Low-Power Digital Filtering Using Approximate Processing

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

Abstract—We present an algorithmic approach to the design of low-power frequency-selective digital filters based on the concepts of adaptive filtering and approximate processing. The proposed approach uses a feedback mechanism in conjunction with well-known implementation structures for finite impulse response (FIR) and infinite impulse response (IIR) digital filters. Our algorithm is designed to reduce the total switched capacitance by dynamically varying the filter order based on signal statistics. A factor of 10 reduction in power consumption over fixed-order filters is demonstrated for the filtering of speech signals.

I. INTRODUCTION

TECHNIQUES for reducing power consumption have become important due to the growing demand for portable multimedia devices. Since digital signal processing is pervasive in such applications, it is useful to consider how algorithmic approaches may be exploited in constructing low-power solutions.

A significant number of DSP functions involve frequency-selective digital filtering in which the goal is to reject one or more frequency bands while keeping the remaining portions of the input spectrum largely unaltered. Examples include lowpass filtering for signal upsampling and downsampling, bandpass filtering for subband coding, and lowpass filtering for frequency-division multiplexing and demultiplexing. The exploration of low-power solutions in these areas is therefore of significant interest.

To first order, the average power consumption, P, of a digital system may be expressed as

$$P = \sum_i N_i C_i V_{dd}^2 f_s \qquad (1)$$

where C_i is the average capacitance switched per operation of type i (corresponding to addition, multiplication, storage, or bus accesses), 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.

Real-time digital filtering is an example of a class of applications in which there is no advantage in exceeding a bounded

J. T. Ludwig is with the Digital Signal Processing Group, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

S. H. Nawab is with the Knowledge-Based Signal Processing Group, Boston University, Boston, MA 02215 USA.

A. P. Chandrakasan is with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

computation rate. For such applications, an architecture-driven voltage scaling approach has previously been developed in which parallel and pipelined architectures can be used to compensate for increased delays at reduced voltages [1]. This strategy can result in supply voltages in the 1 to 1.5 V range by using conventional CMOS technology. Power supply voltages can be further scaled using reduced threshold devices. Circuits operating at power supply voltages as low as 70 mV (at 300 K) and 27 mV (at 77 K) have been demonstrated [2] [3].

Once the power supply voltage is scaled to the lowest possible level, the goal is to minimize the switched capacitance at all levels of the design abstraction. At the logic level, for example, modules can be shut down at a very low level based on signal values [4]. Arithmetic structures (e.g., ripple carry versus carry select) can also be optimized to reduce transition activity [5]. Architectural techniques include optimizing the sequencing of operations to minimize transition activity, avoiding time-multiplexed architectures which destroy signal correlations, using balanced paths to minimize glitching transitions, etc. At the algorithmic level, the computational complexity or the data representation can be optimized for low power [6].

Another approach to reduce the switched capacitance is to lower N_i. Efforts have been made to minimize N_i by intelligent choice of algorithm, given a particular signal processing task [7]. In the case of conventional filter design, the filter order is fixed based on worst case signal statistics, which is inefficient if the worst case seldom occurs. More flexibility may be incorporated by using adaptive filtering algorithms, which are characterized by their ability to dynamically adjust the processing to the data by employing feedback mechanisms. In this paper, we illustrate how adaptive filtering concepts may be exploited to develop low-power implementations for digital filtering.

Adaptive filtering algorithms have generally been used to dynamically change the values of the filter coefficients, while maintaining a fixed filter order [8]. In contrast, our approach involves the dynamic adjustment of the filter order. This approach leads to filtering solutions in which the stopband energy in the filter output may be kept below a specified threshold while using as small a filter order as possible. Since power consumption is proportional to filter order, our approach achieves power reduction with respect to a fixed-order filter whose output is similarly guaranteed to have stopband energy below the specified threshold. Power reduction is achieved by dynamically minimizing the order of the digital filter.

