An Integrated 200-MHz 3.3-V BiCMOS Class-IV Partial-Response Analog Viterbi Decoder

Mohammad Hossein Shakiba, *Student Member, IEEE,* David A. Johns, *Senior Member, IEEE,* and Kenneth W. Martin, *Fellow, IEEE*

Abstract—Analog Viterbi decoders have recently been shown to be viable alternatives to their digital counterparts. In fact, a commercial analog class-IV partial-response sequence detector for magnetic read channels has already been reported. Analog decoders offer the advantages of reduced power and size primarily due to the elimination of the A/D. The analog Viterbi decoder described here is less complex and more robust compared to other reported realizations. The decoder is based on a new derivation of the difference-metric algorithm which is developed from an analog implementation perspective. This has resulted in a decrease in hardware complexity thereby making an analog approach more attractive for today's demanding high-speed, low-power, and small-size applications, such as magnetic disk-drive storage systems. The decoder was fabricated in a 0.8-μm BiCMOS process. It consists of two time-interleaved dicodes and the interleaving circuitry. The decoder was tested at up to 100 MS/s. However, since each dicode was also tested at this speed, the class-IV decoder should be capable of operating at 200 MS/s. Direct experiments at this speed were not possible due to the test equipment limitations. The chip consumes 30 mW from a 3.3-V power supply and occupies a core area of 0.5 mm^2.

Index Terms— Analog integrated circuits, analog Viterbi decoder, magnetic recording, MLSD, partial-response signaling, PRML, PRS, Viterbi decoder.

I. INTRODUCTION

PARTIAL-RESPONSE signaling (PRS) [1] is a signaling scheme first proposed for data communication [2], [3]. A PRS system introduces a controlled amount of intersymbol interference (ISI) to the signal before the signal is transmitted. This controlled ISI is then removed at the receiver. By relaxing the condition of zero ISI, certain beneficial effects can be attained through convenient spectral shaping. Two examples of these effects are providing more similarity between the spectrum of the transmitted signal and the frequency response of the channel, and realizing minimum-bandwidth transmission systems in practice.

The operation of a PRS system can be modeled by a finite impulse response (FIR) filter. The transfer function of the filter, expressed in terms of a time-step delay D, is known as the coding polynomial. Two commonly used factors of coding polynomials are $1 - D$ and $1 + D$. These two factors,
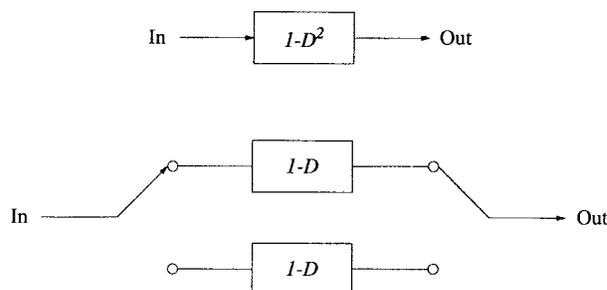
Fig. 1. Time-interleaved structure of a class-IV PRS system.

namely dicode and duobinary, create often-desirable spectral nulls at dc and $f_s/2$, respectively. Combining these two factors results in the class-IV system with the coding polynomial $1 - D^2$. In addition to the usefulness of the spectral shaping attained from this signaling scheme, it is also attractive from an implementation point of view. A class-IV system can be built by time-interleaving two independent dicodes [4]. This decomposition is particularly useful at high speeds, as, in addition to reducing the complexity, it also reduces the speed of each dicode to half the symbol rate. Fig. 1 illustrates the time-interleaved decomposition concept.

Beside data communication, PRS is receiving considerable attention in the magnetic-storage area. It has been shown that the read signal of a saturated magnetic-recording system resembles a partial-response signal [5]. Although more complicated partial-response schemes have been proposed, a class-IV scheme appears to provide a good compromise between the density of the storage device and the complexity of the detector.

PRS is a multilevel signaling scheme and exhibits a loss in the performance if conventional symbol-by-symbol detection is used. However, this loss can be combatted if a more complicated detection scheme is employed. It has been shown that maximum-likelihood sequence detection (MLSD) leads to the optimum performance because it fully exploits the redundancy introduced by the level coding [6], [7]. MLSD is usually realized by the Viterbi algorithm (VA) [8], [9].

The basic idea behind Viterbi detection is to consider the received sequence as a finite-state discrete-time Markov process contaminated by memoryless noise. A trellis diagram is conceptually constructed by unwrapping the state diagram in time. The detector assigns a metric to each branch of the trellis, proportional to the error signal (usually mean-square error) between the received value and the ideal signal resulting from
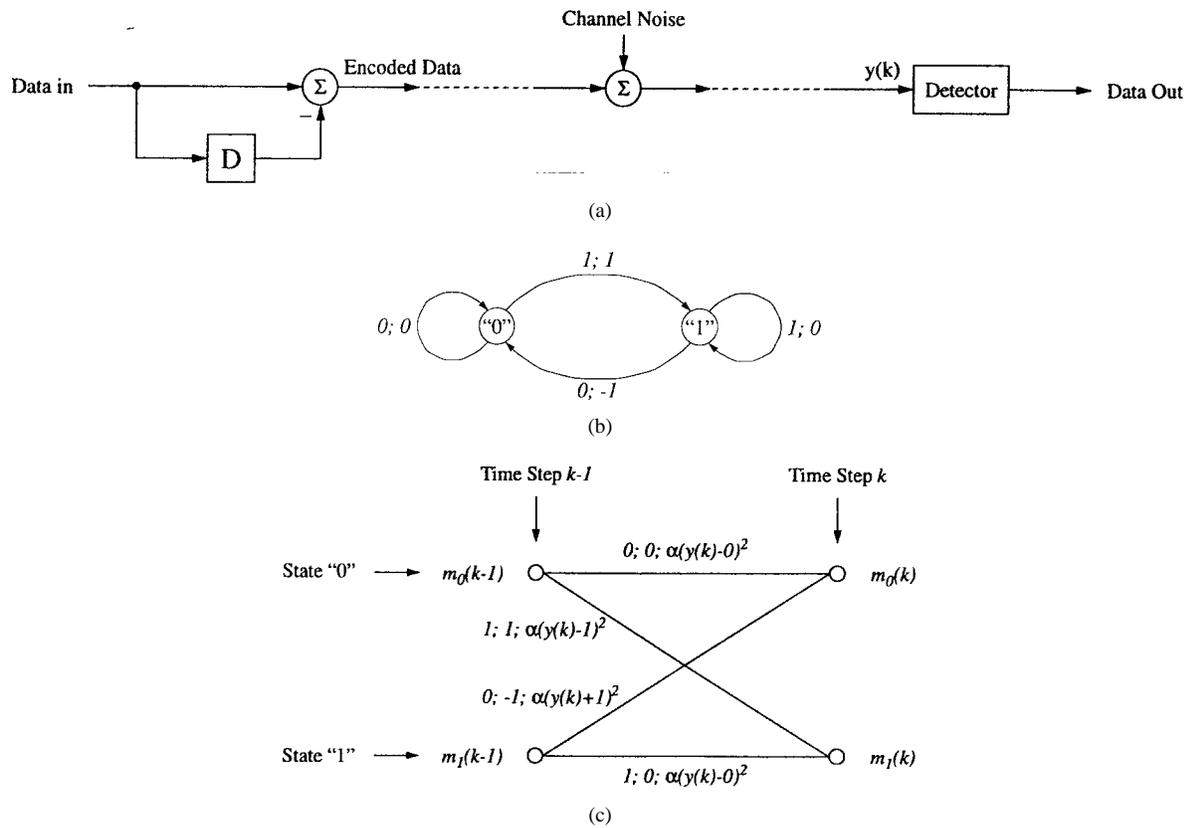
Fig. 2. A dicode system: (a) system model, (b) state diagram, and (c) trellis diagram.

that transition. The maximum-likelihood sequence is the one which results in the minimum accumulated error throughout the trellis. This approach is algorithmic in the sense that at each time step, and for each one of the states of the trellis, the state metric, defined to be the accumulated error signal, is calculated using the previous state metrics and the branch metrics at that time step. In addition to the state metrics, the paths along which these optimum metrics have been obtained are also saved. A block of digital memory can be used to save the required information. Following the literature, we shall refer to this memory as path memory and its contents as survivor sequences.

