Freescale Semiconductor
Application Note

AN2598
Rev. 1, 11/2004

Network Echo Cancellers and Freescale Solutions Using the StarCore™ SC140 Core

by Roman A. Dyba, Perry P. He, and Lúcio F. C. Pessoa

This application note surveys selected echo cancellation techniques for use in TDM telephony/packet telephony applications. An echo canceller is a system encompassing many interconnected sub-blocks incorporating DSP algorithms and control operations. The intended audience includes DSP professionals who are designing and/or evaluating solutions for the Public Switching Telephone Network (PSTN), packet telephony, and terminals related to these applications. Field application engineers who are defining and deploying DSP solutions and technical marketing professionals working in related fields can use this document as a reference.

CONTENTS

1 Echo Occurrence in Telephone Networks 2
2 Echo Canceller Architecture 7
3 Network Delays 10
4 Requirements for Echo Cancellers....................... 13
5 Special Network Configurations 18
6 Special Configurations... 20
7 DSP Algorithms In Echo Canceller Applications. 22
8 Miscellaneous Echo Canceller Features 33
9 Freescale Packet Telephony Echo Cancellers....... 34
10 Conclusions... 39
11 References... 40

1 Echo Occurrence in Telephone Networks

Echoes occur in telephone networks because of impedance mismatches of network elements and electro-acoustical coupling within voice terminals (telephone sets). A line/network echo canceller controls/cancels echoes caused by impedance mismatches. **Figure 1** illustrates sources of echoes on the access network side. Echo canceller performance in a telephone network, either a TDM or packet telephony network, has a substantial impact on the overall voice quality, which affects the entire quality of service for voice networks. Therefore, the echo canceller is the single most important element in telephony networks. By developing state-of-the-art echo cancellation solutions, carriers and network operators can safeguard the quality of modern telephone networks.

The (a) section of **Figure 1** illustrates the Line Echo Canceller (LEC) residing in the access network where echoes occur. Other voice processing modules, such as Automatic Level Control (ALC) and Noise Reduction (NR), are shown merely for illustration purposes. On the terminal side, both acoustic and electrical echo can occur. If the voice terminal is a speaker telephone set with an open acoustical path from the speaker to the microphone, the effect of acoustical echo is combined with the effect of the network echo. Because there is only a miniscule delay between the electrical signal and its echo, the speaker/listener is exposed to an effect known as *side tone*. Typically, instead of mounting a LEC at the terminal, side tone circuitry is placed to control the amount of the returned energy. The (b) section of **Figure 1** shows examples of echo canceller signals representing the following:

- Voice from the PSTN/packet network.
- Reflection of voice returning to the PSTN/packet network in the absence of a LEC (or when the LEC is inactive).
- Residual reflection of voice returning to the PSTN/packet network when there is an active LEC.

1.1 Hybrid Circuits

A chief cause of echo in telephone networks is the presence of certain network interface elements, called hybrid circuits, situated between four-wire and two-wire connections. A hybrid circuit converts a four-wire physical interface (with two separate signal paths, one for the transmit direction and another one for the receive direction) into a two-wire physical interface that provides an electrical link for signals traveling in both directions at the same time. Hybrid circuits used to be implemented as analog circuits composed of transformers and other lumped-constant elements. **Figure 2** shows a simplified analog model of a hybrid circuit. Z_B denotes a balancing network representing a compromise non-loaded, loaded, or 900 Ω + 2.16 μF balancing impedance. Z_{TR} represents the impedance of T/R wires terminated with a telephone set in on-hook or off-hook states.

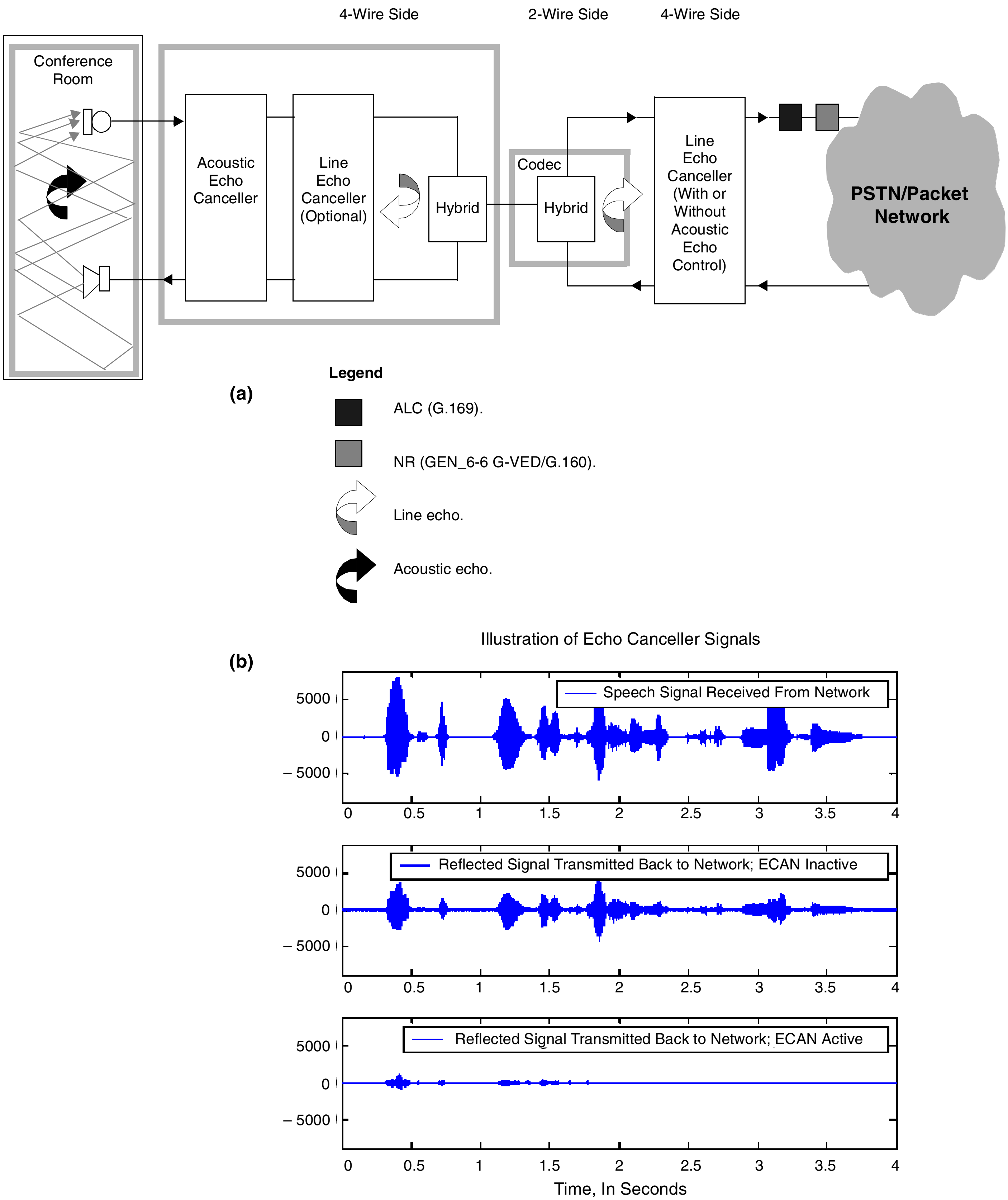


Figure 1. Voice Part of Access Network Where Echoes Can Occur

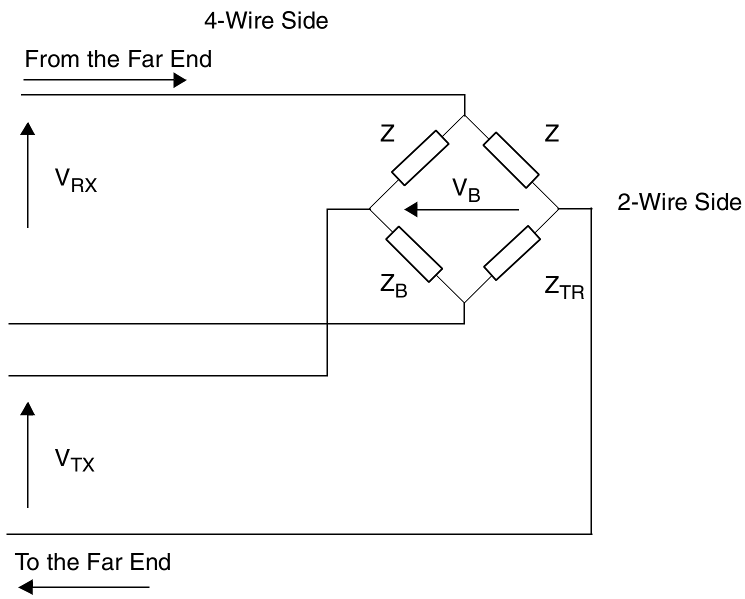


Figure 2. Simplified Analog Model of the Hybrid Circuit

Figure 3 presents an example of practical rollout of the Z_{TR}. The T/R line is represented as a transmission line (using a lumped-constant electrical analogy) terminated with a telephone set impedance Z_{TEL}. For details, refer to Appendix B. Analysis of the hybrid circuit with combined Z_{TR} and Z_{TEL} indicates that many factors contribute to the imbalance of the hybrid bridge, and they all affect the echo characteristics. In **Figure 3**, R_{line}, G_{line}, and C_{line} represent lumped-constant line resistance, conductance, inductance, capacitance, respectively, per a line segment length unit. This representation is an approximation, which is adequate for short lines. For longer T/R lines, more lumped-constant elements must be added to represent a T/R line adequately as a transmission line (see **Appendix B**). Current implementations of the hybrid circuits include mixed-signal (analog and digital) technology. The hybrid circuit is usually part of a G.711-compliant coder-decoder (codec) device, often implemented as an ASIC or as part of a SOC solution. Regardless of the technology used, the basic function of the hybrid remains mostly intact and its analog model (as presented in **Figure 2**) adequately illustrates the mechanism of electrical signal reflection occurring in such a circuit.

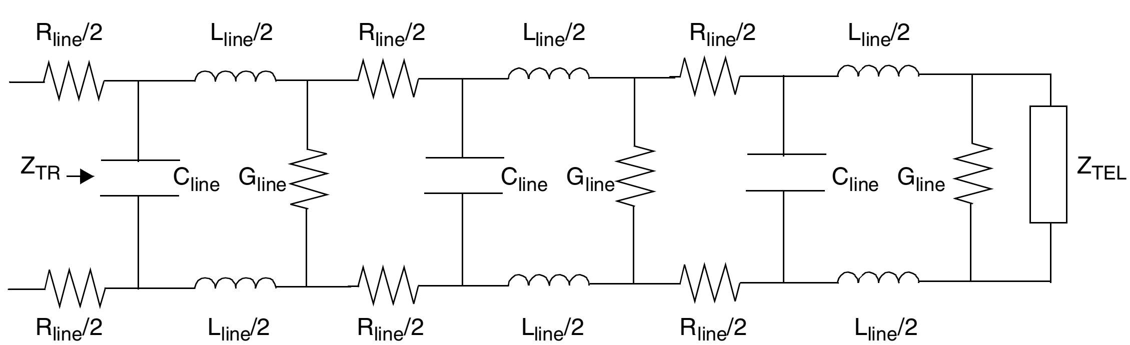


Figure 3. Simple Non-Loaded T/R Line With Terminating Telephone Set Impedance

The echo signal is represented by V_{TX}, as shown in **Figure 2**. When the perfect balance of the hybrid circuit (a form of Wheatstone bridge) is achieved, V_B, which is equal to V_{TX}, has a value of zero. Thus, it implies there is no echo signal. Since one arm of the bridge (Z_{TR}) cannot be accurately determined in advance, the electrical balance is almost never achieved. Even with a careful selection of Z_B, there is a residual lack of balance, which results in a non-zero signal V_{TX}.

To minimize V_{TX}, a set of compromise balance impedances was proposed. Although proper selection of the compromise balance network reduces V_{TX}, it never eliminates it because subscriber loops are transmission lines with equivalent circuits that are more complex than those of compromise balance networks. Also, telephone set equivalent circuits vary. In a line terminated with a couple of telephone sets, these variations contribute to hybrid imbalance. Moreover, the position of a telephone handset is reflected (through electro-acoustical coupling) in impedance Z_{TEL}, which, in turn, can contribute to echo energy and its variations during a telephone call.

1.2 Echo Return Loss (ERL)

- The average V_{TX} in North American network connections, as expressed in a logarithmic scale, is equivalent to 11 dB [21]. The average magnitude of reflection is expressed as echo return loss (ERL), with ERL defined as ERL = 20 log(V_{RXrms}/V_{TXrms}) [10]. Some networks may present much higher ERL, reaching 18 dB or more. However, there are telephone/packetized voice networks with ERL = 8 dB or less. The ITU-T G.168 Recommendation specifies that in no test should the ERL be smaller than 6 dB.
- An analysis of the Z_{TR} (**Figure 3**) indicates that any departure from previously achieved partial balancing is an obvious reason for an increase in echo energy. Such a change in partial balancing may be caused by switching from one telephone set to another with a different input impedance. When the telephone set is in the off-hook state, even a change in acoustical coupling between the mouthpiece (microphone) and earpiece (earphone) may result in a noticeable change in the echo. Other contributors to echo changes include a bridged line or call forwarding. The longer the subscriber line, the smaller the effect of the telephone set impedance on the Z_{TR}.
- Although different configurations of a T/R twisted pair contribute to additional signal reflections created within the T/R pair treated as a transmission line, these reflections are not significant from the viewpoint of the line echo canceller. Even for very long subscriber lines, which, in some countries, may reach the length of 30 kft, the electrical signal delays are negligible. Obviously, the Z_{TR} are different for different twisted pairs, and they affect the reflections from the hybrid circuit.
- There are two main reasons why echo may become noticeable in voice networks, namely, long round-trip signal delay and a small ERL. According to published studies, the echo signal is negligible when ERL is 55 dB or more, regardless of the round trip delay, or when the round trip delay is smaller than 20 ms, regardless of the ERL (more accurate data is illustrated in **Figure 11** on page 12).
- For a typical ERL, the echo signal becomes noticeable as a distinct reflected signal of the speech if the delay between the outgoing signal and the reflected signal is greater than 24 ms – 32 ms. Otherwise, the reflected signal contributes only to a change of telephone handset side tones. Although echo may affect the overall quality of a telephone connection, the signal reflected from the hybrid is far from being perceived as an echo signal and is therefore not annoying or considered as a distortion.

1.3 Echo Cancellers in PSTN

Figure 4 depicts a telephone network with emphasis on the echo of the far-end speech signal. According to the G.168 Recommendation, the far end is the side of the telephone connection that generates a speech signal traveling through the network and reflects from the other end hybrid.

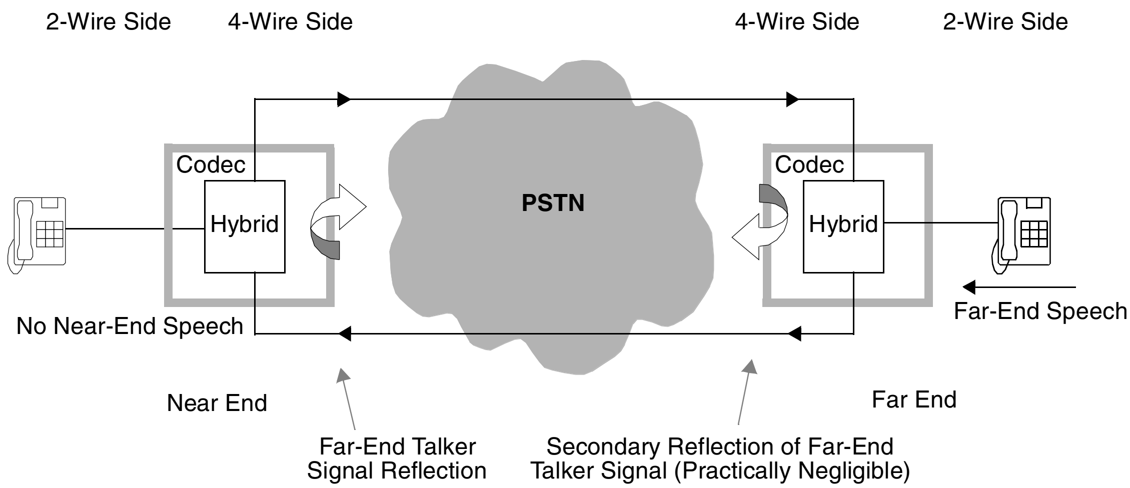


Figure 4. Echo Created by an Imbalance of the Near-End Hybrid

Traditionally, echo cancellers have been deployed on the long-haul trunks of the PSTN. **Figure 5** shows a simplified diagram illustrating two echo cancellers at the two sides of the long-haul telephone connection. One echo canceller is cancelling a reflection at the A end of the telephone connection and the other is cancelling a reflection at the B end.

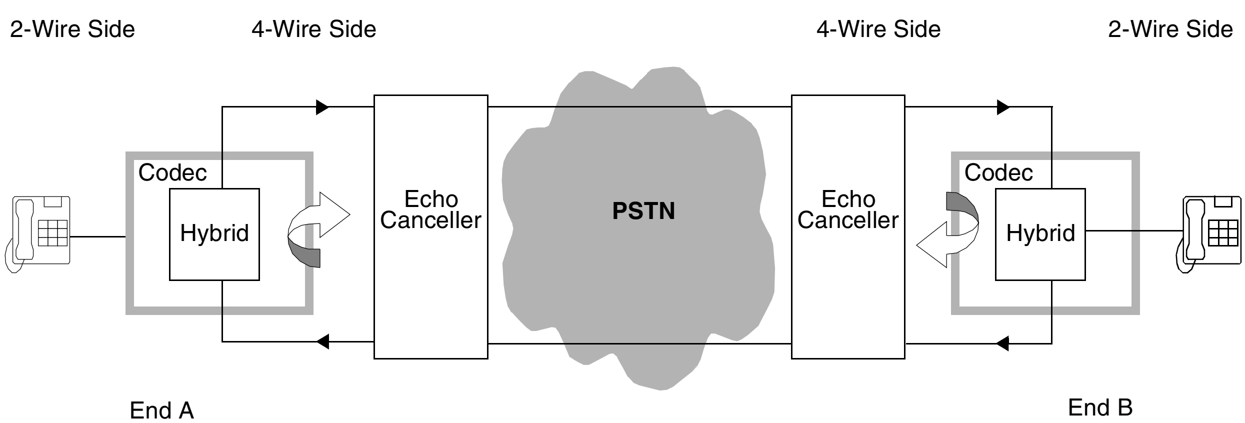


Figure 5. Placement of Echo Cancellers at Two Ends of the Telephone Connection

1.4 Echo Cancellers in Packet Telephony Networks

The role of echo cancellers in packet telephony networks is similar to their role in the PSTN. However, the requirements for packet network-based echo cancellers far more stringent for several reasons, including larger voice signal delays and larger signal distortions. **Figure 6** depicts a packet network with emphasis on the position of the echo cancellers. One echo canceller is cancelling a reflection at the near end of the telephone connection and the other is cancelling a reflection at the far end.