The idea of dynamically reducing cost (in our case, power consumption) while maintaining a desired level of output quality (in our case, stopband energy in the filter output) emanates from the concept of approximate processing in computer science [9]. While approximate processing concepts may be used to describe a variety of existing techniques in digital signal processing (DSP), communications, and other areas, there has recently been progress in formally using these concepts to develop new DSP techniques [10]–[12]. Since our adaptive filtering technique falls into this category, we refer to our approach as adaptive approximate filtering, or simply approximate filtering.

II. DIGITAL FILTERING TRADE-OFFS

A frequency-selective digital filter may have either a finite impulse response (FIR) or an infinite impulse response (IIR). It is well known that IIR filters use fewer taps than FIR filters in order to provide the same amount of attenuation in the stopband region. However, IIR filters introduce nonlinear frequency dispersion in the output signals which is unacceptable in some applications. For such cases, it is desirable to use symmetric FIR filters because of their linear phase characteristic.

An important family of symmetric FIR filters corresponds to the symmetric windowing of the impulse responses of corresponding ideal filters. For example, a lowpass filter of this type has an impulse response given by [13]

$$h[n] = w[n]\frac{\sin \omega_c n}{\pi n} \qquad (2)$$

where $w[n]$ is a symmetric N-point window. This filter has cutoff frequency ω_c and may be implemented using a tapped delay line with N taps. For the purposes of this paper, we refer to such a filter as having order N. In Fig. 1, we display the frequency response magnitudes for three different values of N when $w[n]$ is a rectangular window and $\omega_c = \pi/2$. It should be observed that the mean attenuation beyond the cutoff frequency ω_c increases with filter order. Furthermore, with respect to a tapped delay-line implementation (see Fig. 2), the taps of the shorter Type I filters are subsets of the taps of the longer Type I filters. This ensures that if the filter order is to be decreased without changing the cutoff frequency, we can simply *power down* portions of the tapped delay line for the higher order filter. The price paid for such powering down is that the stopband attenuation of the filter decreases.

Butterworth IIR filters are commonly used for performing frequency-selective filtering in applications where frequency dispersion is tolerable. The frequency response magnitudes of such filters do not suffer from the ripples which can be seen in the frequency response magnitudes for FIR filters. These IIR filters are commonly implemented as cascade interconnections of second-order sections, each of which consists of five multiplies and four delays, as shown in Fig. 3. Also in Fig. 3 is an illustration of a cascade structure for an eighth-order IIR filter as the cascade of four second-order sections. For the purposes of this paper, we consider the order of a Butterworth IIR filter to be equal to twice the number of second-order
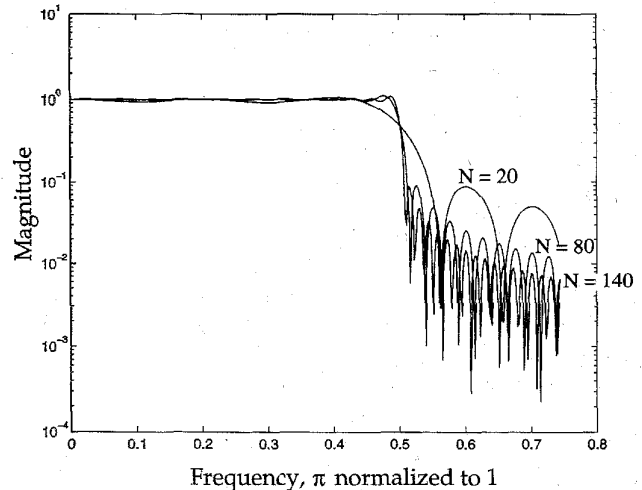

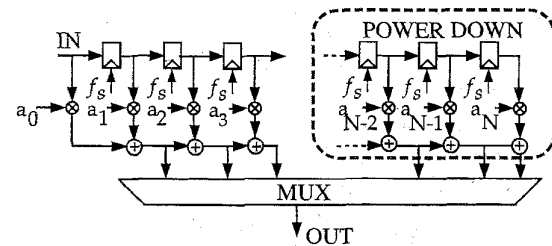Fig. 1. Frequency response magnitudes for FIR filters of orders $N = 20, 80$, and 140.

