

# A MEMORY REDUCTION SCHEME FOR MULTI-CHANNEL ECHO CANCELLER IMPLEMENTATION

*Chang Y. Choo<sup>1</sup> and Hammam Elabd*

RealChip Communications  
Sunnyvale, California 94085

## ABSTRACT

One of the critical resources of the multi-channel echo canceller is the memory that stores both tapped delay lines of voice data and filter tap coefficients. In this paper, a simple variable-length/run-length coding scheme for reducing coefficient memory for multi-channel echo cancellers is proposed. We also describe the corresponding memory system architecture. Simulations based on a bit-accurate Matlab model show that the proposed scheme is effective.

## 1. INTRODUCTION

Recently, the subject of echo cancellation gained renewed interests among the DSP algorithm and architecture researchers, due to the emerging technologies in the areas of voice over IP (VoIP) and wireless communications. The current goals of research on echo cancellation include more efficient algorithm and multi-channel echo cancellation.

In the areas of carrier-class gateways and wireless base stations, the high-performance echo cancellers executing tens to hundreds of billion instructions per second are required in order to handle hundreds to thousands of channels, each with long echo tail length up to 128 ms. Such high performance can only be achieved by using large number of MACs (multiplier-and-accumulator) in parallel [1].

The requirement for memory, as well as MACs, is significantly high in the high-performance echo cancellers. For example, for an echo canceller which processes 672 channels (i.e., DS3), each with 64 ms echo tail length, the size of the data memory is close 3 Mb and that of the coefficient memory twice as much, assuming G.711 based 8-bit data and 16-bit coefficients. This level of memory requirement poses a significant challenge both in DSP processor

based implementation and in ASIC/FPGA based implementation. In the latter approach in particular, the MACs and the memory modules should be tightly integrated on, typically, a single chip. As a result, memory modules and MACs compete for the limited area, and thus reducing the required memory size can become a major design objective.

In the echo cancellers using adaptive FIR filter algorithms such as LMS and NLMS [3,4], filter coefficients are usually updated at every input data sample. Coefficients and input data are stored in separate memory space, as both are accessed simultaneously. Data memory cannot be reduced because they usually are already in compressed format, such as G.711. On the other hand, coefficient memory can be compressed for its size reduction.

No significant research has been done or published on the memory architecture of the high-performance echo cancellers. It is partly because it was not until recently that the high-performance echo cancellers were required, such as in carrier-class voice-over-IP equipment. While recent high-end DSP VLIW processors (e.g., TI C64x, Centillium CT-GWC2256, and Silicon Spice CALISTO) contain increasingly more number of MACs (see Table 1, for example), FPGAs as well as ASICs can have much more MACs, limited only by the chip area.

Moreover, the MACs in the FPGA/ASIC are scalable to any number of bit widths (e.g., 12-bit, 24-bit, 41-bit, and so forth), while the MACs in DSP processors are only scalable to 8-bit, 16-bit, and 32-bit typically. This bit-level scalability allows the DSP system designers to utilize the available resources in more cost-effective manner. The same advantage of the bit-level scalability can be applied to the memory system of the high-performance echo cancellers.

In this paper, we present a scheme to reduce the coefficient memory. After briefly describing typical echo canceller algorithms in Section 2,

---

<sup>1</sup> Also, with San Jose State Univ., San Jose, CA.

we show that the performance of the echo canceller is dependent on the coefficient bit width in the following section. We also show that the coefficient bit width change gradually, as the cancellation time elapses. From these observations, we propose a simple scheme to reduce the coefficient memory that exploits the behavior of the filter coefficients. In Section 4, we show several bit-accurate simulation results to support the proposed scheme. We conclude this paper by suggesting future research directions in the last section.

## 2. ECHO CANCELLATION: ALGORITHMS AND MEMORY ORGANIZATION

For echo cancellation, LMS and RLS FIR filter algorithm are typically used as described in the following.

Consider a subset of input samples

$$\mathbf{x}(n) = (x(n), x(n-1), \dots, x(n-k-1))$$

and the desired outputs

$$\mathbf{d}(n) = (d(1), d(2), \dots, d(n-k-1)).$$

We suppose the output can be modeled as a FIR filter, i.e.,

$$y(n) = \mathbf{H}^T(n) \bullet \mathbf{x}(n),$$

where  $\mathbf{H}$  is the weight vector of  $k$  elements. We define the squared error function

$$e^2(\mathbf{H}) = \sum (d(i) - \mathbf{H}(i)^T \bullet \mathbf{x}(i))^2$$

The gradient vector  $\mathbf{G}(n)$  of the squared error function  $e^2(\mathbf{H})$  evaluated at the point  $\mathbf{H}(n)$  is:

$$\begin{aligned} \mathbf{G}(n) &= \partial e^2(\mathbf{H}) / \partial \mathbf{H} \\ &= -2 \sum (d(i) - \mathbf{H}(i)^T \bullet \mathbf{x}(i)) \mathbf{x}(i). \end{aligned}$$