Although the VA has been traditionally implemented in the digital domain, high-speed, small size, and low-power constraints have motivated researchers to look for analog realizations. Analog Viterbi decoders have demonstrated many advantages over digital realizations [10]–[13], and today's state-of-the-art partial-response read channel often employs an analog detector in its processor core. In an analog implementation, savings are mainly due to the elimination of the A/D, which usually turns out to be a large and power-hungry block at high speeds. This paper describes an integrated analog Viterbi decoder for class-IV partial-response signals. The decoder is based on a new derivation of the difference-metric Viterbi algorithm, to be described in this paper. Here, each dicode decoder has an input-interleaved structure (in addition to time-interleaving two dicodes to realize the class-IV decoder) which eliminates analog feedback and thereby

substantially increases the speed of the overall circuitry. Furthermore, it is less complex and more robust with respect to circuit imperfections than other reported analog integrated decoders. It was fabricated in a 0.8-μm BiCMOS process and consumes 30 mW of power from a 3.3-V single power supply. The decoder should be capable of operating at up to 200 MS/s, since each individual dicode was tested at 100 MS/s. Direct tests of the class-IV decoder were limited to 100 MS/s due to our test equipment limitations. Each dicode decoder consists of a fully differential analog processing core and a digital path memory. The interleaving and de-interleaving circuits are also included on the chip. The areas occupied by the analog and digital parts are only 0.06 and 0.1 mm^2 per dicode. The total core area of the chip is 0.5 mm^2.

II. VITERBI DETECTION OF CLASS-IV SIGNALS: THE DIFFERENCE-METRIC ALGORITHM

A binary PRS system with an Nth order coding polynomial has 2^N states. Consequently, a class-IV system results in a four-state trellis diagram. However, by interleaving two independent dicodes, two identical two-state trellis diagrams can be used to represent the operation of the system. In this paper, we mainly focus on a dicode sequence detector. The final class-IV decoder is constructed by time-interleaving two such decoders.

Fig. 2 illustrates a simplified dicode communication system, the encoder state diagram, and one step of the trellis diagram used by the sequence detector. Without loss of generality, it

is assumed that the combination of the band-limiting transmit filter, the channel, and the noise-reduction receive filter (not shown in the figure) acts as a Nyquist filter such that the ISI is exclusively determined by the FIR filter in the transmitter.[1] In calculating the branch metrics, the mean-square error criterion is used. This minimizes the Euclidean distance throughout the detected sequence and results in the optimum performance in the case of additive white Gaussian noise [7]. The VA applied to the two-state trellis yields the following update equations:

$$m_0(k) = \min \left\{ m_0(k-1) + \alpha y^2(k), \right.$$
$$\left. m_1(k-1) + \alpha(y(k)+1)^2 \right\}$$
$$m_1(k) = \min \left\{ m_1(k-1) + \alpha y^2(k), \right.$$
$$\left. m_1(k-1) + \alpha(y(k)-1)^2 \right\}. \quad (1)$$

Here, $y(k)$ represents the received signal, $m_i(j)$ denotes the state-metric (accumulated error) of state i at time step j, and α is an arbitrary positive scaling factor.[2]

From (1) it can be seen that adding equal amounts to all of the four involved terms does not affect the algorithm outcomes. By defining the difference-metric signal as

$$\Delta m = m_0 - m_1 \quad (2)$$

cancelling the common terms, and subtracting $m_1(k-1)$ from the above expressions one concludes that the update mechanism can equivalently take place by updating only the difference signal given by (2) instead of the individual state metrics. Furthermore, from the four combinations of two min functions, only three are possible. As a result, the above VA reduces to (3), shown at the bottom of the page. The graphs shown in (3) indicate how the path memory should be updated. As an example, if $\Delta m(k-1) < y(k) - 0.5$, the survivor sequence of state "0" simply extends by a "0," whereas the survivor sequence of state "1" consists of the previous sequence of state "0" extended by a "1."

The above simplified algorithm was first proposed in [4], is named the difference-metric algorithm, and has been further examined for magnetic-recording applications in a digital realization [14]. The first reported integrated analog implementation [15] did not fully exploit the algorithm, as the difference signal was obtained by subtracting the state metrics, hence not eliminating the need for calculating the individual metrics. It was shown in [11] that the exact difference-metric algorithm is in fact well-suited for an analog realization and leads to a very efficient and fast structure. In this structure, the threshold levels of the detector were adaptively updated in a feedback

---

[1] Note that in many cases, such as saturated magnetic recording, the coding operation is in fact done by the channel. Nevertheless, the above model is still applicable.

[2] For a negative scaling factor, the min functions should be replaced with max functions.
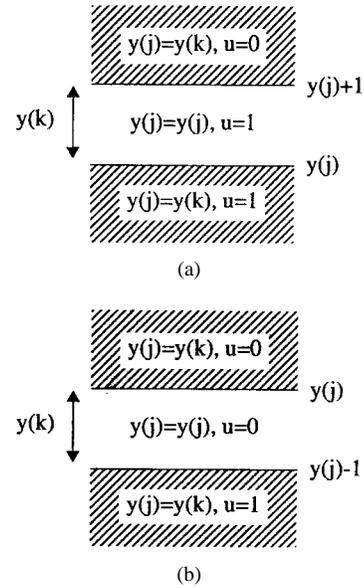
Fig. 3. A graphical sketch of the input-interleaved algorithm: (a) $u = 1$ and (b) $u = 0$.

loop. The outcome is equivalent to dynamically setting the threshold levels [16], but with a different implementation technique. Our implementation here is based on the approach taken in [11], however, with a major improvement in the speed of operation. This improvement is achieved by employing a new derivation of the difference-metric algorithm intended for an analog realization. This algorithm, referred to as the "*input-interleaved algorithm*," is described in the following section.

III. THE INPUT-INTERLEAVED ALGORITHM

A closer look at the recursion given by (3) reveals that $\Delta m(k-1)$ is equal to a dc-shifted version of a previously sampled input signal. Specifically, if this sample is denoted by $y(j)$, then

$$\Delta m(k-1) = y(j) \pm 0.5 \quad (4)$$

which, combined with (3), leads to the possible update equations, (5), shown at the bottom of the next page.

By defining u as a dc offset which can take one of the two values of one and zero (corresponding to the two alternatives), the above expressions can be combined. Also, note that as long as $y(j)$ and u are known, there is no need to calculate the difference signal. Iterations can equivalently proceed as shown in (6) at the bottom of the next page.

Expression (6) simply states that whenever $y(k)$ is in between the threshold levels, no update is required and the previous values of $y(j)$ and u should be retained. However,

$$\Delta m(k) = \begin{cases} y(k) - 0.5, & \Delta m \ (k-1) < y(k) - 0.5 \\ \Delta m(k-1), & y(k) - 0.5 < \Delta m \ (k-1) < y(k) + 0.5 \\ y(k) + 0.5, & y(k) + 0.5 < \Delta m \ (k-1) \end{cases} \quad (3)$$

when $y(k)$ falls outside this region, the previously-sampled input signal should be updated to the current input and the dc offset should be set either to zero or one depending on $y(k)$ being more than the upper or less than the lower threshold level, respectively. This is graphically sketched in Fig. 3. Fig. 4 shows a typical dicode signal and the trajectories of the thresholds. The threshold levels adapt themselves based on the history of the signal such that the noisy signal is successfully sliced.

IV. THE INPUT-INTERLEAVED ARCHITECTURE

The input-interleaved algorithm can be implemented by the block diagram shown in Fig. 5. Note that a master analog circuit [Fig. 5(a)] processes the input signal, whereas a slave digital path memory [Fig. 5(b)] stores the decisions made by the processor. The front-end of the processor contains two S/H's. While the input signal is being sampled and stored by one S/H, the previous input sample is held by the other S/H. The connections between these S/H's resemble an interleaved structure, giving rise to the name "*input-interleaved*" for the algorithm and the architecture that realizes it. "*Input-interleaved*" is chosen to differentiate the concept from what "*interleaving*" traditionally implies, that is interleaving in time by periodically alternating two such dicode decoders. Here, within each dicode, the digital feedback pulses determine the input port to which the signal should be directed. These pulses are not necessarily present in every clock cycle, nevertheless, they redirect the input from one port to the other each time a pulse is generated.

