# Echo Cancellation in Speech and Data Transmission

DAVID G. MESSERSCHMITT, FELLOW, IEEE

*Abstract* —This tutorial paper reviews echo cancellation techniques as applied to both voice and data transmission applications. The echo control problem in telephone voice transmission is described, and the several measures taken to counteract echo are outlined. The problem of full-duplex data transmission over two-wire lines is described, with application to both digital subscriber loop and voiceband data transmission. The unique characteristics of each of these applications, more advanced methods recently proposed, and approaches to implementation are described.

## I. INTRODUCTION

E CHO cancellation is an application of adaptive filtering technology to the control of echo in the telephone network. It has been studied in the research lab for two decades, and due to the advances in microelectronics technology is now finding widespread practical application. This tutorial paper describes the echo control problems addressed by echo cancellation in the context of both voice and data transmission, the echo cancellation algorithms which are used in these applications, and the implementation of echo cancellers.

In Section II, the application of echo cancellation to speech transmission is described, and in Section III applications in full-duplex data transmission are discussed. Section IV outlines the algorithms used for adaptation of echo cancellers in all these applications, and evaluates their performance. Finally, in Section V more recently proposed advanced techniques in echo cancellation are summarized.

## II. ECHO CANCELLATION IN SPEECH TRANSMISSION

The telephone network generates echos at points internal to and near the end of a telephone connection. Echo cancellation is one means used to combat this echo for both speech and data transmission. The requirements for speech and data are quite different, so this section will concentrate on speech and Section III will deal with data. See [1] for another excellent reference on this topic.

A starting point on echo terminology is given In Fig. 1. In Fig. 1(a), a simplified telephone connection is shown.
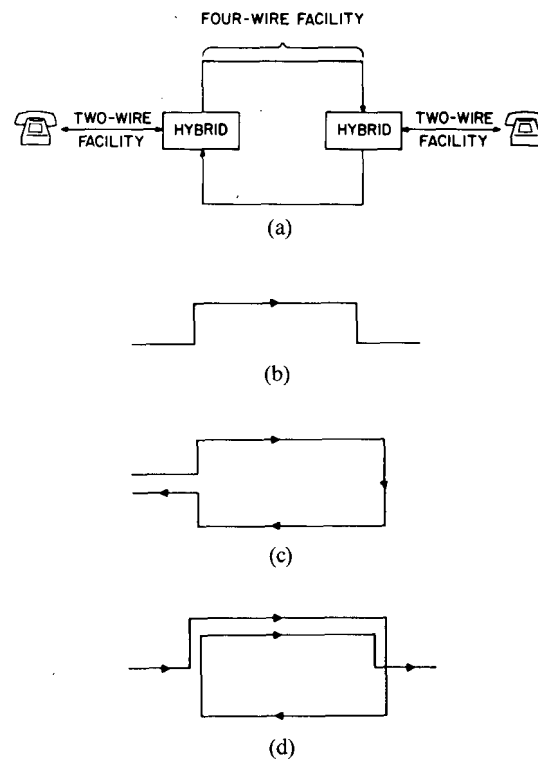
Fig. 1. Sources of echo in the telephone network. (a) Simplified telephone connection. (b) Talker speech path. (c) Talker echo. (d) Listener echo.

This connection is typical in that it contains two-wire segments on the ends (the subscriber loops and possibly some portion of the local network), in which both directions of transmission are carried on a single wire pair. The center of the connection is four wire, in which the two directions of transmission are segregated on physically different facilities (typical of carrier systems).

There is a potential feedback loop around the four-wire portion of the connection, and without sufficient loss in this path there is degradation of the transmission or in extreme cases oscillation (called singing). The hybrid is a device which provides a large loss around this loop without affecting the loss in the two talker speech paths. The remainder of Fig. 1 illustrates more graphically the function of this hybrid. One of the two talker speech paths is shown in Fig. 1(b). In order that this path not have a large attenuation, it is necessary for the hybrid not to have an appreciable attenuation between its two-wire and either
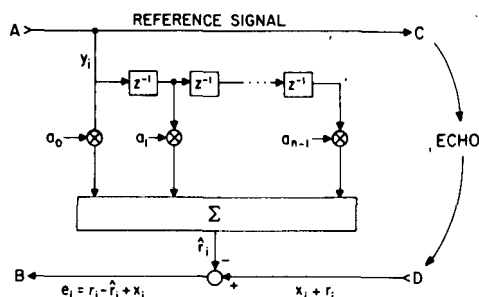
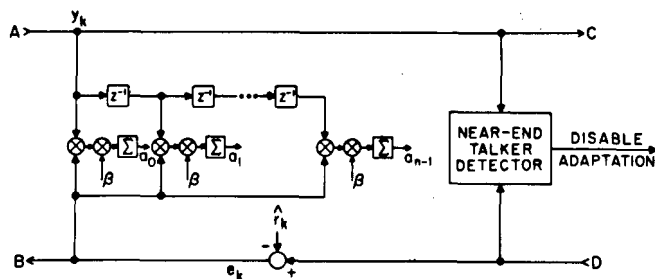Fig. 2. Principle of the echo canceller for one direction of transmission.



Fig. 3. Adaptation mechanism for the echo canceller.

four-wire port. There are two distinct echo mechanisms shown in Fig. 1(c) and (d), "Talker echo" results in the talker hearing a delayed version of his or her own speech, while in "listener echo" the listener hears a delayed version of the talker's speech.

On connections with small propagation delay, the subjective impairment from this echo can be controlled by inserting loss in the connection since the echo suffers this loss twice or three times whereas the talker path suffers this loss only once. For roundtrip delays over about 40 ms, it is necessary to use echo suppressors [1]. For very long delays, such as would be experienced on a satellite connection, the echo canceller is used [2].

The principle of the echo canceller for only one direction of transmission is shown in Fig. 2. The canceller is placed in the four-wire path near the origin of the echo. One direction of transmission is from port $A$ to $C$ (the "reference signal" $y_i$ for the canceller), and the other direction from port $D$ to $B$. The reference signal passing through the echo channel results in the echo signal $r_i$, which together with a near-end talker signal $x_i$ appears on port $D$. Included in $x_i$ is any thermal or quantization noise. A replica of the echo $\hat{r}_i$ is generated by applying the reference signal to a transversal filter (tapped delay line), and this replica is subtracted from the signal on port $D$ to yield the cancellation error signal $e_i$ (which also contains the near-end talker signal).

The $n - 1$ filter coefficients $a_0, \cdots, a_{n-1}$ are caused to adapt to the echo transfer function (minimize the cancellation error) using the circuitry of Fig. 3. While there are more sophisticated adaptation algorithms, as discussed in Section V, this simple algorithm (described in Section IV-A) is adequate for most echo cancellation applications. The residual cancellation error $e_k$ is used to drive the
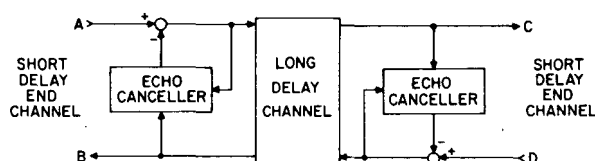


Fig. 4. Split echo canceller configuration for two directions.

canceller adaptation by inferring from this error the appropriate correction to the transversal filter coefficients so as to reduce this error. Specifically, this error is cross correlated with the successive delays of the reference signal. The summation box is an accumulator, the output of which is the corresponding filter coefficient. The correction to the accumulated value is scaled by a step size $\beta$ that controls the speed of adaptation and the asymptotic error in a manner discussed in Section IV-A.

The error signal $e_k$ contains a component of the near-end talker signal $x_k$ in addition to the residual echo cancellation error. The effect of this on the cancellation is naturally of some concern. As long as $y_k$ is uncorrelated with $x_k$, which should be the case, this near-end talker will not affect the asymptotic mean value of the filter coefficients. However, the asymptotic variation in the filter coefficients about this mean will be increased substantially in the presence of a near-end talker due to the introduction of another (large) stochastic component in the adaptation. This effect can be compensated by choosing a very small speed of adaptation (small $\beta$), but it is generally preferable to employ a near-end talker detector as shown. The adaptation proceeds only when a there is no detectable near-end talker signal.

Of course, in practice it is desirable to cancel the echos in both directions of the connection. For this purpose two adaptive cancellers are necessary, as shown in Fig. 4, where one cancels the echo from each end of the connection. The near-end talker for one of the cancellers is the far-end talker for the other. In each case, the near-end talker is the "closest" talker, and the far-end talker is the talker generating the echo which is being cancelled. It is desirable to position these two "halves" of the canceller in a split configuration, as shown in the figure where the bulk of the delay in the four-wire portion of the connection is in the middle. The reason is that number of coefficients in the echo cancellation filter is directly related to the delay of the channel between where the echo canceller is located and the hybrid which generates the echo. In the split configuration the largest delay is not in the echo path of either half of the canceller, and hence the number of coefficients is minimized. Typically cancellers in this configuration require only 128 or 256 coefficients, whereas the number of coefficients required to accommodate a satellite connection in the end link would be impractically large.

