LOW POWER FILTERING USING APPROXIMATE PROCESSING FOR DSP APPLICATIONS

Jeffrey T. Ludwig, S. Hamid Nawab, Anantha Chandrakasan

MIT Research Laboratory of Electronics, Cambridge, MA 02139

ABSTRACT

The growing demand for portable multimedia devices has placed increased importance on low power solutions for DSP tasks such as filtering and source coding. An approach to power reduction in digital CMOS filter design using approximate processing is presented. This involves adaptively reducing the number of operations switched per sample based on signal statistics. Speech filtering examples demonstrate that power consumption can be reduced over conventional solutions by an *order of magnitude* for wireless applications.

1. Introduction

A key issue in the design of portable devices is to minimize the total power consumption in order to maximize the run time while minimizing battery size [1]. To first order, the average power required to perform a signal processing task is:

$$P = \sum_i N_i C_i V_{dd}^2 f_s, \tag{1}$$

where C_i is the average capacitance switched per operation of type i (corresponding to addition, multiplication, storage, etc.), N_i is the number of operations of type i performed per sample, V_{dd} is the operating supply voltage, and f_s is the sample frequency. For real-time signal processing applications, a key attribute is that the sample frequency is fixed and therefore there is no advantage to making the hardware computation faster than f_s. This has been exploited to reduce power consumption by lowering V_{dd} at the cost of increased chip area without loss in throughput [2]. Another approach to reduce power consumption is to lower C_i. Some considerations for this include transistor sizing, choice of circuit style, logic level shut down [3], choice of circuit topology [4], and exploiting naturally occuring signal correlations [1]. A third approach for power reduction is to lower N_i. Efforts have been made to minimize N_i by intelligent choice of algorithm, given a particular task [5]. The focus of this paper is to use approximate processing [6] [7] to *dynamically* reduce the number of operations performed per sample period.

This work was sponsored in part by the Department of the Navy, Office of the Chief of Naval Research, contract number N00014-93-1-0686 as part of the Advanced Research Projects Agency's RASSP program.

2. The Filtering Problem

Finite impulse response (FIR) filters are often used in applications where the goal is to extract from a signal certain frequency components while rejecting others. For example, in order to receive a single channel of a frequency division multiplexed (FDM) signal, the frequency band containing the channel of interest is passed by a frequency-selective filter while other bands are attenuated. Suppose the signal to be filtered is the sum of a desired signal, $s[n]$, and an interference or noise component, $w[n]$:

$$x[n] = s[n] + w[n]. \tag{2}$$

Many contexts arise in which the signals $s[n]$ and $w[n]$ occupy largely disjoint frequency bands. If it were possible to *cost-effectively measure the strength of the interference*, $w[n]$, from observation of $x[n]$, we could determine how much stopband attenuation the filter should have at any particular time. When the energy in $w[n]$ increases, it is desirable to increase the stopband attenuation of the filter. This can be accomplished by using a longer FIR filter. Conversely, the filter may be made shorter when the energy in $w[n]$ decreases. It is also desirable that the

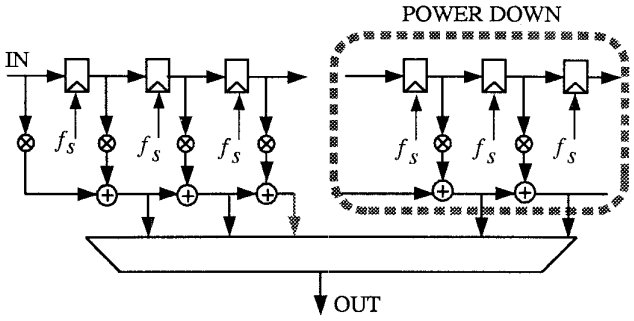

Figure 1: Illustration of dynamic tapped delay line

taps of the shorter filters (with less stopband attenuation) always be subsets of the taps of the longer filters. This ensures that when the filter length is made shorter on a sample period, we can simply *power down* a segment of the full tapped delay line as shown in Fig. 1. Powering down of the higher order taps has the effect of reducing the switched capacitance at the cost of decreasing the attenuation in the stopband. Assuming that the delay

line is implemented using SRAM, even the data shifting operation of the higher order taps can be eliminated through appropriate addressing schemes. We illustrate the reduction in the stopband attenuation for different filter lengths in Fig. 2. Each filter here belongs to a family of FIR lowpass filters which are truncated versions of a single ideal lowpass filter. As mentioned earlier, the

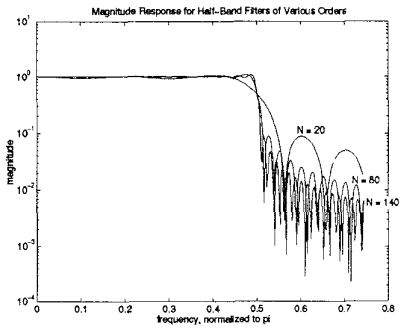

Figure 2: FIR filters of lengths $N = 20, 80$ and 140

key to being able to adaptively adjust the filter length is to somehow obtain a measure of the strength of the interference $w[n]$ in a cost-effective manner – i.e., using only a small number of operations (preferably only additions). In the next section we present two classes of approximate filtering applications in which it is possible to obtain a measure of interference strength at essentially negligible cost.

