples $y_k$ are assumed to be zero for $k < 0$.

In order to permit tracking of slow time variations, a decay factor $w$ with a value slightly less than unity may be introduced. The resulting exponentially windowed RLS algorithm minimizes

$$\sum_{k=0}^{n} w^{n-k} |x_k - \boldsymbol{y}_k^T \boldsymbol{c}_n|^2. \tag{78}$$

The minimizing vector $\boldsymbol{c}_n$ is the solution of the discrete-time Wiener–Hopf equation obtained by setting the derivative of (78) with respect to $\boldsymbol{c}_n$ to zero. Thus we have the solution

$$\boldsymbol{c}_n = \boldsymbol{A}_n^{-1} \boldsymbol{\alpha}_n$$

where the $N \times N$ estimated covariance matrix is given by

$$\boldsymbol{A}_n = \sum_{k=0}^{n} w^{n-k} \boldsymbol{y}_k^* \boldsymbol{y}_k^T + \delta \boldsymbol{I} = w \boldsymbol{A}_{n-1} + \boldsymbol{y}_n^* \boldsymbol{y}_n^T \tag{79}$$

and the $N$-element estimated cross-correlation vector is given by

$$\boldsymbol{\alpha}_n = \sum_{k=0}^{n} w^{n-k} \boldsymbol{y}_k^* x_k = w \boldsymbol{\alpha}_{n-1} + \boldsymbol{y}_n^* x_k.$$

The parameter $\delta$ is selected as a small positive number to ensure that $\boldsymbol{A}_n$ is nonsingular. The matrix $\boldsymbol{A}_n$ and vector $\boldsymbol{\alpha}_n$ are akin to the statistical autocorrelation matrix and cross-correlation vector encountered in the LMS analysis. However, in this case $\boldsymbol{A}_n$ is not a Toeplitz matrix even for $T$-spaced equalizers.

It can be shown [87] that given $\boldsymbol{c}_{n-1}$, the coefficient vector for time $n$ can be generated recursively according to

$$\boldsymbol{c}_n = \boldsymbol{c}_{n-1} + \boldsymbol{k}_n e_n \tag{80}$$

where $e_n = x_n - \boldsymbol{y}_n^T \boldsymbol{c}_{n-1}$ is the equalizer output error and

$$\boldsymbol{k}_n = \boldsymbol{A}_n^{-1} \boldsymbol{y}_n^* \tag{81}$$

is the Kalman gain vector. The presence of the inverse estimated covariance matrix in (81) explains the insensitivity of the rate of convergence of the RLS algorithms to the channel characteristics.

In the Kalman algorithm [41], the inverse matrix $\boldsymbol{P}_n = \boldsymbol{A}_n^{-1}$ and the Kalman gain vector are computed recursively according to

$$\boldsymbol{k}_n = \boldsymbol{P}_{n-1} \boldsymbol{y}_n^* / \left[ w + \boldsymbol{y}_n^T \boldsymbol{P}_{n-1} \boldsymbol{y}_n^* \right]$$

and

$$\boldsymbol{P}_n = \left[ \boldsymbol{P}_{n-1} - \boldsymbol{k}_n \boldsymbol{y}_n^T \boldsymbol{P}_{n-1} \right] / w. \tag{82}$$

The order of $N^2$ complexity of this algorithm is due to the explicit recursive computation of $\boldsymbol{P}_n$. This computation is also susceptible to roundoff noise.

Two other variations of the prewindowed and exponentially windowed least squares criteria are the "growing memory" covariance and "sliding window" covariance performance indices defined as

$$\sum_{k=N-1}^{n} w^{k-n} |x_k - \boldsymbol{y}_k^T \boldsymbol{c}_n|^2$$

and

$$\sum_{k=n-L+1}^{n} |x_k - \boldsymbol{y}_k^T \boldsymbol{c}_n|^2$$

respectively, where $N$ is the number of equalizer coefficients and $L$ is the fixed length of the sliding window. The sliding window method is equivalent to a block least squares approach, where the block is shifted by one sample and a new optimum coefficient vector is determined each sample time.

*2) The Fast Kalman Algorithm:* Consider the input vector $\boldsymbol{y}_{n-1}$ at time $n - 1$ for a $T$-spaced equalizer of length $N$. The vector $\boldsymbol{y}_n$ at time $n$ is obtained by shifting the elements of $\boldsymbol{y}_{n-1}$ by one, discarding the oldest sample $y_{n-N}$, and adding a new sample $y_n$. This shifting property is exploited by using least square linear prediction. Thus an efficient recursive algorithm can be derived [20] for updating the Kalman gain vector $\boldsymbol{k}_n$ without explicit computation of the inverse matrix $\boldsymbol{P}_n$. Here we shall briefly examine the role of forward and backward prediction in the fast Kalman algorithm. See [20] for a more detailed derivation generalized to fractionally spaced and decision-feedback equalizers.

Let $\boldsymbol{F}_{n-1}$ be a vector of $N$ forward predictor coefficients

which minimizes the weighted sum of squares of the forward prediction error $f_n$ between the new input sample $y_n$ and a prediction based on the vector $y_{n-1}$. That is, $F_{n-1}$ minimizes

$$\sum_{k=0}^{n} w^{n-k}|f_k|^2$$

where

$$f_k = y_k - F_{n-1}^T y_{k-1}. \tag{83}$$

The least squares forward predictor coefficients can be updated recursively according to

$$F_n = F_{n-1} + k_{n-1} f_n. \tag{84}$$

Similarly, the vector $B_{n-1}$ of $N$ backward predictor coefficients permits prediction of the old discarded sample $y_{n-N}$ given the vector $y_n$. Thus we have the backward error