Fig. 2. Tapped delay line of an FIR filter structure, and the powering down concept To preserve phase linearity, powering down must be applied at both ends of the structure.
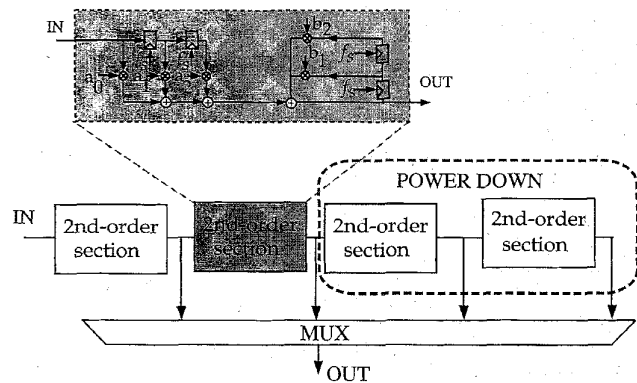

Fig. 3. Cascade implementation of an IIR filter structure. The detail of one of the second-order sections is shown.

sections in its cascade implementation. An interesting property of IIR Butterworth filters is that if the second-order sections are appropriately ordered, one may sequentially power down the later second-order sections and effectively decrease the net stopban attenuation of the filter.

III. ADAPTIVE APPROXIMATE FILTERING

In this section we present the details of our approximate processing approach to low-power frequency-selective filter-

ing. As discussed earlier, frequency-selective filters are used in applications where the goal is to extract certain frequency components from a signal while rejecting others. Suppose a signal, $x[n]$, consists of a passband component, $x_p[n]$, and a stopband component, $x_s[n]$. That is,

$$x[n] = x_p[n] + x_s[n]. \tag{3}$$

If it were possible to *cost-effectively* measure the strength of the stopband component, $x_s[n]$, from observation of $x[n]$, we could determine how much stopband attenuation is needed at any particular time. When the energy in $x_s[n]$ increases, it is desirable to increase the stopband attenuation of the filter. This can be accomplished by using a higher-order filter. Conversely, the filter order may be lowered when the energy in $x_s[n]$ decreases. We have developed a practical technique, based upon adaptive filtering principles, for dynamically estimating the energy fluctuations in the stopband component, $x_s[n]$, and using them to adjust the order of a frequency-selective FIR or IIR filter. As described in the previous section, the decrease in filter order enables the *powering down* of various segments of the filter structure. 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 FIR delay line is implemented using SRAM, even the data shifting operation of the higher order taps can be eliminated through appropriate addressing schemes.

Our overall technique is depicted in Fig. 4. The quantity $d[n]$, which represents the energy differential between the input and the output, is obtained as

$$d[n] = E_x[n] - E_y[n] \tag{4}$$

where

$$E_x[n] = \frac{1}{L} \sum_{k=0}^{L-1} |x[n-k]|^2 \tag{5}$$

and

$$E_y[n] = \frac{1}{L} \sum_{k=0}^{L-1} |y[n-k]|^2. \tag{6}$$

The filter order for sample period n, Order $[n]$, is updated at each sample period. One approach for the update process is to choose Order $[n]$ to be the smallest positive integer which guarantees that the stopband energy, $Q[n]$, of the output signal will be maintained below a specified threshold γ. Assuming that the stopband portion of the input spectrum is essentially flat,[1] the stopband energy in the output can be estimated as

$$Q[n] = \alpha d[n] E_{SB}[\text{Order}[n]] \tag{7}$$

where α is a proportionality constant, and $E_{SB}[k]$ represents the stopband energy in the frequency response, $H_k(\omega)$, of the kth order filter. That is,

$$E_{SB}[k] = \frac{1}{2\pi} \int_{SB} |H_k(\omega)|^2 \, d\omega \tag{8}$$

[1] In practice we have found that this flatness constraint may be relaxed considerably without detrimental effects.
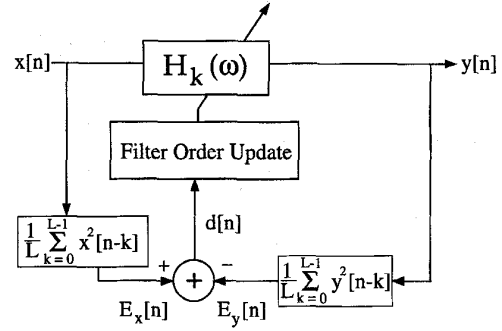