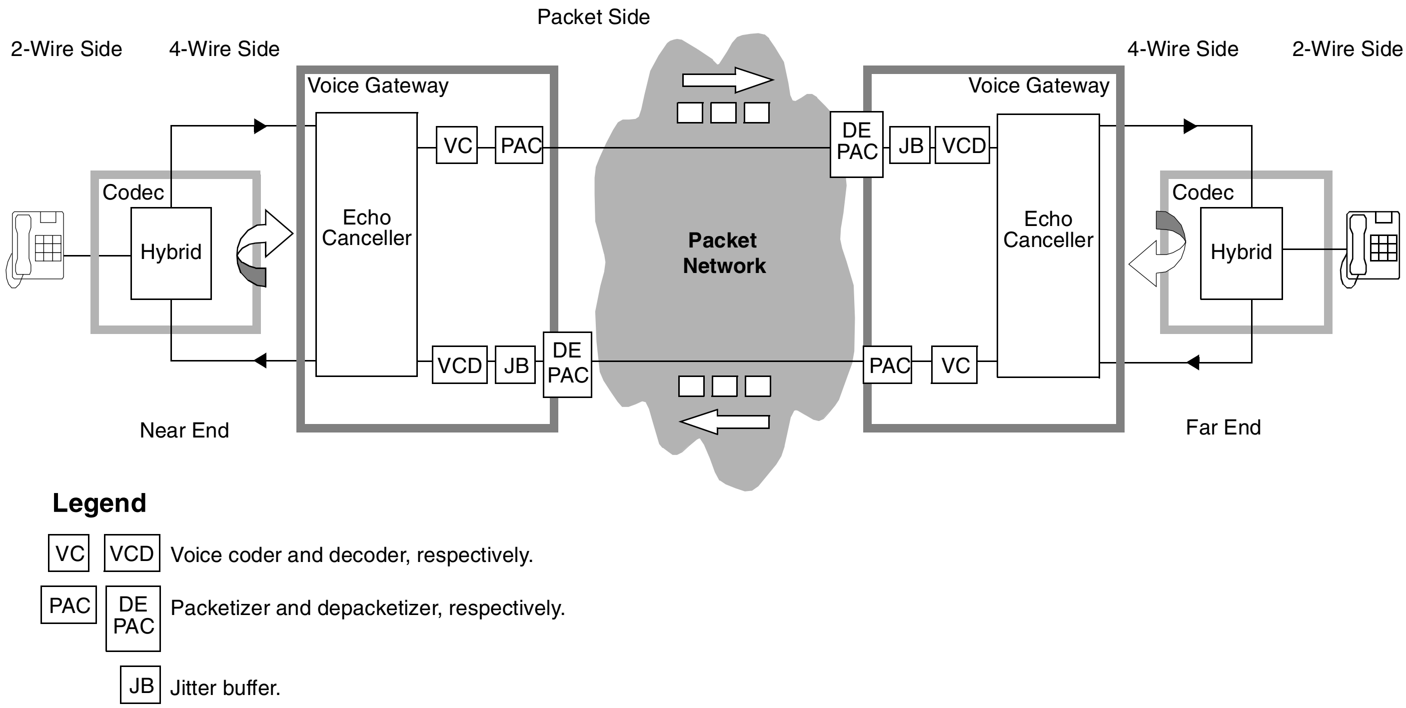


Figure 6. Echo Cancellers Installed at Two Ends of a Packet Telephony Connection

2 Echo Canceller Architecture

For a general view of an echo canceller and related signal notation, refer to **Figure 7**. The adopted signal notation is in line with that of most references, particularly with ITU-T G.168 Recommendations [3–5].

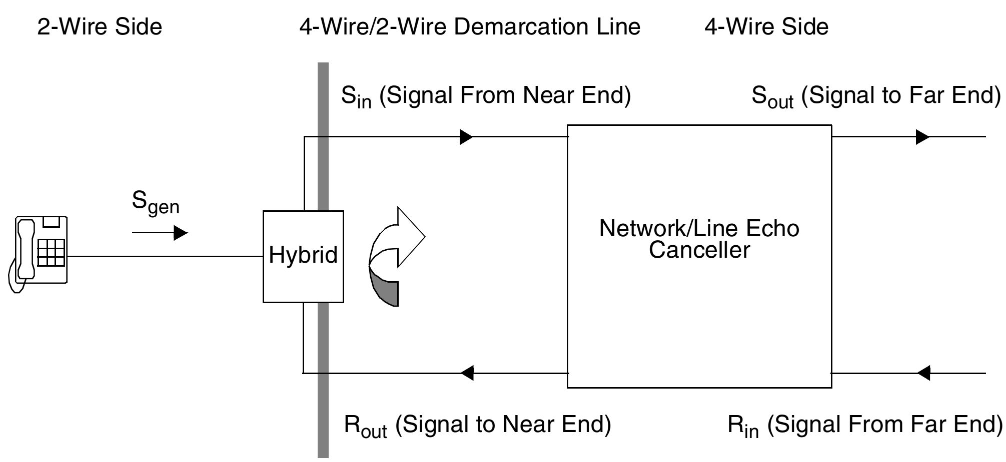


Figure 7. Network Echo Canceller as a Four-port Black Box

In addition to G.168-promoted terminology, a telecommunication convention is used for referencing two out of four signals:

- FE signal is related to the Rin signal, and these signal names can be used interchangeably.
- NE signal is related to the Sin signal, and these signal names can be used interchangeably.

FE pertains to the far-end speaker/listener or related speech signal. NE pertains to the near-end speaker or related speech signal and can thus refer to G.168 S_{gen}. The basic functionality of any G.168-compliant network/line echo canceller must include the following features:

- Adaptive filter.
- Double-talk detector or, alternatively, a near-end voice activity detector with a different functionality but a similar role in controlling the adaptive filter operations.
- Nonlinear processor, unless the adaptive filter itself can meet all requirements for the convergence depth.
- Optional comfort noise generator (typically, a part of the nonlinear processor, NLP).
- Signaling tone detectors, which can be architecturally positioned inside an echo canceller system or outside of it.
- Control block and respective interface allowing for freezing/unfreezing/resetting adaptive filter coefficients and performing other control functions.

Figure 8 shows a white box view of the major sub-blocks in a typical echo canceller. High-pass filters (HPFs) are optional.

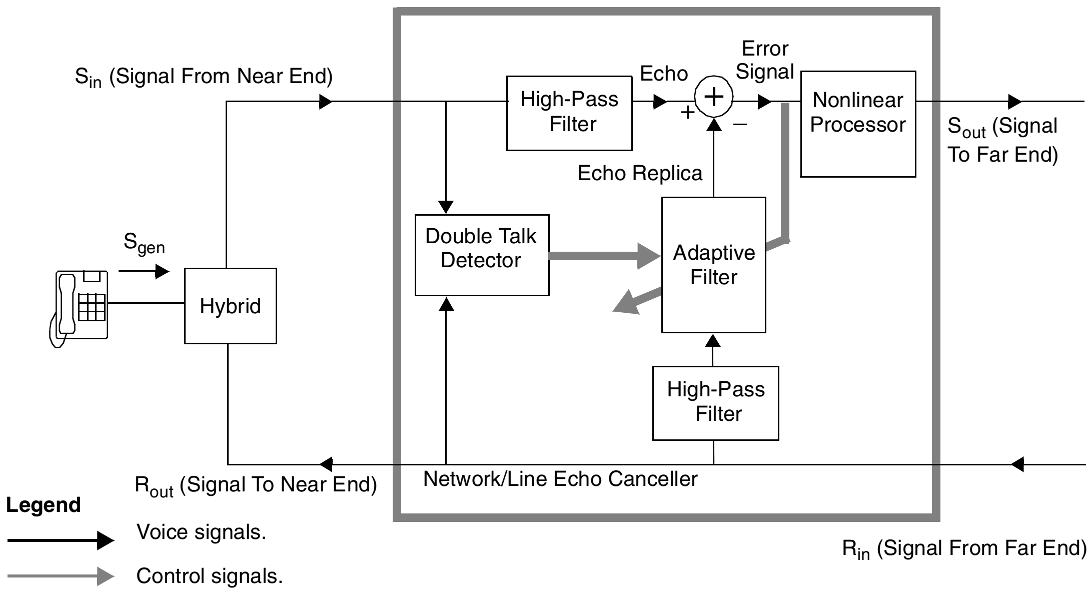


Figure 8. Simplified “White Box” View of a Typical Echo Canceller

2.1 Other Voice Processing Elements

Many network elements may interact with a modern digital echo canceller. They include older elements controlling the echo (for example, G.164 echo suppressors), voice compression elements (G.726 ADPCM voice codecs, G.729 CELP voice codecs, and others), and digital loss pads. Even if they operate as intended, these elements may contribute to lower performance of echo cancellers. There are no specific remedies to deal effectively with such nonlinear elements other than to provide an adequate performance margin during the design/evaluation phases of echo canceller development.

2.2 Hardware-Based Echo Cancellers

Hardware-based echo cancellers include FPGA-based echo cancellers and ASIC echo cancellers. When hardware-based echo cancellers are applied to specific applications, it is important to consider channel density versus silicon real estate (which can indirectly relate to the price per channel) versus dissipated power and other strictly hardware-related issues. Echo canceller algorithms in hardware-based systems usually differ from those in software running on DSPs or other specialized processors.

2.3 Software-Based Echo Cancellers

Software-based echo cancellers are application functions and modules that coexist with other applications related to digital telephony or packet telephony within a software framework. Because of the high demand for signal processing operations, particularly in arithmetic computations, these functions or modules run mostly on fixed-point DSP processors, less frequently on RISC processors, and even less frequently on general-purpose processors.

A typical measure of the DSP-related application computational efficiency is Million Instructions Per Second (MIPS). Most DSP instructions execute in one CPU cycle. As a measure of CPU workload, MIPS are established as one measure of practical algorithm efficiency for short instruction word architectures (that is, non-pipelined processor architectures). In the pipelining/parallel processor architectures, such as Very Long Instruction Word (VLIW) and Single Instruction Multiple Data (SIMD), several instructions can execute in one clock cycle. Therefore, a measure called Million Cycles Per Second (MCPS) is also used. For the SC140 core, CPU load is always measured in MCPS. For floating-point processors, a more appropriate measure of the CPU/ALU usage is Million Floating-Point Operations Per Second (MFLOPS). This document does not discuss echo cancellers running on floating-point processors.

Typically, CPU usage for a software-based line echo canceller during the initial convergence and reconvergence periods is in the range of 5–10 MIPS. CPU usage typically drops to 2–4 MIPS after the initial convergence is achieved and the echo canceller is only monitoring the state of convergence and amount of residual echo signal. These numbers refer to 32 ms NLMS-based echo cancellers with a moderately rich feature set. Some echo canceller vendors report that their products meet the minimum CPU MIPS consumption target of 1.5 MIPS (or even less) during the fully converged state. There is insufficient data on the lab and field performance evaluation of these echo cancellers. It is conceivable that a small, low cost echo path coverage echo canceller with a limited feature set might consume 1.2–1.4 MIPS. Such an echo canceller could hardly compete with high-quality and fully G.168-compliant echo canceller products. Among the means for decreasing CPU MIPS consumption during the converged state are the following:

- Sub-rate processing of support functions, such as energy calculation, Echo Return Loss Enhancement (ERLE) monitoring, and so on.
- Matching the filter length with actual echo tail length.
- Using ERLE (or equivalent) monitoring mode.

2.4 Choosing an Echo Canceller for an Application

When selecting an echo canceller for an application, you should consider the following:

- Objective performance measurements, in particular, G.168-series of ITU-T Recommendation test results with information on the pass margin.
- Subjective evaluation results.
- General specification characteristics.
- Channel density or equivalent measuring points.

- Scalability, particularly for the software solutions.
- Implementation criteria and specific platform requirements.

Even very well designed and properly evaluated echo cancellers may not perform adequately in some network configurations or in some markets. For example, an echo canceller placed into a network that creates signal echo path delays larger than the echo tail length coverage offered by the echo canceller will not perform adequately. An echo canceller deployed in a market in which the conversational patterns differ significantly from the ones the echo canceller was designed for may generate customer complaints. Therefore, it is important to gather initial information on the specific application before selecting an echo canceller. Only very few available models of high-performance echo cancellers are suitable for unusual or extreme applications. Even these echo cancellers may manifest unwanted behavior in certain network situations.

3 Network Delays

Line/network echo cancellers normally control/cancel signal reflections at the near end of the network connection. Echo canceller products to control echoes at the far end are also available. However, far-end echoes should be controlled by respective local echo cancellers that are deployed at local central offices/PBXs or the closest regional or gateway offices. The deployment of far-end echo cancellers is justified only in abnormal situations when echo cancellers perform very poorly, or malfunction, or simply are not enabled. Combining near-end and far-end echo cancellers into one homogenous system is not recommended because the performance, as measured by convergence speed and depth, may suffer. Instead, a cascade arrangement of two echo cancellers is more appropriate for such a special situation. There is a gray area of network configurations in which echo cancellers deal with relatively large echo path delays (greater than 24 ms). These configurations are very rare. The two chief components of echo path delay are pure delay, sometimes informally called bulk delay, and echo-path dispersion time. **Figure 9** illustrates these two components with a graph of the impulse response as captured using Rin as an input and Sin as an output.

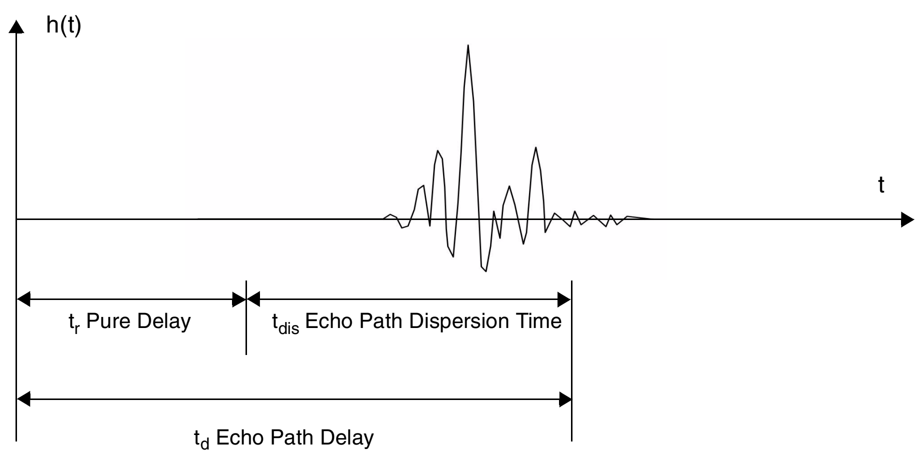


Figure 9. Example of an Impulse Response Using Rin/Sin Signals

Echo path dispersion time combines the hybrid dispersion time related to the first signal reflection and hybrid dispersion times associated with the second, third, and higher reflections. According to a recent report on North America telephone connection surveys, the largest percentage of dispersion time is between 5–7 ms. Only two calls out of 101 long-distance calls had a dispersion time between 11–12 ms. Reference [26] provides additional supporting evidence. Note that G.168-2000/2002 contains an echo path model (model number 4) spreading over a 16 ms segment.

3.1 G.168 Echo Path Models

Table 1 lists the models published by ITU-T G.168-2002. The models were developed mostly through statistical analysis of PSTN network connections in North America. European studies (France Telecom) have confirmed that echo return paths recorded during the study have characteristics equivalent to those of Models 1, 5, and 6 [5, 23].

Table 1. ITU-T G.168-2002 Models

Model Number	Number of Points	Time Span[1]	Characteristics	Comments
1	64	8 ms	Single reflection	
2	96	12 ms	Single reflection	
3	96	12 ms	Single reflection	
4	128	16 ms	Single reflection	Considered especially suitable for testing the echo canceller ability against the echo path delay coverage.
5	96	12 ms	Single reflection	
6	120	15 ms	Single reflection	
7	96	12 ms	Double reflection	Considered one of the most challenging hybrids to pass convergence speed and depth tests.
8	99	12.375 ms	Double reflection	Properties similar to those of model 7.
NOTES: 1.Time span is equivalent to the duration of the echo path delay as illustrated in **Figure 9**.				

3.2 Multiple Windowing

The term "multiple windowing" refers to hypothetical echo path delay characteristics manifested by a non-overlapped train of single reflections. The impulse response of the echo path black box consists of time segments with primary signal reflection and secondary (second, third, and so on) reflections separated by time segments with zeroes. These reflections are thought to be produced by different hybrid circuits distributed on the voice call path. **Figure 10** illustrates a hypothetical case of echo path delays containing three windows. Neither ITU-T G.168 2000/2002 nor any other known documents report practical examples of impulse responses representing echo path delays consisting of separate (non-overlapping) individual reflection signals. Although effort has been dedicated to tracing cases of multiple windowing in public and private voice networks, results have been inconclusive and no statistics have been offered to date.

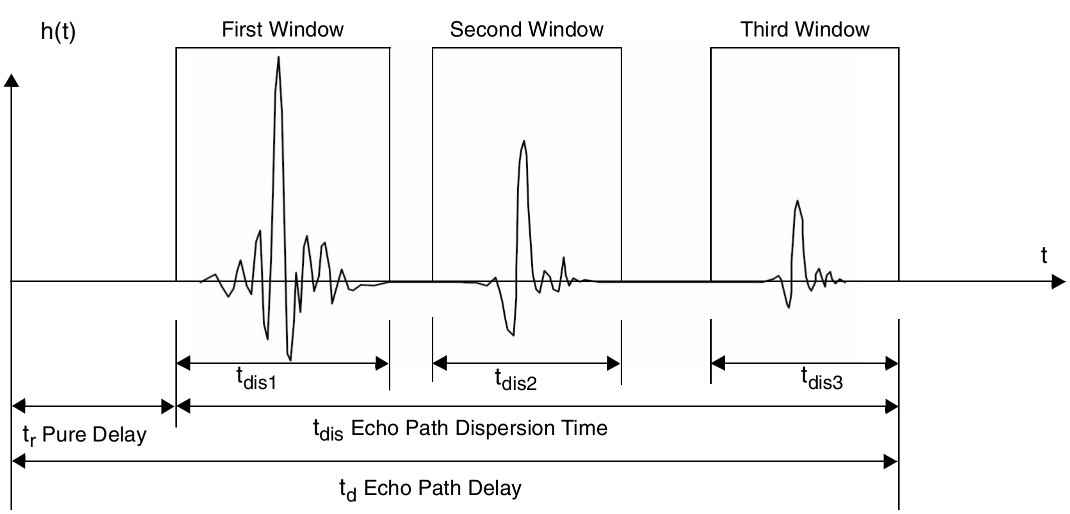


Figure 10. Hypothetical Case of Three Non-overlapping Reflection Windows

3.3 Delay Perception Threshold

Numerous studies, including [4, 20, 21], indicate that reflections (talker echo) can become noticeable and may become objectionable when the delay between two telephone sets is greater than 16 ms, which is the equivalent of a 32 ms round-trip delay. **Figure 11** shows how perception of echo transmission delay and ERL are related. G.131 requires the use of echo cancellation for one-way delays greater than 25 ms. For details, consult ITU-T G.131 and G.114 [17, 19]. Echo canceller treatment of smaller delays can be beneficial, too, because it can positively contribute to voice quality by better controlling the side tone performance. Echo and side-tone perceptual thresholds are the subject of applied psychological acoustics, and their details are published in specialized literature.

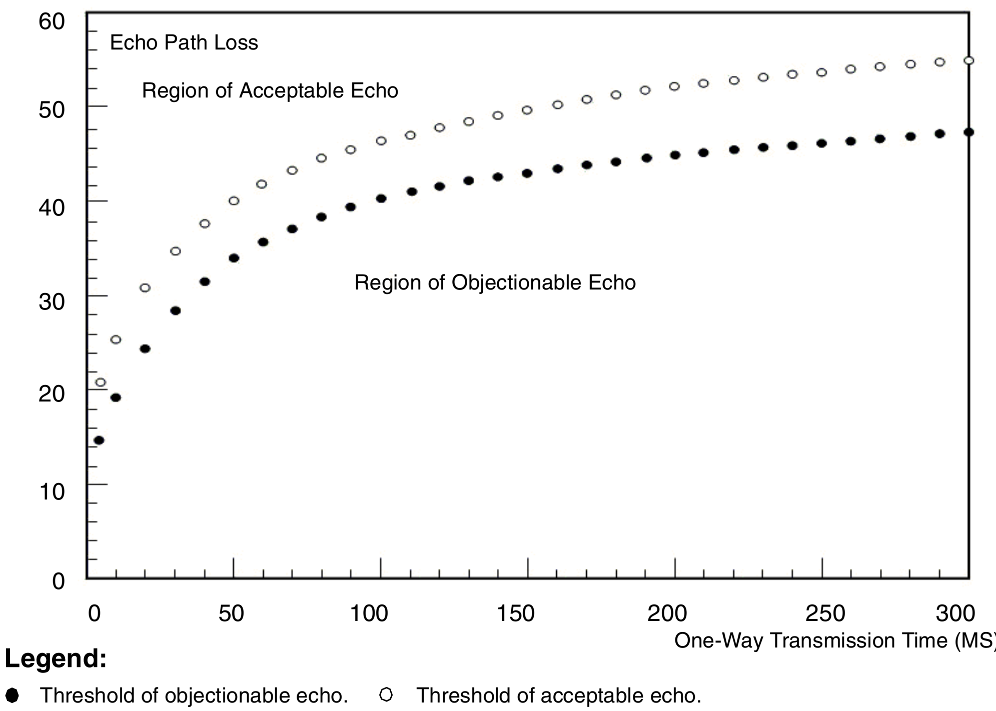


Figure 11. Talker Echo Tolerance as a Function of Transmission Delay

3.4 Examples of Network Element Delay

Table 2 summarizes delays (and dispersion times, as applicable) for a variety of network elements.