3. Approximate Filtering

The first class of applications, the weak-signal case, is characterized by extremely low SNR at the input. The second application, demultiplexing of FDM signals, is characterized by an input SNR which fluctuates between high (essentially no interference) and moderately low (interference comparable to desired signal) values.

3.1. Weak-Signal Case

When the SNR is very low, the input signal $x[n]$ is approximately equal to the interference signal $w[n]$. We can therefore use $x[n]$ to measure the strength of $w[n]$. In particular, the measure we have used for this purpose is given by:

$$M_x[n] = \frac{1}{L} \sum_{m=n-L+1}^{n} |x[m]|, \tag{3}$$

We selected this measure because of its direct proportionality to the local energy in $x[n]$ and because its computation does not require any multiplication operations. Furthermore, only two additions are required to obtain $M_x[n+1]$ from $M_x[n]$. The choice the parameter L involves a tradeoff between suppression of sensitivity to local fluctuations and preservation of time-varying nature of the signal energy. When the value of L is less than the maximum filter length, then there is no extra storage required to compute $M_x[n]$. When $M_x[n]$ is known for a particular time, we may use this value to estimate the length of the filter which would provide the requisite amount of attenuation in the stopband. The family of filters we choose from at each time are truncated versions of an ideal lowpass filter. In particular, if the filter at time n is to be of length N, its mth tap is given by:

$$h_n[m] = \begin{cases} \frac{\sin \omega_c m}{\pi m} & 0 \le m \le N-1 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

The one question that now remains is how we map a value of $M_x[n]$ to a particular filter length. For addressing this question, we first assume that the maximum acceptable noise level in the output is given *a priori* to be $M_{\max}$. We also note that the relative heights of the various sidelobes of each FIR filter shown in Fig. 2 remain constant regardless of the length of the filter. The first 31 sidelobes provide monotonically increasing attenuation ranging from 20 to 40 dB. We now pick a "design frequency" ω_d which is larger than the cutoff frequency ω_c of the lowpass filter. We also choose the design frequency so that most of the interference energy is at frequencies greater than the design frequency. It is possible to vary the filter attenuation at ω_d by changing the filter length to place the sidelobe peak which provides the required amount of attenuation at ω_d. The kth sidelobe peak occurs at frequency ω_k, where

$$\omega_k = w_c + \frac{2\pi}{N}k. \tag{5}$$

The relative heights of the peaks of the sidelobes may be calculated in advance and placed in a lookup table. For a given input signal strength, $M_x[n_o]$, the value of k is selected from the lookup table to ensure that $M_x[n_o]$ is attenuated to the level $M_{\max}$. The value of N is then chosen to be the integer which best satisfies the following constraint:

$$\omega_d = w_c + \frac{2\pi}{N}k \tag{6}$$

Satisfying this criterion assures that all interferences at or above frequency w_d will be attenuated sufficiently to guarantee $M_y[n] \le M_{\max}$.

Our overall strategy for approximate filtering in weak-signal cases is depicted in Fig. 3. For each time n, the value of $M_x[n]$ is used to adaptively adjust the length of the impulse response $h_n[m]$ for time n. It should be noted that $M_x[n]$ is recursively updated by a process which requires only 2 additions per sample.

To demonstrate the effectiveness of our approach to weak-signal filtering, we present two illustrative examples. The top graph in Fig. 4 shows a waveform corresponding to a 2 sec. (16K samples) speech utterance,

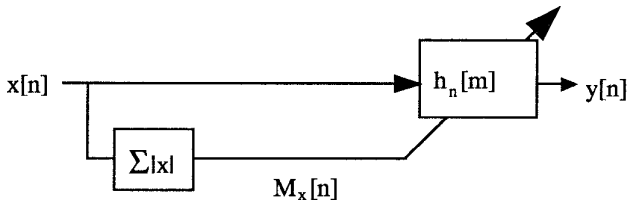

Figure 3: Overview of weak-signal filtering strategy

"the chef made lots of stew." The middle plot shows the speech with a synthetically-generated tonal interference component at $w_t = 0.65\pi$ rad/sample, whose amplitude linearly increases over time. The filtering strategy depicted in Fig. 3 was then applied to the corrupted speech.