Mathematically, the transversal filter of Fig. 2 is a direct realization of a finite impulse response (FIR)

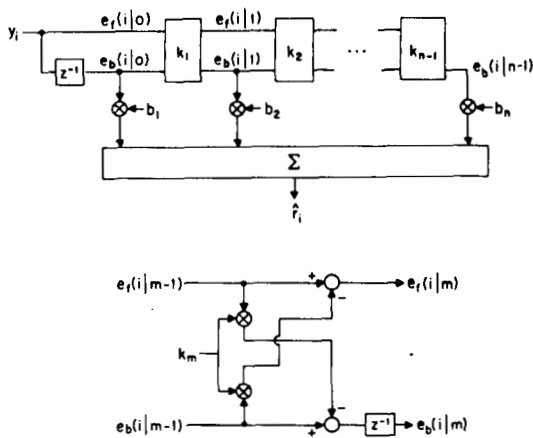$$r_k = \sum_{m=0}^{n-1} a_m y_{k-m} \qquad (2.1)$$

Fig. 5. Lattice filter echo canceller.



Fig. 6. Application of digital subscriber loop.

where there are $n$ filter coefficients $a_0, \cdots, a_{n-1}$. Of course there are many filter realizations other than the transversal filter with the same response. One of these alternatives of particular importance is the lattice structure shown in Fig. 5 [3], [4]. The lattice filter structure consists first of all of a set of $n-1$ stages with internal coefficients $k_j, 1 < j \leqslant n-1$, which are commonly called the *reflection* or *PARCOR* coefficients. The signals at the output of the $m$th stage, $e_f(1|m)$ and $e_b(i|m)$, are called, respectively, the forward and backward prediction error of order $m$. For purposes of realization of the echo canceller, the salient property of these prediction errors is that when the reflection coefficients are chosen appropriately, the successive backward prediction errors are uncorrelated. The so-called order update equations for the prediction errors are given by

$$e_f(i|m) = e_f(i|m-1) - k_m e_b(i|m-1) \tag{2.2}$$

$$e_f(i|0) = y_i \tag{2.3}$$

$$e_b(i|m) = e_b(i-1|m-1) - k_m e_f(i-1|m-1) \tag{2.4}$$

$$e_b(i|0) = y_{i-1}. \tag{2.5}$$

The echo replica is generated by forming a weighted linear combination of the backward prediction errors of successive orders with weights $b_1, \cdots, b_n$, given by

$$r_i = \sum_{m=1}^{n} b_m e_b(i|m-1). \tag{2.6}$$

One way to think of this structure is as a transversal filter where the delay elements have been replaced by the more complicated lattice stages. The purpose of the lattice stages are to decorrelate the input signal, so that the final replica is formed from a sum of uncorrelated signals. The reflection coefficients are therefore chosen irrespective of the desired echo replica on the basis of the reference signal input statistics alone. The transfer function of the overall filter is controlled by the choice of the $b_1, \cdots, b_n$.

The advantage of the lattice structure arises from its greater speed of adaptation, as discussed in Section V-B.
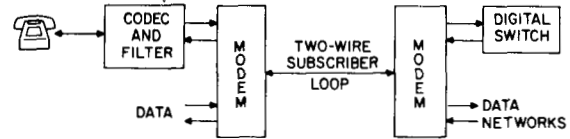
## III. ECHO CANCELLATION IN DATA TRANSMISSION

When it is attempted to transmit data signals through the network, the same echoes are encountered as in speech transmission. When half-duplex data transmission is used, echoes present no problem since there is no receiver on the transmitting end to be affected by the echo. In full-duplex transmission, where the data signals are transmitted in both directions simultaneously, echoes from the data signal transmitted in one direction can interfere with the data signal flowing in the opposite direction, unless these two data signals are in nonoverlapping frequency bands.

Full-duplex transmission over a common media has arisen in two important applications. In both these applications, the need for a common media arises because the public telephone network typically only provides a two-wire connection to each customer premise. (This is because of the high cost of copper wire, and the large percentage of the telephone network investment in this facility.)

The first application, Fig. 6, is digital transmission on the subscriber loop, in which the basic voice service as well as enhanced data services are provided over the two-wire subscriber loop. Total bit rates for this application that have been proposed are 80 and 144 kbits/s in each direction where the latter rate includes provision for two voice/data channels at 64 kbits/s each plus a data channel at 16 kbits/s, and the first alternative allows only a single 64 kbits/s voice/data channel. This digital subscriber loop capability is an important element of the emerging integrated services digital network (ISDN), in which integrated voice and data services will be provided to the customer over a common facility. As shown in Fig. 6, voice transmission requires a codec and filter to perform the analog-to-digital and digital-to-analog conversion on the customer premises, together with the modem for transmitting the full-duplex data stream over the two-wire subscriber loop. Any data signals to be accommodated are connected directly to the modem. The central office end of the loop has another full-duplex modem, with connections to the digital central office switch for voice or circuit-switched data transmission, and to data networks for packet switched data transport capability.

The second application for full-duplex data transmission is in voiceband data transmission where the basic customer interface to the network is often the same two-wire subscriber loop as shown in Fig. 7. In this case, however, the transmission link is usually more complicated, due to the possible presence of four-wire trunk facilities in the middle of the connection. The situation can be further complicated by the presence of two-wire toll switches, allowing inter-
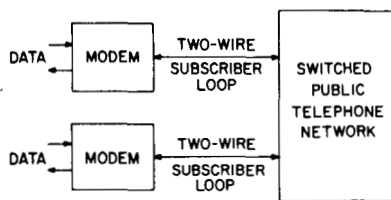
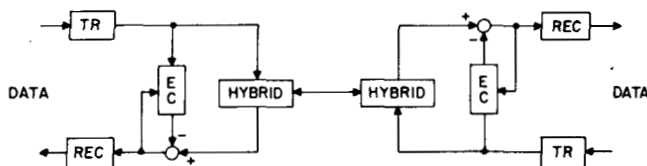Fig. 7. Application of full-duplex voiceband data modems.



Fig. 8. Echo cancellation for full-duplex data transmission.

mediate four-two-four wire conversions internal to the network.

The two applications differ substantially in the types of problems which must be overcome. For the digital subscriber loop, the transmission medium is fairly ideal, consisting of cable pairs with a relatively wide bandwidth. The biggest complication is the presence in some countries of bridged taps—open-circuited wire pairs bridged onto the main line. The voiceband data modem, while requiring a lower speed of transmission, encounters many more impairments. In addition to the severe bandlimiting when carrier facilities are used, there are problems with noise, nonlinearities, and sometimes even frequency offset. Another difference is that the subscriber loop can use baseband transmission, while the voiceband data set always uses passband transmission; that is, modulates a carrier with the data stream.

One approach to full-duplex data transmission is the use of an echo canceller to isolate the two directions of transmission as shown in Fig. 8. There is a transmitter and receiver on each end of the connection, and a hybrid is used to provide a virtual four-wire connection between the transmitter on each end and the receiver on the opposite end. Unfortunately, there is some feedthrough the hybrid of the high-level signal from the transmitter into the local receiver. The hybrid depends on knowledge of the two-wire impedance for complete isolation, and in practice no better than 10 dB or so of attenuation through the hybrid can be guaranteed with a single compromise termination.

In both subscriber loop and voiceband data transmission, the loss of the channel from one transmitter to the receiver on the other end can be as high as 40–50 dB. This implies that in the worst case, the undesired transmitter local feedthrough (echo) signal can be 30–40 dB higher in level than the data signal from the far end (assuming both transmitted signals are at the same level). Since signal to interference ratios on the order of 20 dB or more are required for reliable data transmission, the echo canceller is added to each end to give an additional attenuation of the undesired feedthrough signal of about 50–60 dB.

## A. Subscriber Loop Digital Transmission

There are two competing methods of providing full-duplex data capability on the subscriber loop, frequency-division multiplexing (FDM) and time-compression multiplexing (TCM) [5]. In spite of this competition, echo cancellation has achieved significant attention because of its ability to provide greater range (distance between subscriber and central office) in some circumstances. This greater range is due to the approximate halving of the transmitted signal bandwidth relative to the alternatives.

The implementation of a data echo canceller differs from the speech echo canceller discussed in Section II in several respects. One difference is that the reference signal for a data echo canceller is the sequence of transmitted data symbols. The fact that these symbols assume only a small number of values simplifies the implementation of the canceller [6]. Another difference is that signals in both directions are present at all times (except perhaps during a training period), making the effect of the near-end signal on adaptation much more significant an issue. A third difference is that the desired degree of cancellation is much larger, making design and particularly implementation much more challenging. Finally, there are significant issues of timing recovery, synchronization, and equalization which interact with the echo cancellation and do not exist in speech cancellation [7].