In the above architecture, the current input and the previously sampled input signals are combined in two branches. A dc offset is also added to only one branch. This branch is determined from the results of the previous iteration. Specifically, adding the offset to the upper branch corresponds to $u = 1$ in (6) and Fig. 3, whereas adding it to the lower branch corresponds to $u = 0$. Two comparators check the

polarity of the resulting signals. This is equivalent to slicing the input signal with the threshold levels specified by (6). The comparator outputs are used to generate the required switching signal S (to switch the dc offset between the two branches) and the toggling signal T (to toggle the input S/H's if an update in $y(j)$ is needed) as well as to update the contents of the path memory.

The update mechanism illustrated in Fig. 3 determines the rule for generating signals S and T. The input S/H's should toggle whenever an update in $y(j)$ is required (i.e., when the input signal exceeds the region between two threshold levels). In Fig. 5(a), it can be verified that if the input signal lies in between these levels none of the comparator outputs will be high and if $y(k)$ exceeds this region then one (and only one) of the comparators will result in a high output. Consequently, a toggling should occur if either one of the outputs is high. This is accomplished by employing a T flip-flop toggled by either one of the comparators. If the dc offset is already added to the upper branch, it should be switched to the lower branch only if $y(k)$ is above the upper threshold, that is, if the output of the upper comparator is high. (Note that the lower comparator output is low.) The offset should only be switched back to the upper branch if $y(k)$ falls below the lower threshold level. In this case, the lower comparator will have a high output and the upper comparator will have a low output. As a result, an SR flip-flop, set by the upper comparator and reset by the lower comparator, can be utilized to switch the dc signal back and forth between two branches. When $y(k)$ lies in the middle decision region, none of the comparator outputs are high, and the SR flip-flop does not change its state.

Note that in Fig. 5(a), in combining the sampled-input signals, a sign change results whenever a toggling occurs. This sign change is compensated by utilizing polarity switches which are controlled by the T flip-flop.

The outputs of the comparators in Fig. 5(a) are also used to update the path memory. The register-exchange method is a common technique for storage of the survivor sequences

$$\Delta m(k) = \begin{cases} y(k) - 0.5, & y(j) + 1 < y \quad (k) \\ y(j) + 0.5, & y(j) < y \quad (k) < y(j) + 1 \\ y(k) + 0.5, & y \quad (k) < y(j) \end{cases} , \quad \Delta m(k-1) = y(j) + 0.5$$

$$\Delta m(k) = \begin{cases} y(k) - 0.5, & y(j) < y \quad (k) \\ y(j) - 0.5, & y(j) - 1 < y \quad (k) < y(j) \\ y(k) + 0.5, & y \quad (k) < y(j) - 1 \end{cases} , \quad \Delta m(k-1) = y(j) = 0.5. \quad (5)$$

$$\begin{cases} y(j) = y(k), u = 0, & y(j) + u < y \quad (k) \\ y(j) = y(j), u = u, & y(j) + u - 1 < y \quad (k) < y(j) + u \\ y(j) = y(k), u = 1, & y \quad (k) < y(j) + u - 1 \end{cases} \quad (6)$$
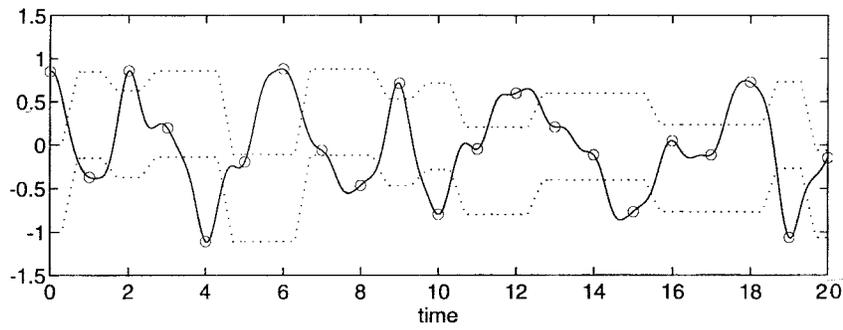
Fig. 4. An example of adaptation of the threshold levels in the input-interleaved algorithm.
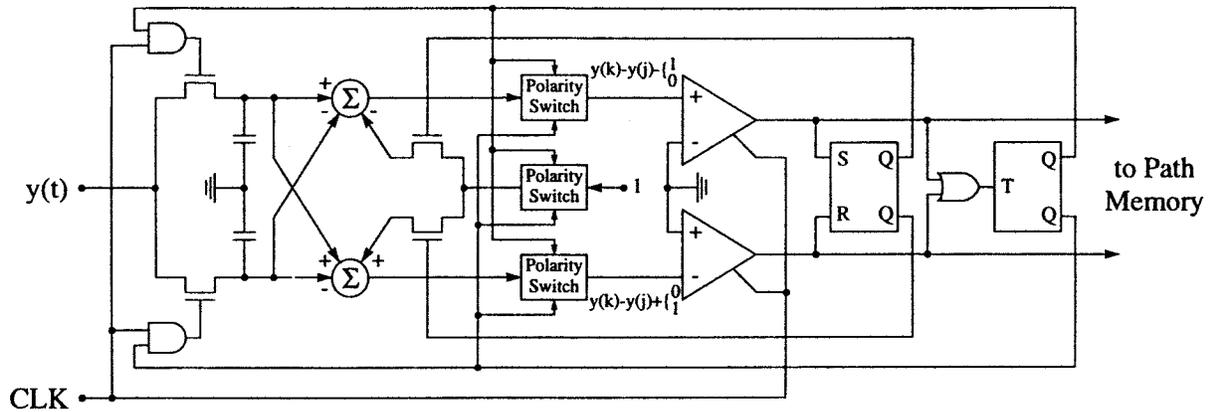
(a)
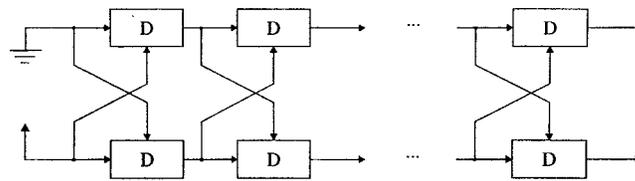
(b)

Fig. 5. The input-interleaved architecture: (a) analog processor and (b) digital path memory.
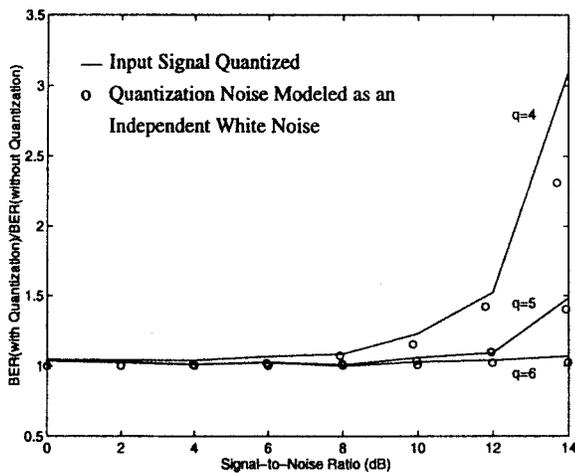
Fig. 6. Performance degradation of the dicode decoder due to quantizing the input signal.
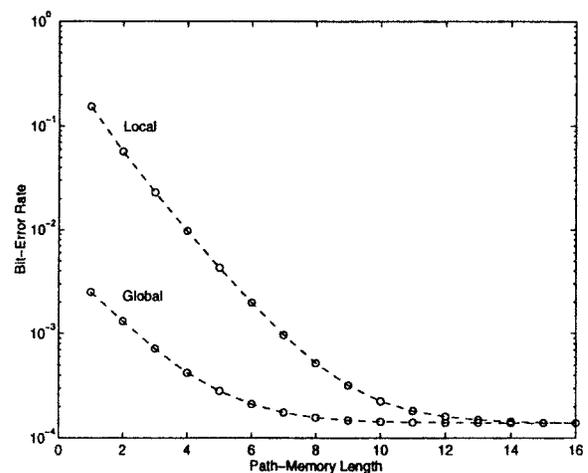
Fig. 7. Effect of employing the local trace-back method on the noise performance of the dicode Viterbi decoder at SNR = 12 dB.