$$b_n = y_{n-N} - B_{n-1}^T y_n \tag{85}$$

and the update equation

$$B_n = B_{n-1} + k_n b_n.$$

The updated Kalman gain vector $k_n$, which is not yet available, can be obtained as follows. Define

$$f_n' = y_n - F_n^T y_{n-1} \tag{86}$$

as the error between $y_n$ and its prediction based on the updated forward predictor. Let

$$E_n = wE_{n-1} + f_n'^* f_n \tag{87}$$

be the estimated exponentially weighted squared prediction error. Then the augmented or extended Kalman gain vector with $N+1$ elements is given by

$$\bar{k}_n = \begin{bmatrix} f_n'^*/E_n \\ \hline k_{n-1} - F_n f_n'^*/E_n \end{bmatrix} = \begin{bmatrix} k_n' \\ \hline \mu_n \end{bmatrix} \tag{88}$$

where the dashed lines indicate partitions of the vector $k_n$. Finally, the updated backward predictor vector and the Kalman gain vector are given by the recursive relationships

$$B_n = \left[B_{n-1} + k_n' b_n\right]/[1 - \mu_n b_n] \tag{89}$$

and

$$k_n = k_n' + B_n \mu_n \tag{90}$$

where $k_n'$ and $\mu_n$ are the $N$-element and scalar partitions, respectively, of the augmented Kalman gain vector defined in (88).

The matrix computations (82) involved in the Kalman algorithm are replaced in the fast Kalman algorithm by the recursions (83) through (90) which use forward and backward predictors to update the Kalman gain vector as a new input sample $y_n$ is received and the oldest sample $y_{n-N}$ is discarded.

A fast exact initialization algorithm for the interval $0 \leqslant n \leqslant N$ given in [10] avoids the choice of a stabilizing $\delta$ in (79) and the resulting suboptimality of the solution at $n = N$.

*3) Adaptive Lattice Algorithms:* The Kalman and fast Kalman algorithms obtain their fast convergence by orthogonalizing the adjustments made to the coefficients of an ordinary linear transversal equalizer. Adaptive lattice (AL) algorithms, on the other hand, use a lattice filter structure to orthogonalize a set of received signal components [127]. The transformed received signal components are then linearly weighted by a set of equalizer coefficients and summed to produce the equalizer output. We shall briefly review the gradient [96] and least squares [87], [97] forms of adaptive lattice algorithms for linear $T$-spaced complex equalizers. See [78] and [52] for generalization of the least squares AL algorithm to fractionally spaced and decision-feedback equalizers.

The structure of an adaptive lattice gradient equalizer is shown in Fig. 26. An $N$-coefficient equalizer uses $N-1$

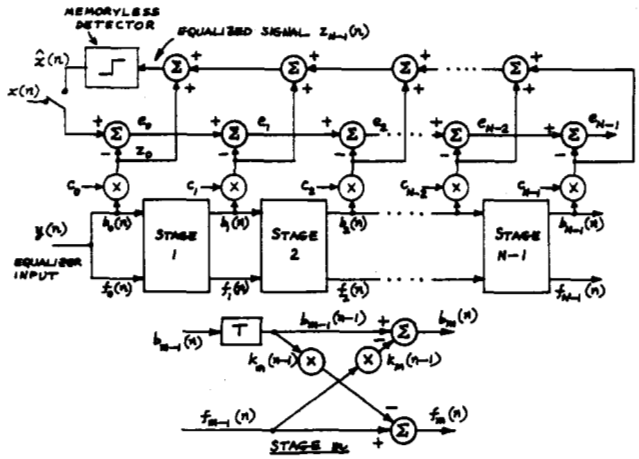


**Fig. 26.** Gradient adaptive lattice equalizer.

lattice filter stages. Each symbol interval a new received sample $y(n)$ enters stage 1. The $m$th stage produces two signals $f_m(n)$ and $b_m(n)$ which are used as inputs by stage $m+1$. These signals correspond to the forward and backward prediction errors, respectively, of $m$th-order forward and backward linear LMS predictors. The two predictors have identical so-called reflection coefficients $k_m$ for the $m$th stage. At time $n$, the prediction errors are updated according to

$$f_0(n) = b_0(n) = y(n) \tag{91}$$

and for $m = 1, \cdots, N-1$

$$f_m(n) = f_{m-1}(n) - k_m(n-1)\, b_{m-1}(n-1) \tag{92}$$

$$b_m(n) = b_{m-1}(n-1) - k_m(n-1)\, f_{m-1}(n). \tag{93}$$

The reflection coefficients are updated to minimize the sum of the mean-square value of the forward and backward prediction errors. That is,

$$k_m(n) = k_m(n-1) + \left[f_{m-1}^*(n)\, b_m(n) + b_{m-1}^*(n-1)\, f_m(n)\right]/\nu_m \tag{94}$$

where

$$\nu_m = w\nu_m(n-1) + |f_{m-1}(n)|^2 + |b_{m-1}(n-1)|^2. \tag{95}$$

The equalizer output for each stage is computed according to

$$z_m(n) = z_{m-1}(n) + c_m(n-1)\, b_m(n), \qquad m = 0, 1, \cdots, N-1. \tag{96}$$

The final output $z_{N-1}(n)$ is used during data mode to compute the receiver decision $x(n)$. Initially, a reference signal $x(n)$ is substituted for the receiver decision in order to compute the error signals

$$e_m(n) = x(n) - z_m(n), \qquad m = 0,1,\cdots,N-1. \quad (97)$$

The last step in the algorithm is the equalizer coefficient update equation

$$c_m(n) = c_m(n-1) - 2e_m(n)\,b_m^*(n)/\nu_m, \qquad m = 0,1,\cdots,N-1. \quad (98)$$

Equations (91)–(98) define the gradient AL equalizer algorithm.