Fig. 4. Overview of approximate filtering strategy.

where SB denotes the stopband region. Since for every sample period this approach requires an expensive search over the stored values of $E_{SB}[k]$, we have designed a more efficient strategy which incrementally updates the most recent filter order. In this case, we estimate the stopband energy in the output as

$$Q[n] = \alpha d[n] E_{SB}[\text{Order}[n-1]]. \tag{9}$$

The decision rule for choosing Order $[n]$ is then given by

$$\text{Order}[n] = \begin{cases} \text{Order}[n-1] + N_0, & Q[n] > \gamma \\ \text{Order}[n-1], & \gamma - \delta \leq Q[n] \leq \gamma \\ \text{Order}[n-1] - N_0, & Q[n] < \gamma - \delta \end{cases} \tag{10}$$

where α, γ, δ, and N_0 are application-specific parameters. It should be noted that the filter order is changed at most by N_0 during each sample period.

The parameters δ and N_0 in (10) control the sensitivity of the time evolution of the filter order. The choice of the parameter L in (5) and (6) involves a trade-off between suppression of sensitivity to local fluctuations and preservation of the possible time-varying nature of the signal energy. For the case of FIR filters, we also observe that when the value of L is less than the maximum filter order, there is no extra storage required to compute $E_x[n]$ beyond that required for the filter implementation. On the other hand, excess storage is always required to update $E_y[n]$.

The arithmetic cost of the update process can be easily shown to involve five multiplications, five additions, one table lookup from a small memory module, and simple control. This cost is roughly equivalent to that of increasing the FIR filter order by five or the IIR filter order by two. This, for example, means that net power savings can be expected in the FIR case if for significant periods of time the dynamic FIR filter order decreases by more than five with respect to the maximum filter order. The overhead of multiplication is reduced to one multiplication instead of five per update if absolute value operations are used to compute $E_x[n]$ instead of magnitude-squared operations.

IV. RESULTS

In the context of FIR filters, we have used simulations of our approximate filtering technique to show that reduction in
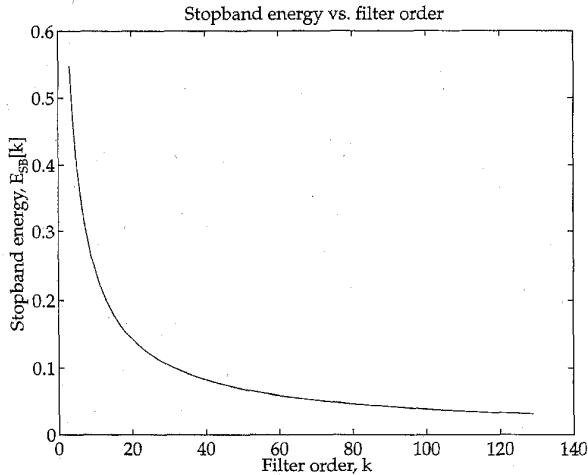
Fig. 5. FIR filter stopband energy, $E_{SB}[k]$ versus filter order, k, for the rectangular window family of FIR filters.

power consumption by an order of magnitude is achieved over fixed-order filter implementations when the stopband energy of the output signal is stipulated to remain below a given threshold γ. The context for most of these simulations is frequency-division demultiplexing of pairs of speech waveforms.

1) The Speech Signals: Each of the speech signals used in our simulations was sampled at 8 KHz and normalized to have maximum amplitude of unity. Each signal corresponds to a complete sentence with negligible silence at its beginning and end.