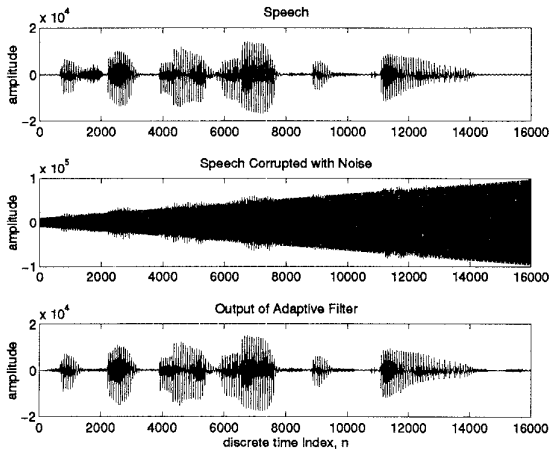

Figure 4: First example's speech waveform, corrupted speech waveform, and processed waveform

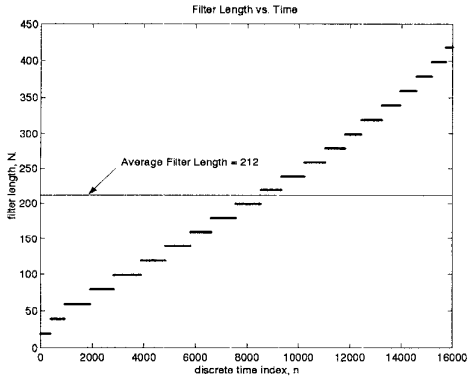

Figure 5: The time evolution of the filter length for the first weak-signal case example

The filtered output is shown in the third graph. The tonal interference has effectively been removed. During the processing of this example, we obtained the variation in filter lengths as depicted in Fig. 5. We see that the filter length increases as the interference amplitude increases, yielding a filtered output signal with a uniformly high SNR. If we had been restricted to a *fixed* filter length strategy, we would have had to pick a filter length, $N_{\max}$, which corresponds to the maximum amount of attenuation needed at the end portion of the ramp interference. In our example, $N_{\max}$ was 439. On the other hand, the average of the filter lengths used by the strategy shown in Fig. 3 turned out to be $N_{\text{ave}} = 212$. This translates to a power savings factor ($N_{\max}/N_{\text{ave}}$) of 2.1. However, even larger power savings can be obtained if the SNR is not always restricted to be very low. This is illustrated in our next example.

The filtering strategy of Fig. 3 was also applied to the speech waveform of the previous example corrupted by a different interference signal. This time the interference was a constant tone during the first third of the speech utterance and zero afterwards. The corrupted speech signal is shown in Fig. 6. The filter orders used during the processing are illustrated in Fig. 7. The resulting power savings is by a factor of more than 3. In fact, even greater savings would be obtained if regions with little or no power were of longer duration.

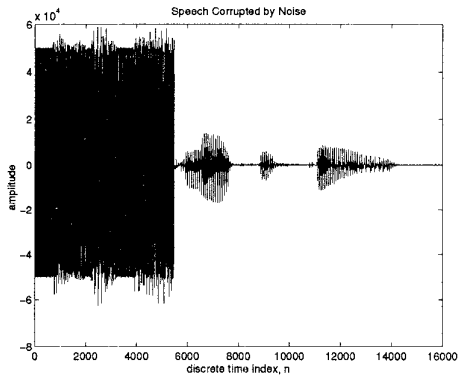

Figure 6: Second example's corrupted speech waveform

3.2. FDM Case

When performing filtering for demultiplexing of received FDM signals, the energy in the bands to be rejected can vary from essentially negligible levels to levels which are comparable to the desired signal. In other words, the low SNR assumption of the weak-signal case is no longer applicable. For such situations, we have developed an approximate filtering strategy which involves the addition of a feedback loop to the weak-signal strategy of Fig. 3. The basic idea is to compare the signal strength at the output of the filter with the signal strength at the input