Setting  $\mathbf{G}$  equal to the null vector and solving for  $\mathbf{H}$ , we obtain the optimal  $\mathbf{H}$

$$\mathbf{H}^* = (\sum d(i) \mathbf{x}(i)) / (\sum \mathbf{x}(i) \mathbf{x}(i)^T)$$

This algorithm is called the least-square (LS) estimation algorithm. The computational complexity of the LS algorithm grows exponentially with the dimension of the weight vector  $\mathbf{H}$ . The RLS (recursive-least-square)

algorithm updates the above inverse auto-correlation matrix in efficient manner.

A less computation-intensive algorithm, the LMS algorithm, simplifies the computation by using  $-e(n)\mathbf{x}(n)$  as an instantaneous estimate of the gradient vector  $\mathbf{G}(n)$ . Thus

$$\begin{aligned} \mathbf{H}(n+1) &= \mathbf{H}(n) - \mathbf{G}(n) \\ &= \mathbf{H}(n) + \mu e(n) \mathbf{x}(n). \end{aligned}$$

This is the well-known LMS algorithm. Here,  $\mu$  is called the convergence factor.  $\mu$  is used to control the speed of convergence. When set too high, the system may diverge. On the other hand, when set too low, the system may converge too slow. In NLMS algorithm, the convergence factor is adjusted based on the power estimate of the input data. The adjustment is achieved by dividing  $\mu$  by the power estimate.

Figure 1 shows a simple architecture for echo canceller. The echo cancellation filter requires data memory (tapped delayed line of voice data for all channels) and coefficient memory.

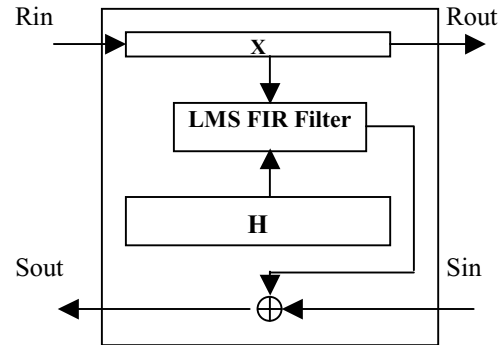


Figure 1: Simple Echo Cancellation Architecture

## 3. A SCHEME FOR REDUCING COEFFICIENT MEMORY

In DSP processor based implementations, the coefficient memory is usually 16 bits wide. However, previous research [1,5] indicates that the coefficient bit width may be reduced while still compliant to the ITU-T G.165/168 echo canceller standard [2].

In addition, we observed as shown in the next section that the coefficient bit width for a particular call session changes dynamically. Typically, during the beginning of a call session, the bit width required for coefficients are

relatively small. As the call time progresses, the bit widths tend to increase logarithmically. Since the beginning and ending times of call sessions across all channels are randomly distributed, the coefficient memory size can be reduced when aggregated in the integrated coefficient memory.

From above observations, we propose a scheme to reduce the aggregate memory requirement for filter coefficients. The scheme involves a simple variable-length/run-length coding. The encoding algorithm is described below:

- a) At each end of coefficient update iteration, the maximum coefficient bit width is determined.
- b) Coefficients for each channel are stored as a block structured as a vector (CH, BW, COEFs), where CH, BW and COEFs are channel number, bit width, and coefficients, respectively.

For decoding, the whole process is reversed.

In DSP/ASIC/FPGA based echo cancellers, the memory is typically byte-addressable. For  $L$  ms echo tail length, the coefficient bit width  $B$  and the number of bytes  $N$  for coefficient storage have the following relations:

$$N = L \cdot B.$$

The above equality shows that in general, each ms of echo tail length requires  $B$  bytes of coefficient memory. Note that in practice,  $L$  is power of 2. The above relation is useful in designing the coefficient encoder and decoder when the channel boundaries are to be determined.

Figure 2 shows an architecture incorporating the scheme described in this section. When the filter block presents the channel number to be processed, the VLD/RLD (variable-length/run-length decoder) brings the corresponding block from the coefficient memory and feed the coefficients to the filter.

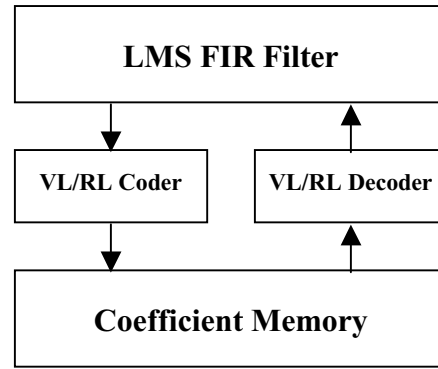
On the other hand, when updated coefficients are to be stored, VLC/RLC (VL/RL coder) encodes the coefficients and stores them in the right location corresponding to the particular channel.