in Viterbi decoders with low number of states [17]. In this method, one shift register with a length equal to the length of the path memory is used for each state. The different shift registers are then interconnected according to the trellis diagram such that the optimum paths along the trellis are directly mapped into digital sequences stored in these registers.
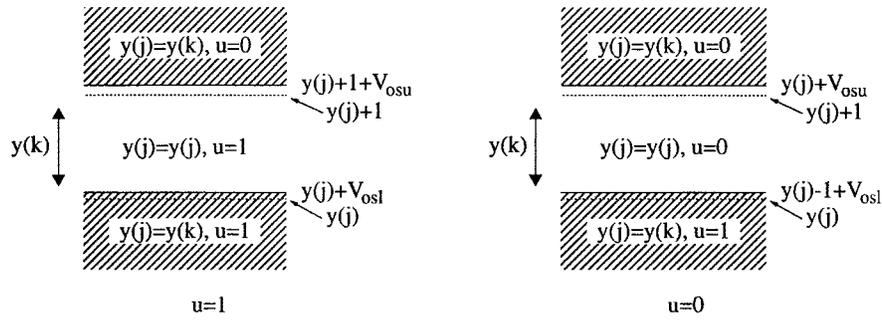
Fig. 8. Effect of comparator offsets on the threshold levels of the decoder.

(a)

(b)
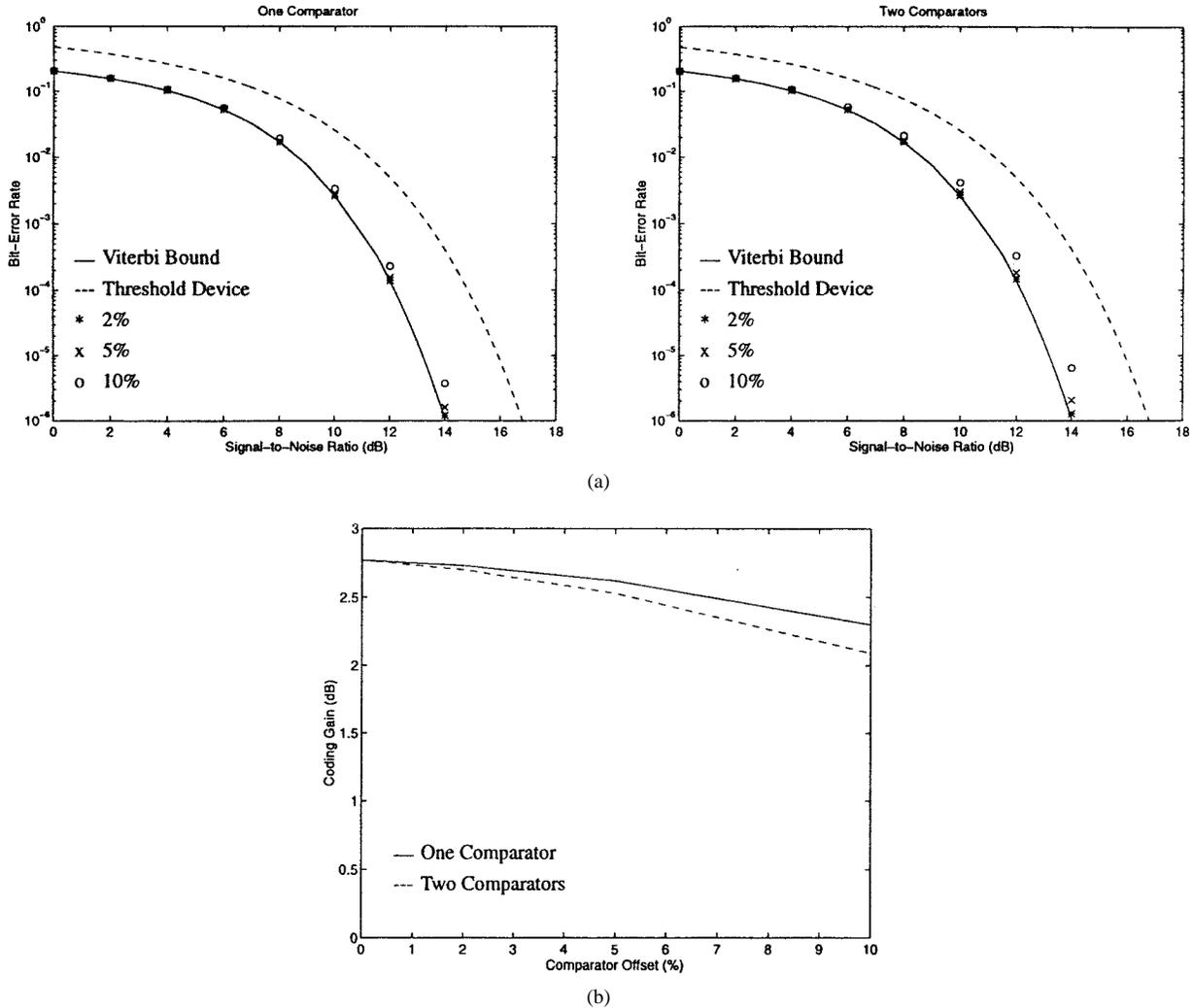
Fig. 9. Performance degradation of the input-interleaved decoder due to the comparator offsets: (a) BER performance and (b) coding gain at BER $= 10^{-6}$.

Applying this method to the dicode decoder results in two interconnected serial/parallel in/out shift registers as shown in Fig. 5(b). The serial/parallel loadings are set by the comparator outputs.

The advantages of the input-interleaved structure presented here to the structures used in other analog realizations can be summarized in its higher speed of operation, increased robustness against circuit imperfections, and simplicity in its fully differential implementation. In contrast to [10], in which analog signals were involved in the feedback paths, in the present structure only digital signals are fed back at the

end of iterations. Absence of analog signals in the feedback path eliminates the need for delays in the analog signals and significantly increases the overall speed. Loop delays are necessary to prevent a destructive feedback while the quantities are updated and were implemented by using master/slave S/H's in [10]. Although these S/H's were eliminated in the adaptive-threshold decoder proposed in [11], still further improvement in speed is achieved by avoiding the need for an intermediate S/H. Reducing the required sampling operations to the minimum of one greatly increases the speed, as they play the major role in this regard [18]. Also, by removing

analog signals from the feedback path in the present structure, an improvement in the robustness to analog imperfections is expected, since the decoder no longer faces accumulation of analog errors in the loop.

V. PRACTICAL IMPERFECTIONS

In a Viterbi decoder, some nonideal effects are structure independent and are present even in digital realizations. Among these, truncating the length of the path memory, quantizing the input signal, and simplifying the trace-back mechanism are usually the most important considerations. To reduce the decoding delay and the amount of memory, detecting is usually started before every transmitted symbol is received. This corresponds to using a truncated path memory and results in a degradation in the noise performance of the decoder. In general, the length of the memory is truncated such that the excess bit-error rate (BER) is negligible compared to the decoder BER [19].

In a digital realization, the limited number of bits used in the binary representation of the signals is another source of imperfection. The effect of this quantization is often considered as an independent additive white noise [20]. Depending on the relative power of this noise to the channel noise, the minimum required number of bits is determined. The simulated BER degradation of the dicode decoder with an input signal limited to ± 1 (peak values of the noiseless encoded signal) and quantized to q bits is plotted in Fig. 6. The results are also shown for the case where the signal is not quantized and the quantization noise is taken into account as an additional independent component in the overall noise. The model becomes more accurate as q increases. From this figure, it can be concluded that a minimum number of 6 b is required at a moderate-to-relatively-high SNR. The required number of bits can be translated to the accuracy needed in the analog realization. A 6-b accuracy is considered moderate and relatively simple analog circuits, hence fast, can be used.