A particularly important constraint imposed by timing recovery considerations is that the sampling rate at the output of the echo canceller must usually be chosen to be a integral multiple of the data symbol rate. The data symbol rate is usually not adequate because the received data signal bandwidth exceeds half the data symbol rate. This implies that the echo canceller has different sampling rates at input and output since the input rate is equal to the symbol rate. To see how this is done, assume that the transmitted signal is

$$s(t) = \sum_m C_m g(t - mT) \qquad (3.1)$$

where $C_m$ is the sequence of transmitted data symbols, $g(t)$ is the transmitted pulse shape, and $T$ is the interval between transmitted data symbols (the baud interval). Suppose this signal is passed through a filter with transfer function $H_e(w)$ representing the echo response. If we denote the response of this filter to $g(t)$ as $h(t)$ and the echo response as $r(t)$, then the latter becomes

$$r(t) = \sum_m C_m h(t - mT). \qquad (3.2)$$

Assuming that this signal is to be reconstructed by the echo canceller with a sampling rate equal to $RT$, where $R$ is an appropriately chosen integer (typically 2 or 4), define

$$r_i(l) = r\left(\left(i + \frac{l}{R}\right)T\right), \qquad 0 \leqslant l \leqslant R - 1 \qquad (3.3)$$

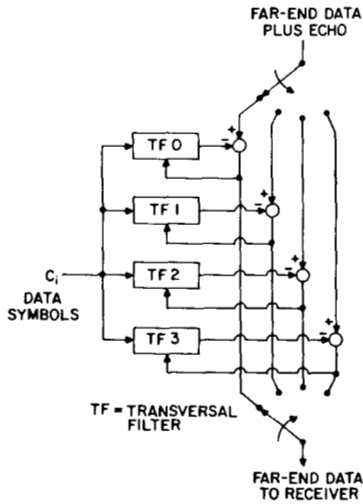where the index $i$ represents the data symbol epoch and $l$ represents the sample from among $R$ samples uniformly

Fig. 9. Interleaved echo cancellers to achieve higher output sampling rate ($R = 4$).
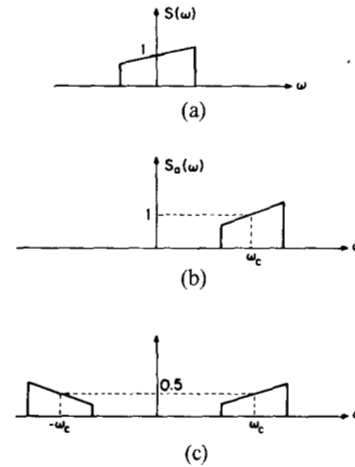


Fig. 10. Spectra of signals for quadrature amplitude modulation (QAM). (a) Baseband signal with complex-valued data symbols. (b) After modulation by complex exponential. (c) After taking real part.

spaced in this epoch. Similarly, define a notation for the samples of the echo pulse response

$$h_i(l) = h\left(\left(i + \frac{l}{R}\right)T\right), \qquad 0 \leqslant l \leqslant R - 1. \quad (3.4)$$

Using this convenient notation, from (3.2) the samples of the received echo become

$$r_i(l) = \sum_m h_m(l) C_{i-m}. \quad (3.5)$$

This relation shows that the samples of the echo can be thought of as $R$ independent echo channels, each channel being driven by the same sequence of data symbols. The impulse response of the $l$th echo channel is $h_i(l)$.

Since the $R$ echo channels are independent, the index $l$ can be dropped. The transversal filter has a finite impulse response, so consider building an echo canceller to generate the replica

$$r_i = \sum_{m=0}^{n-1} C_{i-m} a_m \quad (3.6)$$

where $a_i$, $0 \leqslant i < n - 1$, are the $n$ filter coefficients of one of the $R$ interleaved transversal filters. This transversal filter generates an FIR approximation to the echo response. Note the analogy to (2.1) where the far-end talker samples of the voice canceller have been replaced by the transmitted data symbols at the canceller reference input.

The $R$ independent echo cancellers can be used in the manner of Fig. 9 (shown for $R = 4$). Each canceller has the same input data sequence at its reference input. The far-end data plus echo signal from the hybrid is decimated to four independent signals at the data symbol rate, and these are independently cancelled and recombined into a single sample stream representing the far-end data signal alone. Each of the cancellation error signals is used to drive the adaptation of the corresponding canceller.

Each canceller can be thought of as adapting to the impulse response of the echo channel sampled at a rate

equal to the symbol rate, but with a particular phase out of $R$ possible phases. These cancellers independently converge, although they do have in common the same input sequence of data symbols, and the presence of multiple interleaved cancellers does not affect the speed of adaptation.

### B. Voiceband Data Transmission

The preceding section described the application of echo cancellation to full-duplex baseband data transmission (the digital subscriber loop); the present section extends this to the passband case typical of voiceband data transmission. Full-duplex voiceband data transmission has been provided for some years at 1.2 kbits/s and below (and more recently at 2.4 kbits/s) using FDM. However, at 4.8 kbits/s and above, frequency separation becomes impractical due to inadequate total bandwidth, and the use of echo cancellation must be considered.

The usual method of voiceband data transmission is quadrature amplitude modulation (QAM), which is briefly reviewed here. In a mathematical formulation of this passband modulation, it is convenient to use complex notation, so consider a baseband data signal identical to (3.1) except that the data symbols are complex

$$C_i = a_i + jb_i \quad (3.7)$$

and the transmitted pulse $g(t)$ is still real-valued. Since this signal is complex-valued, it has an asymmetrical spectrum as illustrated in Fig. 10(a). If this complex data signal is modulated up to passband by multiplying by a complex exponential with frequency $\omega_c$, the carrier frequency, the spectrum is shifted by $\omega_c$ radians as illustrated in Fig. 10(b). If the time domain equivalent is called $s_a(t)$, this complex valued signal is given by

$$s_a(t) = \sum_m C_m g(t - mT) e^{j\omega_c t}. \quad (3.8)$$

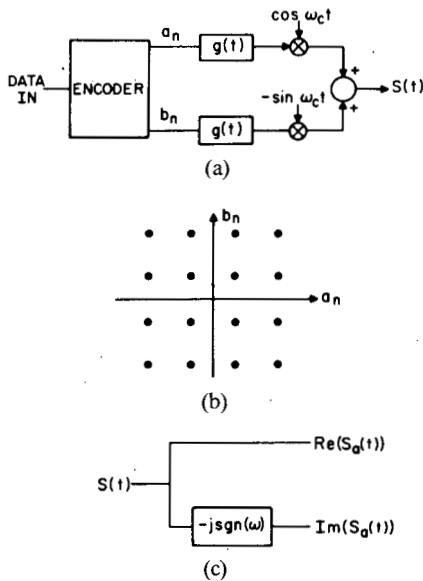Finally, to make the transmitted signal real-valued, simply

(a)

(b)

(c)

Fig. 11.  Principle of quadrature amplitude modulation (QAM). (a) QAM modulator. (b) 16 point signal constellation. (c) Phase splitter in receiver.
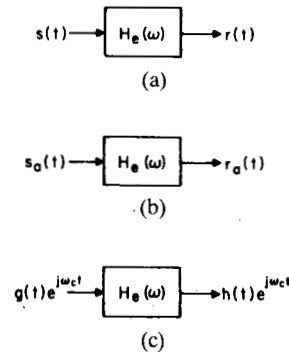


(a)

(b)

(c)

Fig. 12.  Response of echo path filter to a single modulated pulse. (a) Transmitted signal and echo. (b) Analytic signals. (c) Isolated pulse response.

take the real part of $s_a(t)$,

$$s(t) = \text{Re}(s_a(t)). \quad (3.9)$$

Since the real part can be written alternatively as

$$s(t) = \frac{s_s(t) + s_a^*(t)}{2} \quad (3.10)$$

and since the Fourier transform of $s_a^*(t)$ is $S_a^*(-\omega)$, the Fourier transform of $s(t)$ is

$$S(\omega) = \frac{S_a(\omega) + S_a^*(-\omega)}{2} \quad (3.11)$$

as illustrated in Fig. 10(c).

For purposes of implementation of the transmitter, it is convenient to expand in terms of (3.7), so that (3.9) becomes

$$s(t) = \sum_m a_m g(t - mT) \cos(\omega t) - \sum_m b_m g(t - mT) \sin(\omega t) \quad (3.12)$$

where the amplitude modulation of two independent data signals by quadrature carriers is evident. A method of generating this modulated signal is illustrated in Fig. 11(a). The input data stream is encoded to yield the $a_k$ and $b_k$ values. There are many ways of doing this, and Fig. 11(b) illustrates one method. In this case, the complex-valued data symbol $C_k$ assumes one of 16 distinct values, with the real part $a_k$ having four values and the imaginary part $b_k$ having four values (each equally spaced). The encoder in this case maps four input bits into the complex-valued data symbol, two of those bits into the real part and two into the imaginary part. The remainder of the transmitter in Fig. 11(a) modulates a pulse $g(t)$ by the real and imaginary parts, modulates them up to passband by multiplying by quadrature carriers, and sums the results.

The time domain signal with only positive frequency components corresponding to Fig. 10(b) is called the "analytic signal." In the receiver processing, it is convenient to recover this complex valued signal prior to further processing. Given the passband QAM modulated signal, the analytic signal can be recovered in the manner illustrated in Fig. 11(c) utilizing a phase splitter. The imaginary part of the analytic signal is simply the output of a 90° phase shift network [transfer function $-j\text{sgn}(\omega)$], also called a Hilbert transform filter. To see this, note that the Fourier transform of the complex-valued signal generated in Fig. 11(c) is

$$S(\omega) + \text{sgn}(\omega)S(\omega) = \begin{cases} 2S(\omega) & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (3.13)$$

and this signal has only positive frequency components.