Table 2. Delays and Dispersion Times

Network Element	Typical Delay	Extreme Delay	Remark
Hybrid circuit with the rest of codec circuitry combined	3 ms	5 ms	Vendor dependant; the delay number is dispersion time.
Time-division switching matrix (part of CO or PBX switching fabric)	0.25 ms	0.5 ms	Vendor dependant
Two-wire subscriber line (non-loaded)	9 µs	10 µs	Reference length of 6 kft (one-way delay)
Two-wire subscriber line (loaded)	25 µs	27 µs	Reference length of 15 kft (one-way delay)
Extended range subscriber line	40 µs	43 µs	Reference length of 24 kft (one-way delay)
Terrestrial connection within North America continent	Any between 0 ms to 24 ms	24 ms	One-way delay (no specific route)

Table 2. Delays and Dispersion Times

Network Element	Typical Delay	Extreme Delay	Remark
Satellite connection (via a geostationary telecommunication satellite): ground station–transponder–ground station	240 ms	N/A	One-way delay (geosynchronous orbit at 35,784 km)
Low-Earth Orbiting (LEO) satellites	4.7 ms	N/A	Example of 1,414 km LEO used for cellular communication; one-way delay Atlantic underwater telecommunication cable (fiber optic cable)
Atlantic underwater telecommunications cable (fiber optic cable)	38 ms	N/A	One-way delay (Columbus III cable project completed in July 1999; optical cable linking Hollywood, Florida with Lisbon, Portugal; Conil, Spain; and Mazara del Vallo, Italy)
Packet telephony gateway	80 ms	200 ms	One-way delay (often beyond acceptable limits). Includes algorithmic delay for voice codecs, processing delay, and jitter buffer delay; vendor dependant

3.5 Conclusions on Network Delay

Both well documented studies and unofficial, unpublished reports indicate that an echo path delay coverage of 24 ms is adequate for almost all known applications within PSTN/packet telephony networks. Only special situations require echo path coverage beyond 24 ms. These situations are probably due to pure delays of relatively large values. An echo canceller equipped with one window of statically or dynamically adjustable length up to 24 ms and a bulk delay estimation mechanism is believed to be a very good compromise solution for most (if not all) PSTN/packet telephony applications. However, echo canceller products often exceed these requirements.

4 Requirements for Echo Cancellers

ITU-T G.165 provided the first set of ITU recommendations for line echo cancellers. Now, G.165 is augmented by a newer set of requirements incorporated into the G.168 Recommendation. ITU-T G.165 Recommendations remain in force.

4.1 ITU-T G.168 Requirements

The ITU-T G.168-series Recommendations specify a set of requirements for a digital network echo canceller, as follows:

- Rapid initial convergence
- Low returned echo level during single talk
- Low divergence during a double talk condition
- Assured double talk detection
- Proper operation during facsimile/modem transmission

To ensure a minimum level of acceptable echo canceller performance, the G.168 series defines a set of objective tests composed of 15 test groups. These tests continue to be redefined, enhanced, or modified, and some are still under study or have optional status.[1] **Table 3** summarizes all tests in the G.168 (2002) Recommendations. Passing these objective tests does not guarantee quality of service. Therefore, it is necessary to conduct subjective tests on the performance of an echo canceller, as well.

Table 3. Tests in the G.168 Recommendations

Test Number	Description	Status	Remarks
1	Steady state residual and returned echo level test	Deleted	The test is incorporated into the Test 2 group.
2A	Convergence test with NLP enabled	Enforced	To ensure rapid initial converge and low ERLE with NLP.
2B	Convergence test with NLP disabled	Enforced	To ensure rapid initial converge and low ERLE without NLP.
2C	Convergence test in the presence of background noise	Enforced	To ensure rapid initial converge and low ERLE with background noise.
3A	Double talk test: double talk with low NE levels	Enforced	To ensure double-talk detection sensitivity and a low false detection rate.
3B	Double talk test: double talk with high NE levels	Enforced	To ensure double-talk detection sensitivity and speed.
3C	Double talk test: double talk under simulated conversation	Enforced	To ensure no artifact during and after double talk.
4	Leak rate test	Enforced	To ensure slow leakage on H registers.
5	Infinite return loss convergence	Enforced	To ensure prevention of unwanted echo when echo path is open.
6	Non-divergence on narrowband signals	Enforced	To ensure convergence for narrowband signals.
7	Stability test	Enforced	To ensure stability for narrowband signals.
8	Nonconvergence on System 5, 6, and 7 in-band signaling and continuity check tones	Optional	To ensure proper transmission of specific signals.
9	Comfort noise test: matching, adjustment down, adjustment up	Enforced	To ensure comfort noise level matching Sin noise level.
10A	Facsimile test: canceller operation on the calling station side	Enforced	To ensure rapid initial convergence and prevent unwanted echo.
10B	Facsimile test: canceller operation on the called station side	Enforced	
10C	Facsimile test: canceller operation on the calling station side during page transmission and page breaks	Enforced	
11	Tandem echo canceller test	Under Study	To test performance of two echo cancellers in parallel.
12	Residual acoustic echo test	Under Study	To test acoustic echo cancellation.
13	Performance with low bit rate coders	Optional	To ensure performance with low bit-rate codecs.
14	Performance with low-speed modems	Enforced	To ensure proper operation with a low-speed modem.
15	Pulse Code Modulation (PCM) offset test	Under Study	To ensure proper operation with the PCM offset on Sin.

1. The latest version of the Recommendation is G.168 06/2004.

4.2 Requirements Beyond G.168

This section discusses requirements that are not part of the ITU-T Recommendations yet are often included in design specifications and/or commercial specifications. Some of these requirements are independent of specific markets, others are very specific.

4.2.1 Double-Talk /Near-end Voice Activity Detection

G.168-series Recommendations provide only general testing guidelines for double-talk detection. To ensure that double-talk detection/near-end voice activity detection results in unnoticeable echo under different conversational scenarios, the design effort should focus on the following areas:

- Reliable and fast detection of near-end voice activity, even for relatively small ERL values.
- Reliable and fast detection of transitions from double-talk to single talk (far end active) state.
- Double-talk detection (DTD) strategy and filter adaptation strategy for conversation scenarios with a very high percentage of double-talk condition (calls beginning with double-talk included).

The DTD is an important element that contributes greatly to the quality of the entire echo canceller system.

4.2.2 NLP Performance

The nonlinear processor, typically combined with a comfort noise generator, is another block of an echo canceller system that affects overall echo canceller quality. Important aspects of the NLP and CNG performance are:

- Selection of signal threshold determining when NLP becomes active.
- Spectral structure and power level of the signal generated by CNG.
- Strategy of activating and deactivating the CNG in relation to background noise variations at the near end.

4.2.3 Echo Tail Coverage

The ITU-T G.168-series poses limited requirements on the echo path delay. Echo path models defined in G.168-2002 do not go beyond 16 ms of the echo path delay (t_d). For a variety of reasons, most of which are unrelated to the technical requirements imposed by specific network configurations, echo canceller users often ask for echo path coverage longer than needed. The typical range of echo path coverage offered by leading echo canceller vendors is 32–64 ms. At the high end, the coverage reaches 128 ms. Only very specialized echo cancellers offer higher coverage. Echo cancellers with extended echo path coverage (beyond 128 ms) require additional hardware (memory) to store longer signal history.

4.2.4 Far-End Echo Coverage

If an echo canceller at one end (A) of the telephone connection is not operational or does not perform adequately, an operator providing echo cancellers at end B for the benefit of end A may consider introducing a far-end echo canceller. **Figure 12** depicts a system with two echo cancellers, a classical one and a far-end one (FEECAN) that is designed to control echoes with very large pure delay. A far-end echo canceller is equipped with a bulk-delay estimator that can determine the bulk delay from 0 ms to 500 ms, or more. In **Figure 12**, S_{out} of echo canceller A becomes R_{in} of echo canceller B; conversely, S_{out} of echo canceller B becomes R_{in} of echo canceller A.

4.2.5 Reconvergence

As echo canceller systems become increasingly sophisticated and network functionality becomes more complex, the reconvergence capability of echo cancellers becomes increasingly important. The following scenarios create a need for reconvergence: adding a new telephone extension during an existing call, call forwarding during an existing call, conference circuit/bridge operation, and voice mail operation.

4.2.6 Adaptation Stalling in Special Situations

Efficient adaptation can be achieved when signal levels are adequate. Adaptation on weak signals is very problematic in most cases. An echo canceller system is supposed to have a safeguard to prevent it from adapting in the presence of weak single talk signals.

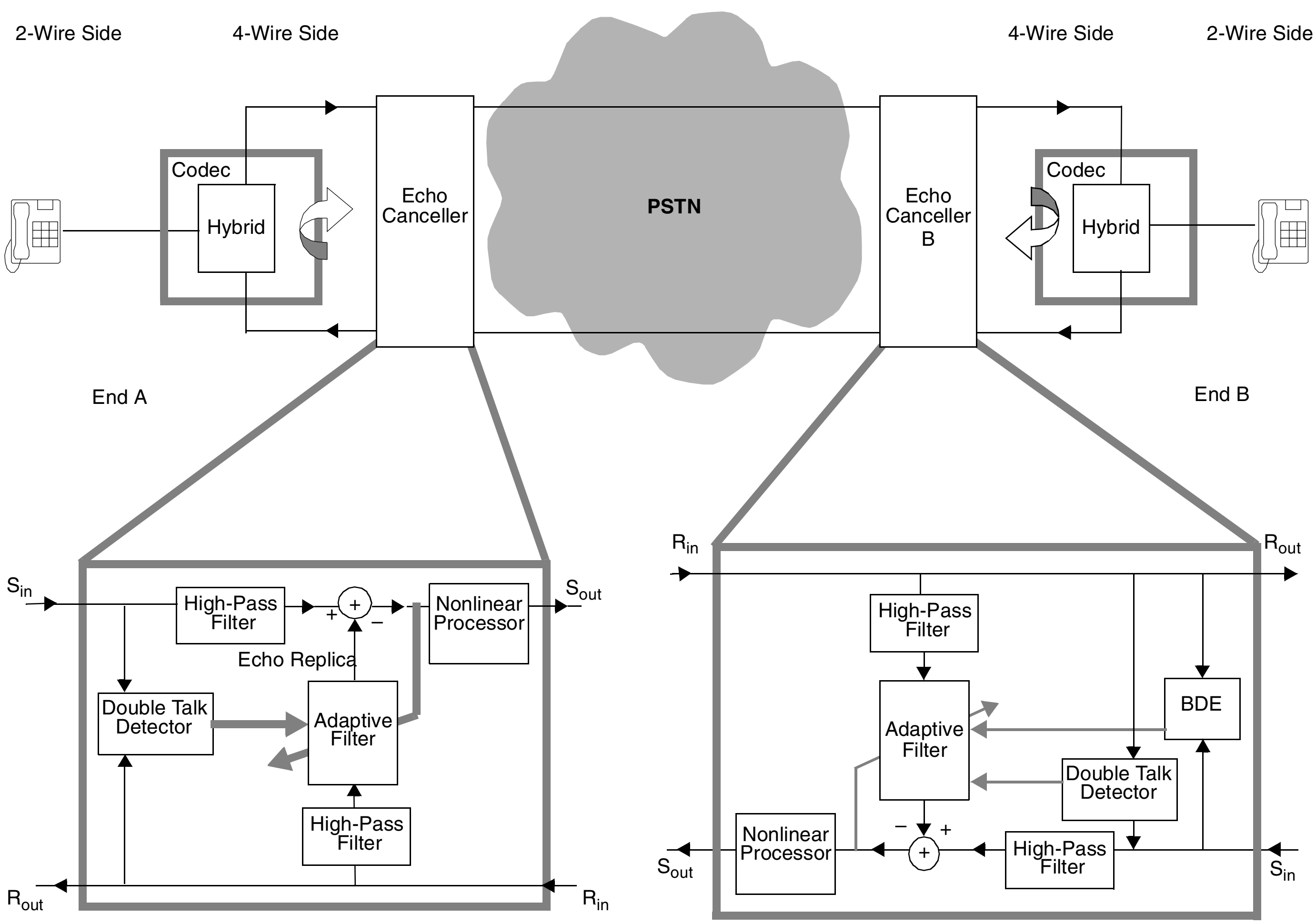


Figure 12. Coexistence of Near-End and Far-End Echo Cancellers

4.2.7 System Stability

A 4-wire connection between two-end hybrid circuits poses a potential problem of instability because it creates a feedback to each end (see **Figure 4,** *Echo Created by an Imbalance of the Near-End Hybrid,* on page 6). If the hybrid circuits are modelled with linear FIR filters, one can analyze stability by finding the poles of the system. The hybrid circuits can be more adequately represented as IIR systems; however, we limit out considerations to FIR systems. This limitation does not impose substantial restrictions because typical IIR systems can be approximated by FIR models. Let $H_{NE}(z)$ and $H_{FE}(z)$ denote the Z-transforms of the near-end and far-end FIR filters, as illustrated in **Figure 13**, where $X_{NE}(z)$ and $Y_{FE}(z)$ ($X_{FE}(z)$ and $Y_{NE}(z)$) denote the Z-transforms of the near-end transmit and far-end receive (far-end transmit and near-end receive) signals, respectively, as follows:

$$\begin{bmatrix} Y_{FE}(z) \\ Y_{NE}(z) \end{bmatrix} = \frac{1}{P(z)} \begin{bmatrix} 1 & H_{NE}(z) \\ H_{FE}(z) & 1 \end{bmatrix} \begin{bmatrix} X_{NE}(z) \\ X_{FE}(z) \end{bmatrix}$$

where $P(z) = 1 - H_{NE}(z)H_{FE}(z)$ is the characteristic polynomial of the closed-loop system. The closed-loop system becomes unstable if any root of P(z) has a magnitude larger than one. In practice, even without echo cancellers, such a system may manifest unstable behavior when the total ERL, represented as a sum of individual ERLs, is close to zero. This instability effect, known in telephony as singing, is well documented [10]. Some stability features of telecommunications systems with echo cancellers are described in the literature [14, 15].

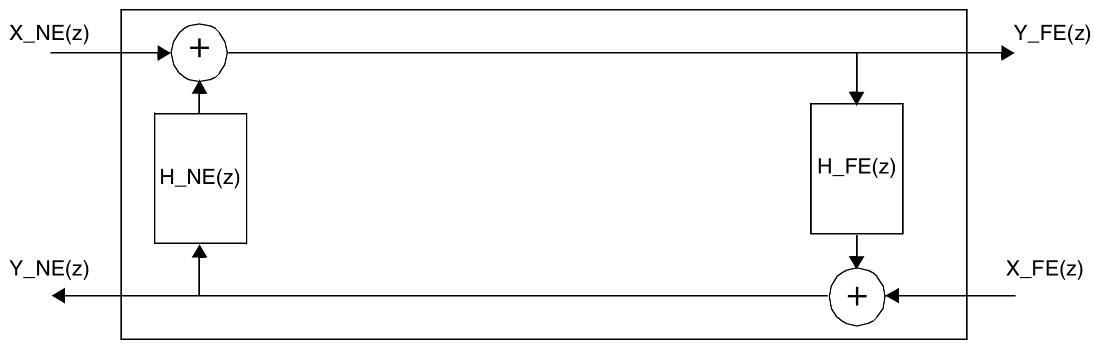


Figure 13. Closed-loop Communications System With Hybrid Circuit at Each End

The inclusion of echo cancellers can be modeled in a similar way (see **Figure 5,** *Placement of Echo Cancellers at Two Ends of the Telephone Connection,* on page 6 and its corresponding model in **Figure 14**), but with $H_{NE}(z)$ and $H_{FE}(z)$ replaced by equivalent FIR filters—the difference between the actual hybrid transmittance function and the adaptive filter generating the replica of the echo signal. The poles of the system do change after every update of the echo canceller filter taps, which poses a potential risk of instability. That is, the presence of two echo cancellers creates an adaptive closed-loop system with inherent ability to become unstable. Therefore, well-designed echo cancellers should have a built-in safeguard mechanism that prevents system instability.

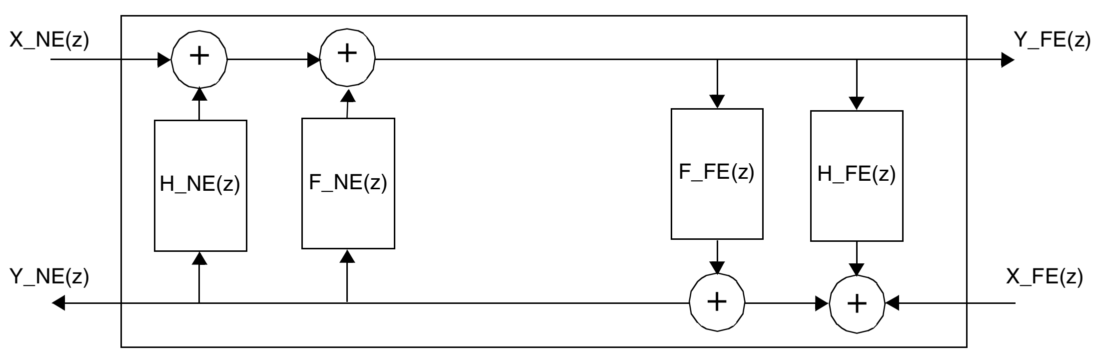


Figure 14. Closed-loop System With a Hybrid Circuit and an Echo Canceller at Each End

4.3 Subjective Evaluation Criteria and Methods

A G.168-compliant echo canceller may not provide adequate quality of service for TDM and packet telephony voice calls because not all scenarios are covered. Therefore, additional evaluation of the echo canceller is required before it is declared an adequately performing product. Subjective evaluation can be performed on-line and off-line. The on-line methods include talking and listening via the telephone connection when one echo canceller is active. The off-line methods include listening to S_{out} speech signals generated by the echo canceller under evaluation fed with a set of given input signals (pre-recorded and pre-processed speech files). These signals are generated by (a) passing pre-recorded speech through R_{in}, while S_{gen} is silence signal, and (b) passing prerecorded speech through the R_{in} port and a prerecorded speech and/or noise through the S_{gen} port. Subjective evaluation methods include the following:

- Initial convergence (rate, or speed).
- Infinite convergence (that is, convergence depth).
- Double talk detection.
- Nonlinear processor—threshold test.
- Nonlinear processor—false recognition test.
- Nonlinear processor—very soft NE speech.
- Near-end background noise matching.
- Mid-call convergence.
- Overload test.
- Free human sound generation on both ends in an unsynchronized manner to mimic a casual conversation—no structure and framework are defined for this test.

The evaluation sessions must be performed in an acoustically controlled environment. Some of the aspects are discussed in the P-series of the ITU-T Recommendations [1, 2].

4.4 Special Tests

Special tests include the following:

- High double-talk content speech signals passed to R_{in} and S_{gen} ports.
- Tests with presence of various noise signals on the S_{gen} side.
- Convergence speed and depth for various speech and noise signals.
- Reconvergence/mid-call convergence tests using natural speech and noise signals.
- System stability tests involving one and two echo cancellers.

5 Special Network Configurations

A Plain Old Telephone Service (POTS) call is relatively easy for an echo canceller to handle. After a replica of the echo path reflection is determined, there is usually no reason to continue adaptation after the full convergence is achieved. However, certain call features, such as call forwarding and conference bridges, may destabilize the echo path and the Z_{TR}. A call forward feature activated during an already established call causes several changes in characteristics of the telephone connection, thus requiring the echo canceller reconvergence. In **Figure 15**, the echo cancellers face a challenging task of detecting hybrid change during an active call. The call forwarding action performed during an already established telephone call results in the following:

- Switching from an old hybrid to a new hybrid.
- Using a new subscriber line.
- Using a new telephone set.
- Introducing new near-end subscriber line delay.

Only echo cancellers equipped with continuous ERLE monitoring/hybrid change detectors can react promptly to the abrupt hybrid change. Note that one solution to the hybrid change detection is a continuous, per-sample (or per short block of voice samples) adaptation. To avoid high CPU consumption, many echo cancellers do not use this approach.