2) Frequency-Division Multiplexing: Each digitized speech waveform was pre-filtered to have a maximum frequency of 1.5 KHz. A guard band of 1 KHz was used in multiplexing a reference speech signal (corresponding to the sentence, "That shirt seems much too long,") with each of the other speech signals. The reference signal always occupied the 0 to 1.5 KHz band, while the other signals always occupied the 2.5 KHz to 4 KHz band.

3) The Demultiplexing: Demultiplexing involves lowpass filtering (cutoff frequency 2 KHz) to isolate the reference speech signal. The approximate filtering technique was used to perform this lowpass filtering for each of the 10 frequency-division multiplexed (FDM) signals. The parameter values in (10) were chosen to be

$$10 \log \gamma = -40 \, \text{dB}, \quad \delta = \frac{\gamma}{10}, \quad N_0 = 2, \quad L = 100. \quad (11)$$

The family of FIR filters used in these simulations corresponds to (2) with $w[n]$ rectangular. The values of $E_{SB}[k]$ for this case are plotted in Fig. 5.

4) Performance: In Table I we have listed various measures obtained for the performance of the approximate filter as it was applied to each FDM signal. The first column contains the sentence number for the stopband component of the input signal. The second and third columns, respectively, list the minimum and maximum filter orders used by the approximate filter in each case. The final column shows the relative power

TABLE I
FILTERING PERFORMANCE FOR DEMODULATING FDM SPEECH

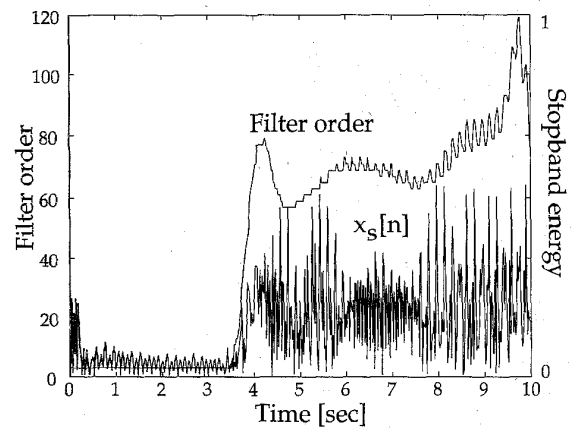| Sentence Number	Minimum Order	Maximum Order	Power Consumption $(P_{\text{fixed}}/P_{\text{adaptive}})$
1	3	99	4.4
2	3	117	6.3
3	3	87	5.6
4	3	93	5.9
5	3	61	4.2
6	3	119	5.9
7	3	125	7.1
8	3	111	7.1
9	3	127	7.1
10	3	107	8.3
Average	3	104.6	5.9

Fig. 6. Evolution of filter order for an FDM example. Two plots are shown in the figure. One shows the filter order as a function of time, while the other shows the stopband energy of the input signal as a function of time.

consumption of the approximate filter with respect to a fixed-order filter which is guaranteed to keep the stopband energy in the output below γ for all times. We observe that our adaptive technique reduces the average power consumption by a factor of 5.9.

To gain further insight into the source for this power reduction, in Fig. 6 we illustrate the nature of the adaptation performed by our technique in the case of one of the FDM signals. One of the curves shows the evolution of the filter order while the other curve shows the energy profile of the stopband signal. Clearly, the variations in filter order roughly follow the energy variations of the stopband signal. In particular, the most power savings is achieved during the silence regions of the stopband signal.