In practice, the phase splitter consists of two filters which are all-pass (have unity magnitude response) and a phase difference of 90°. In addition, it would be common to implement the phase splitter in discrete time.

The passband echo canceller is similar to the baseband case except that a complex-valued data sequence and the complex-valued output of the phase splitter are input to the canceller rather than a real-valued sequence, and the filter coefficients are also complex-valued [8]. Consider the response of the echo transfer function [denoted by $H_e(\omega)$ as in the baseband case] to the transmitted data signal given by (3.9). In fact, since the analytic signal of this echo response will be generated by the phase splitter at the receiver input, what is really desired is the analytic signal of this echo signal. Examining Fig. 12, if the response of the echo filter to the transmitted signal is denoted $r(t)$ as shown in Fig. 12(a), and the corresponding analytic signal is $r_a(t)$, then as shown in Fig. 12(b) $r_a(t)$ is the response of the echo filter to the transmitted analytic signal $s_a(t)$. This is because the impulse response of the echo filter is real-valued. Determining the response to the analytic transmitted signal is simple if the response to a single isolated analytic pulse can be determined, as illustrated in Fig. 12(c). The output is expressed in terms of another analytic signal where $h(t)$ is the equivalent baseband pulse which returns through the echo path. Taking the Fourier transform of the input and output of the echo path filter in Fig. 12(c),
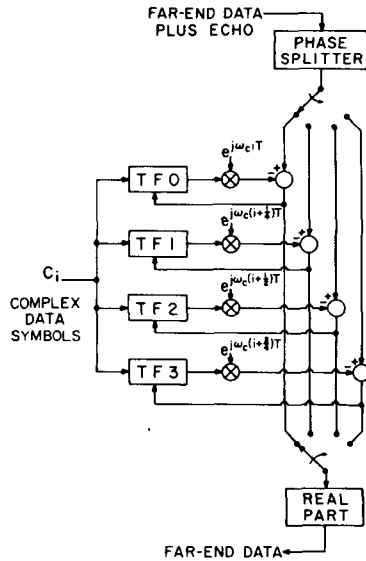
Fig. 13. Interleaved complex baseband echo cancellers in a passband data transmission system ($R = 4$).

$$G(\omega - \omega_c)H_e(\omega) = H(\omega - \omega_c) \qquad (3.14)$$

or equivalently,

$$H(\omega) = G(\omega)H_e(\omega + \omega_c). \qquad (3.15)$$

This demonstrates that the baseband output pulse can be obtained by shifting the echo transfer function in the vicinity of the carrier frequency down to dc, as might be expected. It also follows from (3.15) that $h(t)$ is complex-valued in general, even though $g(t)$ is real-valued.

The analytic echo signal follows from superposition as

$$r(t) = \sum_m C_m h(t - mT) e^{j\omega_c t} \qquad (3.16)$$

very much in analogy to (3.2). As in the baseband case, the echo canceller must be implemented in discrete time, and in order to be able to reconstruct the echo signal the sampling rate must be an integer $R$ times as high as the data symbol rate, where $R$ is an appropriate integer on the order of 2 or 4. Defining $r_{a,i}(l)$ and $h_i(l)$ as in (3.3) and (3.4), a relation similar to (3.6) is obtained,

$$r_{a,i}(l) = \left( \sum_m C_m h_{i-m}(l) \right) e^{j\omega_c(i+(l/R))T}. \qquad (3.17)$$

A transversal filter FIR approximation to this echo becomes

$$r_{a,i} = \left( \sum_{m=0}^{n-1} a_m C_{i-m} \right) e^{j\omega_c(i+(l/R))T} \qquad (3.18)$$

where the $l$ has again been suppressed and the filter coefficients $a_m$ are complex-valued.

An illustration of this approach for $R = 4$ is shown in Fig. 13. The same complex data symbols $C_i$ are applied to each of the transversal filters. These transversal filters are operating at baseband, have complex-valued filter coefficients, and the outputs are modulated up to the carrier

frequency $\omega_c$ prior to cancellation at passband. The phase of the modulation carriers differs for each of the interleaved cancellers. The output of the hybrid, containing the undesired echo signal, is first converted to the analytic signal with a phase splitter, and then decimated by 4. The four interleaved echo cancellations are performed, and then the signals are recombined into a single canceller error signal containing the far-end data signal. The real value of this signal is taken to yield a passband representation of the far-end data signal. Typically, this signal would be applied to an interpolation filter and then resampled synchronously with the far-end data since it is desired that the data signals in the two directions be asynchronous.

## IV. Design Considerations in Echo Cancellation

The design of an echo canceller involves many considerations, such as the speed of adaptation, the effect of near- and far-end signals, the impact of signal levels and spectra, and the impact of nonlinearity. Many of these considerations are outlined in this section [8]-[11].

### A. Adaptation Algorithm

There are two important measures of performance for an echo canceller adaptation algorithm: the speed of adaptation and the accuracy of the cancellation after adaptation. Generally there is a tradeoff between these two measures: for a particular class of adaptation algorithm, as the speed of adaptation is increased the accuracy of the transfer function after adaptation becomes poorer. This tradeoff is fundamental since a longer averaging time is necessary to increase asymptotic accuracy, but slows the rate of convergence.

Most often the motivation for adapting an echo canceller is that the transfer function of the echo is not known in advance. It is also probable that the echo transfer function is changing with time, although in most cases the change will be quite slow (say in response to temperature changes of the transmission facilities). In most instances the accuracy of the final cancellation of the echo is a critical design factor. Although the ability of the canceller to rapidly track a changing echo response is usually not important, the speed of initial adaptation to the echo response from an arbitrary initial condition is often important. For a voice echo canceller, where the canceller is not dedicated to a particular subscriber, the adaptation must occur on each new call, and during the adaptation period the voice quality is degraded. For a data echo canceller, the adaptation is a part of the training or initialization period during which useful data transmission cannot occur, and there is, therefore, motivation to speed adaptation. However, speed of adaptation in echo cancellers is usually not so critical that the simple stochastic gradient algorithm cannot give adequate performance. Nevertheless, some approaches to speeding up the adaptation are discussed in Section V.

It is convenient to define a vector notation for the filter coefficients

$$a^T \equiv [a_0, a_1, \cdots, a_{n-1}], \tag{4.1}$$

where $T$ denotes transpose, and the reference input samples,

$$y_i^T \equiv [y_i, y_{i-1}, \cdots, y_{i-n+1}]. \tag{4.2}$$

If the impulse response of the echo channel is $h_i, 0 \leqslant k < \infty$, then

$$h^T = [h_0, h_1, \cdots, h_{n-1}] \tag{4.3}$$

is defined as a vector of the first $n$ of these samples. The error signal for a voice echo canceller as shown in Fig. 2 can be written as

$$\begin{aligned} e_k &= \sum_{m=0}^{\infty} h_m y_{k-m} - \sum_{m=0}^{n-1} a_m y_{k-m} + x_k \\ &= (h - a) y_k + v_k \end{aligned} \tag{4.4}$$

where

$$v_k = \sum_{m=n}^{\infty} h_m y_{k-m} + x_k \tag{4.5}$$

is the residual uncancellable echo corresponding to echo delays which exceed the number of coefficients in the transversal filter, plus the noise and near-end talker signal. For the baseband data echo canceller, (4.4) is still valid where the reference data symbols $C_{i-m}$ replace $y_{i-m}$. It should be noted that the two terms in (4.4) are correlated when the reference signal samples are correlated.

For purposes of analyzing a given adaptation algorithm, it is often necessary to assume that the reference signal $y_i$ and near-end talker $v_i$ are jointly wide-sense stationary. For this case, define

$$p = E[v_i y_i] \tag{4.6}$$

$$\Phi = E[y_i y_i^T] = \begin{bmatrix} R_0 & R_1 & \cdots & R_{n-1} \\ R_1 & & \ddots & \vdots \\ \vdots & & & \vdots \\ R_{n-1} & & \cdots & R_0 \end{bmatrix} \tag{4.7}$$

where

$$R_j = E[y_i y_{i+j}] \tag{4.8}$$

are the autocorrelation coefficients of the reference process. It is readily shown that the filter coefficient vector which minimizes the mean-square error $E[e_i^2]$ under this assumption is

$$a_{\text{opt}} = h + \Phi^{-1} p. \tag{4.9}$$

Note that in general the optimum filter coefficient vector is not equal to the first $n$ samples of the echo impulse response. However, if we assume that the reference samples are mutually uncorrelated, and that $x_k$ is uncorrelated with the reference samples, then we get

$$p = 0 \tag{4.10}$$

$$\Phi = R_0 I \tag{4.11}$$

where $I$ is the identity matrix. In this case the optimum filter coefficient vector is equal to the echo impulse response

$$a_{\text{opt}} = h. \tag{4.12}$$

Another condition under which this is true is when the echo impulse response is truncated to $n$ coefficients or less. In practice, $n$ will be chosen sufficiently large for this latter condition to be nearly valid.