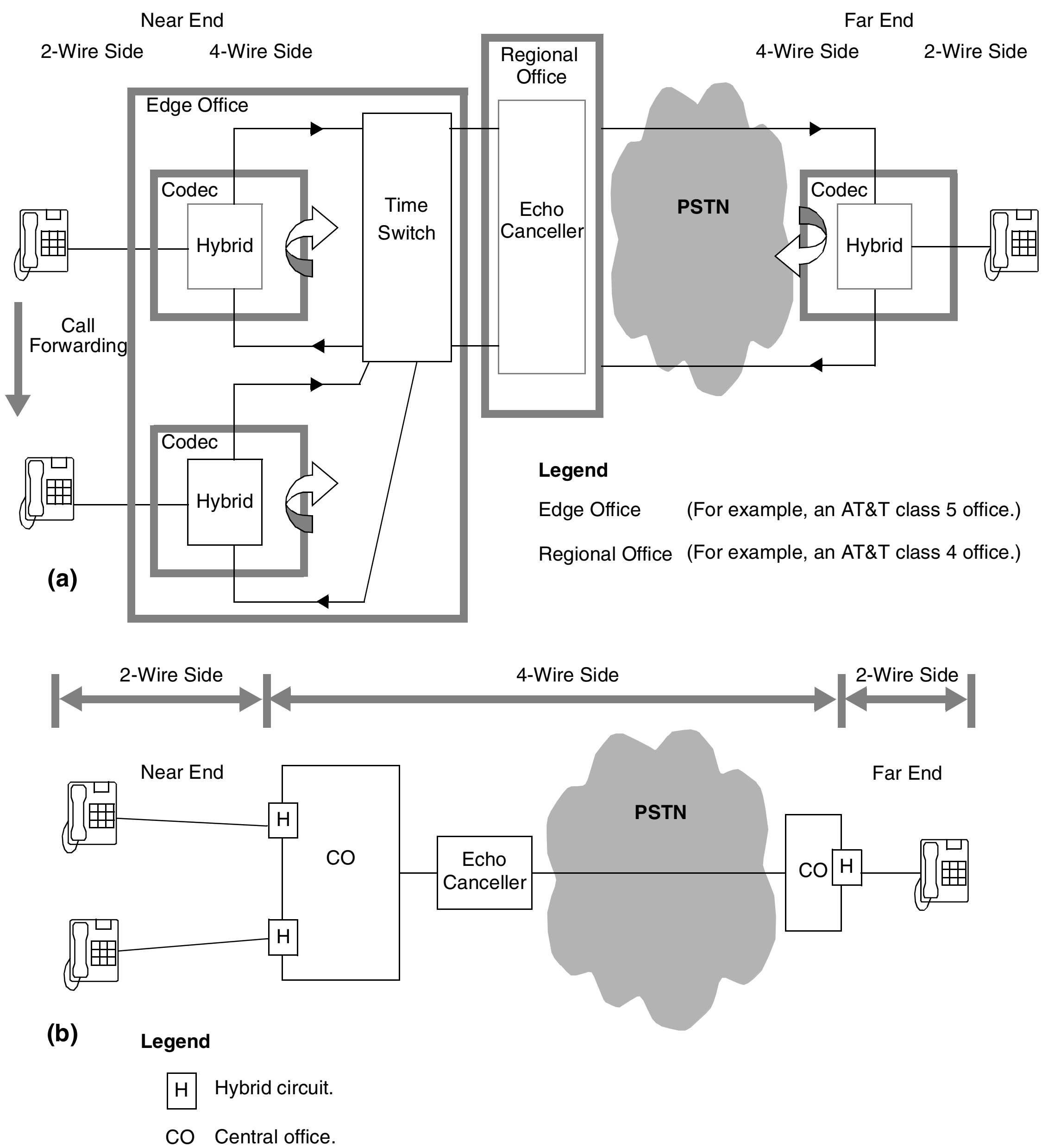


Figure 15. Call Forwarding Scenarios

Conference circuits (or bridges) pose a challenging environment for echo cancellers. Since control of voice path distribution through a conference circuit varies dynamically, the echo canceller must deal with ever changing characteristics of echo path delay. Therefore, to perform adequately, it must continuously operate in adaptation mode. **Figure 16** illustrates an example of network topology to accommodate a conference circuit.

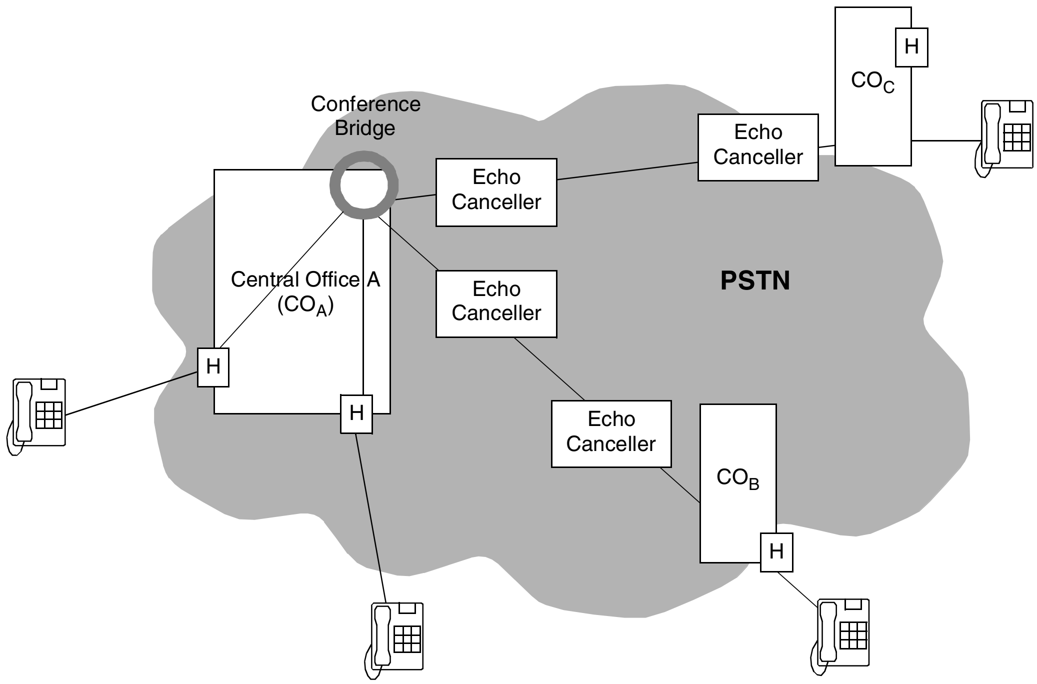


Figure 16. Echo Cancellers Provisioned on Trunks Attached to CO_A, CO_B, and CO_C

6 Special Configurations

In today's networks, there are call configurations that may produce multiple reflections, resulting in a contiguous pulse as opposed to non-overlapping multiple echoes. Therefore, a design approach to echo cancellers handling multiple reflections typically includes an adaptive filter of longer coverage of echo paths. **Figure 17** shows a network connection for a bridged call in which only one echo canceller is deployed. In such a configuration, multiple reflections may occur. In **Figure 17** (a), there is no echo canceller between offices CO_B and CO_C. Depending on the distance between CO_B and CO_C, the second reflection can be merged with the first, or, if distance is very large, the second reflection may arrive after the first reflection has already disappeared. Here, an echo treatment using a multi-window sparse approach is appropriate. The multiple reflections do not occur in **Figure 17** (b) because the echo cancellers are properly provisioned.

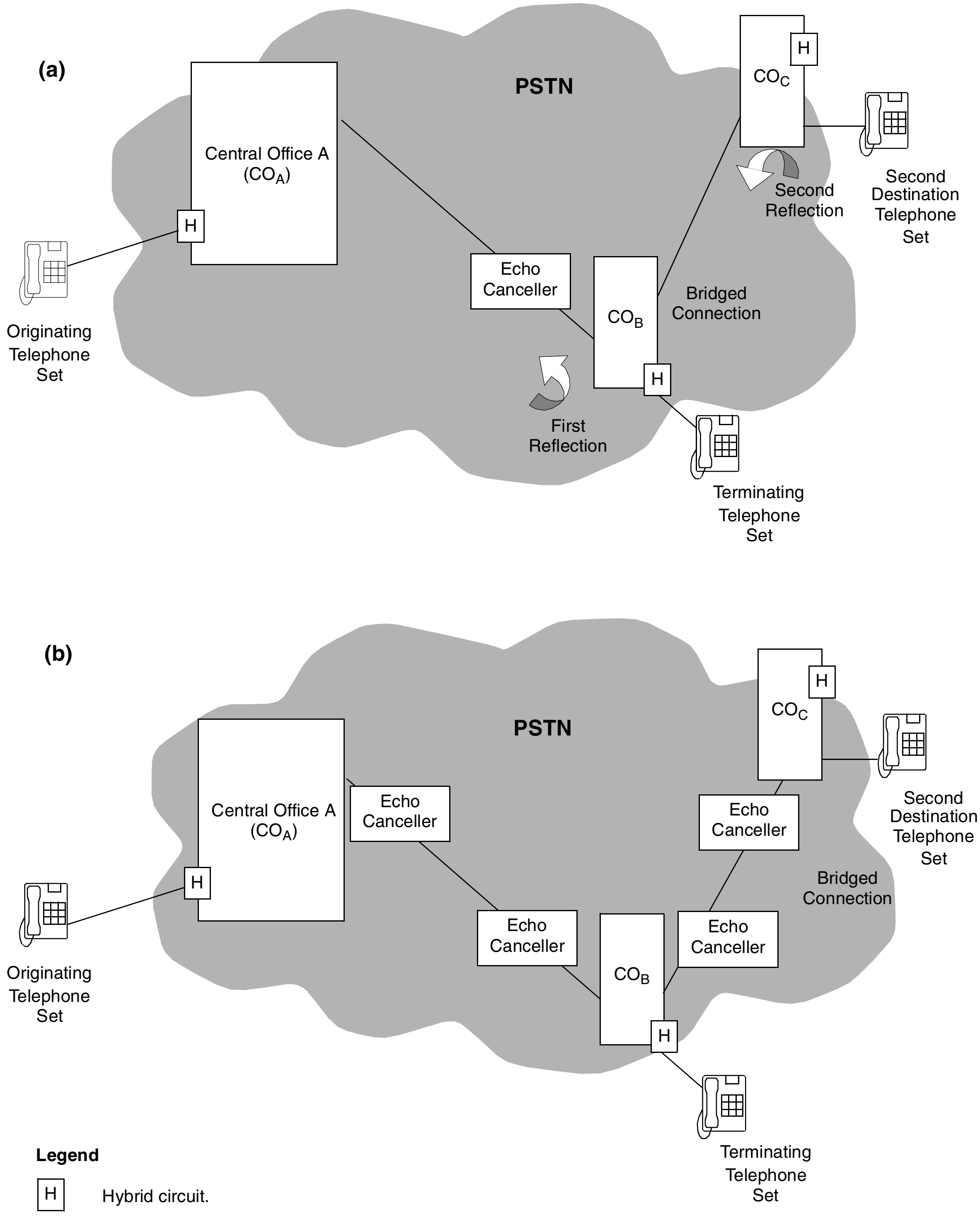


Figure 17. Reflections in a Hypothetical Bridged Telephone Call

7 DSP Algorithms In Echo Canceller Applications

The following DSP algorithms are considered in this section:

- Least Mean Squares (LMS)
- Normalized Least Mean Squares (NLMS)
- Proportionate NLMS (PNLMS)
- Block NLMS
- Transform Domain LMS
- Affine Projection
- Sub-Band NLMS
- Double-Talk Detector Block
- Geigel and Modified Geigel

7.1 Least Mean Squares (LMS)

Figure 18 shows an adaptive filter based on the structure of a linear FIR filter [11, 12]; *N* is the filter length.

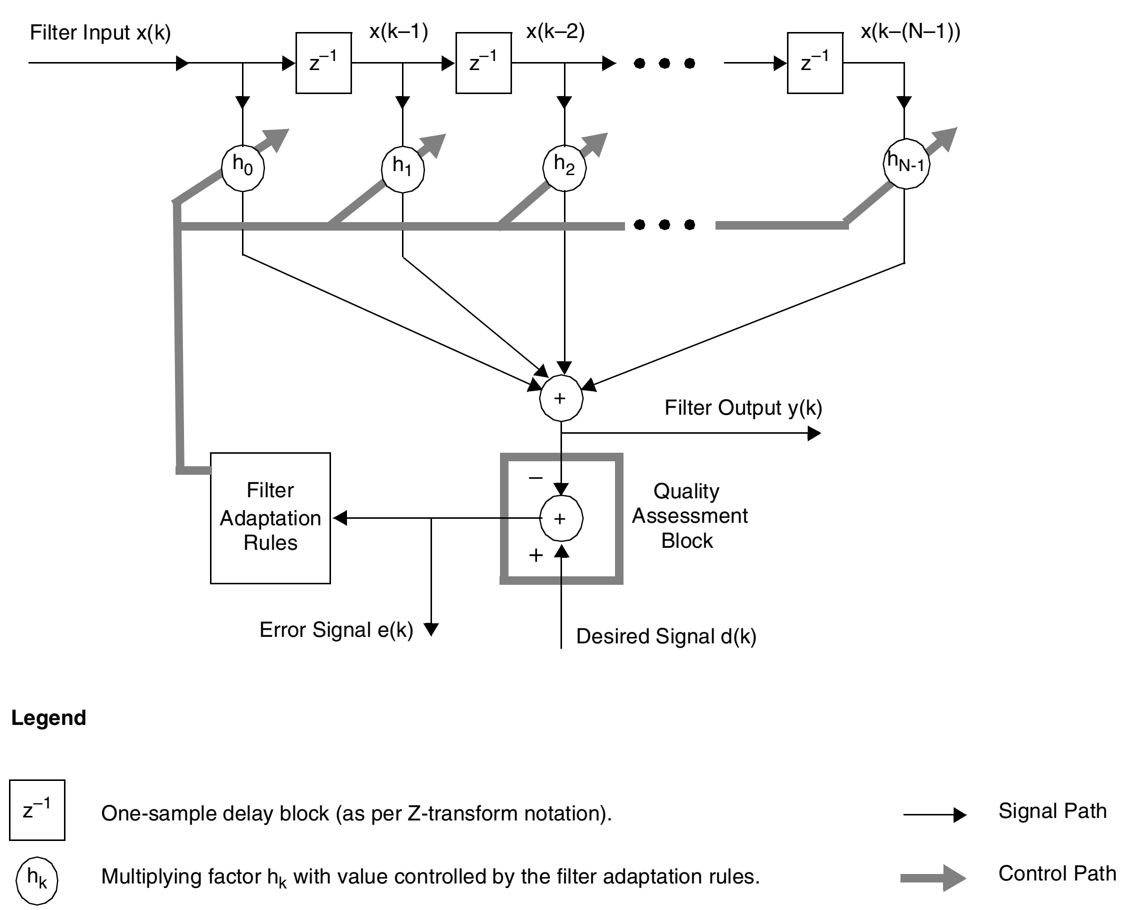


Figure 18. Adaptive Filter Controlled by an Error Signal

For network/line echo cancellers, the association of generic signals x(k), y(k), e(k) and d(k), as shown in **Figure 18**, with specific ones applicable to a G.168-series echo canceller[2] is as follows:

- x(k), $R_{in}(k)$
- y(k), Replica of $S_{in}(k)$
- e(k), Signal entering the NLP block
- d(k), $S_{in}(k)$

The following elements differentiate adaptation algorithms:

- Filter structure and type
- Quality assessment rules
- Filter adaptation rules

Most echo cancellation algorithms use FIR filters, as illustrated in **Figure 18**. IIR filters are also used in adaptation algorithms, although some reports indicate that the advantages of IIR adaptive filters are not compelling, [30]. In applications other than telecommunications, there have been attempts to use nonlinear filters. One classical nonlinear approach is based on the Volterra series, [11, 41]. IIR and nonlinear filters are not covered in this review.

For an FIR filter, the output is as follows:

$$y(k) = \sum_{i=0}^{N-1} h_i \bullet x(k-i)$$

The following notation concisely describes the LMS algorithm:

$$\mathbf{H} = [h_0, h_1, \ldots, h_{N-1}]^T$$

and

$$\mathbf{X}(k) = [x(k), x(k-1), \ldots, x(k-N+1)]^T$$

The output signal is represented as a product of two vectors, **X** and **H**, as follows:

$$y(k) = \mathbf{X}^T(k) \cdot \mathbf{H}$$

or, in an equivalent form, as follows:

$$y(k) = \mathbf{H}^T \cdot \mathbf{X}(k)$$

Generally speaking, the core of the adaptation problem is the strategy of finding the best impulse response **H**, as measured using an L^2 distance between the filter output y(k) and the desired signal d(k):

$$J = \sum_k \left| d(k) - y(k) \right|^2$$

When J achieves the minimum value, the respective **H** generates the best replica of the desired signal. In an echo canceller application, the best **H** results in minimizing the echo. In other words, the object of the adaptive filter algorithm is to find the impulse response that mimics the response of the hybrid circuit. During adaptation, the consecutive vectors **H** are supposed to converge to the desired best form, $\mathbf{H}_{conv}$.

2. See **Figure 8,** *Simplified "White Box" View of a Typical Echo Canceller,* on page 8.

The consecutive vectors approaching $\mathbf{H}_{conv}$ are denoted as $\mathbf{H}(m)$, where m (m=0,1,2, . . .) is a consecutive iteration. For a given m, the modified vector expression for H(m) is as follows:

$$\mathbf{H}(m) = [h_0(m), h_1(m), . . . , h_{N-1}(m)]^T$$

The process of minimizing J through adaptive changes or iterations $\mathbf{H}(m)$ can be achieved via the Least Mean Squares (LMS) algorithm, popularized by Widrow [11, 12, 24]. The algorithm can be written, for kth iteration, as follows:

$$y(k) = \mathbf{H}^T(k) \cdot \mathbf{X}(k) \qquad \text{Filter Output}$$

$$e(k) = d(k) - y(k) \qquad \text{Error Signal}$$

$$\mathbf{H}(k+1) = \mathbf{H}(k) + \mu \cdot e(k) \cdot X(k) \qquad \text{Adaptation Formula}$$

where μ is the step size of the LMS algorithm and k is a sample number that is also the iteration number. Because the sample number coincides with the iteration number, the algorithm is said to perform a per-sample adaptation. Many other versions of the LMS algorithm have been developed. One of them is a per-block version of adaptation, which has become popular because of its computational processing efficiency.

Because of its simplicity, relatively low cost in terms of the number of additions and multiplications, and its relatively good convergence properties, the LMS algorithm is often used in practical applications. One of the prominent applications of the LMS algorithm is the LMS-based adaptive filter for network/line echo cancellers. As the LMS algorithm equations indicate, the algorithm is implemented in three main computational steps: the output of the transversal (FIR) filter, the error, and the updated coefficients of the FIR filter.

7.2 Normalized LMS Adaptation (NLMS)

The LMS algorithm uses μ, a small constant that determines, among other things, the speed of the algorithm convergence. One practical problem of using μ is to ensure that μ does not become large enough in relation to the input signal energy to cause the algorithm to diverge. A normalized LMS adaptation algorithm (NLMS) is a straightforward version of the classical LMS algorithm in which a new adaptation constant is selected through normalization of the original constant with respect to the signal power. Although it requires slightly more computational effort than its LMS prototype, the NLMS algorithm is widely used in network/line echo cancellers because of its simplicity and acceptable convergence speed and depth. The algorithm governing equations are given as follows:

$$y(k) = \mathbf{H}^T(k) \cdot \mathbf{X}(k) \qquad \text{Filter Output}$$

$$e(k) = d(k) - y(k) \qquad \text{Error Signal}$$

$$\mathbf{H}(k+1) = \mathbf{H}(k) + \frac{\mu \cdot e(k) \cdot \mathbf{X}(k)}{\gamma + \mathbf{X}^T(k) \cdot \mathbf{X}(k)} \qquad \text{Adaptation Formula}$$

where

$$\frac{\mu}{\gamma + \mathbf{X}^T(k) \cdot \mathbf{X}(k)}$$

is the NLMS algorithm step size, γ is a protection term to ensure that the update term in the adaptation formula does not become excessively large when $\mathbf{X}^T(k) \cdot \mathbf{X}(k)$ temporarily becomes small. As the governing equations indicate, four major computational steps are required to implement the algorithm:

- Output of the transversal (FIR) filter
- Error
- Far-end signal (that is, input signal) power
- Updated coefficients of the FIR filter.

A closer examination of the adaptation formula demonstrates that $\mathbf{X}^T(k) \cdot \mathbf{X}(k)$ (which is a short-term energy estimate of the input signal x(k)) can be efficiently computed using the recursive approach or using single-pole IIR filtering in an appropriate fashion.