One important property of the lattice structure is that as the reflection coefficients converge, the backward prediction errors $b_m(n)$, $m = 0,1,\cdots,N-1$, form a vector $\boldsymbol{b}(n)$ of orthogonal signal components, i.e.,

$$E\left[b_m(n)\,b_j(n)\right] = 0, \qquad \text{for } j \neq m.$$

The vector $\boldsymbol{b}(n)$ is a transformed version of the received vector $\boldsymbol{y}(n)$ with elements $y(n), y(n-1),\cdots,y(n-N+1)$. This transformation is performed by an $N \times N$ lower triangular matrix $\boldsymbol{L}$ according to $\boldsymbol{b}(n) = \boldsymbol{L}\boldsymbol{y}(n)$, where $\boldsymbol{L}$ is formed by the backward predictor coefficients of order $m, m = 1,\cdots,N-1$. The prediction error $b_{m+1}(n)$ is given by

$$b_{m+1}(n) = y(n-m) - \sum_{j=1}^{m} B_m(j)\,y(n-m+j)$$

where the predictor coefficients $B$ of order $m$ and lower can all be derived from the first $m$ reflection coefficients $k_j$, $j = 1,\cdots,m$.

The lower triangular form of the above transformation permits a simple way to increase the length or order of the lattice equalizer since the existing prediction errors and equalizer coefficients remain unchanged when another stage is added. The prediction errors at the $m$th stage are not functions of the reflection coefficients at succeeding stages as can be seen from Fig. 26, (92), and (93). The lattice equalizer is, therefore, order-recursive as well as time-recursive.

A computationally complex (requiring larger number of computations) but faster converging least squares form of the AL equalizer results when the performance index or cost function to be minimized is the exponentially windowed sum of squared errors given in (78) instead of the MSE. The structure of the least squares AL equalizer is shown in Fig. 27. Note that the lattice coefficients for forward and backward prediction for any of the lattice stages are no longer equal, each being independently updated to minimize the weighted sum of squared forward and backward prediction errors, respectively. The least squares AL algorithm for a $T$-spaced complex equalizer is summarized below.

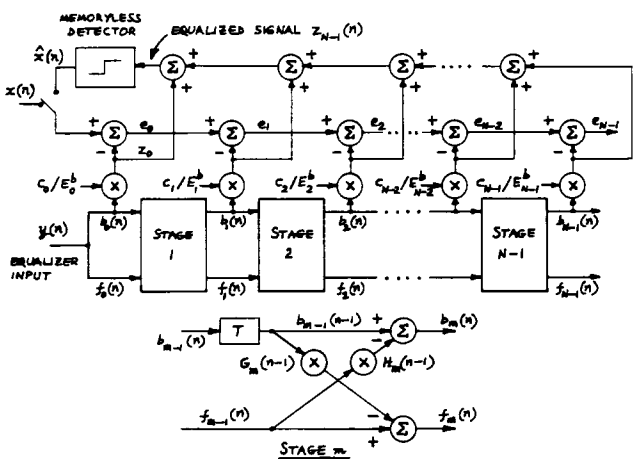


**Fig. 27.** Least squares adaptive lattice equalizer.

At time $n$, the inputs to the first lattice stage are set to the newly received sample, i.e.,

$$f_0(n) = b_0(n) = y(n) \quad (99)$$

and

$$E_0^f(n) = E_0^b(n) = wE_0^f(n-1) + y^*(n)\,y(n) \quad (100)$$

where $E_m^f(n)$ and $E_m^b(n)$ are the estimated sum of the squared forward and backward prediction errors, respectively, at stage $m$. Next, the order updates are performed for $m = 1,\cdots,N-1$

$$K_m(n) = wK_m(n-1) + t_m(n-1)\,f_{m-1}(n) \quad (101)$$

the forward prediction errors

$$f_m(n) = f_{m-1}(n) - G_m(n-1)\,b_{m-1}(n-1) \quad (102)$$

the backward prediction error

$$b_m(n) = b_{m-1}(n-1) - H_m(n-1)\,f_{m-1}(n) \quad (103)$$

the lattice coefficient for forward prediction

$$G_m(n) = K_m(n)/E_{m-1}^b(n-1) \quad (104)$$

the lattice coefficient for backward prediction

$$H_m(n) = K_m^*(n)/E_{m-1}^f(n) \quad (105)$$

and

$$E_m^f(n) = E_{m-1}^f(n) - G_m(n)\,K_m^*(n) \quad (106)$$

$$E_m^b(n) = E_{m-1}^b(n) - H_m(n)\,K_m(n) \quad (107)$$

$$t_m(n) = \left[1 - \gamma_{m-1}(n)\right] b_{m-1}^*(n) \quad (108)$$

$$\gamma_m(n) = \gamma_{m-1}(n) + |t_m(n)|^2/E_{m-1}^b(n). \quad (109)$$

Now the equalizer output can be computed according to

$$z_m(n) = z_{m-1}(n) + \left[c_m(n-1)/E_{m-1}^b(n-1)\right] b_{m-1}(n), \qquad m = 0,1,\cdots,N-1. \quad (110)$$

The final equalized signal is given by $z_{N-1}(n)$. The error signals are computed and the corresponding equalizer coefficient updates are performed for $m = 0,1,\cdots,N-1$ according to

$$e_m(n) = x(n) - z_m(n) \quad (111)$$

and

$$c_m(n) = wc_m(n-1) + t_m(n)\,e_{m-1}(n). \quad (112)$$

Equations (99) through (112) define the least squares AL algorithm.

*4) Complexity and Numerical Stability:* In the preceding sections, we have presented an overview of three basic forms of recursive least squares equalization algorithms. Fast RLS algorithms are still being actively studied to reduce computational complexity, specially for "multichannel" (fractionally spaced and decision-feedback) equalizers, and to improve stability when limited precision arithmetic is used. Some of the recent results are reported in [10], [11], and [52].