Another source of nonideality, which is not specific to analog realizations, is related to the trace-back mechanism. In many cases, to simplify the processing, one state of the Viterbi decoder is arbitrarily chosen and its corresponding optimum sequence is traced back to obtain the decoded data. Compared to the global trace-back method, in which the optimum sequence of the state with the minimum accumulated error is traced back, this local trace-back technique results in a degradation in the BER, since the selected sequence may not yet have been merged with the actual optimum sequence. Apparently, this degradation can be arbitrarily reduced by increasing the depth of the path memory. This fact is demonstrated in Fig. 7, which shows that the BER of the dicode decoder with a local trace-back method approaches its minimum value as the length of the path memory in increased. In the cases where this increase in the decoding delay is not a critical issue, the local trace-back method might be preferable, since increasing the length of the path memory is straightforward.

Analog realizations usually suffer from dc offsets, mismatches, and charge injections. To examine the sensitivity
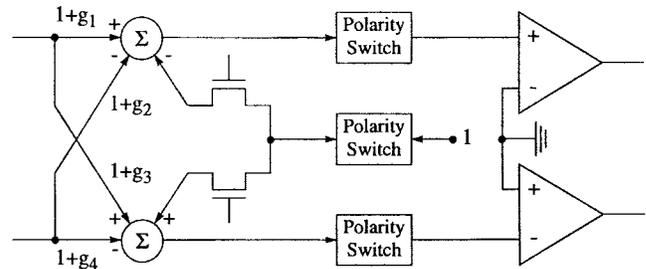
Fig. 10. Gain deviation factors used to quantify mismatches.
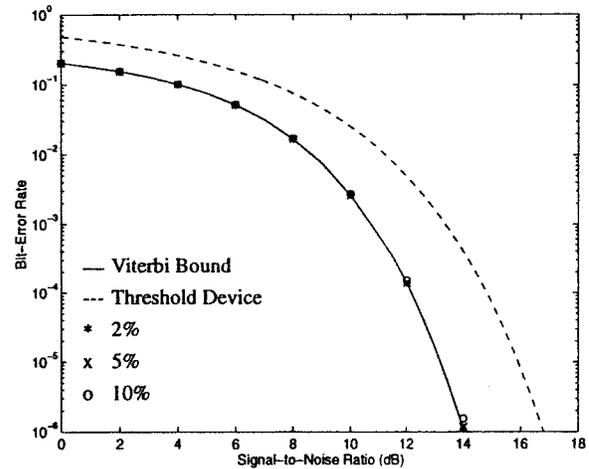
(a)
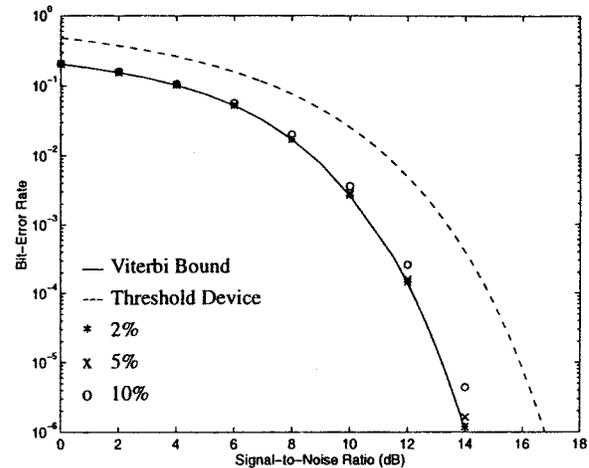
(b)

Fig. 11. Effects of gain mismatches on the performance of the input-interleaved decoder: (a) one gain factor and (b) all of the gain factors (worst case).

of the proposed structure to analog imperfections, the input-interleaved structure was simulated in the presence of these impairments. In what follows, major sources of errors are considered and performance degradation is evaluated. In our evaluations, the amounts of impairments may have been exaggerated. This is to illustrate the robustness of the decoder.

An offset introduced by one of the comparators in Fig. 5(a) can be translated to a shift in the threshold level realized by that comparator. Fig. 8 illustrates the concept when offsets
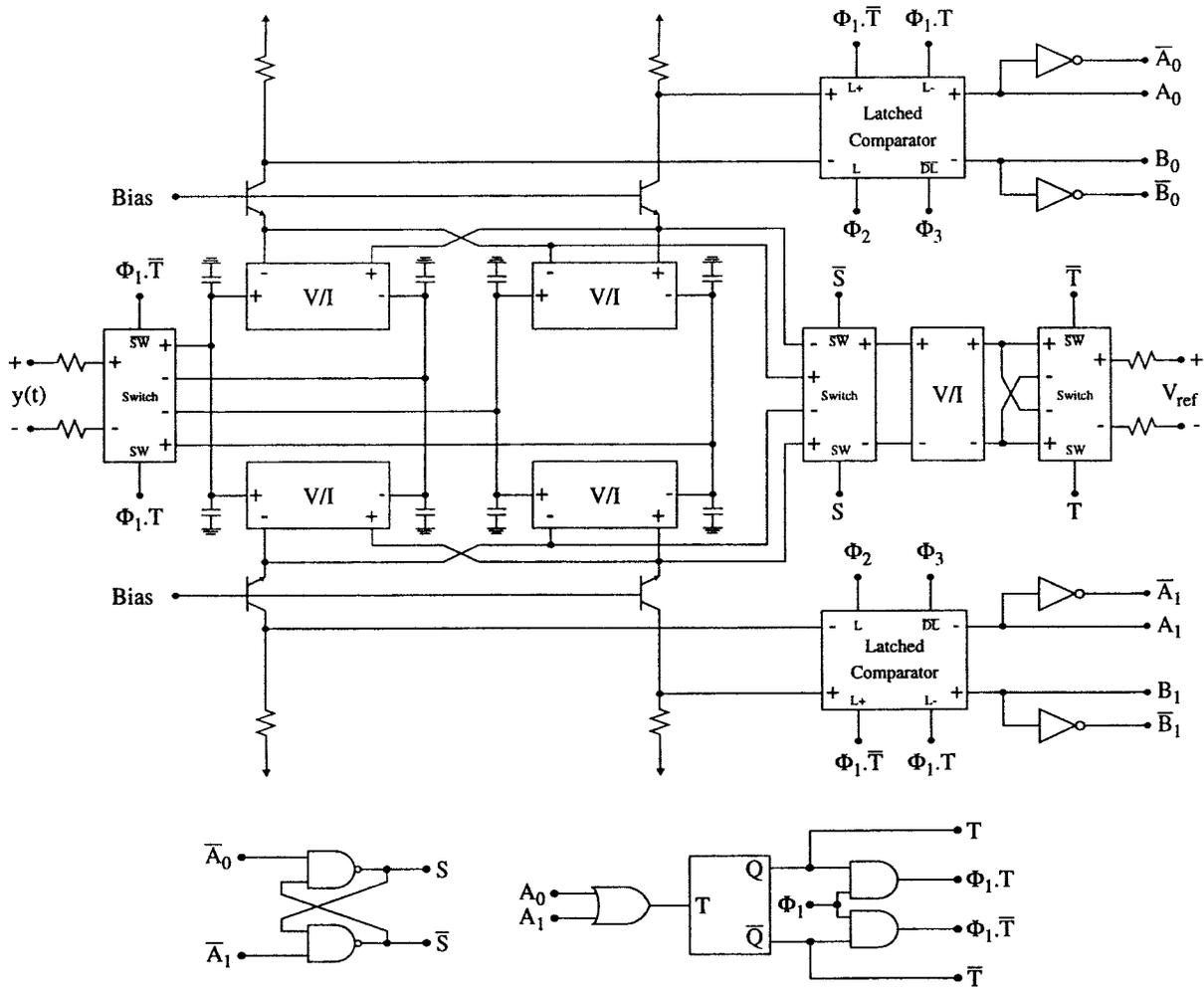
Fig. 12. Analog signal processor of the input-interleaved decoder. Note that each holding capacitor of the S/H's is replaced with two parallel capacitors to emphasize the fully differential structure of the circuit. This is also the case in the layout.
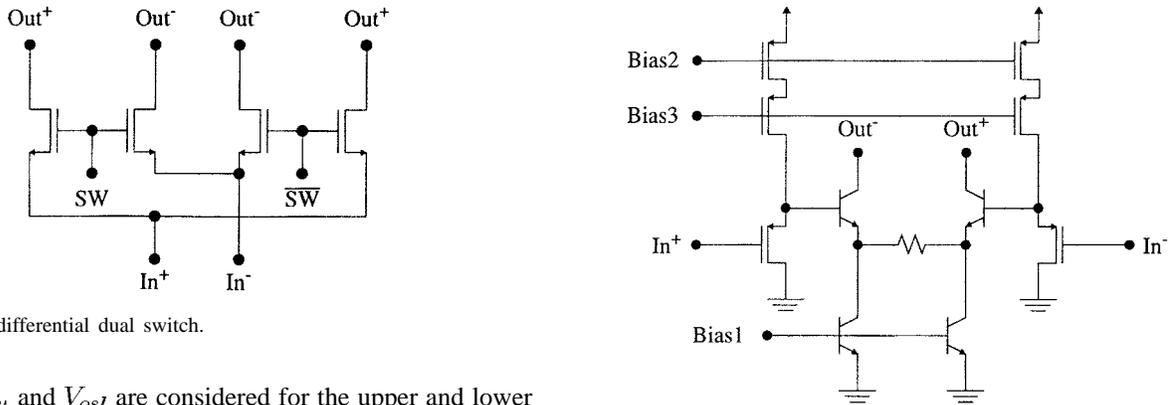
Fig. 13. The differential dual switch.