7.2.1 NLMS with Adaptive Step Size

The main distinguishing feature of the NLMS with Adaptive Step Size algorithm is a variable step size used during the adaptation process. A classical prototype of the NLMS algorithm is the Gradient Adaptive Lattice (GAL) algorithm [11], which offers very similar convergence properties. Specific strategies for making the step size variable are subjects of IP and are not addressed in this document.

7.2.2 PNLMS

The proportionate normalized least-mean-squares (PNLMS) adaptation algorithm yields faster convergence time than the NLMS adaptation algorithm. In both cases, the convergence depth is very similar. PNLMS differs from NLMS in that the input signal energy is distributed unevenly over the *N* tabs of the adaptive FIR filter. While the governing equations for PNLMS remain largely intact, the update term is modified to include the **G**(k) factor representing the energy distribution over the filter tabs so that the PNLMS adaptation formula becomes as follows:

$$\mathbf{H}(k+1) = \mathbf{H}(k) + \frac{\mu \cdot \mathbf{G}(k) \cdot e(k) \cdot \mathbf{X}(k)}{\gamma + \mathbf{X}^T(k) \cdot \mathbf{X}(k)} \qquad \text{Adaptation Formula}$$

where **G**(k) is a diagonal matrix for energy distribution that is determined by analyzing successive iterations of **H**(k)). Equations for **G**(k) are not elaborated here. Because of a significant computational penalty imposed by the modified adaptation formula combined with the generation of matrix **G**(k), the PNLMS algorithm appears to be most suitable for implementation in an ASIC technology [25].

7.3 Block NLMS

Some signal processing applications require adaptive filters to have lengths that exceed a few hundreds taps. Block processing of data by (N)LMS leads to algorithm modifications frequently earmarked as block LMS (BLMS) or block NLMS (BNLMS). For the BLMS, the adaptive filter coefficients are updated every block of data (of length L) instead of every sample. Thus, instead of using the formula for sample LMS adaptation in the following form:

$$\mathbf{H}(k+1) = \mathbf{H}(k) + \mu \cdot e(k) \cdot \mathbf{x}(k) \qquad \text{LMS Adaptation Formula}$$

where k denotes the sample number, $\mathbf{x}(k) = [x(k), x(k-1), \ldots, x(k-N+1)]^T$, $\mathbf{H}(k) = [h_0(k), h_1(k), \ldots, h_{N-1}(k)]^T$, we use the following formula[3]:

$$\mathbf{H}(n+1) = \mathbf{H}(n) + \mu_B \cdot \frac{1}{L} \cdot \sum_{i=0}^{L-1} e(nL+i) \cdot \mathbf{x}(nL+i) \qquad \text{BLMS Adaptation Formula}$$

where L is the block length, μ_B is the block size parameter, and *n* is the block index. For the computation of the error e(nL+ i) = d(nL+1) – y(nL+i), i=0,1, . . . ,L – 1, the output y(nL+ i) = $\mathbf{H}^T(n)\mathbf{x}(nL+i)$ is calculated using the coefficient update **H**(n) from the previous block. The **H**(n) in the per-block version differs from the classical **H**(n) in which adaptation occurs every sample. Therefore, one might introduce a separate symbol for adaptation coefficients. In this document, the number of symbols is minimized, though precision of description may suffer. With slightly different definitions of signals used in the block adaptation formula, we can simplify otherwise more complicated notation. Therefore, we define **X**(n) as a new matrix using **x**(n) columns, as follows:

$$\mathbf{X}(n) = [\mathbf{x}(nL), \mathbf{x}(nL+1), \ldots, \mathbf{x}(nL+L-1)]^T$$

The column vectors such as **d**(n), **y**(n) and **e**(n) are defined as follows:

$$\mathbf{d}(n) = [d(nL), d(nL+1), \ldots, d(nL+L-1)]^T$$

$$\mathbf{y}(n) = [y(nL), y(nL+1), \ldots, y(nL+L-1)]^T$$

$$\mathbf{e}(n) = [e(nL), e(nL+1), \ldots, e(nL+L-1)]^T$$

Note that

$$\mathbf{y}(n) = \mathbf{X}(n)\mathbf{H}(n)$$

and

$$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{y}(n)$$

Also note that

$$\sum_{i=0}^{L-1} e(nL+i) \cdot \mathbf{x}(nL+i) = \mathbf{X}^T(n) \cdot \mathbf{e}(n)$$

Thus, by substituting the preceding identity in the expression for block adaptation, we arrive at a compact formula to the block adaptation:

$$\mathbf{H}(n+1) = \mathbf{H}(n) + \frac{\mu_B}{L} \cdot \mathbf{X}^T(n) \cdot \mathbf{e}(n) \qquad \text{BLMS Adaptation Formula}$$

Although the formulas for LMS and BLMS look similar, the difference between them is substantial: the BLMS operates on blocks so that the adaptation process requires much less computational effort than the corresponding LMS adaptation process. As simulation results (not included) demonstrate, the convergence rate and depth of the BLMS algorithm are very close to those of the LMS algorithm, if L is not too large. Therefore, BLMS is an attractive alternative to LMS because it offers reduced computational complexity with little, if any, reduced performance. The adaptation formula for the BNLMS algorithm can be easily derived using a similar matrix representation technique. The normalization can be performed using a common step-size parameter (one normalization value per block of length L), or the normalization can be performed for each element of **H**(n), a technique known as step-normalization [34].

3. Note that the input vector **x**(k) corresponds to vector **X**(k) used in **Section 7.1**, *Least Mean Squares (LMS)*).

7.4 Transform Domain LMS

Most time domain-based LMS algorithms manifest convergence slow downs when they are exposed to highly self-correlated signals such as human speech. The frequency domain can be used to decorrelate the input signal of the LMS algorithm. One approach is to apply a Fourier series decomposition to the input signal **x**(k) that yields the coefficients for each frequency component. The base Fourier series functions are defined as follows:

$$\{\exp(jn\omega_o t)\},\ \text{where } j = \sqrt{-1}$$

These functions form an orthogonal set, so the different frequency components should be totally uncorrelated. After each frequency component is weighted and normalized with the inverse of its power, a Fourier synthesis is performed, and the synthesized signal is subtracted from the desired signal d(k). The weights can be adjusted via the LMS algorithm. This approach leads to more general transform domain LMS (TDLMS) algorithms. TDLMS algorithms form an interesting and practical class of adaptive algorithms addressing the concerns related to the correlated input signals. Not only the Fourier transform but any other orthogonal transform can be used to decorrelate the input signals [34–36]. Adaptive filters incorporating transform domain algorithms are frequently referred to as transform domain adaptive filters (TDAF). TDAFs implement the adaptation operations in the given transform domain, as opposed to classical time domain adaptive filters, where the adaptation is performed using signal samples. **Figure 19** shows two TDAF configurations. In the (a) portion, the error signal is computed in the time domain; in (b), the error is computed in the transform domain. Details of these two basic configurations are not exposed. T denotes an orthogonal transform; T^{-1} is the inverse transform.

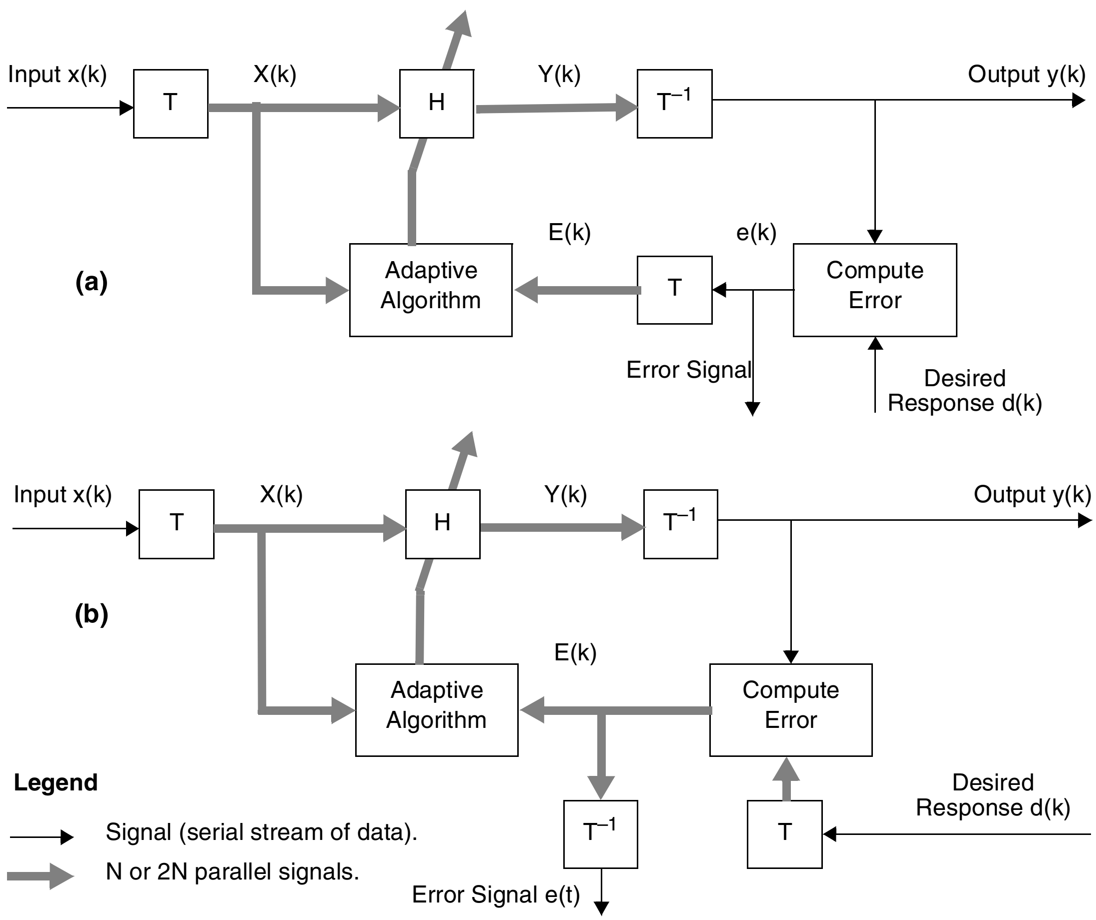


Figure 19. Two TDAF Configurations

Although there are infinite possible choices for the transformations used in TDAF (and specifically for TDLMS algorithms), only a few transforms have been adopted in practical applications. Main features of these transforms are (a) orthogonality and (b) adaptability to fast (that is, computationally inexpensive) implementations. For TDAF applications, the most frequently used orthogonal transforms are:

- Complex Discrete Fourier Transform (DFT)
- Real Discrete Fourier Transform (RDFT)
- Discrete Hartley Transform (DHT)
- Discrete Cosine Transform (DCT)
- Walsh-Hadamard Transform (WHT).

Like the classical time domain adaptive filters, the TDAF can be implemented in two different methods:

- *Per-sample*. Uses sliding window transforms.
- *Per-block*. Processes an entire block of the input data.

Both methods encompass several TDAF architectures and their respective algorithms.[4] This discussion is limited to one example of TDAF. Let us first review the Frequency Domain NLMS (FDNLMS) algorithm, which uses a sliding window Discrete Fourier Transform (DFT). The block version of the FDNLMS algorithm can be derived from the block version of the time domain NLMS algorithm described in **Section 7.3**. The governing formula for the FDNLMS adaptation algorithm is very similar to the formula for the LMS algorithm, except that some variables are represented in the frequency domain. Thus, the recursion formula for the FDNLMS algorithm can be presented in a vector form as follows:

$$\mathbf{H}_T(k+1) = \mathbf{H}_T(k) + \mu \cdot \underline{\mathbf{D}}^{-1} \cdot e(k) \cdot \mathbf{x}_T(k) \qquad \text{DFNLMS Adaptation Formula}$$

where **D** is an estimate of a diagonal matrix **D**, which is defined as follows:

$$\mathbf{D} = \begin{bmatrix} E\{x^2_{T,0}(k)\} & & & 0 \\ & \cdot & & \\ & & \cdot & \\ 0 & & \cdot & E\{x^2_{T,N-1}(k)\} \end{bmatrix}$$

where the subscript T refers to the DFT (or FFT), although it can be generalized for any of the preceding transforms; μ refers to the step-size adaptation parameter. Error e(k) can be defined exactly the same way as for the time domain LMS, that is, e(k) = d(k) – y(k). The presence of $\underline{\mathbf{D}}$ in the DFNLMS adaptation formula is equivalent to using different step-size parameters at various taps of the filter.

Each step-size parameter is chosen proportional to the inverse of the power of its associated input signal. The transformation **T** converts the set of the input sample vector $\mathbf{x}(k) = [x(k), x(k-1), \ldots, x(k-N+1)]^T$ to their transform-domain vector $\mathbf{x}_T(k) = [x_{T,0}(k), x_{T,1}(k), \ldots, x_{T,N-1}(k)]^T$, (that is, $\mathbf{x}_T = \mathbf{T}\{\mathbf{x}\}$). The filter output is obtained according to the following equation.

$$y(k) = \mathbf{H}_T^T \cdot \mathbf{x}_T(k)$$

where $\mathbf{H}_T(k) = [H_{T,0}(k) \quad H_{T,1}(k) \ldots \quad H_{T,N-1}(k)]^T$. Although $\mathbf{x}_T(k)$ is in the transform domain, the filter output is in the time domain. **Figure 20** illustrates the algorithm architecture.

4. For a comprehensive review of this large class of transform domain adaptive filters, consult [35]. Many important and interesting details on transform domain algorithms can be found in [11, 34].

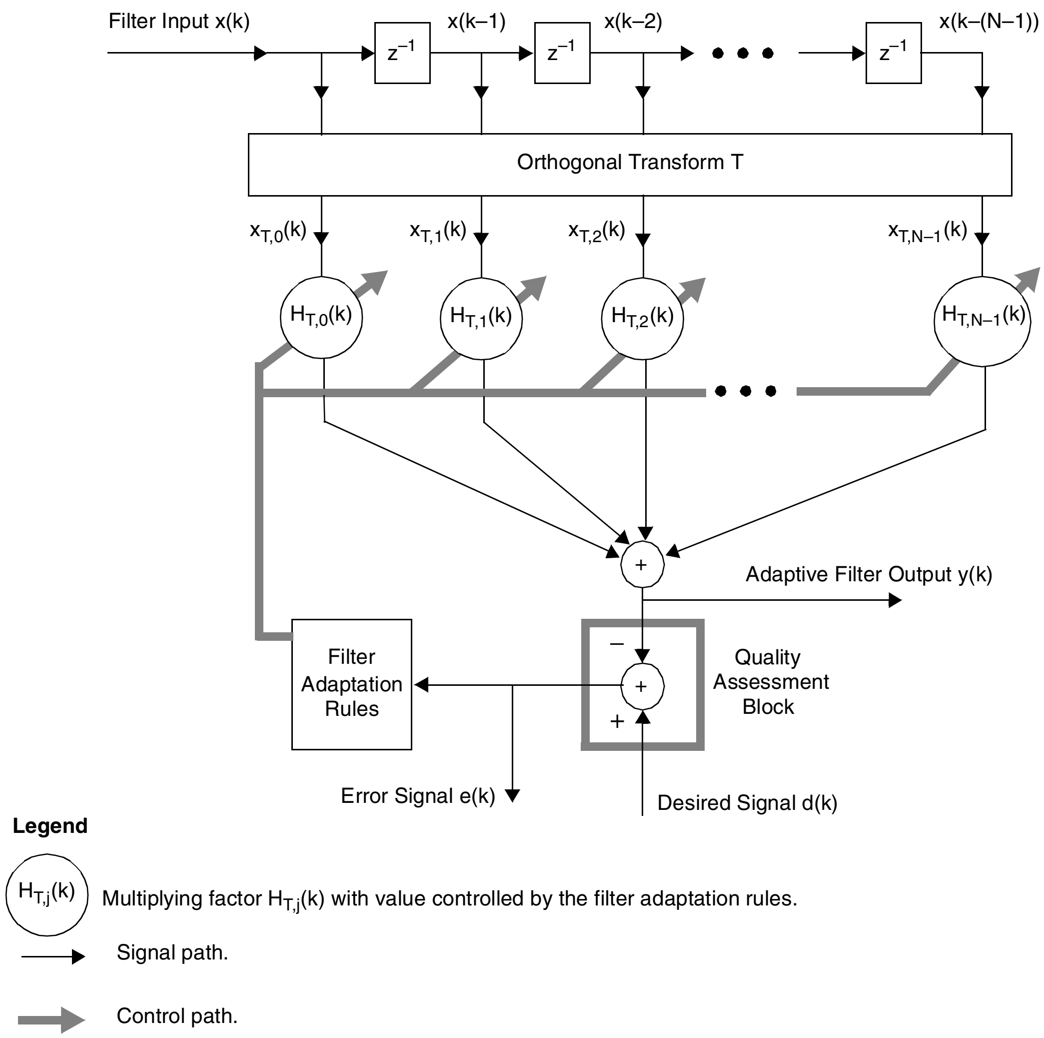


Figure 20. An Exemplary Architecture of a Transform Domain LMS Adaptive Filter

7.5 Affine Projection

The affine projection (AP) algorithm has relatively moderate computational complexity, which is only moderately higher than in the NLMS, and its convergence speed is significantly faster than that of the NLMS algorithm. Affine projection algorithms are being vigorously researched, and several more advanced versions have already been published, including a fast AP (FAP) algorithm and a sub-band FAP [28, 29]. The AP algorithm is a generalization of the NLMS adaptive filtering algorithm. Using notation similar to that in previous sections of this document, one can define the AP algorithm as follows [28]:

$$\underline{e}_n = \underline{S}_n - \mathbf{X}_n{}^T\, \mathbf{h}_{n-1}$$

$$\varepsilon_n = [\mathbf{X}_n{}^T\, \mathbf{X}_n + \delta\mathbf{I}]^{-1}\, \underline{e}_n$$

$$\underline{h}_n = \underline{h}_{n-1} + \mu\mathbf{X}_n\varepsilon_n$$

The excitation signal matrix $\mathbf{X}_n$ is L by N, and it has the following structure:

$$\mathbf{X}_n = [\underline{\mathbf{x}}_n, \underline{\mathbf{x}}_{n-1}, \ldots, \mathbf{x}_{n-(N-1)}]$$

where the vectors $\underline{\mathbf{x}}_n = [x_n, x_{n-1}, \ldots, x_{n-(L-1)}]^T$ are transposed sample vectors of length L. The adaptive filter tap weight vector is $\underline{\mathbf{h}}_n = [h_{0,n}, h_{1,n}, \ldots, h_{L-1,n}]^T$, where $h_{i,n}$ is the i^{th} tap at sample time n. The vector $\underline{e}_n$ is of length N and represents the residual echo (that is, adaptation error). N is a projection parameter, (N = 1, 2, . . . , L –1). The N-length vector, $\underline{s}_n$, is the system output consisting, in general, of the response to the excitation $\mathbf{X}_n$ combined with an additive system noise $\underline{n}_n$.

$$\underline{s}_n = \mathbf{X}_n^T \underline{\mathbf{h}}_{EP} + \underline{n}_n$$

where $\mathbf{h}_{EP}$ is the echo path impulse response. The scalar δ is the regularization parameter for the sample autocorrelation matrix inverse used in the formula for ε_n. The step-size parameter, μ, is the relaxation factor. Like the NLMS, the AP algorithm is stable for $0 = \mu < 2$. If N is set to 1, the preceding formulas reduce to the ones representing the NLMS algorithm.

7.6 Sub-Band (N)LMS

Adaptive filtering in sub-bands is motivated by the following well-known problems occurring in full-band adaptive filtering:

- Convergence and tracking (that is, reconvergence) can be slow if the input signal correlation matrix is ill conditioned (a typical case for the speech input).
- RLS and AP algorithms assure fast convergence and fast tracking, but high numerical precision is required, and the algorithms may become numerically unstable.
- Even for NLMS-based echo cancellers, high numerical precision may be required if the adaptive filter length is large.
- High order full-band adaptive filters are computationally expensive. Therefore, echo cancellers that employ them become bottlenecks in providing competitive system solutions with high channel density.

The sub-band approach is intended as a reasonable compromise in addressing these problems. The filter bank is used to split the input signals into several frequency bands. Within these frequency sub-bands, several independently operating adaptive filters perform classical adaptation. Splitting the signals into sub-bands accomplishes two tasks:

- Signal whitening (or decorelating) to improve convergence and tracking speed.
- Reducing sampling rate within the sub-bands to save computational efforts and to shorten the filter length of the sub-band adaptive filters. This measure, in turn, leads to relaxed arithmetic precision requirements.