Accurate counts of the number of multiplications, additions/subtractions, and divisions are hard to summarize due to the large number of variations of the RLS algorithms

which have been reported in the literature. In the table below, the number of operations (multiplications and divisions) required per iteration is listed for $T$- and $T/2$-spaced transversal equalizers of span $NT$ seconds for the LMS gradient, Kalman, fast Kalman/fast transversal, and RLS lattice algorithms. In each case, the smallest number of operations is given from the complexity estimates reported in [78], [10], and [52].

| Algorithm | Number of Operations per Iteration | |
|---|---|---|
| | $T$ Equalizer | $T/2$ Equalizer |
| LMS gradient | $2N$ | $4N$ |
| Kalman | $2N^2 + 5N$ | $8N^2 + 10N$ |
| Fast Kalman/fast transversal | $7N + 14$ | $24N + 45$ |
| RLS lattice | $15N - 11$ | $46N$ |

The fast Kalman is the most efficient type of RLS algorithm. However, compared to the LMS gradient algorithm, the fast Kalman algorithm is still about four times as complex for $T$ equalizers and six times for $T/2$ equalizers. The RLS lattice algorithm is still in contention due to its better numerical stability and order-recursive structure, despite a two-fold increase in computational complexity over the fast Kalman algorithm.

The discussion on RLS algorithms would not be complete without a comment on the numerical problems associated with these algorithms in steady-state operation. Simulation studies have reported the tendency of RLS algorithms implemented with finite precision to become unstable and the adaptive filter coefficients to diverge [10], [46], [52], [78], [126]. This is due to the long-term accumulation of finite precision errors. Among the different types of RLS algorithms, the fast Kalman or fast transversal type algorithms are the most prone to instability [46], [78]. In [78] instability was reported to occur when an exponential weighting factor $w < 1$ was used for the fast Kalman algorithm implemented with single-precision floating-point arithmetic. The Kalman and RLS lattice algorithms did not show this instability. A sequential processing dual-channel version of the RLS lattice algorithm for a DFE is reported to be stable even for fixed-point arithmetic with 10- to 12-bit accuracy [52]. However, in [10] an "unnormalized" RLS lattice algorithm is shown to become unstable. Normalized versions of fast transversal [10] and RLS lattice [127] algorithms are more stable but both require square roots, the lattice type having greater computational complexity.

The stability of RLS algorithms can be improved [10], [46], [126] by modification of the least squares criterion, and periodic reinitialization of the algorithm to avoid precision error buildup. The modified criterion takes into account the squared magnitude of the difference of the filter coefficients from their initial (or restart) values. The rationale for this so-called soft constraint [10] is the same as for the stochastic gradient algorithm with tap leakage discussed in Section III-C.

For a short transition period following each reinitialization, while the RLS algorithm is reconverging, an auxiliary LMS adaptive filter is used to compute outputs. Results of four variations of the periodic restart procedure for an adaptive decision feedback equalizer are given in [126].

## V. Concluding Remarks

Adaptive equalization and the more general field of adaptive filtering have been areas of active research and development for more than two decades. It is, therefore, tempting to state that no substantial further work remains to be done. However, this has not been the case in the last decade despite how mature the field appeared in 1973 [59]. In fact, a number of the topics covered in this paper, e.g., fractionally spaced equalizers, decision-aided ISI cancellation, and fast recursive least squares algorithms, were not yet fully understood or were yet to be discovered. Of course, tremendous strides have since been made in implementation technology which have spawned new applications, e.g., digital subscriber loops, and pushed existing applications toward their limits, e.g., 256-QAM digital radios and voice-band modems with rates approaching 19.2 kbits/s. Programmable digital signal processors now permit implementation of ever more sophisticated and computationally complex algorithms; and so the study and research must continue—in new directions. There is still more work to be done in adaptive equalization of nonlinearities with memory and in equalizer algorithms for coded modulation systems. However, the emphasis has already shifted from adaptive equalization theory toward the more general theory and applications of adaptive filters, and toward structures and implementation technologies which are uniquely suited to particular applications.


## Acknowledgment

The author wishes to thank G. D. Forney, Jr., for his encouragement and guidance in preparing this paper. Thanks are also due to a number of friends and colleagues who have reviewed and commented on the paper.


## References and Bibliography


[1] K. Abend and B. D. Fritchman, "Statistical detection for communication channels with intersymbol interference," *Proc. IEEE*, vol. 58, pp. 779–785, May 1970.

[2] O. Agazzi, D. A. Hodges, and D. G. Messerechmitt, "Large scale integration of hybrid-method digital subscriber loops," *IEEE Trans. Commun.*, vol. COM-30, pp. 2095–2108, Sept. 1982.

[3] M. E. Austin, "Decision-feedback equalization for digital communication over dispersive channels," MIT Lincoln Lab., Lexington, MA, Tech. Rep. 437, Aug. 1967.

[4] C. A. Belfiore and J. H. Park, Jr., "Decision feedback equalization," *Proc. IEEE*, vol. 67, pp. 1143–1156, Aug. 1979.

[5] E. Biglieri, A. Gersho, R. D. Gitlin, and T. L. Lim, "Adaptive cancellation of nonlinear intersymbol interference for voiceband data transmission," *IEEE J. Selected Areas Commun.*, vol. SAC-2, pp. 765–777, Sept. 1984.

[6] D. M. Brady, "An adaptive coherent diversity receiver for data transmission through dispersive media," in *Proc. 1970 IEEE Int. Conf. Commun.*, pp. 21–35 to 21–39, June 1970.

[7] C. Caraiscos and B. Liu, "A roundoff error analysis of the LMS adaptive algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, pp. 34–41, Feb. 1984.

[8] R. W. Chang, "A new equalizer structure for fast start-up digital communication," *Bell Syst. Tech. J.*, vol. 50, pp. 1969–2014, July–Aug. 1971.