The most widely used practical algorithm for adaptation of an echo canceller is the stochastic gradient algorithm (sometimes known as the LMS algorithm). It is based on the following idea. If the ensemble statistics of the reference and near-end talker processes were known, the matrix inversion of (4.9) could be avoided by using an iterative gradient algorithm. In this approach, successive corrections to $a$ are made proportional to the negative of the gradient of $E[e_i^2]$ with respect to $a$. Since this gradient is in the direction of maximum rate of increase of $E[e_i^2]$, subtracting this gradient from $a$ should reduce the error. It can be shown that this algorithm does converge to the minimum of (4.9) as long as the constant multiplying the gradient is small enough.

An expression for the gradient is

$$\nabla_a \{ E[e_i^2] \} = -2 E[e_i y_i]. \tag{4.13}$$

The troublesome part of this expression is the expectation operator since the underlying statistics are not known in practice. The principle behind the stochastic gradient algorithm is to ignore the expectation operator. The quantity which is left, while it is random, has an expected value equal to the desired gradient. Thus, it is an unbiased estimate of the gradient. This "noisy" or "stochastic" gradient is substituted for the actual gradient, resulting in the so-called stochastic gradient algorithm

$$a_i = a_{i-1} - \frac{\beta}{2} \nabla_a [e_i^2] \tag{4.14}$$

or

$$\begin{aligned} a_i &= a_i + \beta e_i y_i \\ &= (I - \beta y_i y_i^T) a_i + \beta (y_i y_i^T h + v_i y_i). \end{aligned} \tag{4.15}$$

The first form of (4.15) is suitable for implementation since all the quantities are known, while the second is used for analysis. The constant $\beta$ is used to control the convergence of the algorithm. In general, making $\beta$ larger speeds convergence, while making $\beta$ smaller reduces the asymptotic error. (It will be found later that there is an optimum $\beta$ from the point of view of convergence rate.)

The nature of adaptation algorithm (4.5) was illustrated in Fig. 3, which is based on (4.15) rewritten for the $j$th component as

$$[a_i]_j = [a_{i-1}]_j + \beta e_i y_{i-j}. \tag{4.17}$$

The convergence properties of this algorithm will be analyzed in two steps in the following two subsections: First the average trajectory of the filter coefficient vector will be considered, and then the fluctuation of the trajectories about this average trajectory will be discussed. In both cases, wide-sense stationarity will be assumed. This is unrealistic for a speech canceller, but nevertheless useful for developing insight.

*1) Average Trajectory of a Coefficient Vector:* If the expectation of the filter coefficient vector in (4.15) is taken,

$$E[a_i] = E[(I - \beta y_i y_i^T)a_{i-1}] + \beta E[y_i y_i^T h + v_i y_i]$$
$$\approx (I - \beta\Phi)Ea_{i-1} + \beta(\Phi h + p). \qquad (4.18)$$

The key approximation that has been made is that the reference signal is uncorrelated with the filter coefficient vector. This approximation, which is necessary to make tractable the stochastic analysis of the gradient algorithm, is fortunately valid for small $\beta$ because of the slow trajectory of $a_i$ which results.

If $a_{opt}$ from (4.9) is subtracted from both sides of (4.18) and if

$$\epsilon_i = E[a_i] - h - \Phi^{-1}p \qquad (4.19)$$

is defined as the error between the actual and optimal coefficient vector, then an iterative equation for the error results,

$$\epsilon_{i+1} = (I - \beta\Phi)\epsilon_i. \qquad (4.20)$$

Iterating this equation,

$$\epsilon_i = (I - \beta\Phi)^i \epsilon_0. \qquad (4.21)$$

In addressing the question of whether this error approaches zero, several important properties of the $\Phi$ matrix are critical. Specifically, it is symmetric, Toeplitz, and positive definite. The latter implies that it has positive real eigenvalues and is invertible. This inverse $\Phi^{-1}$ is also a symmetric matrix. $\Phi$ can be written in the form

$$\Phi = V\Lambda V^T \qquad (4.22)$$

where $\Lambda$ is a diagonal matrix of eigenvalues of $\Phi$

$$\Lambda = \text{diag}[\lambda_1, \cdots, \lambda_n] \qquad (4.23)$$

and $V$ is an orthonormal matrix,

$$VV^T = I \qquad (4.24)$$

whose $j$th column is the eigenvector of $\Phi$ associated with the $j$th eigenvalue. Since $\Phi$ is assumed to be positive definite, the eigenvalues are positive real-valued.

The matrix in (4.21) can be written in the form

$$(I - \beta\Phi)^i = (VV^T - \beta V\Lambda V^T)^i$$
$$= (V(I - \beta\Lambda)V^T)^i$$
$$= V(I - \beta\Lambda)^i V^T. \qquad (4.25)$$

Thus, the vector $\epsilon_i$ obeys a trajectory which is the sum of $n$ modes, the $j$th of which is proportional to $(1 - \beta\lambda_j)^i$. Assume for simplicity also that all the eigenvalues are distinct, and order them for smallest $\lambda_{min}$ to largest $\lambda_{max}$. It follows that $\epsilon(j)$ decays exponentially to zero as long as

$$0 < \beta < \frac{2}{\lambda_{max}} \qquad (4.26)$$

which is the condition for each of the factors to be less than unity. Convergence of the average trajectory is thus assured if $\beta$ is sufficiently small. Conversely, an excessively large value for $\beta$ leads to instability in the form of an exponentially growing error.

For a fixed $\beta$, the speed of convergence of the algorithm is determined by the smallest eigenvalue. Specifically, the slowest converging mode is proportional to $(1 - \beta\lambda_{min})^i$. If $\beta$ is chosen to be near its maximum value, say

$$\beta = \frac{1}{\lambda_{max}} \qquad (4.27)$$

then the slowest converging mode is proportional to

$$\left(1 - \frac{\lambda_{min}}{\lambda_{max}}\right)^i. \qquad (4.28)$$

The ratio of largest to smallest eigenvalue is thus seen to be of fundamental importance; it is called the *eigenvalue spread*. The eigenvalue spread has a minimum value of unity, and can be arbitrarily large. The larger the eigenvalue spread of the autocorrelation matrix, the slower the convergence of the filter coefficients.

It is instructive to relate the eigenvalue spread to the power spectral density of the reference random process. It is a classical result of Toeplitz form theory that

$$\min_\omega S(\omega) < \lambda_j < \max_\omega S(\omega) \qquad (4.29)$$

where $S(\omega)$ is the power spectral density. While the eigenvalues depend on the order of the matrix $n$ as $n \to \infty$ the maximum eigenvalue

$$\lambda_{max} \to \max_\omega S(\omega) \qquad (4.30)$$

and the minimum eigenvalue

$$\lambda_{min} \to \min_\omega S(\omega). \qquad (4.31)$$

It follows that spectra of the reference signal which result in slow convergence of the average trajectory are those for which the ratio of the maximum to minimum spectrum is large, and spectra which are almost flat (have an eigenvalue spread near unity) result in fast convergence.

Intuitively, the reason for slow convergence of the average trajectory for some spectra is the interaction among the adaptation of the different coefficients. As the successive reference samples become more correlated, there is more undesirable interaction among the coefficient adaptations.

Since the modes of convergence of the stochastic gradi-

ent algorithm are all of the form of $\gamma^i$ where $\gamma$ is a positive real number less than unity and $i$ is the iteration number, the error in decibels can be determined by taking the logarithm of the square,

$$10\log_{10}(\gamma^{2i}) = (10\log_{10}(\gamma^2))i \qquad (4.32)$$

and thus the error expressed in decibels decreases linearly with iteration number (the constant factors multiplying these exponentially decaying terms give a constant factor in decibels). The convergence of a gradient algorithm is thus often expressed in units of dB/s., which is the number of decibels of decrease in the error power per second.

Unfortunately, the convergence of the mean filter coefficient vector does not mean that any particular coefficient vector trajectory itself converges to the optimum, but only that the average of all trajectories converges to the optimum. In fact, the coefficient vector does not converge to the optimum. To see this, observe that even after convergence of the coefficient vector in the mean-value sense, difference equation (4.15) still has a stochastic driving term, and therefore the coefficient vector continues to fluctuate about the optimum coefficient vector randomly. The larger the value of the step size $\beta$, the larger this fluctuation.

It is therefore appropriate to study the fluctuation of the filter coefficient vector about its mean after convergence. This is done in the next section.

*2) Fluctuation of Trajectories About the Average:* The last section studied the convergence of the average filter coefficient trajectory; this section studies the fluctuation of a given trajectory about this average. To greatly simplify this analysis it will be assumed that the reference signal samples are uncorrelated (the reference signal is white). This assumption is not valid for speech signals, but the analysis nevertheless gives useful insight. This assumption is usually valid for a data echo canceller.