Figure 21 presents a general outline of an example sub-band (N)LMS method. Practical implementations of sub-band echo cancellers incorporate several sub-band adaptive filters.

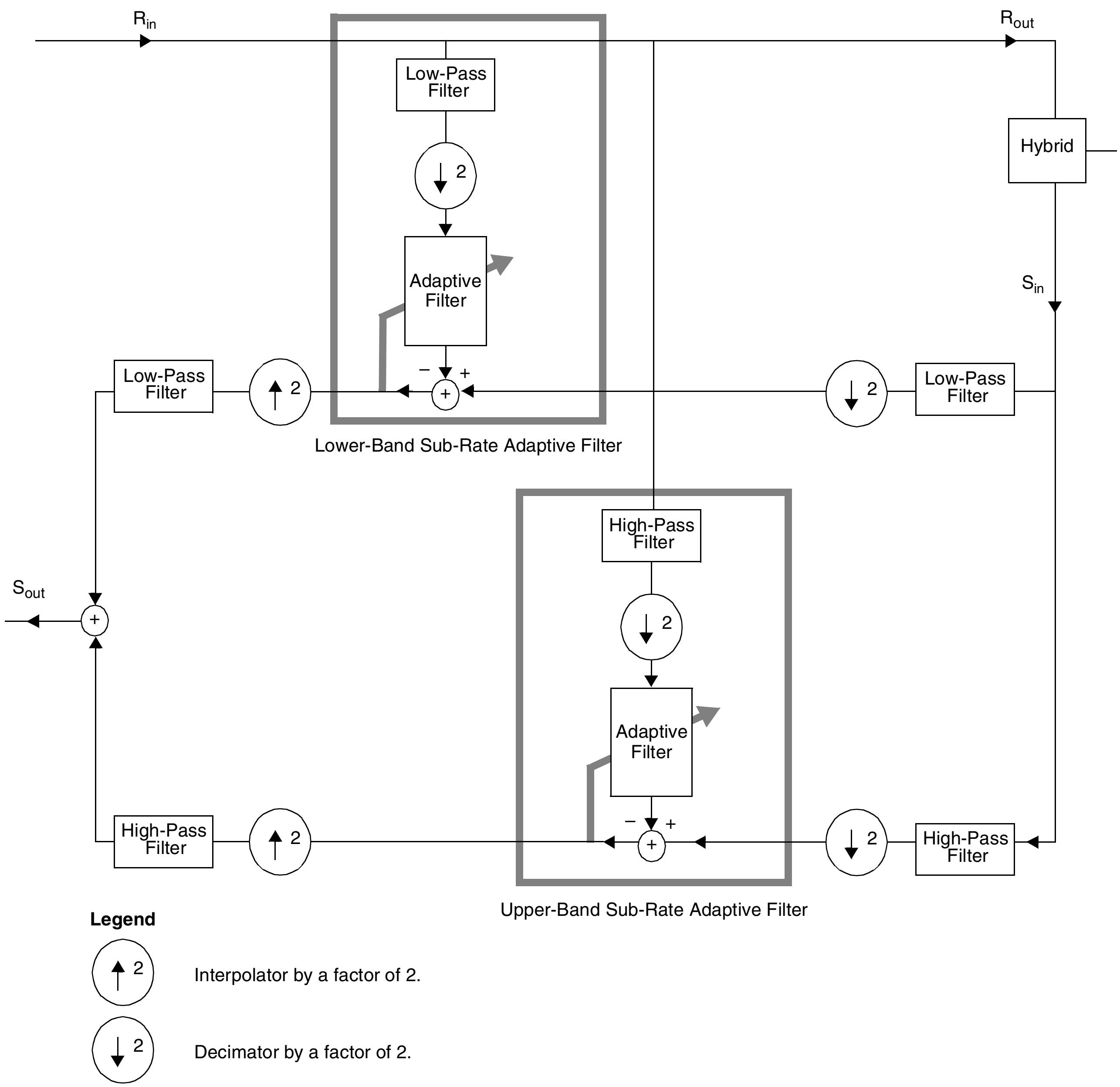


Figure 21. Generalized View of a Sub-Band Echo Canceller With Two Sub-Bands

Through signal spectrum partitioning, individual echo cancellers operating on signals of narrower spectrums converge faster. LMS adaptation processing in separate sub-bands increases convergence speed since the adaptation size in each sub-band can be better matched to the energy of the input signal in that band [27]. Implementation costs of sub-band echo cancellers include pass-band filtering and multiple adaptation processes. These costs are partly offset through sub-rate processing [16, 27]. However, subsampling processes often introduce aliasing, particularly when typical FIR filters of relatively low cost are used for band partitioning. These and related aspects of sub-band adaptive filtering are covered in detail in [29], which also includes an extensive list of related references. The paper concludes that selection of filters for band partitioning is important to echo cancellation performance. The paper focuses on the use of an all-pass polyphase IIR filter, concluding that extensive simulations performed with both noise signals and speech demonstrate the potential usefulness of these filters in practical implementations.

7.7 Double Talk Detector Block

The Double Talk Detector function controls adaptive filter behavior during periods when the near-end signal, also known as the double talk signal, reaches a significant level. The main purpose of the Double Talk Detector is to avoid adaptation whenever the double talk signal is present. Otherwise, the adaptation process leads to incorrect adaptive filter coefficients. Double-Talk Detector/Near-End Voice Activity Detectors are echo canceller system sub-blocks whose performance can have a significant impact on voice quality.

7.7.1 Geigel and Modified Geigel

When the NE speaker talks, and the respective speech signal levels observed at S_{in} are above a predefined threshold, double talk occurs. An NE voice detector operating on the NE signal can effectively detect double talk status. A commonly used algorithm originated by A. A. Geigel, [12], compares $S_{in}(n)$ with a short-term history of $R_{out}(n)$, as follows:

$$|S_{in}(n)| \geq \frac{1}{2} \cdot \max\{|R_{out}(n)|, |R_{out}(n-1)|, \ldots, |R_{out}(n-N)|\}$$

where N is the FIR adaptive filter order. When an active status of double talk is declared, the filter adaptation process is frozen. The factor of one half in the preceding inequality is based on the assumption that ERL is at least 6 dB. A closer analysis of the Geigel algorithm leads to the following observations:

- It does not reliably flag the double talk status if the ERL threshold does meet the 6 dB condition.
- It cannot be directly adopted to non-zero pure delay network connections.
- It has limited features allowing for tuning up.
- It manifests poor performance in a noisy environment.

A more reliable version of the algorithm uses short-term power estimates of the respective signals. Typically, the short term-power estimates are calculated using single pole IIR filters or arithmetical averages with FIR filters applied to instantaneous powers of signals under consideration.

7.7.2 Cross Correlation-Based Method

The cross correlation method estimates the cross correlation function R and its maximum value. Specifically, it estimates the following function of R(k):

$$R(k) = \frac{E\{R_{out}(n) \cdot S_{in}(n+k)\}}{\sqrt{E\{R_{out}(n) \cdot R_{out}(n)\}} \cdot \sqrt{E\{S_{in}(n) \cdot S_{in}(n)\}}}$$

If, for certain k:

$|R(k)| > DTCT$

where DTCT is a double-talk correlation threshold and $E\{\cdot\}$ is an expectancy operator that can be associated with long-time averaging, the double-talk condition is declared. Since human speech manifests short-term correlation properties, this approach requires some tuning before it can be used in an echo canceller product. Several variants of this approach are based on correlating the R_{out} and S_{in} signals.

7.7.3 Intelligent Double Talk Detector

The intelligent double talk detector is a non-Geigel approach that operates on adaptive mechanisms. Because its details constitute IP, they are not described in this document.

8 Miscellaneous Echo Canceller Features

Three important aspects of an echo canceller are tone and signal saturation detection, control, and debugging features.

8.1 Tone and Signal Saturation Detection Features

Echo canceller detection features usually include the following:

- *Tone Detection*. An echo canceller must react properly to special signaling and control tones. Standard ITU-T C5, C6, and C7 signaling tone detectors disable an echo canceller, thus allowing for compatibility with network in-band signaling. G.164/G.165 and G.168 echo canceller disable tone detectors allow for in-band control of echo cancellers. Tone detectors can be a part of the echo canceller system, or they can be placed outside the system.
- *Phase discontinuity detection*. This feature is used for G.165/G.168 echo canceller phase reversal tone detection.
- *Saturation detection*. An echo canceller may encounter signals that are distorted due to saturation (clipping effect). Signal saturation is an abnormal condition beyond echo canceller control. Nevertheless, an echo canceller design should be robust enough to behave predictably when such a condition occurs. Since the signal spectrum can be significantly distorted, the adaptation process may lead to quite different adaptive filter coefficients than a similar signal without clipping distortions. To prevent adaptation during this abnormal condition, some echo cancellers are equipped with signal saturation detectors, which temporarily disable the adaptation process.

8.2 Control Features

An echo canceller is equipped with a variety of control/set up options, depending on the particular design. However, certain control options are mandatory, as stated in the G.165/G.168 Recommendation:

- Adaptive filter coefficient reset.
- Coefficient freeze/unfreeze.
- Three combinations of NLP/CNG on/off (NLP off CNG on is not a meaningful combination).

8.3 Debugging Features

Special network configurations, particularly private networks, may present unexpected echo formation. It is desired if an echo canceller system is equipped with special debugging features allowing for storing segments of PCM samples corresponding to R_{in} and S_{in}. Also, a facility allowing filter coefficients to be stored at a desired time during the progress of the telephone call is helpful when dealing with special network configurations. Other debugging features are available in some echo canceller products. Since these are specific and often IP-related features, this document does not discuss them.

9 Freescale Packet Telephony Echo Cancellers

Freescale echo cancellation solutions include a set of network echo cancellers to provide carrier-class echo cancellation service in various telecommunications markets. There are two echo cancellers designs:

- *Sparse*. Employs a dynamically positioned window to cover a desired echo tail length. A Freescale sparse echo canceller uses a 24 ms moving window covering a 128 ms echo path delay, which is sufficient for all the practical applications in near-end echo cancellation, as described in **Section 2**.
- *Non-Sparse*. Applies a full-length adaptive filter to cover a specific echo tail length. It uses full-window adaptive filter algorithms to cover a 24 ms echo path delay. The non-sparse design provides a solution with higher channel density and lower resource usage for applications with relatively short echo tail length coverage.

The echo cancellers are based on NLMS adaptive algorithms equipped with the proprietary enhancements in some key elements, such as adaptive filter, echo canceller controller, bulk delay estimator, double talk detector, and nonlinear processor, as shown in **Figure 22**.

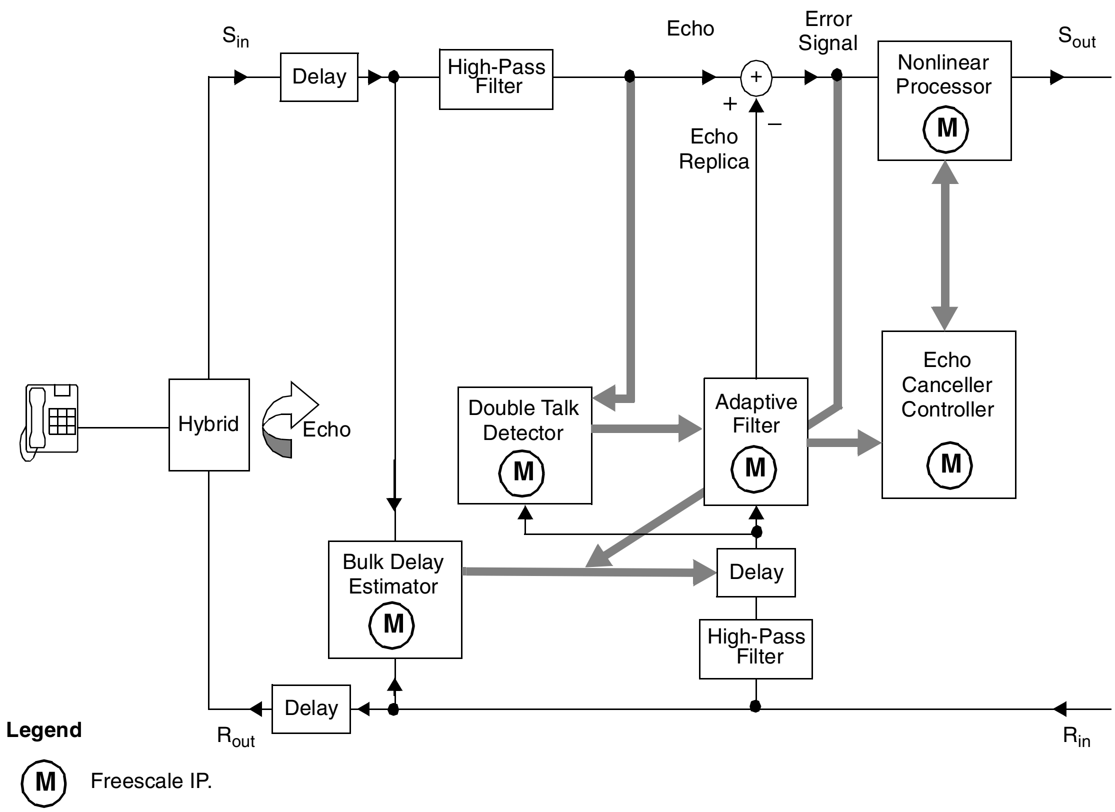


Figure 22. Functional Blocks of Freescale Echo Canceller and Freescale IP Coverage

Freescale echo cancellers have the following major features:

- ITU-T G.168 (2000/2002) compliant.
- High channel density (low processing load with 2.6 MCPS on average).
- Rapid initial convergence rate (echo cancelled within 50 ms).
- Deep and stable infinite convergence depth.
- Intelligent near-end talk single detection mechanism.

- Adaptive nonlinear processor with optional comfort noise matching.
- Innovative mechanism for near-end background noise handling.
- Effective background processing for divergence prevention.
- Configurable full-window and sparse echo cancellers covering different echo path delays.
- Fast middle-call convergence, that is, reconvergence following a hybrid change in the middle of a telephone call.
- Special mechanisms for maintaining system stability.

In addition to ITU-T G.168 tests, the echo cancellers are systematically evaluated with Freescale proprietary voice quality evaluation techniques and benchmarked against the echo cancellers from several market leaders. **Figure 22** illustrates the major functional components of the Freescale echo cancellers. The sparse monitor and bulk delay estimator are for the sparse echo canceller only.

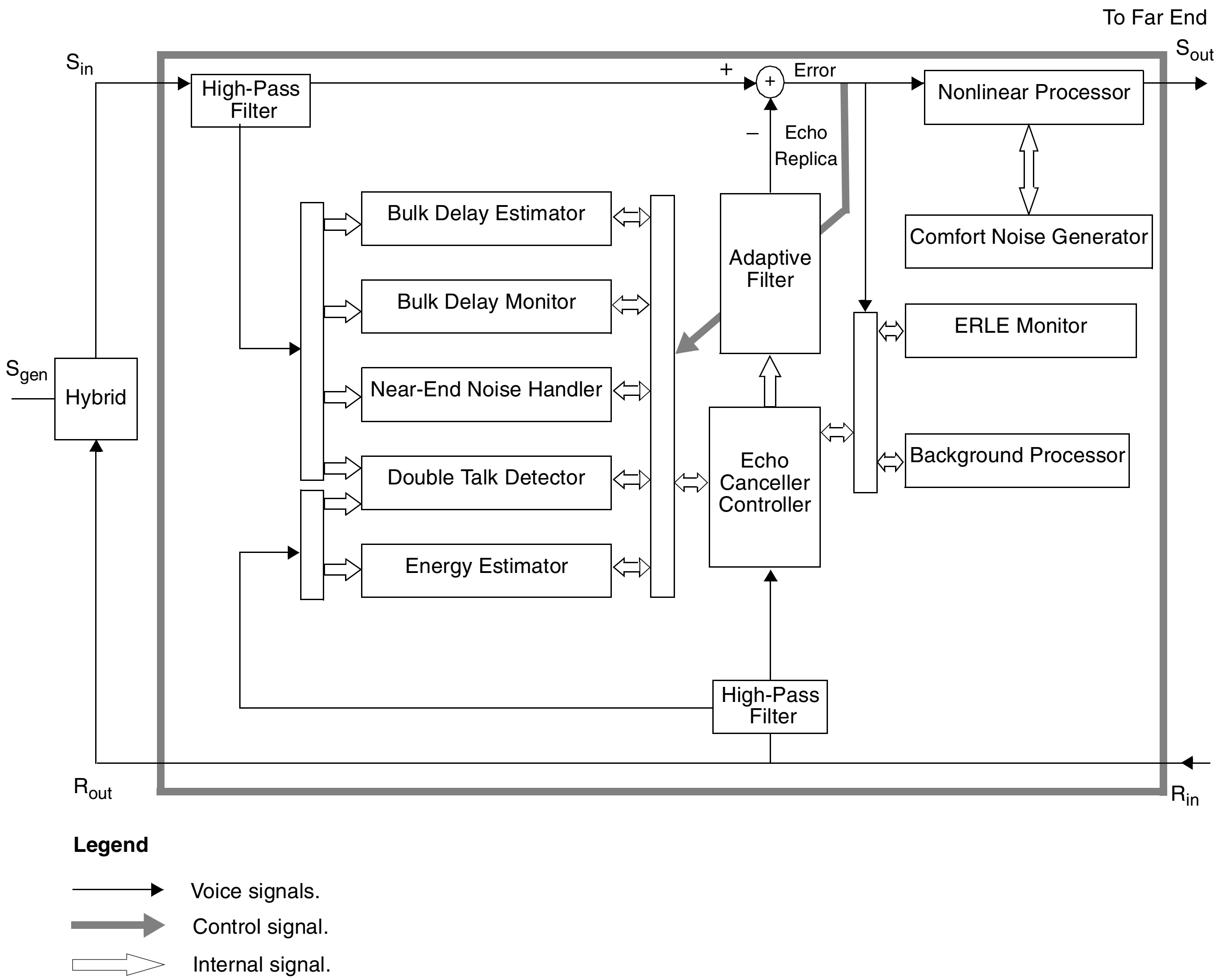


Figure 23. Functional Components of a Freescale Echo Canceller

Table 4 describes each of the components depicted in **Figure 23**. Further detailed descriptions, including algorithms, are available in a separate document under NDA. Additional details of some solutions can be found in [32, 33].

Table 4. Functional Components of Freescale Echo Cancellers

Component	Description
High-Pass Filter (HPF)	There are two low cost, high performance high-pass filters on both the receiving and sending paths of the echo canceller. The filters remove DC component from R_{in} and S_{in}, respectively.
Adaptive Filter	An NLMS-based FIR filter that produces an estimated echo replica of the R_{in} signal. The filter coefficients are updated on the basis of an estimated residual echo obtained from the error estimator.
Error Estimator	Calculates error signals (that is., residual echo) based on the echo replica and S_{in} signals. The adaptive process uses the error signal to generate the echo replica.
Bulk Delay Estimator	A sub-rate process that estimates the pure delay of the echo path to determine the location of the sparse window for the adaptive filter.
Bulk Delay Monitor	A process closely related to the bulk delay estimator that, at a sub-rate sampling frequency, monitors the change of the pure delay of the echo path.
ERLE Monitor	Monitors ERLE for the filter update and other processes.
Energy Estimator	Uses innovative algorithms for R_{in}, S_{in}, and S_{out} energy estimation. This process improves performance of double talk detection, nonlinear processing, and comfort noise generation.
Double Talk Detector	The echo cancellers apply an intelligent double talk detection algorithm for faster and more accurate double talk detection than is available through conventional double talk detection algorithms, such as Geigel or Geigel-derivative algorithms. This intelligent algorithm is very sensitive, so it can detect double talk at an early stage to prevent divergence. It reduces the rate of false detection, thus speeding up the adaptation process, if necessary.
Nonlinear Processor	Eliminates residual echo, if it is below a certain level, by replacing it with either comfort noise when the comfort noise option is on or silence when the comfort noise option is off.
Comfort Noise Generator	An optional process that generates pink-like comfort noise to match the level of background noise.
Background Processor	Controls the filter coefficient adaptation, the double talk detection threshold, and so on.
Near-End Noise Handler	The echo cancellers implement an innovative technology for near-end noise handling to avoid noticeable or annoying switching of the perceived background noise when the near-end background noise level is relatively high.