[9] R. W. Chang and E. Y. Ho, "On fast start-up data communication systems using pseudo-random training sequences," *Bell Syst. Tech. J.*, vol. 51, pp. 2013–2027, Nov. 1972.

[10] J. M. Cioffi and T. Kailath, "Fast, recursive-least-squares, transversal filters for adaptive filtering," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, pp. 304–337, Apr. 1984.

[11] J. M. Cioffi and T. Kailath, "An efficient exact-least-squares fractionally spaced equalizer using intersymbol interpolation," *IEEE J. Selected Areas Commun.*, vol. SAC-2, pp. 743–756, Sept. 1984.

[12] D. L. Duttweiler, J. E. Mazo, and D. G. Messerschmitt, "An upper bound on the error probability in decision-feedback equalization," *IEEE Trans. Inform. Theory*, vol. IT-20, pp.

490–497, July 1974.
[13] D. L. Duttweiler, "Adaptive filter performance with nonlinearities in the correlation multiplier," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-30, pp. 578–586, Aug. 1982.
[14] T. Ericson, "Structure of optimum receiving filters in data transmission systems," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-17, pp. 352–353, May 1971.
[15] D. D. Falconer and G. J. Foschini, "Theory of minimum mean-square-error QAM system employing decision feedback equalization," *Bell Syst. Tech. J.*, vol. 53, pp. 1821–1849, Nov. 1973.
[16] D. D. Falconer, "Jointly adaptive equalization and carrier recovery in two-dimensional digital communication systems," *Bell Syst. Tech. J.*, vol. 55, pp. 317–334, Mar. 1976.
[17] D. D. Falconer and F. R. Magee, Jr., "Adaptive channel memory truncation for maximum likelihood sequence estimation," *Bell Syst. Tech. J.*, vol. 52, pp. 1541–1562, Nov. 1973.
[18] _____, "Evaluation of decision feedback equalization and Viterbi algorithm detection for voiceband data transmission–Parts I and II," *IEEE Trans. Commun.*, vol. COM-24, pp. 1130–1139, Oct. 1976, and pp. 1238–1245, Nov. 1976.
[19] D. D. Falconer and J. Salz, "Optimal reception of digital data over the Gaussian channel with unknown delay and phase jitter," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 117–126, Jan. 1977.
[20] D. D. Falconer and L. Ljung, "Application of fast Kalman estimation to adaptive equalization," *IEEE Trans. Commun.*, vol. COM-26, pp. 1439–1446, Oct. 1978.
[21] D. D. Falconer, "Adaptive equalization of channel nonlinearities in QAM data transmission systems," *Bell Syst. Tech. J.*, vol. 57, pp. 2589–2611, Sept. 1978.
[22] _____, "Adaptive reference echo cancellation," *IEEE Trans. Commun.*, vol. COM-30, pp. 2083–2094, Sept. 1982.
[23] G. L. Fenderson, J. W. Parker, P. D. Quigley, S. R. Shepard, and C. A. Siller, Jr., "Adaptive transversal equalization of multipath propagation for 16-QAM, 90-Mb/s digital radio," *AT& T Bell Lab. Tech. J.*, vol. 63, pp. 1447–1463, Oct. 1984.
[24] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363–378, May 1972.
[25] _____, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, Mar. 1973.
[26] G. D. Forney, S. U. H. Qureshi, and C. K. Miller, "Multipoint networks: Advances in modem design and control," in *Nat. Telecom. Conf. Rec.*, pp. 50-1-1 to 50-1-4, Dec. 1976.
[27] G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, 'Efficient modulation for band-limited channels," *IEEE J. Selected Areas Commun.*, vol. SAC-2, pp. 632–647, Sept. 1984.
[28] G. J. Foschini and J. Salz, "Digital communications over fading radio channels," *Bell Syst. Tech. J.*, vol. 62, pp. 429–459, Feb. 1983.
[29] L. E. Franks, Ed., *Data Communication.* Stroudsburg, PA: Dowden, Hutchinson and Ross, 1974.
[30] S. A. Fredricsson, "Joint optimization of transmitter and receiver filters in digital PAM systems with a Viterbi detector," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 200–2210, Mar. 1976.
[31] D. A. George, 'Matched filters for interfering signals," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-11, pp. 153–154, Jan. 1965.
[32] D. A. George, R. R. Bowen, and J. R. Storey, "An adaptive decision feedback equalizer," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 281–293, June 1971.
[33] A. Gersho, "Adaptive equalization of highly dispersive channels," *Bell Syst. Tech. J.*, vol. 48, pp. 55–70, Jan. 1969.
[34] A. Gersho and T. L. Lim, "Adaptive cancellation of intersymbol interference for data transmission," *Bell Syst. Tech. J.*, vol. 60, pp. 1997–2021, Nov. 1981.