Calculating the norm of the difference between the filter coefficient vector and the optimum vector (which is $h$ for this case), taking the expectation, and assuming that $v_i$ is zero-mean and uncorrelated with $y_i$, and that the reference signal and the filter coefficient vector are uncorrelated,

$$E|a_i - h|^2 = E\left[(a_{i-1} - h)^T E(I - \beta y_i y_i^T)^2 (a_{i-1} - h)\right] + \beta^2 \sigma_v^2 nR_0. \qquad (4.33)$$

The final quantity to calculate,

$$E(I - \beta y_i y_i^T)^2 = I - 2\beta R_0 I + \beta^2 E\left[(y_i^T y_i)(y_i y_i^T)\right] \qquad (4.34)$$

depends on fourth-order statistics of the reference signal. This fourth moment can be evaluated simply for two cases: the reference signal white and Gaussian, and the reference signal a binary-valued independent identically distributed sequence of data symbols. For each case, the answer is

$$E\left[(y_i^T y_i)(y_i y_i^T)\right] = nR_0^2 I \qquad (4.35)$$

(for the Gaussian case $n$ is replaced by $n + 2$, which is nearly the same for large $n$). Substituting (4.35) back into (4.33),

$$E|a_i - h|^2 = (1 - 2\beta R_0 + \beta^2 nR_0^2)E|a_{i-1} - h|^2 + \beta^2 nR_0 \sigma_v^2. \qquad (4.36)$$

This simple recursion shows that for small $\beta$ the mean-square filter coefficient error approaches an asymptotic value of

$$E|a_i - h|^2 \rightarrow \frac{\beta n}{2 - \beta nR_0}\sigma_v^2 \qquad (4.37)$$

or the rms error per coefficient is

$$\frac{E|a_i - h|^2}{n} \approx \frac{\beta \sigma_v^2}{2}. \qquad (4.38)$$

This illustrates that the filter coefficient vector can be made more accurate asymptotically by reducing $\beta$, which also slows adaptation, or by keeping $\sigma_v^2$ small during periods of adaptation. The latter is accomplished for a voice canceller by disabling adaptation during periods of significant near-end talker energy.

The error approaches the asymptote of (4.37) as

$$(1 - 2\beta R_0 + \beta^2 nR_0^2)^i \qquad (4.39)$$

and hence only approaches this asymptote if the quantity in parentheses is less than unity in magnitude. This in turn requires that

$$0 < \beta < \frac{2}{nR_0}. \qquad (4.40)$$

It is instructive to define the time constant of the convergence of the mean-square filter coefficient vector error as the $\tau$ such that

$$(1 - 2\beta R_0 + \beta^2 nR_0^2)^\tau = \frac{1}{e}. \qquad (4.41)$$

When $\beta$ is very small, this can be solved for $\tau$ as

$$\tau \approx \frac{1}{2\beta R_0} \qquad (4.42)$$

iterations.

In comparing (4.40) to (4.26), note that for the simpler white reference signal case considered here, all the eigenvalues of $\Phi$ are equal to $R_0$. Hence, (4.40) is considerably more stringent than (4.26), particularly for large $n$. Further, while (4.25) implies that $\beta$ should be chosen as large as possible for fastest convergence (subject to stability constraints), (4.39) has an optimum $\beta$ for fastest convergence given by

$$\beta = \frac{1}{nR_0} \qquad (4.43)$$

(half the largest $\beta$), with a resultant convergence as

$$\left(1-\frac{1}{n}\right)^i \tag{4.44}$$

or a time constant of

$$\tau \approx \frac{n}{2}. \tag{4.45}$$

The preceding gives the mean-square error in the filter coefficient. This quantity can also be related to the mean-square echo residual. The latter is given by

$$E\left[e_i^2\right] = R_0 E|a_i - h|^2 + \sigma_v^2 \tag{4.46}$$

assuming again that $p = 0$ and $\Phi = R_0 I$. The second term is the minimum mean-square error which would be possible if the correct constant filter coefficient vector was used. The first term therefore represents an "excess" mean-square error which is due to the adaptation algorithm. When the step size of (4.43) which results in maximum convergence rate is used, (4.46) becomes

$$E\left[e_i^2\right] = 2\sigma_v^2 \tag{4.47}$$

and the excess mean-square error is equal to the variance of the uncancellable signal.

Several properties of this convergence should be noted. First, the $\beta$ which should be chosen to keep the excess mean-square error small is considerably smaller than would be predicted by (4.26), which relates only to the convergence of the mean-value of the filter coefficient vector. Second, the fastest convergence of the mean-square error depends on the number of coefficients in the echo canceller, as the number of coefficients $n$ increases, the fastest convergence slows down. Finally, the optimum step size for fastest convergence in (4.43) is inversely proportional to the reference signal power, $R_0$. Thus, to properly choose $\beta$, $R_0$ should be known, and if $R_0$ is actually varying slowly with time (as in speech cancellers), best performance requires that $\beta$ track this variation. This is considered in Section IV-C.

### B. Deterministic Theory of Convergence

There are several shortcomings of the convergence analysis just performed. The assumption of wide-sense stationarity is invalid for speech, and the approximation of uncorrelated coefficient vector and reference signal is of doubtful validity as $\beta$ gets large.

There is also a fundamental question of what happens when the input signal does not cover the whole band up to half the sampling rate (an extreme case would occur when the reference signal is a sinusoid). Then from (4.31) the minimum eigenvalue approaches zero and the eigenvalue spread approaches infinity. Does this imply that the average coefficient trajectory does not approach the echo impulse response? Indeed it does since there are many coefficient vectors which will result in complete cancellation under this condition, each of these vectors having a different transfer function in the frequency band where there is no reference signal power.

There is an important deterministic theory of canceller adaptation which is able to rigorously derive upper and lower bounds on coefficient vector error for a given input reference signal given waveform [1], [12]. This theory gives qualitatively similar results to the stochastic analysis previously described, but it is very comforting to be able to state conditions under which convergence is guaranteed. In particular, this theory assumes a "mixing condition" to be satisfied in order for convergence of the coefficient vector to the region of the actual echo impulse response. This mixing condition is analogous to the reference signal having power over the entire band up to half the sampling rate.

### C. Modifications to the Adaptation Algorithm

There are several useful modifications which can be made to the adaptation algorithm to improve performance or simplify implementation which are described in this section.

A useful modification which is often made is to replace the error signal $e_i$ by the sign of this signal $\text{sgn}(e_i)$. The motivation for doing this is simplification of the hardware: the adaptation equation of (4.17) becomes

$$[a_i]_j = [a_{i-1}]_j + \beta \text{sgn}(e_i) y_{i-j} \tag{4.48}$$

where the only multiplication is by $\beta$ or $-\beta$. This limited multiplication is particularly simple to implement if $\beta$ is chosen to be a power of $2^{-1}$ in a digital implementation.

One would expect intuitively that the value of the cross correlation would not be affected in a major way by using the sign of the error. It has been shown [13] that the convergence of the algorithm is not compromised by this simplification, but that the speed of adaptation is reduced somewhat. In fact, it is shown more generally that any nonlinearity in the multiplication adversely affects convergence to some extent.

Another common modification to the stochastic gradient algorithm is to normalize the step size to eliminate an undesirable dependence of speed of convergence on input signal power. This dependence can be seen from (4.42) where the time constant of adaptation is inversely proportional to $R_0$. Further, if $\beta$ is kept constant and the signal power is increased, eventually from (4.40) the algorithm becomes unstable. This is particularly a problem in a speech canceller where the input signal power varies considerably. From (4.43), it is desirable to normalize $\beta$ by an estimate of the reference signal power, as in

$$\beta_i = \frac{a}{\sigma_i^2 + b} \tag{4.49}$$

where $\beta_i$ is the step size at sample $i$, $a$, and $b$ are some appropriately chosen constants, and $\sigma_i^2$ is an estimate of the reference signal power at time $i$. The purpose of the $b$ in the denominator is to prevent $\beta_i$ from getting too large (causing instability) when the input signal power gets very small, and is obviated if adaptation is disabled under this condition.

Another modification to the stochastic gradient algorithm is required for the adaptation of a passband canceller [8]. For this case, an algorithm analogous to (4.15) is

$$a_i = a_{i-1} + \beta e_i C_i^* e^{-j\omega_c(i+(l/R))T} \qquad (4.50)$$

where $e_i$ is the analytic cancellation error and $C_i^*$ is the conjugate of a vector of complex data symbols.

## V. RECENT ADVANCES IN ECHO CANCELLATION

The major portion of recent work in echo cancellation has focused on three areas: the extension to nonlinear echo models, speedup of the adaptation, and implementation. This section reviews this recent activity.

### A. Nonlinear Echo Cancellation

The basic echo cancellation model presented earlier was capable, within the constraints of a finite impulse response, of exactly cancelling only an echo which is a linear function of the reference signal. Some older work in speech cancellers and some more recent work in data cancellers has extended the adaptive echo canceller to nonlinear echo generation phenomena [14], [15]. This extension is not critical to speech echo cancellers because the actual echo generation mechanisms, while not necessarily precisely linear, are close enough to linear that a linear canceller is able to meet the cancellation objectives. In data transmission, on the other hand, the objectives for degree of cancellation are sufficiently ambitious that nonlinear echo generation phenomena are of importance. The primary sources of nonlinearity in data cancellers are the data converters in the implementation of the canceller itself, as well as transmitted pulse asymmetry and saturation of transformer magnetic media. While these mechanisms result in a very small degree of nonlinearity, they are nevertheless of importance when the objective is 50–60 dB of cancellation.