Fig. 14. Circuit realization of the V/I.

equal to V_{osu} and V_{osl} are considered for the upper and lower comparators, respectively. As can be seen from this figure, the offsets do not affect the performance if the input signal does not lie in the regions between the original threshold levels and the shifted levels. Otherwise, an error equal to either V_{osu} or V_{osl} will occur in updating the threshold levels. Although the effect can be modeled as a noise added to the input signal, the fact that this additional noise is neither Gaussian nor uncorrelated makes simulations more appealing. The BER performance is plotted in Fig. 9(a). The Viterbi bound and the performance of a fixed-threshold detector are also included in

the figure for comparison. From these plots, the achievable coding gain of the decoder in the presence of comparator offsets can be obtained. This is further illustrated in Fig. 9(b) for a BER of 10^{-6}. Low sensitivity to comparator offsets shows that for reasonable signal amplitudes (on the order of a fraction of a volt) simple and fast comparators without offset cancelation techniques can be employed.
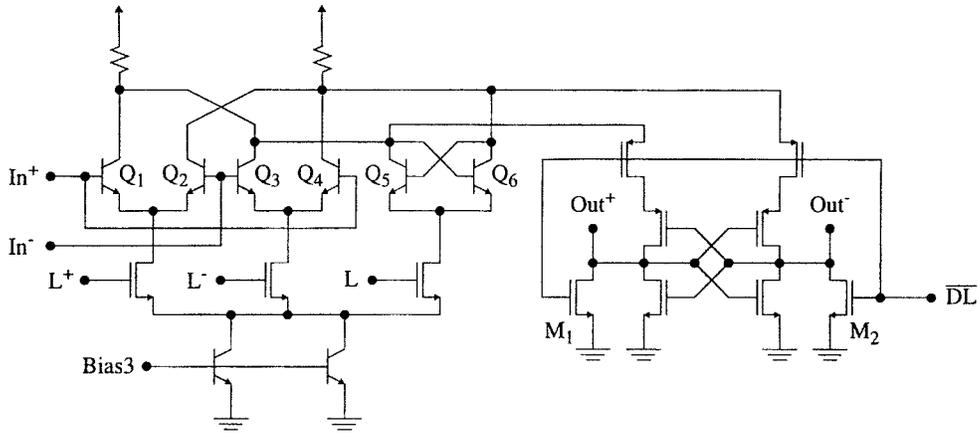
Fig. 15. The latched-comparator circuit.
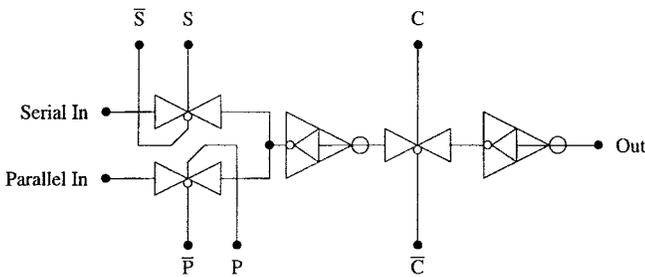
Fig. 16. Multiplexed-input D flip-flop used to construct the path memory.

From Fig. 5(a) it can be seen that offsets produced at the outputs of the combiners are equivalent to offsets in their corresponding comparators. Consequently, the previous results can be directly applied.

Gain mismatches result if the relative weights at the inputs of the combiners deviate from their nominal values. However, two sets of weights, corresponding to two combiners, can be scaled independently without affecting the performance. The effects of gain mismatches can be quantified by gain deviation factors shown in Fig. 10. Due to the symmetry in its input stage, the input-interleaved structure has identical sensitivities to all of the gain deviation factors. Fig. 11 depicts the sensitivity to a single factor. It also shows the overall effect of gain mismatches when all of these deviations are present. From the different combinations, the worst case is illustrated.

Since the reference voltage is switched to only one of the two branches in Fig. 5(a) at any time, any deviation from its value can directly be mapped to an equivalent offset in the corresponding comparator. As a result, a certain change in the reference voltage has a similar impact on the performance as an equal amount of offset in one comparator has.

The above sensitivity is also applicable to the cases where the input signal undergoes an unwanted attenuation or amplification. This is because the decoder is only sensitive to the relative amplitudes of the input signal and the reference voltage. In fact, the reference voltage should be scaled based on the amplitude of the input signal, which is often set by an automatic gain control (AGC) stage in practice.

In general, the S/H's used in the analog decoder contribute to the errors by partly injecting their channel charges and

clock signals to the stored voltages. Fortunately, equal errors introduced by these S/H's will be cancelled out in the combiners. As a result, only signal-dependent terms of the injected voltages may degrade the performance. However, if the input signal fluctuation is small, this degradation can be neglected altogether. This is the case in our implemented decoder, where the peak-to-peak value of the input is only a fraction of a volt compared to the full-swing control signals. Charge injection and clock-feedthrough are further rejected by employing a fully differential circuit in our realization.

VI. CIRCUIT IMPLEMENTATION

Fig. 12 shows a circuit-level block diagram of the implemented input-interleaved analog processor [Fig. 5(a)]. All signals are differential to combat destructive effects such as common-mode noise and S/H errors. The input S/H's, consisting of a differential dual switch connected to holding capacitors, store the present and the previous input signals. These signals are converted to currents and combined with appropriate polarities by passing the currents through resistors via cascode transistors. A dc voltage, obtained from an off-chip differential reference, is also converted to current, adequately switched, and added to one of the parallel branches.

The resulting differential voltages are applied to two latched comparators which decide on the polarity of these signals. Comparison results are used to update the path memory and also to generate the toggling T and the switching S signals. These signals are fed back to possibly update the previously sampled input signal by toggling the input interleaved S/H's and the dc offset signal by switching it from one branch to the other.

Φ_1 to Φ_3 are different phases of a clock signal. These phases are obtained from a master clock by a simple circuit which is discussed later in this section.