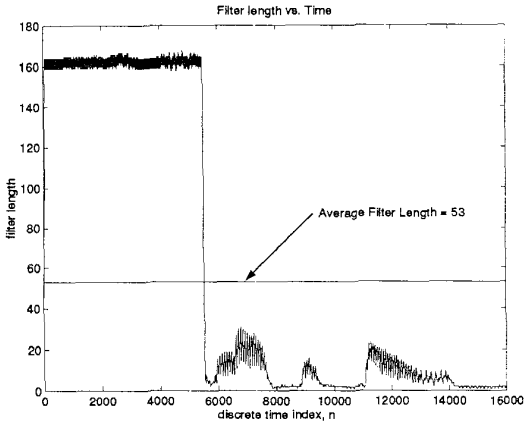

Figure 7: The time evolution of the filter length for the second weak-signal case example

to help in adjusting the filter length for the next sample. As the amount of noise rejected by the filter increases, the difference in the strengths of the input and output signals also increases. It follows that increasing (positive) values of $d[n] = M_y[n] - M_x[n]$ indicate increasing strength of the interference $w[n]$. For the sake of brevity, our description of the system architecture is restricted to the case where only two filter lengths are used. The more general case will be reported in later publications.

Our FDM filtering strategy is depicted in Fig. 8. The filter $h[n]$ is selected to have high stopband attenuation. This filter is meant to be used when the strength of $w[n]$ is comparable to the strength of $s[n]$. Otherwise, an identity filter is to be used. To begin with, the FIR filter $h[n]$

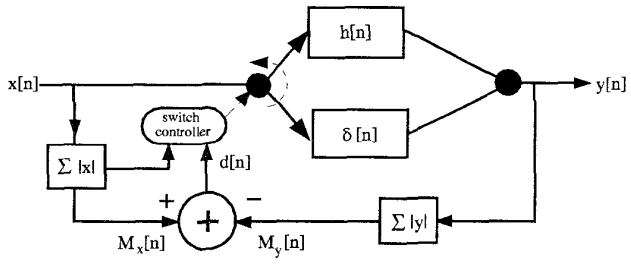

Figure 8: Overview of FDM Filtering Strategy

is **on** for at least B_{on} samples. After the lag period B_{on}, the filter will be turned **off** only if $d[n]$ becomes smaller than γ_2. When the filter $h[n]$ is **off**, it remains **off** for a lag period of at least B_{off} samples. After the lag period B_{off}, the filter is turned **on** only if the input signal energy exceeds γ_1. The parameters $B_{\text{on}}, B_{\text{off}}, \gamma_1$, and γ_2 are selected in an application-dependent manner.

We have successfully applied the strategy of Fig. 8 to separate a low-frequency FDM channel from a high-frequency one. Each channel consisted of a different speech utterance by a different speaker. Both speech sentences were of comparable loudness levels. The waveform of the channel to be rejected is shown in Fig. 9. The rectangular boxes superimposed on that waveform indicate the time regions where the filter $h[n]$ was in the **on** position. Clearly the boxes correspond mostly to those times when this out-of-band speech waveform has significant energy. Since longer periods of speech communication will include significantly longer periods of silence, one can expect power savings of an order of magnitude.

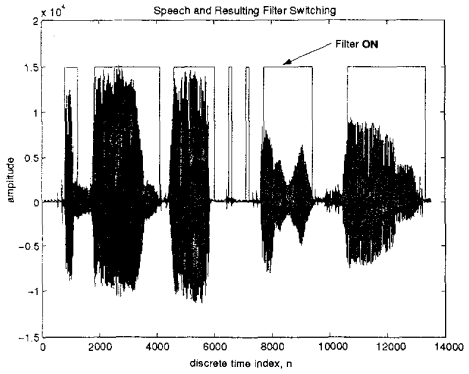

Figure 9: Actual filter cycling and underlying interference

4. References

[1] A. P. Chandrakasan, et. al., "Design of Portable Systems," invited tutorial paper presented at the Custom Integrated Circuits Conference, May 1994.

[2] A. P. Chandrakasan, S. Sheng, and R. W. Brodersen, "Low-power digital CMOS design," IEEE Journal of Solid State Circuits, pp. 473-484, April 1992.

[3] M. Alidina, et. al., "Precomputation-Based Sequential Logic Optimization for Low Power," 1994 International Workshop on Low-power Design, April 1994, pp. 57-62.

[4] T. Callaway and E. Swartzlander, Jr., "Optimizing Arithmetic Elements for Signal Processing," VLSI Signal Processing, V, pp. 91-100, IEEE Special Publications, 1992.

[5] B. M. Gordon and T. Meng, "A Low Power Subband Video Decoder Architecture." International Conference on Acoustics, Speech, and Signal Processing, April, 1994.

[6] V. R. Lesser, J. Pavlin, and E. Durfee, "Approximate Processing in Real-Time Problem Solving," AI Magazine, pp. 49-61, Spring, 1988.

[7] S. H. Nawab and E. Dorken, "A Framework for Quality Versus Efficiency Tradeoffs in STFT Analysis." IEEE Transactions on Signal Processing, April 1995.