The method of extending echo cancellation to nonlinear echo mechanisms that has been proposed is the use of the Volterra expansion. This expansion is capable of representing any echo mechanism which is time invariant. That is, the nonlinearity can have memory, such as in the hysteresis in a magnetic medium, but the nature of the nonlinearity cannot change with time. Of course, since the canceller is adaptive, the nonlinearity can change slowly in time, as long as the adaptation mechanism can keep up. For a data canceller, the number of coefficients required for a Volterra expansion is finite even for perfect cancellation because of the finite number of possible transmitted signals [15]. This assumes as in the linear case that the echo can be represented as a (nonlinear) function of a finite number of past transmitted data symbols.

### B. Speedup of Adaptation

Where faster adaptation than is afforded by the stochastic gradient algorithm is desired, there are several techniques for speeding adaptation.

In almost all echo cancellation applications, high speed of convergence is desirable for initial acquisition, but is not necessary for subsequent tracking since the echo transfer function changes very slowly. Thus, a large constant of adaptation $\beta$ is desirable during acquisition, but is not necessary following reasonable convergence. A natural algorithm to consider is to use a relatively large $\beta$ initially, and then "gear-shift" to a smaller $\beta$ after convergence is achieved.

A more radical means for increasing adaptation rate is to use a lattice filter in place of the transversal filter. As was detailed in Section IV-A-2), the transversal filter algorithm convergence suffers when the reference samples are highly correlated. In effect the lattice filter does an adaptive prefiltering of the reference signal prior to the adaptation to the echo transfer function. This prefilter serves the purpose of whitening the reference signal, thereby speeding the convergence of the canceller.

The extension of the stochastic gradient algorithm to the lattice filter is straightforward. The simplest, but not necessarily best performing, approach will be described here (see [16], [17] for more details). Referring to (2.6), if the derivative of $e_i^2$ with respect to the coefficient $b_m$ is taken, a stochastic gradient algorithm becomes

$$[b_i]_m \doteq [b_{i-1}]_m + \beta e_i e_b(i|m-1) \qquad (5.1)$$

which is identical to the transversal filter algorithm of (4.17) except for the substitution of $e_b(i|m-1)$ for the reference signal $y_{i-m}$. The adaptation of the reflection coefficients can proceed by taking the derivative of $e_f^2(i|m)$ with respect to $k_m$, yielding the stochastic gradient algorithm

$$k_m(i) = \left(1 - \beta e_b^2(i|m-1)\right)k_m(i-1) \\ + \beta e_f(i|m-1)e_b(i|m). \qquad (5.2)$$

Note that the adaptation of each lattice stage is based on the input signals to that stage. Therefore, the algorithm converges sequentially by stage: the adaptation of the first stage is necessary before the adaptation of the second state, etc. Since each stage requires only a scalar adaptation, the speed of adaptation is essentially independent of eigenvalue spread. The adaptation serves to minimize the forward prediction error, which it turns out is equivalent to making the successive orders of backward prediction error uncorrelated. The latter in turn speeds the adaptation of the $b_m$ coefficients since it eliminates the interaction among those coefficient adaptations. Experience has shown that for input signals with a large eigenvalue spread, the overall adaptation of the lattice filter is considerably faster than the adaptation of the transversal filter.

For the data canceller, the data symbols are usually approximately uncorrelated. This is particularly true in the presence of scrambling, which "breaks up" probable highly correlated patterns such as marking sequences. While the primary purpose of scrambling is to aid timing recovery, it is also of benefit to canceller adaptation. Thus, for data cancellers, there is little benefit to using the lattice filter.

For speech cancellers, on the other hand, the signal samples are usually highly correlated. Unfortunately, however, the exact nature of the correlation changes fairly rapidly with time. The reflection coefficients of the lattice filter therefore have to continually readapt to the spectrum of the reference speech signal. As the reflection coefficients adapt, the $b_m$ coefficients also have to readapt to maintain the same overall transfer function. This illustrates a disadvantage of the lattice filter for echo cancellation applications: while the transversal filter coefficients are largely independent of the reference signal statistics, and depend mostly on the impulse response of the echo channel, the lattice filter coefficients depend strongly on the reference signal statistics and must continually readapt as those statistics change.

In cases where the lattice filter is of little benefit or even faster convergence is desired, another class of least-square (LS) adaptation algorithms closely related to the Kalman filter in control theory can be used. These algorithms are based on the minimization of the least-square cancellation error over the choice of adaptive filter parameters, with a weighting function decreasing exponentially into the past to give the algorithm finite memory. A simple example is appropriate for illustration. Define a squared error at time $i$ as

$$\epsilon_i = \sum_{j=0}^{i} \alpha^j \left( r_{i-j} - a_i^T y_{i-j} \right)^2. \tag{5.3}$$

Note that the filter coefficient vector at iteration $i$, $a_i$, is used in forming the error for the past, and the error is minimized by the choice of this vector. The adaptive nature of the solution is enabled by the exponentially decreasing weighting into the past, resulting in greater weight for more recent reference samples. Defining

$$\Phi_i = \sum_{j=0}^{i} \alpha^j y_{i-j} y_{i-j}^T$$

$$= \alpha \Phi_{i-1} + y_i y_i^T \tag{5.4}$$

which is an estimate of the reference signal correlation matrix (4.7), and also defining the cancellation error

$$e_i = r_i - a_{i-1}^T y_i \tag{5.5}$$

then it is readily shown that the vector $a_i$ which minimizes (5.3) can be defined iteratively as

$$a_i = a_{i-1} + \Phi_i^{-1} e_i y_i. \tag{5.6}$$

Comparing this to the stochastic gradient algorithm of (4.15), the only difference is the replacement of $\beta$ with $\Phi_i^{-1}$. This replacement, which substantially increases the complexity of the algorithm, also speeds up adaptation substantially for reference signals with a large eigenvalue spread. This matrix adaptation constant can be thought of again as uncoupling the adaptation of the individual components of the filter coefficient vector.

There are many versions of the LS algorithms, including
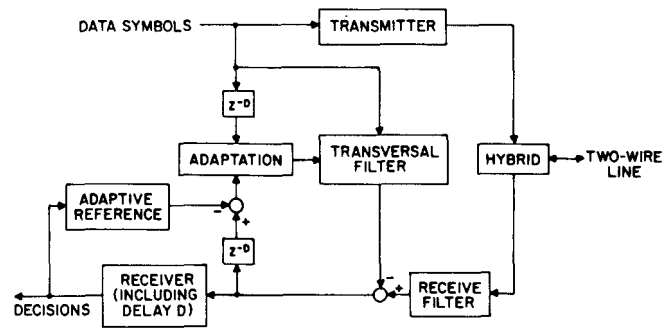


Fig. 14. Baseband adaptive reference echo canceller.

those based on both a transversal filter [18] and lattice filter realization [19]. Neglecting finite precision effects, the performance is identical since the quantity being minimized in each case is identical. There are also versions of the LS algorithms which have much reduced computational requirements relative to the simple algorithm just displayed [19].

The LS algorithms are particularly effective in speeding adaptation for the data canceller case where the data symbols can be chosen during a training period to assist in the canceller adaptation. In this case, the reference signal algebraic properties become much more important than the stochastic properties, and it has been shown that the mean values of the filter coefficients of a canceller based on least squares can adapt in $n$ data symbols for an $n$-coefficient canceller [20]. Furthermore, it has been shown that the least-squares algorithm can be virtually as simple as the stochastic gradient algorithm for a reference signal which is chosen to be a pseudorandom sequence [21]. This sequence is also particularly simple to generate.

In data echo cancellation, a significant factor slowing adaptation is the far-end data signal. This suggests yet another way of speeding adaptation: the data signal can be adaptively removed from the cancellation error in a decision-directed fashion [22] in an approach called an "adaptive reference" canceller. This is illustrated in Fig. 14 where the adaptive reference is attached to the output of the receiver. It forms a linear combination of the current and past receiver decisions to form a replica of the far-end data signal appearing in the cancellation error signal. This replica is subtracted to yield a new error signal which drives the echo canceller adaptation algorithm. Because the far-end data signal has been removed from the error, $\sigma_v^2$ in (4.37) is made smaller and the adaptation constant $\beta$ can be chosen to be larger for a given excess mean-square cancellation error. This in turn speeds adaptation.

Because the receiver decision circuit typically includes delay, a compensating delay is inserted in the other inputs to the canceller adaptation algorithm. In effect the entire adaptation operates on a delayed basis.

Of course there is a problem getting started, when the cancellation may not be adequate to support a reasonable receiver error rate. This problem is solved by either starting out with a known training sequence (not requiring receiver decisions), or by disabling the adaptive reference and using
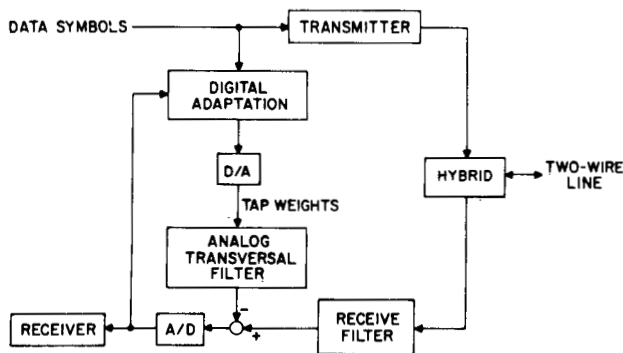
Fig. 15.   Implementation of baseband data echo canceller insensitive to data converter nonlinearity.