5) Speech Communication Implications: Longer periods of speech communication generally include significantly larger fractions of silence periods than an individual sentence. To factor this into our analysis, we repeated our simulations while inserting additional silence at the end of each speech signal. The average (over all 10 cases) of the relative power consumption is displayed in Fig. 7 as a function of the silence duration relative to the duration of the entire signal. As expected, the power reduction improves as the relative amount of silence is increased.
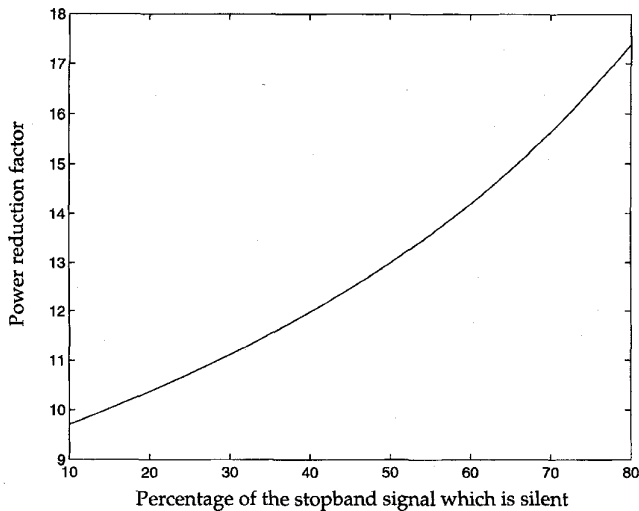
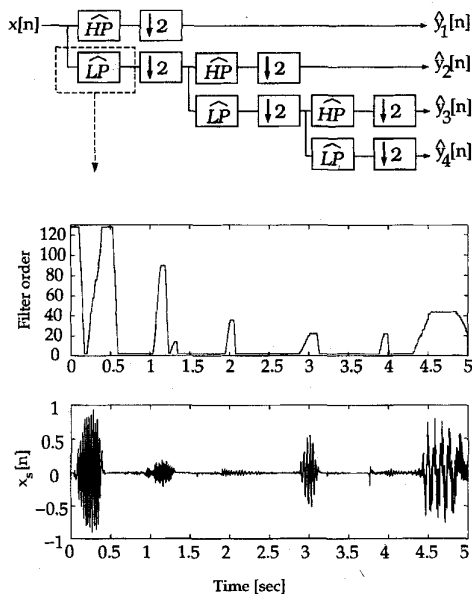Fig. 7. Filter performance versus percentage silence in stopband signal.

Fig. 8. Filter order evolution for the approximate filtering subband decomposition example. The top plot shows the filter order as a function of time, which tracks the input's stopband component $x_s[n]$, which is shown in the bottom plot.

6) Subband Coding: Data compression techniques for voice signals often use a binary tree-structured filterbank of highpass and lowpass filters, as depicted at the top of Fig. 8. Each of these filters may be implemented using the proposed approximate filtering technique. To illustrate the potential for power savings in the first stage of the subband decomposition, an approximate FIR lowpass filter was applied to a speech signal, $x[n]$, corresponding to the sentence, "That shirt seems much too long." The time-varying FIR filter order used by our technique is shown in the top plot of Fig. 8. The bottom plot in Fig. 8 shows the input's stopband component, $x_s[n]$, to demonstrate that the filter order roughly tracks the stopband energy of the input signal.

V. CONCLUSIONS

An algorithm-based approach has been presented for obtaining low-power implementations of important classes of IIR and FIR digital filters. In this approach, adaptive filtering and approximate processing concepts are combined to design digital filters which have the important property that the filter order can be dynamically varied in accordance with the stopband energy of the input signal. Simulations of the proposed technique using a variety of speech signals have shown that our approach offers significant power savings over standard fixed-order implementations. Finally, we note that while we illustrated our proposed technique in the context of lowpass filtering applications, it is equally applicable to other types of frequency-selective filtering.