#### 4. SIMULATION RESULTS

We implemented a bit-accurate simulation model of the multi-channel echo canceller using Matlab.

The model implemented a NLMS algorithm. We ran several simulations to confirm the effectiveness of the proposed memory reduction scheme.

We first confirmed that the coefficient bit width can be reduced from usually 16 bits while not violating the ITU-T G.165/168 standard requirement. For example, Figure 3 on the next page shows how the echo cancellation performance, in terms of steady-state residual echo, varies as the coefficient bit width changes. We found that the coefficient bit width of 12 bits still satisfies the G.165/168 requirements, although more bits will further enhance the system performance.



**Figure 2: Coefficient Memory System**

We also observed that, as shown in Figure 4 on the next page, the coefficient bit width grows in logarithmic manner during a call session. It is interesting to note that often coefficients do not need full bit width it was assigned initially. For example, even though 12 bits are assigned initially, when steady state was reached, the maximum bit width was only 9 bits in the case of  $L_{rin} = -15$  dBm0. Measuring statistically the temporal behavior of coefficient bit widths should provide the further insight in this area.

#### 5. CONCLUSIONS

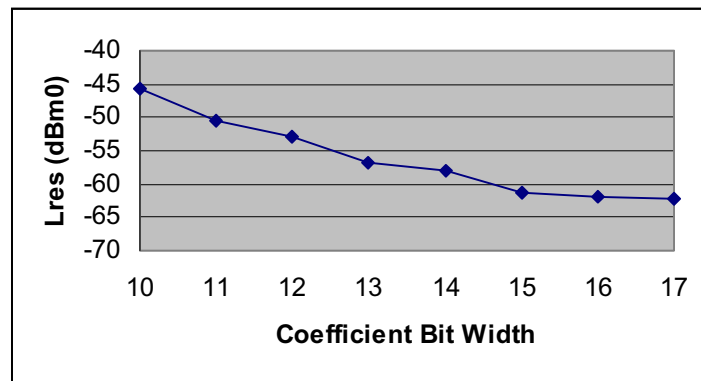
In this paper, a simple VLC/RLC based scheme for reducing coefficient memory for multi-channel echo cancellers was proposed. We also described the corresponding memory system. Simulation results show that the proposed scheme is effective.

Future research activities may take the following directions:

- Variable-length COEFs not only across iterations, but also within each iteration.
- Efficient scheme to combine zero or near-zero coefficients.

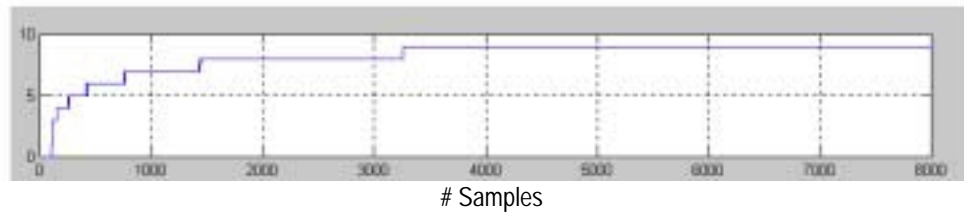
## REFERENCES

1. Chang Choo, "Designing a High-Performance Echo Canceller for Voice-over-IP Applications," DSP Engineering, vol.2, no.2, Spring 2000.
2. ITU-T, G.168: Digital Network Echo Cancellers, Apr. 1997.
3. Texas Instruments, Digital Voice Canceller with a TMS32020, User Information.
4. B. Widrow and S.D. Stearns, Adaptive Signal Processing, Prentice-Hall, 1985.
5. Zhaohong Zhang and Gunter Schmer, "Analysis of Filter Coefficient Precision on LMS Algorithm Performance for G.165/G.168 Echo Cancellation," Texas Instruments Application Report SPRA561, Feb. 2000 (C6000 App).



**Figure 3: Echo canceller performance on various coefficient bit widths (Lrin=-15dBm0, ERL=9dB, Lres measured after 8000 samples)**

Coeff. Bit Width



**Figure 4: Progressive growth of coefficient bit width (same condition as in Fig. 3; 12-bit coefficient)**

DSP	C54x	C62x	ADSP21xx	SHARC	ADSP-TS
# of 16-Bit MACs	1-2	2	1	2	8
Data Memory	16-128KB	64-512KB	16-48KB	512KB	768KB

**Table 1: Number of 16-bit MACs and data memory size for some DSP processors.**