Based on the above circuit block diagram, and by utilizing a register-exchange path memory, a Viterbi decoder was designed. The chip contains two input-interleaved dicode decoders which were internally time interleaved to decode a class-IV partial-response signal. In what follows, the different building blocks are explained in more detail.
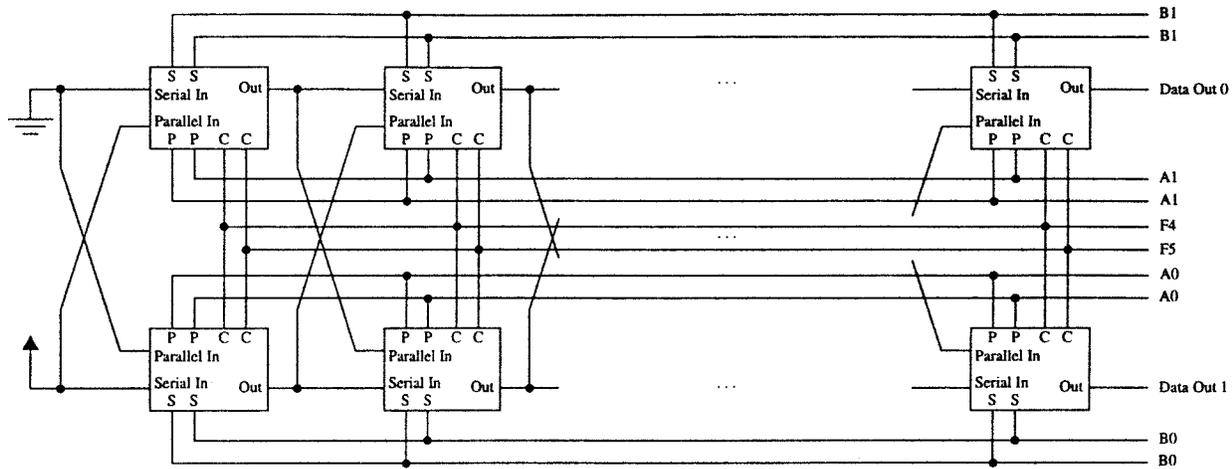
Fig. 17.   Path-memory implementation based on the register-exchange method and by utilizing the multiplexed-input D flip-flops shown in Fig. 16.
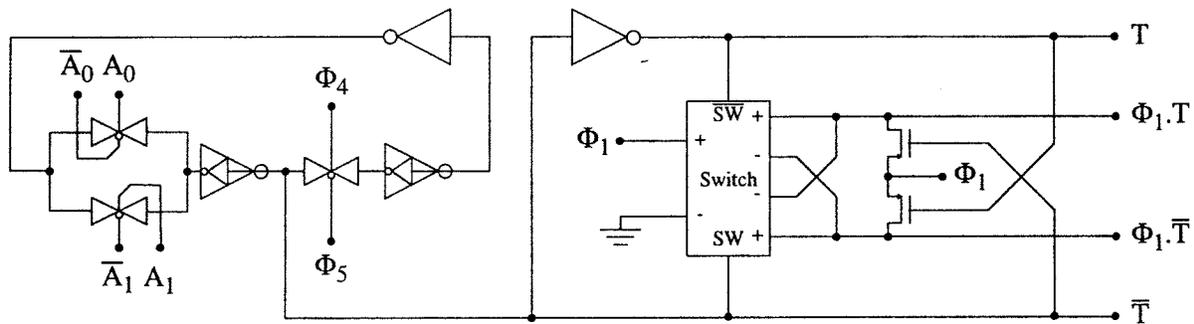
Fig. 18.   Circuit realization of the T flip-flop and required gates used to generate the toggling signals.
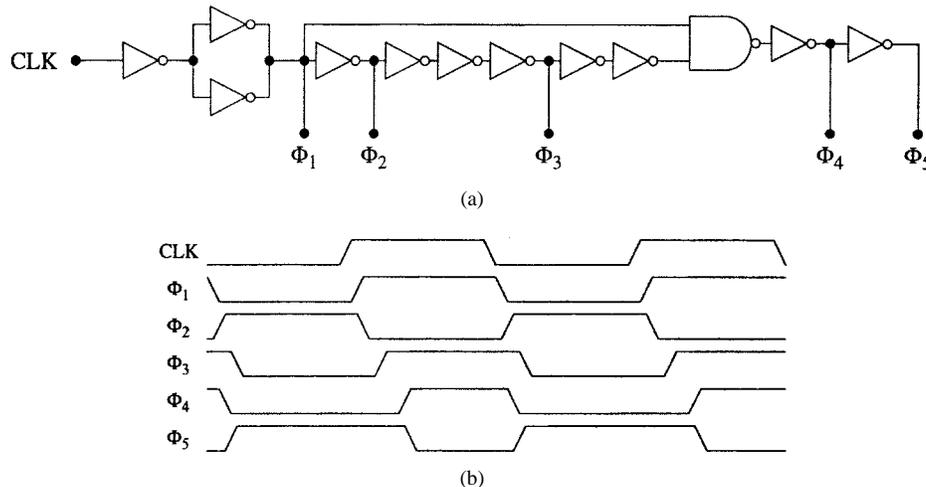
Fig. 19.   Generating different clock phases from a single-phase clock: (a) circuit schematic and (b) timing diagram.

A differential dual switch, shown in Fig. 13, was used in a variety of locations. This switch consists of four NMOS transistors and has one differential input and two differential outputs. Two complementary digital signals determine which output the input signal should be directed to. This switch was used to implement the input S/H's, to switch the offset signal back and forth between two branches, and to serve as a polarity switch for the reference voltage, all in a differential manner.

Also, the switch was employed to perform AND functions, as will be described below.