9.1 Benefits of StarCore

Echo cancellers, especially carrier-class echo cancellers, can be computationally intensive. A software-based echo canceller running on a DSP in a packet telephony voice solution normally adds a good portion, even a major portion, of the processing load to the DSP. Therefore, the echo canceller has a major impact not only on voice quality but also on channel density and price per channel. The challenge in developing an echo canceller is to achieve high voice quality while minimizing the processing load (MCPS). The Freescale StarCore™ architecture provides a superb platform to meet this challenge. The StarCore SC140 core is equipped with four arithmetic logic units (ALUs) and two address arithmetic units (AAUs), as shown in **Figure 24**. The SC140 core can execute 6 instructions, including 4 arithmetic/logic and 2 data movement instructions, in only one cycle. For echo cancellation applications, the SC140 core can be extremely effective, especially in handling the FIR and coefficient updates. For example, it can perform the following tasks in one cycle:

1. Load a set of data to be processed for the next cycle.
2. Process a set of data loaded during the last cycle.

The instructions to execute can be the same, such as four MACs, or different, such as two MACs, one shift, and one logic AND.

3. Restore a set of data already processed during the last cycle.

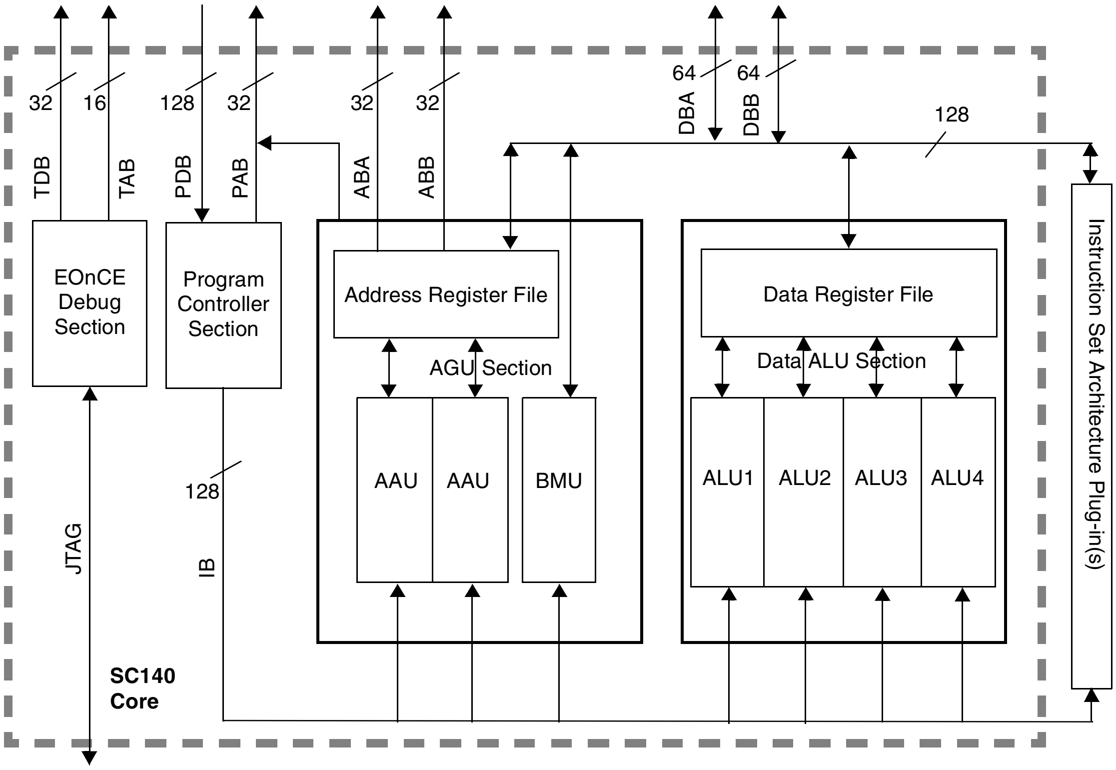


Figure 24. StarCore SC140 Core Block Diagram

Figure 25 shows examples to demonstrate the efficiency and flexibility of the SC140 core.

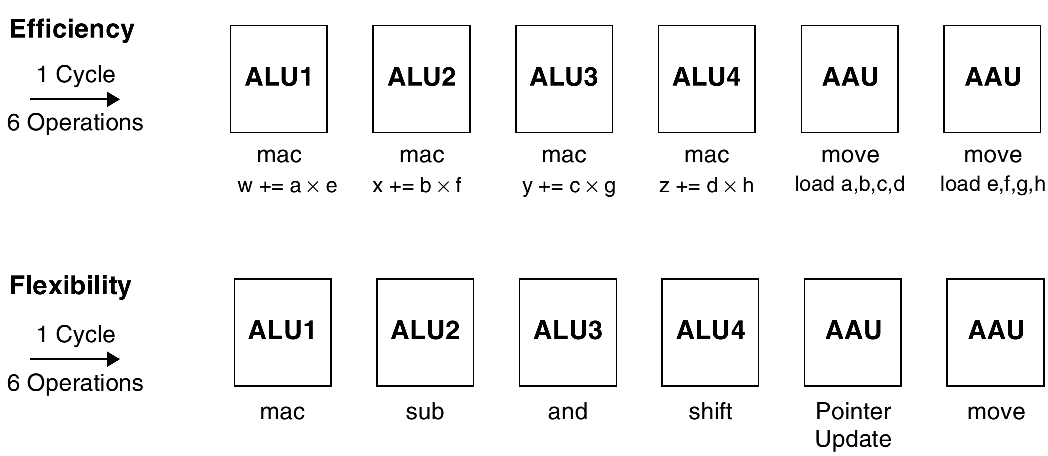


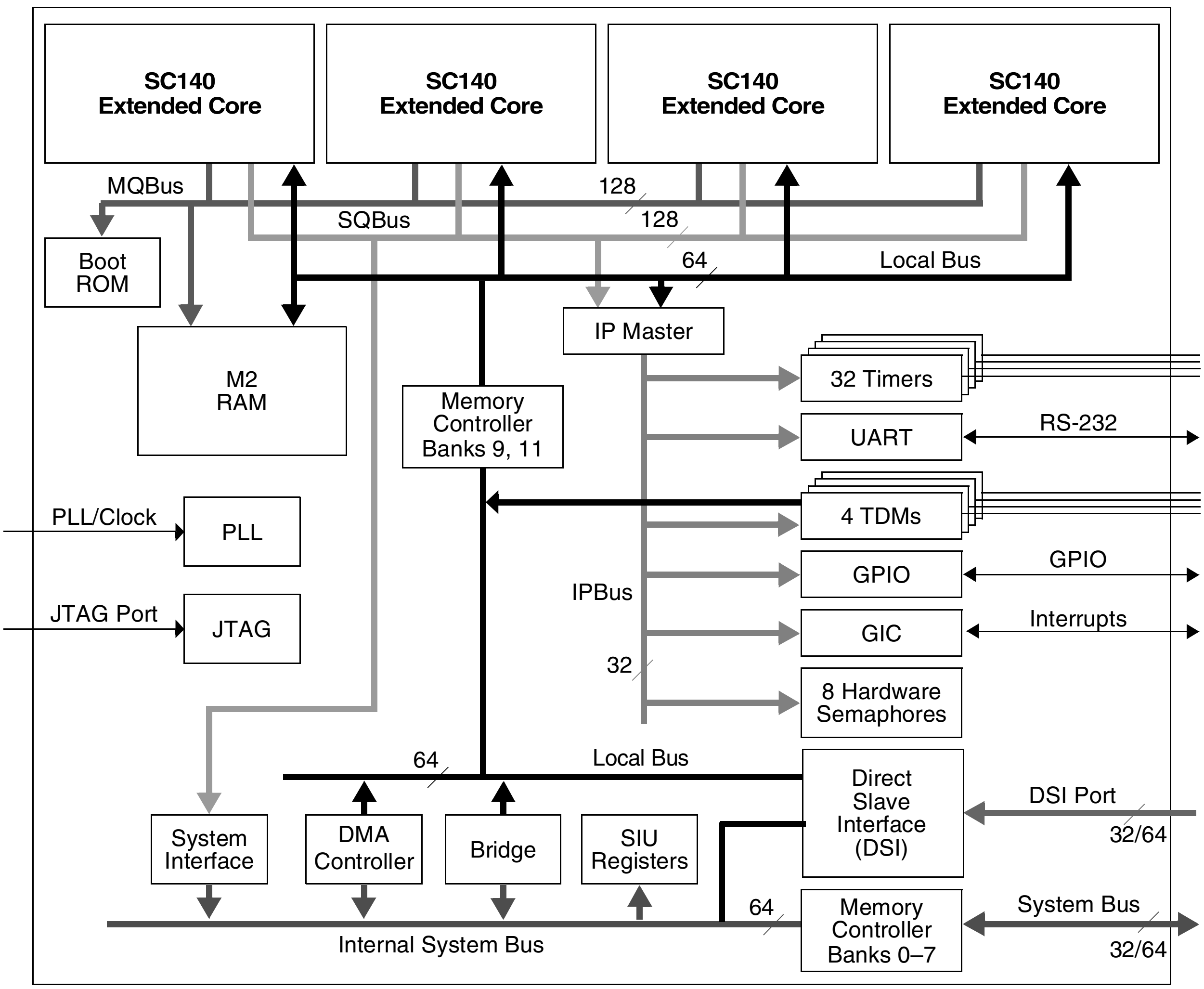
Figure 25. SC140 Efficiency and Flexibility

9.1.1 MSC8101/MSC8102

The Freescale MSC8101 and MSC8102 devices are both based on the SC140 core. The MSC8101 device is equipped with one SC140 core, a programmable communications processor module (CPM), and a 60x-compatible bus interface. This combination of features offers advanced signal processing performance, flexible network connectivity, and seamless system integration. The MSC8102 device is equipped with four SC140 cores (see **Figure 26**), thus offering a DSP-Farm-on-a-Chip level of performance integration. The MC8102 is well suited for computationally intensive DSP applications, such as echo cancellation. Both the MSC8101 and MSC8102 devices are suitable for a wide range of channel density applications. Assuming that both devices run at 300 MHz, with a specific application, such as an echo canceller plus G.711 and some framework (requiring a 3 MCPS processing load), the channel density on each device can be estimated as follows:

- The number of channels supported by one MSC8101 for the application is 100 (=300/3).
- The number of channels supported by one MSC8102 for the application is 400 (=300 · 4/3).

These estimations are based on the processing load (MCPS) only. In a practical system design, we must also take into account the memory usage, data transmission bandwidth, and other factors.



Note: The arrows show the direction from which the transfer originates.

Figure 26. Block Diagram of Freescale MSC8102 DSP Device

9.1.2 Freescale Echo Canceller Evaluations

Freescale packet telephony echo cancellers have been evaluated in three major categories: G.168 tests, voice quality test, and special tests. The G.168 tests include Tests 2–7 and Tests 9–10. Test 1 is combined with Test 2 in the G.168 2000/2002 versions. The remaining tests are optional or under study. Each test includes sub-tests. Refer to G.168 (2000/2002) for specific test conditions and requirements [3–5]. The Freescale StarCore echo cancellers have been extensively tested against G.168 (2000/2002) on all eight G.168 recommended hybrid models, in a combination with different pure echo path delays within the echo span coverage for which the echo canceller is designed. A complete suite of G.168 tests, as implemented by Freescale, exceeds 3000 individual test runs.

The voice quality tests (also called subjective evaluations) are regarded as even more important than the G.168 objective tests. Subjective evaluation methods follow G.168 (2000/2002) subjective testing guidance, with some additional techniques developed by Freescale. The basic set-up requires two persons (far-end and near-end talkers) to talk over a pair of phones on a simulated telephone network. The echo canceller under evaluation is placed into the network at the near-end and cancels the echo perceived by the far-end talker. Both talkers are placed in separate quiet rooms, and the near-end noise level in the near-end quiet room is controlled. The subjective tests focus on the following aspects of the echo canceller:

- Initial convergence rate
- Finite and stable convergence depth
- Double talk detection
- Near-end background noise handling
- Mid-call convergence

In addition to the subjective tests, Freescale has conducted special echo canceller tests to evaluate echo canceller performance under special and/or harsh scenarios to minimize undesirable echo canceller behavior and thereby minimize field failures when echo cancellers are deployed. The special tests normally treat the echo canceller as a black box. Predefined R_{in} and S_{in} files are fed via file I/O to the echo canceller, and S_{out} is recorded for detailed analysis. The predefined R_{in} and S_{in} files include laboratory-generated tone and speech files with various signal levels, double talk patterns and near-end background noises, and field-recorded voice signals Freescale has collected.

10 Conclusions

The future of echo cancellation technology will probably evolve around two application areas:

- Inexpensive echo cancellation solutions covering typical system and market requirements that offer robust performance and high voice quality not only for a typical networking environment but also for dynamically changing network configurations. These solutions are to predominate for software-based echo cancellers for which computational complexity is a decisive factor for mass deployment;
- More sophisticated echo cancellation solutions for which computational complexity is not a primary concern. These solutions include hardware-based systems in which echo cancellation is performed on dedicated silicon devices.

Echo cancellation standards have evolved significantly in recent years. The current ITU-T G.168 2002 standard is not the last word on echo cancellation testing.[5] There is additional interest in pushing the echo cancellation testing methodology further to include voice enhancement devices [37], which are already being standardized. As well, there are groups interested in exploring advanced testing methodologies using complex test signals [38, 39].

5. A new revision, G.168- 06/2004, is already available.

Echo cancellation algorithms, which include adaptive filter algorithms as well as other algorithms supporting the entire echo canceller system, will continue to evolve to meet the increasing demand for perfecting voice quality. LMS family of adaptive filter algorithms will most likely continue to dominate in line/network echo canceller applications. Speech quality and voice quality evaluation methods that correlate with subjective evolutions will improve and expand to cover echo cancellation. The ITU-T Recommendation entitled *Standard Perceptual Evaluation of Speech Quality (PESQ)* [40] is a standardization step in this direction.

11 References

[1] ITU-T Recommendation P800, *Methods for Subjective Determination of Transmission Quality*, August 1996.

[2] ITU-T Recommendation P.831, *Subjective Performance Evaluation of Network Echo Cancellers*, December 1998.

[3] ITU-T G.168, *Digital Network Echo Cancellers* (04/1997).

[4] ITU-T G.168, *Digital Network Echo Cancellers* (04/2000).

[5] ITU-T G.168, *Digital Network Echo Cancellers* (06/2002).

[6] Superior Algorithms for Optimal Audio Quality Assessment, SwissQual, The Network Quality Integrators (06/06/2001, http://www. swissqual.com).

[7] Audio Quality algorithms for the new and emerging demand, SwissQual, The Network Quality Integrators (06/06/2001, http://www. swissqual.com).

[8] H. Yasukawa, et al., "Characteristics of Acoustic Echo Cancellers Using Sub-Band Sampling and Decorrelation Methods," *Electronics Letters*, vol.24, no. 16, August 4, 1988.

[9] ITU-T Com 12-def-E, August 2001, *Initial Draft of New Recommendation G.ECPLAN.*

[10] IEEE Std 743-1995. *IEEE Standard Equipment Requirements and Measurement Techniques for Analog Transmission Parameters for Telecommunications*. Approved 12 December 1995, IEEE Standards Board.

[11] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 1996.

[12] J. R. Treichler et al. *Theory and Design of Adaptive Filters*. Prentice Hall, 2001.

[13] F. P. Duffy, G. K. McNees, I. Nasell, and T. W. Thatcher, Jr., "Echo Performance of Toll Telephone Connections in the United States," *Bell System Technical Journ*al, 1974.

[14] W. A. Sethares et al., "Analysis of Bursting in Telephony Loops with Adaptive Hybrids," *IEEE 1988 International Conference on Acoustics, Speech, and Signal Processing*, New York, April, 1988.

[15] W. A. Sethares et al., "Bursting in Adaptive Hybrids," *IEEE Transactions on Communication*, Vol.37, No.8 August 1989, pp.791–799.

[16] N, J. Fliege, *Multirate Digital Signal Processing: Multirate Systems*. John Wiley & Sons, 1999.

[17] ITU-T Recommendation G.131, *Control of Talker Echo* (1996).

[18] K. El Malki, et al., Echo Cancellation for Voice Over ATM, The University of Sheffield (UK), IEEE 1999.

[19] ITU-T Recommendation G.114, *One-way Transmission Time* (1996).

[20] ITU-T Recommendation G.107, *The E-model, a Computational Model for Use in Transmission Planning* (1998).

[21] ITU-T Recommendation G.108, *The Application of the E-model: A Planning Guide* (09/1999).

[22] ITU-T G.109, *Definition of Categories of Speech Transmission Quality* (1999).

[23] *ITU-T Hybrid Measurements in France Telecom Network*, GEN EC-9 (01/30/01).

[24] B. Widrow and M. Hoff, Jr., "Adaptive Switching Circuits," *IRE WESCON Conv. Record*, pt. 4, 1960, pp.96–104.

[25] D.L. Duttweiler, "Proportionate Normalized Least-Mean-Squares Adaptation in Echo Cancellers," *IEEE Trans. On Speech and Audio Processing*, Vol.8, No. 5, Sep. 2000, pp. 508–518.

[26] K.C. Ho, et al, "New Sampling of Echo Paths in North American Networks," *IEEE International Conference on Acoustics., Speech, Signal Processing*, June 2000, pp.809–812.

[27] A, Gilloire and M. Vitterli, "Adaptive Filtering in Subbands with Critical Sampling: Analysis, Experiments, and Application to Acoustic Echo Cancellation," *IEEE Transactions on Signal Processing*, Vol 40, No. 8, Aug 1992, pp. 1862–1975.

[28] S.L. Gay and S. Tavathia, "The Fast Affine Projection Algorithm," I*nternational Conference on Acoustics, Speech, and Signal Processing*. *ICASSP-95*, Vol. 5, 9-12 May 1995, pp. 3023–3026.

[29] Q.G. Liu et al., "On the Use of a Modified Fast Affine Projection Algorithm in Subbands for Acoustic Echo Cancellation," *Proceedings of the 1996 IEEE Digital Signal Processing Workshop*, pp. 354–357.

[30] Liavas, A.P.; Regalia, P.A., "Acoustic Echo Cancellation: Do IIR Models Offer Better Modeling Capabilities Than Their FIR Counterparts?," IEEE Transactions on Signal Processing, Vol. 46, Issue: 9, Sept. 1998, pp. 2499–2504.

[31] Naylor, P.A.; Tanrikulu, O.; Constantinides, A.G., 'Subband Adaptive Filtering for Acoustic Echo Control Using Allpass Polyphase IIR Filterbanks, " *IEEE Transactions on Speech and Audio Processing*, Volume: 6 Issue: 2 , March 1998, pp. 143–155.

[32] Pessoa, L.F.C., Dyba, R.A., He, P.P., "Joint Audio Power Estimation and Tone Indication in Network Echo Cancellers," *International Signal Processing Conference*, March 31–April 3, 2003, Dallas, Texas.

[33] Dyba, R.A., Pessoa, L.F.C., He, P.P., "Network Echo Cancellers with Parallel Architecture Using Full Rate and Sub Rate Adaptive Filters", *International Signal Processing Conference*, March 31–April 3, 2003, Dallas, Texas.

[34] B. Farhang-Boroujeny. *Adaptive Filters: Theory and Applications*, John Wiley & Sons, 1998.

[35] Shynk, J. J., "Frequency-Domain and Multirate Adaptive Filtering," *IEEE Signal Processing Magazine*, January 1992; pp.14–37.

[36] Narayan, S.S., Peterson, A.M., and Narasimha, M.J., "Transform Domain LMS Algorithm," *IEEE TRANSACTIONS ASSP, ASSP-31*, No.3, June 1983, pp. 609–615.

[37] ITU-T G.VED Draft Recommendation G.VED/G.160, *Voice Enhancement Devices*, Draft No.8, January 28, 2003.

[38] ITU-T P.501, *Test Signals for Use in Telephonometry* (05/2000).

[39] ITU-T P.502, *Objective Test Methods for Speech Communication Systems Using Complex Test Signals* (05/2000).