[35] R. D. Gitlin, E. Y. Ho, and J. E. Mazo, "Passband equalization of differentially phase-modulated data signals," *Bell Syst. Tech. J.*, vol. 52, pp. 219–238, February 1973.
[36] R. D. Gitlin, J. E. Mazo, and M. G. Taylor, "On the design of gradient algorithms for digitally implemented adaptive filters," *IEEE Trans. Circuit Theory*, vol. CT-20, pp. 125–136, Mar. 1973.
[37] R. D. Gitlin and F. R. Magee, Jr., "Self-orthogonalizing algorithms for accelerated covergence of adaptive equalizers," *IEEE Trans. Commun.*, vol. COM-25, pp. 666–672, July 1977.
[38] R. D. Gitlin and S. B. Weinstein, "On the required tap-weight precision for digitally-implemented adaptive mean-squared equalizers," *Bell Syst. Tech. J.*, vol. 58, pp. 301–321, Feb. 1979.
[39] _____, "Fractionally-spaced equalization: An improved digital transversal equalizer," *Bell Syst. Tech. J.*, vol. 60, pp. 275–296, Feb. 1981.
[40] R. D. Gitlin, H. C. Meadors, Jr., and S. B. Weinstein, "The tap-leakage algorithm: An algorithm for the stable operation of a digitally implemented, fractionally-spaced adaptive equalizer," *Bell Syst. Tech. J.*, vol. 61, pp. 1817–1939, Oct. 1982.
[41] D. N. Godard, "Channel equalization using a Kalman filter for fast data transmission," *IBM J. Res. Develop.*, vol. 18, pp. 267–273, May 1974.
[42] _____, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867–1875, Nov. 1980.
[43] _____, "A 9600 bit/s modem for multipoint communication systems," in *Nat. Telecomm. Conf. Rec.* (New Orleans, LA, Dec. 1981), pp. B3.3.1–B3.3.5.
[44] N. Grenander and G. Szego, *Toeplitz Forms and Their Application.* Berkeley, CA: Univ. Calif. Press, 1958.
[45] L. Guidoux, "Egaliseur autoadaptif a double echantillonnage," *L'Onde Electrique*, vol. 55, pp. 9–13, Jan. 1975.
[46] M. L. Honig and D. G. Messerschmitt, *Adaptive Filters; Structures, Algorithms, and Applications.* Boston, MA: Kluwer Academic Pub., 1984.
[47] C. R. Johnson, Jr., "Adaptive IIR filtering: Current results and open issues," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 237–250, Mar. 1984.
[48] P. Kabal and S. Pasupathy, "Partial-response signaling," *IEEE Trans. Commun.*, vol. COM-23, pp. 921–934, Sept. 1975.
[49] H. Kobayashi, "Simultaneous adaptive estimation and decision algorithm for carrier modulated data transmission systems," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 268–280, June 1971.
[50] E. R. Kretzmer, "Binary data communication by partial response transmission," in *1965 ICC Conf. Rec.*, pp. 451–456; also, "Generalization of a technique for binary data communication," *IEEE Trans. Commun. Technol.*, vol. COM-14, pp. 67–68, Feb. 1966.
[51] W. U. Lee and F. S. Hill, "A maximum likelihood sequence estimator with decision feedback equalization," *IEEE Trans. Commun. Technol.*, vol. COM-25, pp. 971–979, Sept. 1977.
[52] F. Ling and J. G. Proakis, "A generalized multichannel least squares lattice algorithm based on sequential processing stages," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, pp. 381–389, Apr. 1984.
[53] H. L. Logan and G. D. Forney, Jr., "A MOS/LSI multiple configuration 9600 b/s data modem," in *Proc. IEEE Int. Conf. Commun.*, pp. 48-7 to 48-12, June 1976.
[54] R. W. Lucky, J. Salz, and E. J. Weldon, Jr., *Principles of Data Communication.* New York: McGraw-Hill, 1968.
[55] R. W. Lucky, "Automatic equalization for digital communication," *Bell Syst. Tech. J.*, vol. 44, pp. 547–588, Apr. 1965.
[56] _____, "Techniques for adaptive equalization of digital communication systems," *Bell Syst. Tech. J.*, vol. 45, pp. 255–286, Feb. 1966.
[57] R. W. Lucky and H. R. Rudin, "An automatic equalizer for general-purpose communication channels," *Bell Syst. Tech. J.*, vol. 46, pp. 2179–2208, Nov. 1967.
[58] R. W. Lucky, "Signal filtering with the transversal equalizer," in *Proc. 7th Ann. Allerton Conf. on Circuits and System Theory*, pp. 792–803, Oct. 1969.
[59] _____, "A survey of the communication theory literature: 1968–1973," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 725–739, Nov. 1973.
[60] O. Macchi and E. Eweda, "Convergence analysis of self-adaptive equalizers," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 161–176, Mar. 1984.
[61] L. R. MacKechnie, "Maximum likelihood receivers for channels having memory," Ph.D. dissertation, Dep. Elec. Eng., Univ. of Notre Dame, Notre Dame, Jan. 1973.
[62] F. R. Magee, Jr., and J. G. Proakis, "Adaptive maximum-likeli-