Fig. 16.   Baseband data nonlinear echo canceller.

a larger $\beta$ initially until the echo is cancelled sufficiently for a reasonable error rate.

### C. Implementation of Echo Cancellation

The widespread application of echo cancellation has recently been stimulated by advances in microelectronics, making the computational requirements of the canceller within the reach of an inexpensive single-ship implementation. Because the potential applications of echo cancellation are numerous, there has been considerable activity in the design of devices for echo cancellation.

The first special-purpose chip for speech cancellation was developed by Bell Laboratories [23]. This device implemented a single 128 coefficient canceller using digital techniques, with an interface to the standard $\mu$-255 format widely used in the digital transmission network. Cleverly, advantage was taken of the nearly logarithmic nature of this companding law to simplify the multiplications required in the implementation of the transversal filter. A later version of this device added the capability to cascade chips to achieve up to four times the number of coefficients [24].

The implementation of a data echo canceller in monolithic form represents special challenges [7], [10]. The major difficulty is in achieving the required accuracy without incorporating expensive off-chip precision components or trimming during manufacture. The source of the problem is that the adaptation must be implemented digitally to achieve the long time constants required for high asymptotic accuracy, but the interface to the transmission medium is inherently analog. This implies the need for high-speed analog-to-digital conversion somewhere between the medium and the adaptation, and this conversion must have an accuracy on the order of 12 bits with perfect linearity. While this accuracy can be achieved with trimming, without trimming special measures must be taken to overcome the nonlinearity of the data converters.

Two solutions have been proposed to this problem. One solution [7] is shown in Fig. 15, in which the transversal filter is implemented in analog circuitry where good linearity can be insured. The adaptation circuitry is digital, ensuring that the long time constants which are necessa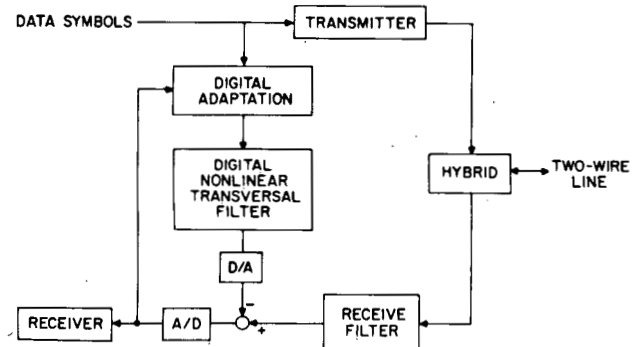ry can be achieved. The D/A conversion occurs at the output of the adaptation circuitry where the filter coefficients are converted to analog. In this case, a mild monotonic nonlinearity can be compensated by the adaptation algorithm.

The second solution [15] is to use the nonlinear cancellation algorithm described earlier. This approach is illustrated in Fig. 16 where the transversal filter is implemented digitally. The nonlinearity in the D/A converter in this case is compensated by extra nonlinear taps in the canceller. This method has the advantage that sources of nonlinearity external to the canceller, such as transmitted pulse asymmetry and transformer saturation, can also be compensated. Finally, the two solutions can be combined: extra nonlinear taps can be added to the analog transversal filter in Fig. 15 for the compensation of any external nonlinearities.

## VI. Conclusions

This paper has described the applications, algorithms, and technology of echo cancellation. Due to advances in microelectronics, which make echo cancellation more economical, broad application of these techniques can be expected. New applications of echo cancellation, such as the elimination of acoustic reverberation, can also be expected.

### Acknowledgment

The author is indebted to many colleagues who have contributed to his understanding of echo cancellation, and particularly Dr. D. Duttweiler, Dr. O. Agazzi, D. Falconer, S. Weinstein, and D. Hodges.

### References

[1] M. Sondhi and D. A. Berkley, "Silencing echos on the telephone network," *Proc. IEEE*, vol. 68, Aug. 1980.
[2] H. G. Suyderhoud, S. J. Campanella, and M. Onufry, "Results and analysis of a worldwide echo cancellation field trial," *COMSAT Tech. Rev.*, vol. 5, p. 253, Fall 1975.
[3] A. H. Gray and J. D. Markel, "Digital lattice and ladder filter synthesis," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, p. 491, 1973.
[4] J. Makhoul, "A class of all-zero lattice digital filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, p. 304, Aug. 1978.
[5] S. V. Ahamed, P. P. Bohn, and N. L. Gottfried, "A tutorial on two-wire digital transmission in the loop plant," *IEEE Trans. Commun.*, vol. COM-29, Nov. 1981.

[6] K. H. Mueller, "A new digital echo canceler for two-wire full-duplex data transmission," *IEEE Trans. Commun.*, vol. COM-24, Sept. 1976.

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

[8] S. B. Weinstein, "A passband data-driven echo canceller for full-duplex transmission on two-wire circuits," *IEEE Trans. Commun.*, vol. COM-25, July 1977.

[9] B. Widrow et al., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, Aug. 1976.

[10] N. A. M. Verhoeckx et al., "Digital echo cancellation for baseband data transmission," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, Dec. 1979.

[11] D. L. Duttweiler, "A twelve-channel digital echo canceller," *IEEE Trans. Commun.*, vol. COM-26, May 1978.

[12] A. Weiss and D. Mitra, "Digital adaptive filters: conditions for convergence, rates of convergence, effects of noise and errors arising from the implementation," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 637–652, Nov. 1979.

[13] D. L. Duttweiler, "Adaptive filter performance with nonlinearities in the correlation multiplier," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, p. 578, Aug. 1982.

[14] E. J. Thomas, "An adaptive echo canceller in a nonideal environment (nonlinear or time variant)," *Bell Syst. Tech. J.*, vol. 50, pp. 2779–2795, Oct. 1971.

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

[16] J. Makhoul, "Stable and efficient lattice methods for linear prediction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, p. 423, Oct. 1977.

[17] E. H. Satorius and S. T. Alexander, "Channel equalization using adaptive lattice algorithms," *IEEE Trans. Commun.*, vol. COM-27, p. 899, June 1979.

[18] D. D. Falconer and L. Ljung, "Application of fast Kalman estimation to adaptive equalization," *IEEE Trans. Commun.*, vol. COM-26, p. 1439, Oct. 1976.

[19] B. Friedlander, "Lattice filters for adaptive processing," *Proc. IEEE*, vol. 70, p. 829, Aug. 1982.

[20] M. S. Mueller, "On the rapid initial convergence of least-squares equalizer adjustment algorithms," *Bell Syst. Tech. J.*, vol. 60, pp. 2345–2358, Dec. 1981.

[21] J. Salz, "On the start-up problem in digital echo cancelers," *Bell Syst. Tech. J.*, vol. 62, p. 1353, July–Aug. 1983.

[22] D. D. Falconer, "Adaptive reference echo cancellation," *IEEE Trans. Commun.*, vol. COM-30, p. 2083, Sept. 1982.

[23] D. L. Duttweiler and Y. S. Chen, "A single-chip VLSI echo canceler," *Bell Syst. Tech. J.*, vol. 59, pp. 149–160, Feb. 1980.

[24] Y. G. Tao, K. D. Kolwicz, C. W. K. Gritton, and D. L. Duttweiler, "A cascadable VLSI echo canceller," this issue, pp. 297–303.

David G. Messerschmitt (S'65–M'68–SM'78–F'82), for a photograph and biography, see this issue p. 281.

# A Cascadable VLSI Echo Canceller

YING G. TAO, KEVIN D. KOLWICZ, C. W. K. GRITTON, AND DONALD L. DUTTWEILER, SENIOR MEMBER, IEEE

*Abstract* — A new single-chip echo canceller has been developed. In addition to offering improved performance over an earlier single-chip echo canceller [14], the new canceller has many desirable features not available before. The most important of these features is that multiples of the chip can be connected in cascade to synthesize very long impulse responses. This paper describes the new features implemented on the chip and characterizes its performance.

## I. INTRODUCTION

TWENTY years ago when echo cancelling was conceived [1], [2], it was little more than a mathematical curiosity. Technological progress made possible first prototypes for field experiment [3]–[8], then designs seriously

proposed for manufacture [9]–[12], and eventually a single-chip canceller [13], [14]. There are now numerous cancellers being marketed, some based on single-chip realizations [15], others mixtures of VLSI and catalog logic, and still others with only catalog logic.

Our own interest is in VLSI realizations. In this paper we describe a new chip that offers new features and improved performance over an earlier VLSI chip of ours. The most important of the features on this second generation chip is that multiples can be connected in cascade to synthesize very long impulse responses. A programmable transversal filter which is cascadable has been announced (Western Digital WD3150) and the idea of cascading an adaptive transversal filter has occurred to others as well [16], [17] but the new chip will be the first fully adaptive transversal filter which is cascadable.

The new echo canceller chip is realized in 3.5 $\mu$ NMOS technology with an area of $344.5 \times 375.5$ mil$^2$. It is packaged