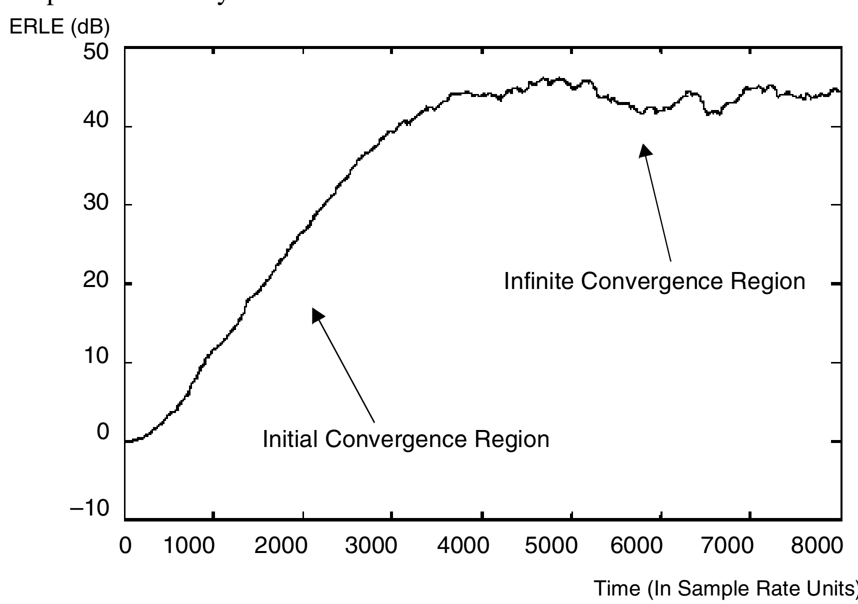
[40] ITU-T P.862, *Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs* (02/2001).

[41] Im, S., Powers, E.J., "A Block LMS Algorithm for Third-Order Frequency-Domain Volterra Filters," *IEEE Signal Processing Letters*, Vol. 4, No. 3, March 1997.

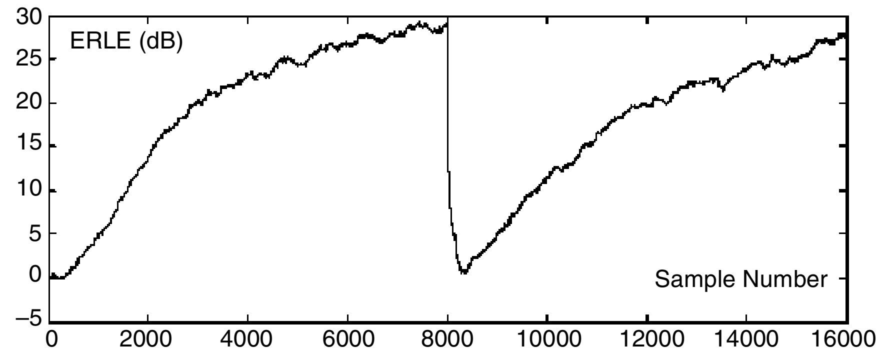
Appendix A:Definitions

Initial Convergence Convergence region for the time period starting at the beginning of the call and lasting until the ERLE reaches 80 percent of its expected value. Typically, the initial convergence lasts 300–800 ms. For very fast echo cancellers, the initial convergence may last only 150–200 ms or even less.

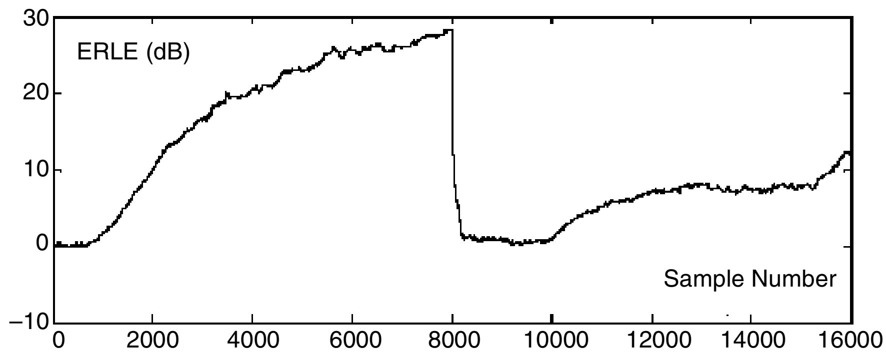
Infinite Convergence Convergence region after the ERLE reaches its asymptotic value. The infinite convergence region starts somewhere between 1 s and 12 s. The ERLE asymptotic value (not in a strictly mathematical sense) may reach 35–40 dB, and if the FE signal characteristics and hybrid characteristics do not change, it remains more or less within +/–5 dB range of its average. ERLE versus time changes depend on the type of echo canceller and on such factors as hybrid dispersion, echo path bulk delay, FE signal level, and echo canceller settings. The ERLE versus time graph included here was produced using a wideband noise, fed into the R_{in} direction, with the spectral density uniformly distributed within the telephony band. The infinite convergence region starts at 500 ms. The G.711 compression and decompression processes affect the infinite convergence performance by several dBs.



Reconvergence The process of readapting to a new echo path. The following figures illustrate ERLE behavior during reconvergence.



(a) Reconvergence of Expected Speed and Depth



(b) Reconvergence of Poor Speed and Depth

Echo Return Loss (ERL) Typically, the average magnitude of signal reflection is expressed as echo return loss (ERL), as defined by the following expression:

$$ERL = 20 \log(V_{RXrms}/V_{TXrms}),$$

where V_{RXrms} and V_{TXrms} represent RMS values of respective signals. For echo cancellation systems, V_{RXrms} and V_{TXrms} relate to R_{out} and S_{in}, respectively. More specific definitions of ERL can be found in IEEE Standard 743 [10]. Often, ERL is used interchangeably with THL. However, THL is a generic term, and it is defined for any harmonic and non-harmonic signals, but the ERL definition is restricted to specific signals, [10]. G.168 adopts ERL for the use of CCS signals.

Echo Return Loss Enhancement (ERLE) One measure of improvement of echo return loss, ERLE is defined by the following expression:

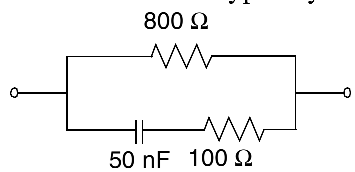
$$ERLE = 20 \log(V_{1rms}/V_{2rms}),$$

where V_{1rms} and V_{2rms} represent RMS values of respective signals. For echo cancellation systems, V_{1rms} and V_{2rms} relate to S_{in} and the error signal, respectively. For NLP put in a disable mode, V_{1rms} and V_{2rms} relate to S_{in} and S_{out}, respectively.

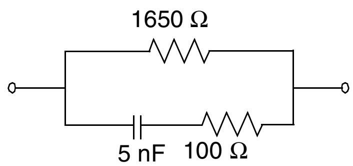
Bulk Delay (Pure Delay) The initial portion of the echo path impulse response, which consists of zeroes or otherwise negligible values. Bulk delay is related to network/link delays between the points where the echo canceller and the hybrid are installed.

Compromise Balance Networks
Lumped-constant circuits to increase ERL through better impedance matching; typically compromise balance networks are part of the codec ASIC and are implemented via switched-capacitor techniques.

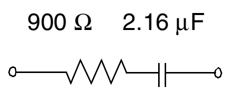
Non-loaded Balance Network
A compromise balance network typically to balance non-loaded subscriber lines.



Loaded Balance Network
A compromise balance network typically to balance non-loaded subscriber lines.

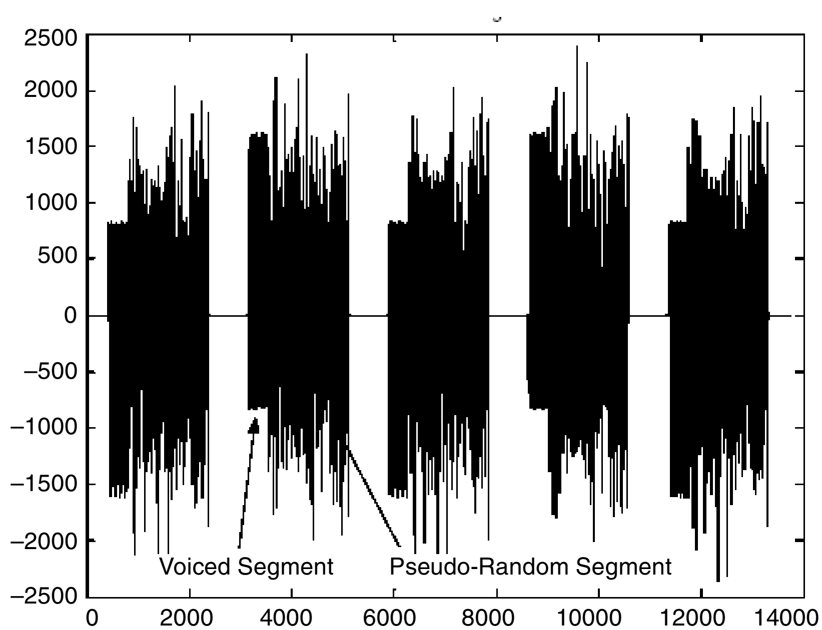


Special 900 Ω + 2.16 µF Network
A special compromise balance network to balance short non-loaded lines.



CSS Signal
An ITU-T G.168-defined signal for objective testing of network echo cancellers that consists of two parts separated by a pause. Each part contains a voiced sound segment and a pseudo noise segment. There are two versions of the CSS signal, one for single talk-type tests (that is, when $S_{gen} = 0$) and another for double-talk tests. The following figure illustrates the CSS waveform for a single-talk test application.

CSS Waveform Segments



Tip and Ring (T/R)
The two wires that constitute the twisted-pair wire connection between the telephone or other types of terminal equipment and the central office, or private branch exchange. This twisted pair is generally referred to as the local loop or subscriber loop.

Appendix B:Hybrid and Line Models

Hybrid and line models are lumped constant analog equivalent circuits that simulate the behavior of the 4-wire-to-2-wire interface. These models include not only the hybrid circuit but also Z_{TR} impedance based on a transmission line model as well as a telephone terminal model. This appendix discusses how these models can be useful tools in simulating the behavior of echo cancellers in a specific network environment. **Figure 27** and **Figure 28** illustrate two fully analog models of the hybrid circuit. **Figure 27** presents a simplified model that includes the impedance Z_{TR} as seen from the terminals 1,1' of the subscriber line and telephone set. This model does not include 4-wire front-end transfer function blocks. **Figure 28** shows a modified model incorporating H_1 and H_2 transfer functions from the receive PCM to the open T/R circuit and from the T/R circuit to the transmit PCM, respectively.

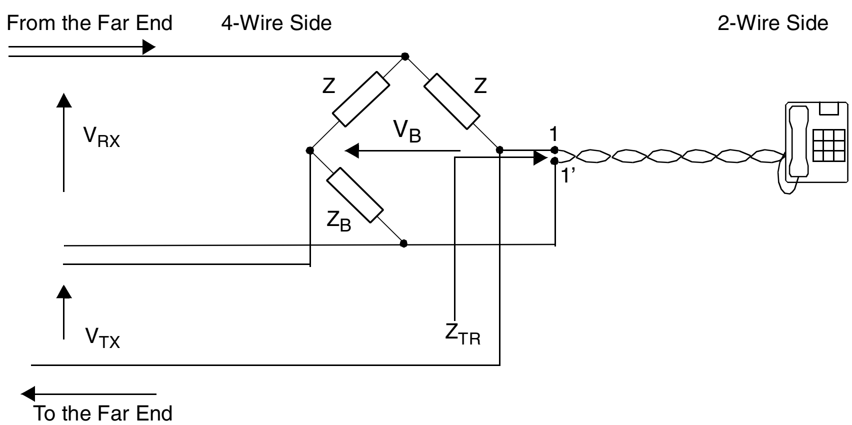


Figure 27. Simplified Model of a Hybrid Circuit

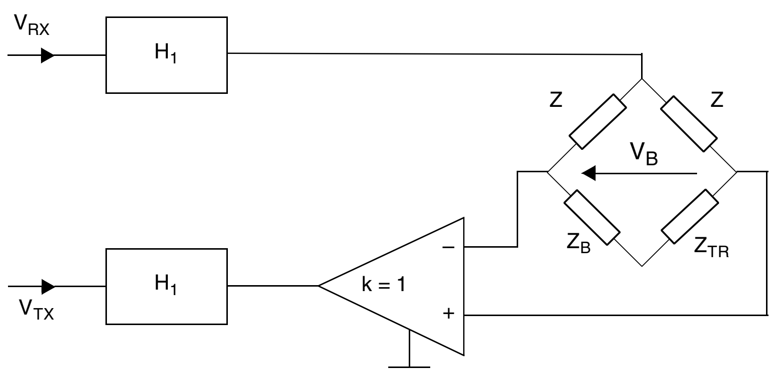


Figure 28. Simplified Model of a Hybrid Circuit that Includes H_1 and H_2 Transfer Functions

Analysis of the circuit in **Figure 28** leads to the following formula for the transhybrid gain (THG = VTX/VRX):

$$THG = H_1 \cdot H_2 \cdot \left(\frac{Z_{TR}}{Z + Z_{TR}} - \frac{Z_B}{Z + Z_B} \right)$$

Because THG = 1/THL, and ERL can be represented as ERL = 20 log(V_{RXrms}/V_{TXrms}), the following equation applies:

$$ERL = -20 \cdot \log\left[H_1 \cdot H_2 \cdot \left(\frac{Z_{TR}}{Z + Z_{TR}} - \frac{Z_B}{Z + Z_B}\right)\right]$$

This equation enables us to analyze and simulate the echo path (dispersion part) of the network. Typically, H_1, H_2, and Z are known for a given line card model. Z_{TR} and Z_{TEL} must be determined separately. A full knowledge of Z_B is required as well. Z_{TR} depends on the configuration of the subscriber line. Z_{TEL} is the input impedance of the telephone set. In the on-hook state, $|Z_{TEL}|$ is relatively large and can be treated as "open." In the off-hook state, Z_{TEL} presents complex impedance for which the specific characteristics depend not only on the electrical characteristics of the telephone set but also on the acousto-electrical characteristics and position of the telephone handset, including its position in relation to the subscriber ear and mouth. A subscriber line is a transmission line that can precisely be described using Maxwell partial differential equations, which, in turn, can be simplified using transmission line equations. These equations can be further simplified to a lumped-constant analog model. A 3kft segment of the telephone subscriber line can be represented as a lumped-constant structure as shown in **Figure 29**.

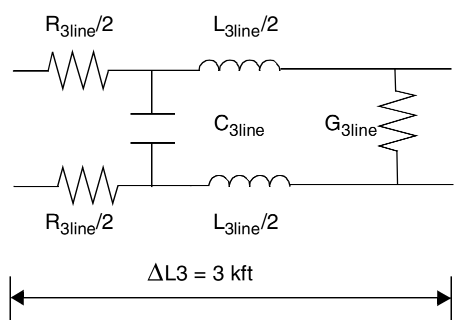


Figure 29. One Equivalent Circuit of a 3 kft Segment of Telephone Subscriber Line

Lumped constants R_{line}, G_{line}, L_{line}, and C_{line} can be calculated as follows:

- $R_{line} = \Delta L_3 \cdot r_{line}$
- $G_{line} = \Delta L_3 \cdot g_{line}$
- $L_{line} = \Delta L_3 \cdot l_{line}$
- $C_{line} = \Delta L_3 \cdot c_{line}$

where r_{line}, g_{line}, l_{line}, c_{line}, are line resistance per a linear unit (kft), line conductance per a linear unit, line inductance per a linear unit, and line capacitance per a linear unit, respectively. The line parameters per a linear unit are functions of physical parameters of the twisted pair of wires. There are several acceptable equivalent circuit topologies, some approximating the line segment better than others. **Figure 30** illustrates some examples.

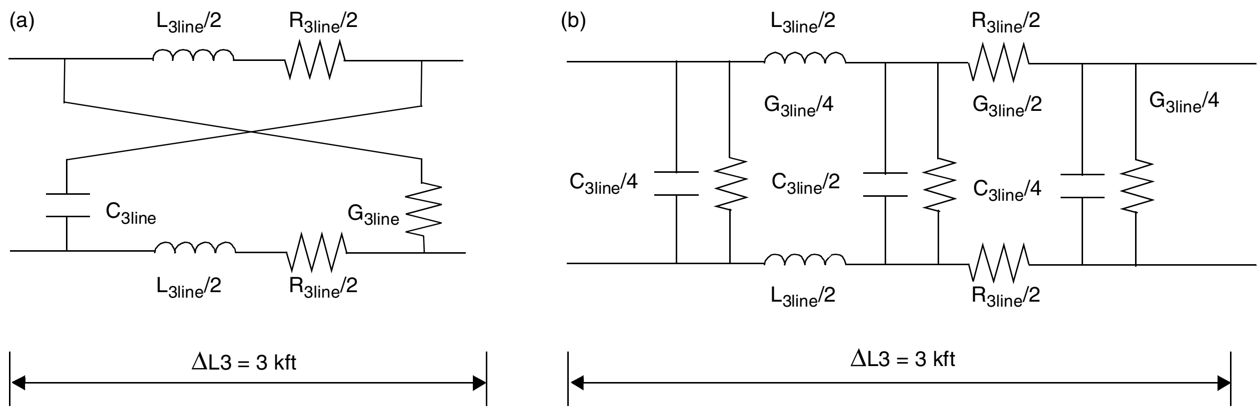


Figure 30. Equivalent Circuits of a 3 kft Segment of Telephone Subscriber Line

Subscriber lines, which are longer than 12 kft, are often equipped with loading coils that boost the amplitude frequency characteristic at a higher frequency range (1500 Hz, 3000 Hz). The loading coils are usually installed at every 6 kft segment. **Figure 31** presents a loading coil circuit model. **Figure 32** shows one of loading coil distributions for a 17 kft subscriber line.

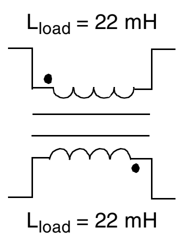


Figure 31. Equivalent Circuit of a Typical Loading Coil

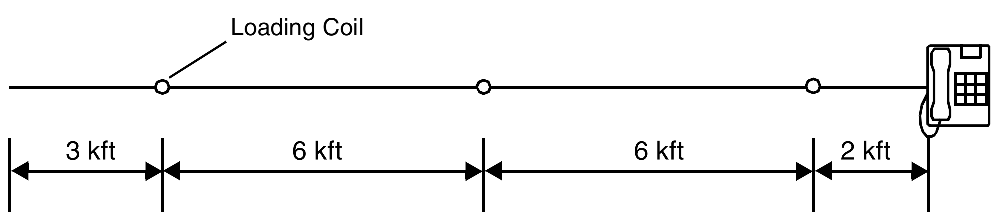


Figure 32. Example Loading Coil Distributions for a Hypothetical 17 kft Subscriber Line

To simulate an impulse response of a hybrid circuit terminated with loaded or non-loaded subscriber line, a knowledge of the physical parameters of the line (distribution of loading coils, bridge taps, physical parameters of the twisted pair of wires, and so on) as well as the electrical parameters of the telephone set is essential.

How to Reach Us:

Home Page:
www.freescale.com

E-mail:
support@freescale.com

USA/Europe or Locations not listed:
Freescale Semiconductor
Technical Information Center, CH370
1300 N. Alma School Road
Chandler, Arizona 85224
+1-800-521-6274 or +1-480-768-2130
support@freescale.com

Europe, Middle East, and Africa:
Freescale Halbleiter Deutschland GMBH
Technical Information Center
Schatzbogen 7
81829 München, Germany
+44 1296 380 456 (English)
+46 8 52200080 (English)
+49 89 92103 559 (German)
+33 1 69 35 48 48 (French)
support@freescale.com

Japan:
Freescale Semiconductor Japan Ltd.
Headquarters
ARCO Tower 15F
1-8-1, Shimo-Meguro, Meguro-ku,
Tokyo 153-0064, Japan
0120 191014 or +81 3 5437 9125
support.japan@freescale.com

Asia/Pacific:
Freescale Semiconductor Hong Kong Ltd.
Technical Information Center
2 Dai King Street
Tai Po Industrial Estate
Tai Po, N.T. Hong Kong
+800 2666 8080

For Literature Requests Only:
Freescale Semiconductor Literature Distribution Center
P.O. Box 5405
Denver, Colorado 80217
1-800-441-2447 or 303-675-2140
Fax: 303-675-2150
LDCForFreescaleSemiconductor@hibbertgroup.com

AN2598
Rev. 1
11/2004