hood sequence estimation for digital signaling in the presence of intersymbol interference," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-19, pp. 120–124, Jan. 1973.
[63] J. Makhoul, "A class of all-zero lattice digital filters: properties and applications," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-26, pp. 304–314, Aug. 1978.
[64] J. E. Mazo, "Optimum timing phase for an infinite equalizer," *Bell Syst. Tech. J.*, vol. 54, pp. 189–201, Jan. 1975.
[65] ———, "On the independence theory of equalizer convergence," *Bell Syst. Tech. J.*, vol. 58, pp. 963–993, May–June 1979.
[66] ———, "Analysis of decision-directed equalizer convergence," *Bell Syst. Tech. J.*, vol. 59, pp. 1857–1876, Dec. 1980.
[67] D. G. Messerschmitt, "A geometric theory of intersymbol interference: Part I," *Bell Syst. Tech. J.*, vol. 52, pp. 1483–1519, Nov. 1973.
[68] ———, "Design of a finite impulse response for the Viterbi algorithm and decision feedback equalizer," in *Proc. IEEE Int. Conf. Communications, ICC-74* (Minneapolis, MN, June 17–19, 1974).
[69] ———, "Echo cancellation in speech and data transmission," *IEEE J. Selected Areas Commun.*, vol. SAC-2, pp. 283–296, Mar. 1984.
[70] A. Milewski, "Periodic sequences with optimal properties for channel estimation and fast start-up equalization," *IBM J. Res. Develop.*, vol. 27, pp. 426–431, Sept. 1983.
[71] P. Monsen, "Feedback equalization for fading dispersive channels," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 56–64, Jan. 1971.
[72] ———, "MMSE equalization of interference on fading diversity channels," *IEEE Trans. Commun.*, vol. COM-32, pp. 5–12, Jan. 1984.
[73] M. Morf, T. Kailath, and L. Ljung, "Fast algorithms for recursive identification," in *Proc. 1976 IEEE Conf. Decision Contr.* (Clearwater Beach, FL, Dec. 1976), pp. 916–921.
[74] M. Morf, A. Vieira, and D. T. Lee, "Ladder forms for identification and speech processing," in *Proc. 1977 IEEE Conf. Decision Contr.* (New Orleans, LA, Dec. 1977), pp. 1074–1078.
[75] K. H. Mueller, "A new, fast-converging mean-square algorithm for adaptive equalizers with partial-response signaling," *Bell Syst. Tech. J.*, vol. 54, pp. 143–153, Jan. 1975.
[76] K. H. Mueller and D. A. Spaulding, "Cyclic equalization—A new rapidly converging equalization technique for synchronous data communication," *Bell Syst. Tech. J.*, vol. 54, pp. 369–406, Feb. 1975.
[77] K. H. Mueller and J. J. Werner, "A hardware efficient passband equalizer structure for data transmission," *IEEE Trans. Commun.*, vol. COM-30, pp. 538–541, Mar. 1982.
[78] M. S. Mueller, "Least-squares algorithms for adaptive equalizers," *Bell Syst. Tech. J.*, vol. 60, pp. 1905–1925, Oct. 1981.
[79] ———, "On the rapid initial convergence of least-squares equalizer adjustment algorithms," *Bell Syst. Tech. J.*, vol. 60, pp. 2345–2358, Dec. 1981.
[80] M. S. Mueller and J. Salz, "A unified theory of data-aided equalization," *Bell Syst. Tech. J.*, vol. 6, pp. 2023–2038, Nov. 1981.
[81] K. Murano, Y. Mochida, F. Amano, and T. Kinoshita, "Multiprocessor architecture for voiceband data processing (application to 9600 bps modem)," in *Proc. IEEE Int. Conf. Commun.*, pp. 37.3.1–37.3.5, June 1979.
[82] T. Murase, K. Morita, and S. Komaki, "200 Mb/s 16-QAM digital radio system with new countermeasure techniques for multipath fading," in *Proc. IEEE Int. Conf. Commun.*, pp. 46.1.1–46.1.5, June 1981.
[83] R. Price, "Nonlinearly feedback-equalized PAM vs. capacity for noisy filter channels," in *Proc. 1972 IEEE Int. Conf. Commun.*, pp. 22-12 to 22-17, June 1972.
[84] J. G. Proakis and J. H. Miller, "An adaptive receiver for digital signaling through channels with intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 484–497, July 1969.
[85] J. G. Proakis, "Adaptive nonlinear filtering techniques for data transmission," in *IEEE Symp. on Adaptive Processes, Decision and Control*, pp. XV.2.1-5, 1970.
[86] ———, "Advances in equalization for intersymbol interference," in *Advances in Communication Systems*, vol. 4, A. J. Viterbi, Ed. New York: Academic Press, 1975, pp. 123–198.
[87] ———, *Digital Communications*. New York: McGraw-Hill, 1983.
[88] S. U. H. Qureshi, "New approaches in adaptive reception of digital signals in the presence of intersymbol interference," Ph.D. dissertation, Univ. of Toronto, Toronto, Ont., Canada, May 1973.
[89] S. U. H. Qureshi and E. E. Newhall, "An adaptive receiver for data transmission over time-dispersive channels," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 448–457, July 1973.
[90] S. U. H. Qureshi, "Fast start-up equalization with periodic training sequences," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 553–563, Sept. 1977.
[91] S. U. H. Qureshi and G. D. Forney, Jr., "Performance and properties of a T/2 equalizer," in *Nat. Telecomm. Conf. Rec.*, Dec. 1977.
[92] W. Rudin, *Real and Complex Analysis*. New York: McGraw-Hill, 1966.
[93] J. Salz, "Optimum mean-square decision feedback equalization," *Bell Syst. Tech. J.*, vol. 52, pp. 1341–1373, Oct. 1973.
[94] ———, "On mean-square decision feedback equalization and timing phase," *IEEE Trans. Commun. Technol.*, vol. COM-25, pp. 1471–1476, Dec. 1977.
[95] Y. Sato, "A method of self-recovering equalization for multilevel amplitude modulation," *IEEE Trans. Commun.*, vol. COM-23, pp. 679–682, June 1975.
[96] E. H. Satorius and S. T. Alexander, "Channel equalization using adaptive lattice algorithms," *IEEE Trans. Commun.* (Concise Paper), vol. COM-27, pp. 899–905, June 1979.
[97] E. H. Satorius and J. D. Pack, "Application of least squares lattice algorithms to adaptive equalization," *IEEE Trans. Commun.*, vol. COM-29, pp. 136–142, Feb. 1981.
[98] C. A. Siller, Jr., "Multipath propagation," *IEEE Commun. Mag.*, vol. 22, pp. 6–15, Feb. 1984.
[99] T. A. Schonhoff and R. Price, "Some bandwidth efficient modulations for digital magnetic recording," in *Proc. IEEE Int. Conf. Communications, ICC-81* (Denver, CO, June 15–18, 1981).
[100] S. Y. Tong, "Dataphone II service: Data set architecture," in *Nat. Telecomm. Conf. Rec.* (New Orleans, LA, Dec. 1981), pp. B.3.2.1–B.3.2.5.
[101] T. Tsuda, Y. Mochida, K. Murano, S. Unagami, H. Gambe, T. Ikezawa, H. Kikuchi, and S. Fujii, "A high performance LSI digital signal processor for communication," in *Proc. 1983 IEEE Int. Conf. Commun.*, pp. A5.6.1–A5.6.5, June 1983.
[102] G. Ungerboeck, "Theory on the speed of convergence in adaptive equalizers for digital communication," *IBM J. Res. Devel.*, vol. 16, pp. 546–555, Nov. 1972.
[103] ———, "Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems," *IEEE Trans. Commun.*, vol. COM-22, pp. 624–636, May 1974.
[104] ———, "Fractional tap-spacing equalizer and consequences for clock recovery in data modems," *IEEE Trans. Commun.*, vol. COM-24, pp. 856–864, Aug. 1976.
[105] ———, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.
[106] P. J. van Gerwen, N. A. M. Verhoeckx, and T. A. C. M. Claasen, "Design considerations for a 144 kbit/s digital unit for the local telephone network," *IEEE J. Selected Areas Commun.*, vol. SAC-2, pp. 314–323, Mar. 1984.
[107] L.-F. Wei, "Rotationally invariant convolutional channel coding with expanded signal space—Part II: Nonlinear codes," *IEEE J. Selected Areas Commun.*, vol. SAC-2, pp. 672–686, Sept. 1984.
[108] S. B. Weinstein, "A baseband data-driven echo canceller for full-duplex transmission on two-wire circuits," *IEEE Trans. Commun.*, vol. COM-25, vol. 654–666, July 1977.
[109] B. Widrow and M. E. Hoff, Jr., "Adaptive switching circuits," in *IRE WESCON Conv. Rec.*, pt. 4, pp. 96–104, Aug. 1960.
[110] B. Widrow, J. R. Glover, Jr., J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, Jr., and R. C. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, pp. 1692–1716, Dec. 1975.
[111] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson,

Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, pp. 1151–1162, Aug. 1976.

[112] M. Yasumoto, T. Enomoto, K. Watanabe, and T. Ishihara, "Single-chip adaptive transversal filter IC employing switched capacitor technology," *IEEE J. Selected Areas Commun.*, vol. SAC-2, pp. 324–333, Mar. 1984.

[113] K. Feher, *Digital Communications: Satellite/Earth Station Engineering*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[114] P. Monsen, "Theoretical and measured performance of a DFE modem on a fading multipath channel," *IEEE Trans. Commun.*, vol. COM-25, pp. 1144–1153, Oct. 1977.

[115] H. Nyquist, "Certain topics in telegraph transmission theory," *Trans. AIEE*, vol. 47, pp. 617–644, Apr. 1928.

[116] J. W. Smith, "The joint optimization of transmitted signal and receiving filter for data transmission systems," *Bell Syst. Tech. J.*, vol. 44, pp. 2363–2392, Dec. 1965.

[117] D. W. Tufts, "Nyquist's problem—The joint optimization of transmitter and receiver in pulse amplitude modulation," *Proc. IEEE*, vol. 53, pp. 248–260, Mar. 1965.

[118] CCITT Recommendation V.27 bis, "4800/2400 bits per second modem with automatic equalizer standardized for use on leased telephone-type circuits," Int. Telegraph and Telephone Consultative Committee, Geneva, Switzerland, 1980.

[119] CCITT Recommendation V.29, "9600 bits per second modem standardized for use on point-to-point 4-wire leased telephone-type circuits," Int. Telegraph and Telephone Consultative Committee, Geneva, Switzerland, 1980.

[120] S. U. H. Qureshi and H. M. Ahmed, "A custom ship set for digital signal processing," in *VLSI Signal Processing*, P. R. Cappello *et al.*, Eds. New York: IEEE PRESS, 1984.

[121] K. J. Wouda, S. J. M. Tol, and W. J. M. Reiknkjens, "An ISDN transmission system with adaptive echo cancelling and decision feedback equalization—A two-chip realization," in *Proc. IEEE Int. Conf. Commun.* (Amsterdam, The Netherlands), pp. 685–690, 1984.

[122] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, pp. 561–580, Apr. 1975.

[123] O. Agazzi, D. G. Messerschmitt, and D. A. Hodges, "Nonlinear echo cancellation of data signals," *IEEE Trans. Commun.*, vol. COM-30, pp. 2421–2433, Nov. 1982.

[124] M. Holte and S. Stueflotten, "A new digital echo canceller for two-wire subscriber lines," *IEEE Trans. Commun.*, vol. COM-29, pp. 1573–1581, Nov. 1981.

[125] V. B. Lawrence and J. J. Werner, "Low-speed data transmission by using high-speed modems," in *Proc. IEEE Globecom* (Atlanta, GA), pp. 677–682, Nov. 1984.

[126] E. Eleftheriou and D. D. Falconer, "Restart methods for stabilizing FRLS adaptive equalizers in digital HF transmission," in *Proc. IEEE Globecom* (Atlanta, GA), pp. 1558–1562, Nov. 1984.

[127] B. Friedlander, "Lattice filters for adaptive processing," *Proc. IEEE*, vol. 70, pp. 829–867, Aug. 1982.

[128] D. L. Lyon, "Envelope-derived timing recovery in QAM and SQAM systems," *IEEE Trans. Commun.*, vol. COM-23, pp. 1327–1331, Nov. 1975.

[129] S. U. H. Qureshi and G. D. Forney, Jr., "A 9.6/16 kb/s speech digitizer," in *Conf. Rec. IEEE Int. Conf. Commun.*, pp. 30–31 to 30–36, June 1975, also in *Waveform Quantization and Coding*, N. S. Jayant, Ed. New York: IEEE PRESS, 1976, pp. 214–219.

[130] G. Picchi and G. Prati, "Self-orthogonalizing adaptive equalization in the discrete frequency domain," *IEEE Trans. Commun.*, vol. COM-32, pp. 371–379, Apr. 1984.