REFERENCES

[1] A. P. Chandrakasan, S. Sheng, and R. W. Brodersen, "Low-power digital CMOS design," *IEEE J. Solid State Circuits*, pp. 473–484, Apr. 1992.
[2] R. M. Swanson and J. D. Meindl, "Ion-implanted complementary MOS transistors in low-voltage circuits," *IEEE J. Solid-State Circuits*, vol. SC-7, no. 2, pp. 146–152, Apr. 1972.
[3] J. B. Burr, "Cryogenic ultra low power CMOS," in *1995 IEEE Symp. on Low Power Electronics*, Oct. 1995, pp. 82–83.
[4] M. Alidina, J. Monteiro, S. Devadas, A. Ghosh, and M. Papaefthymiou, "Precomputation-based sequential logic optimization for low power," in *1994 Int. Workshop on Low-Power Design*, Apr 1994, pp. 57–62.
[5] T. Callaway and E. Swartzlander, Jr., "Optimizing arithmetic elements for signal processing," in *VLSI Signal Processing, V*, pp. 91–100, *IEEE Special Publications*, 1992.
[6] A. Chandrakasan and R. Brodersen, *Low Power Digital CMOS Design*. Norwell, MA: Kluwer, July 1995.
[7] B. M. Gordon and T. Meng, "A low power subband video decoder architecture," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, Apr. 1994, pp. II-409–412.
[8] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1991.
[9] V. R. Lesser, J. Pavlin, and E. Durfee, "Approximate processing in real-time problem solving," *AI Mag.*, pp. 49–61, Spring, 1988.
[10] S. H. Nawab and E. Dorken, "A framework for quality versus efficiency tradeoffs in STFT Analysis," *IEEE Trans. Signal Processing*, pp. 998–1001, Apr. 1995.
[11] S. H. Nawab and J. Winograd, "Approximate signal processing using incremental refinement and deadline-based algorithms," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Apr. 1995, Detroit, MI, pp. 2857–2860.
[12] J. T. Ludwig, S. H. Nawab, and A. Chandrakasan, "Low power filtering using approximate processing for DSP applications," in *Proc. Custom Integrated Circuits Conf. (CICC)*, Santa Clara, CA. May, 1995, pp. 185–188.
[13] A. Oppenheim and R. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

Jeffrey T. Ludwig received the S.B. degree in aeronautics and astronautics in 1991 and the S.M. degree in electrical engineering in 1993, both from the Massachusetts Institute of Technology, Cambridge, MA.

He is currently a graduate student in the Digital Signal Processing Group of the Research Laboratory of Electronics at MIT, pursuing a Ph.D. in electrical engineering. His research interests are in digital signal processing and its applications.

S. Hamid Nawab received the S.B., S.M., and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology in 1977, 1979, and 1982, respectively.

He is currently an Associate Professor in the Department of Electrical, Computer, and Systems Engineering at Boston University. He has held visiting professorships in electrical engineering at MIT (1994–95) and in computer science at University of Massachusetts at Amherst (1988–89). His research primarily involves the exploration of new algorithms and architectures for digital and knowledge-based signal processing. He is co-editor of the book, *Symbolic and Knowledge-based Signal Processing* (Prentice-Hall, 1992). He also joins A. V. Oppenheim and A. S. Willsky on the forthcoming second edition of their Prentice-Hall text on *Signals and Systems*.

Dr. Nawab is the winner of the 1988 Paper Award from the IEEE Signal Processing Society for his paper entitled "Direction Determination of Wideband Signals." He is also the recipient of the 1993 Metcalf Award for Excellence in Teaching at Boston University.

Anantha P. Chandrakasan received the B.S, M.S., and Ph.D. degrees in electrical engineering and computer sciences from the University of California, Berkeley, in 1989, 1990, and 1994, respectively.

Since September 1994, he has been the analog devices career development Assistant Professor of Electrical Engineering at the Massachusetts Institute of Technology, Cambridge. His research interests include the ultra low power implementation of custom and programmable digital signal processors, wireless sensors and multimedia devices, emerging technologies, and CAD tools for VLSI. He is a co-author of the book titled *Low Power Digital CMOS Design* (Kluwer).

Dr. Chandrakasan has received the NSF Career Development Award and the IBM Faculty Development Award. He received the IEEE Communications Society 1993 Best Tutorial Paper Award for the *IEEE Communications Magazine* paper titled, "A Portable Multimedia Terminal."