Depicted in Fig. 14 is a degenerated bipolar junction transistor (BJT) differential pair used to realize the voltage-to-current converter (V/I). In addition to increasing the linearity, resistive degeneration reduces gain mismatches, since voltage gains are dominantly set by resistor ratios. Mismatches between biasing current sources also contribute to the offsets. By
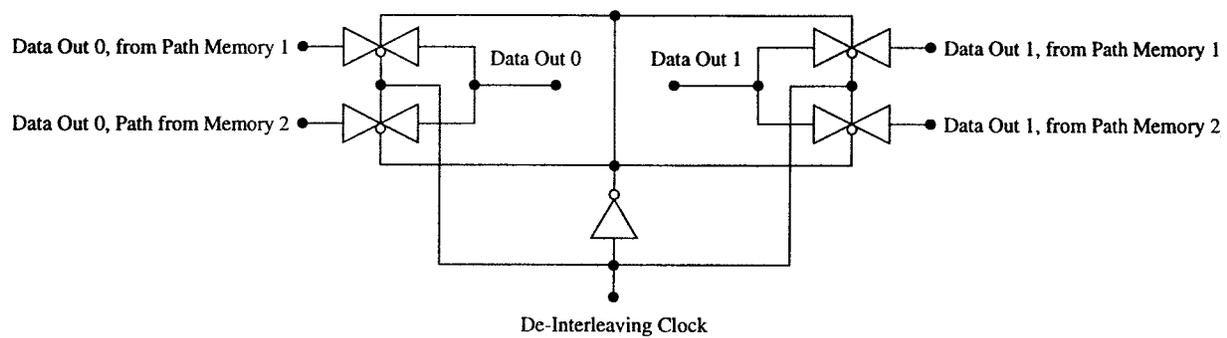
Fig. 20. The de-interleaving multiplexers.
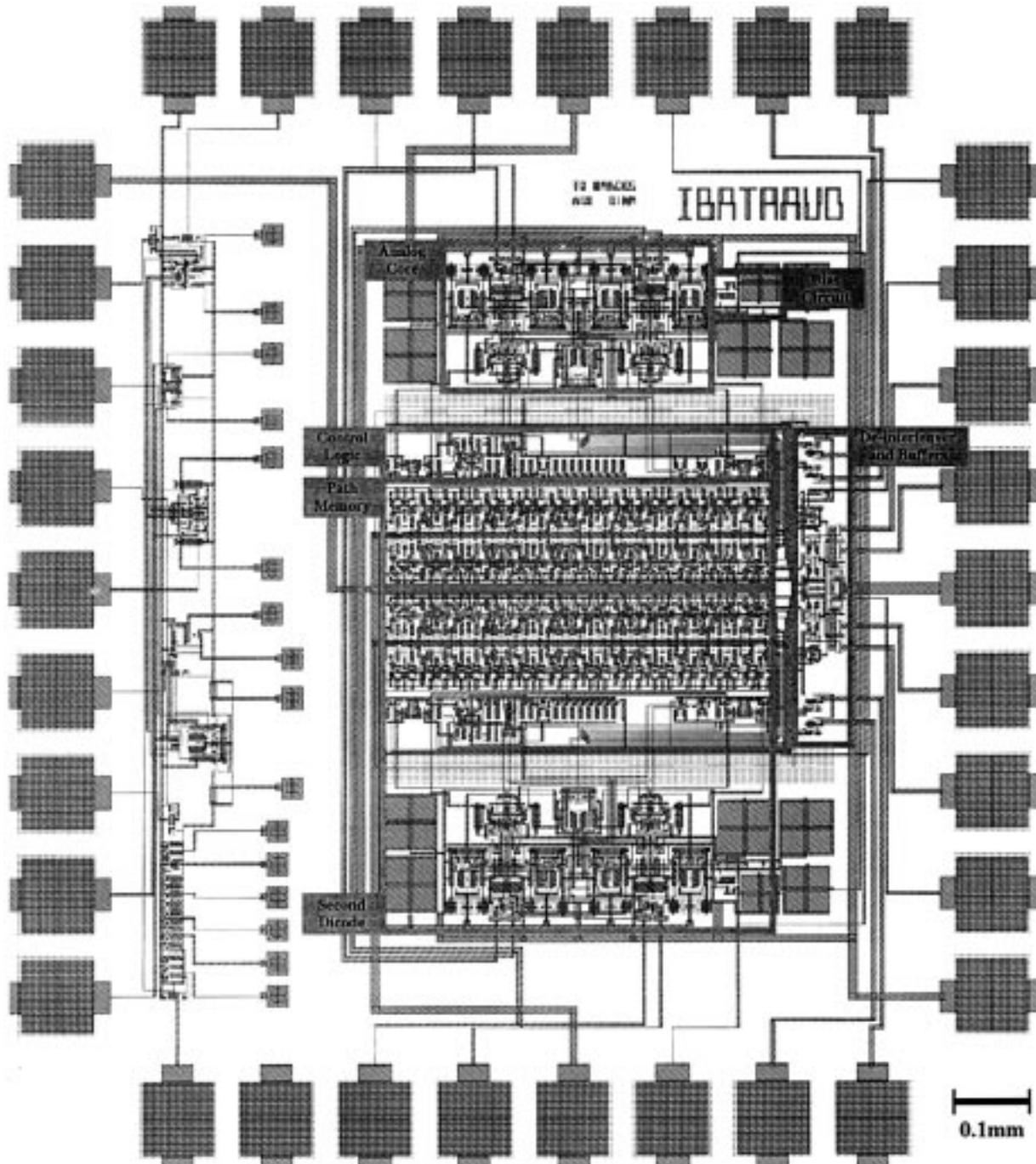
Fig. 21. Layout of the class-IV analog Viterbi decoder, fabricated in a 0.8-μm BiCMOS process.
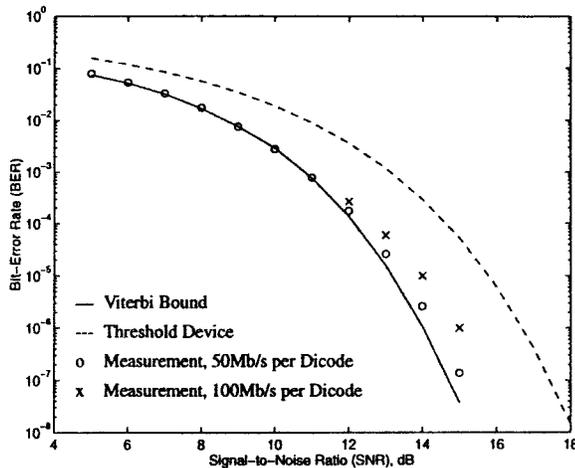
Fig. 22.   Measured BER performance of the decoder.

employing BJT current sources and careful layout design, these mismatches are kept low. Source follower input stages provide the required high input impedances as well as the necessary level shifts. As a result of these level shifts, the on resistances of the input switches are minimized by reducing the input common-mode to near ground.

Analysis shows that, in general, CMOS latches exhibit more offsets compared to bipolar latches [21]. Offset can be greatly reduced by utilizing a low-offset high-gain preamplifier before the latch. In a CMOS realization, large size transistors and/or offset cancelation techniques can help to overcome the offset, however, the speed of operation will be reduced. On the other hand, a bipolar latch has a lower offset and permits a smaller gain in the preamplifier, resulting in a correspondingly faster response. However, bipolar comparators do not have rail-to-rail output swings, required in many applications. All of the above advantages can be attained in a BiCMOS process. The basic idea is to obtain a low input-referred offset voltage by amplifying the signal with a high-gain, wide-band, and low-offset bipolar preamplifier prior to applying it to a CMOS latch [22]. The availability of bipolar transistors can be further appreciated if a bipolar latch is interposed between the preamplifier and the CMOS output latch [21]. This relaxes the constraints on the preamplifier and particularly helps in high-speed and low-power designs. In fact, it has been concluded that to minimize the power-delay product, the amplification required in a comparator is best obtained by means of regeneration [23].

In the design presented here, as shown in Fig. 15, the differential signal is first amplified by activating one of the differential pairs of Q_1 and Q_2 or Q_3 and Q_4. Further amplification is done by incorporating Q_5 and Q_6 in a positive feedback configuration. Regeneration initiates at the beginning of the latch phase L. Slightly after this positive feedback is started and a large-enough signal is developed, a CMOS latch is activated to produce a rail-to-rail swing output signal. This also makes the occurrence of metastability extremely unlikely, particularly within the accuracy of the Viterbi decoder [20]. The CMOS latch is controlled by a delayed version of the latch

signal L. Both of the regenerative stages will be reset during the next amplifying phase. Two cross-coupled differential pairs in the preamplifier provide the capability of reversing the polarity of the signal. This capability allows us to compensate for the sign changes, mentioned in Section IV, by biasing one of the differential pairs at a time.

As shown in Fig. 5(b), the path memory is composed of two interconnected strings of D flip-flops. Serial and parallel load capabilities are provided by using a 2-to-1 multiplexer at the input of each latch. Fig. 16 depicts the circuit. In this circuit, a dynamic latch is converted to a static latch by means of small feedback inverters. Large driving capability for these inverters should be avoided, since it prevents the new data from overwriting the old data. This was achieved by employing small transistors in the feedback inverter.

The path memory consists of 2×12 multiplexed-input D flip-flops, utilized in the structure illustrated in Fig. 5(b). Fig. 17 depicts the result. In this figure, $A_0, \overline{A_0}, B_0, \overline{B_0}, A_1, \overline{A_1}, B_1,$ and $\overline{B_1}$ are outputs of the latched comparators shown in Fig. 12. Based on the decision regions sliced by these comparators, either a serial/parallel, a serial/serial, or a parallel/serial loading occurs in the contents of the upper/lower (corresponding to state "0/1") shift registers. From both of the outputs of each comparator, which are initially pulled down to ground by transistors M_1 and M_2 in Fig. 15, one and only one will be set during the latch phase. The positive transitions are used to perform the loadings, which become complete at phase Φ_4. Either one of two outputs of the last flip-flops in two chains can be treated as the decoded data in our local trace-back method.

The T flip-flop, shown in Fig. 12, generates the toggling signal to control the input S/H's. This flip-flop is constructed from the D latch (Fig. 16) by feeding its inverted output back to the inputs. Two loading controls implement the required OR function at the input. The final toggling signals are derived by the use of AND gates. A fast AND gate is implemented by adding two transistors to the switch shown in Fig. 13. Fig. 18 illustrates the complete toggling circuit.

The different clock phases were obtained by the clock generator circuit shown in Fig. 19(a). This circuit accepts a single-phase clock at its input and generates the appropriate phases Φ_1 to Φ_5 addressed in the previous figures. In the clock generator circuit, the required delays are obtained through the use of inverter gates. Fig. 19(b) depicts a sample timing diagram.

Time interleaving at the input is accomplished by applying the class-IV signal to both of the dicode decoders and using complementary phases for the second dicode. In our case, the complementary phases were simply obtained by a second clock generator similar to Fig. 19(a) with the input inverter replaced with an on-chip RC low-pass circuit. The RC time-constant was chosen to accommodate for the delay of the eliminated inverter. Since the delay of this inverter was only 0.15 ns, the on-chip RC circuit was expected to compensate for the delay to a first-order approximation, with no major concern regarding mismatches and process variation.

De-interleaving is done by two 2-to-1 multiplexers. Each multiplexer combines two corresponding outputs of the path
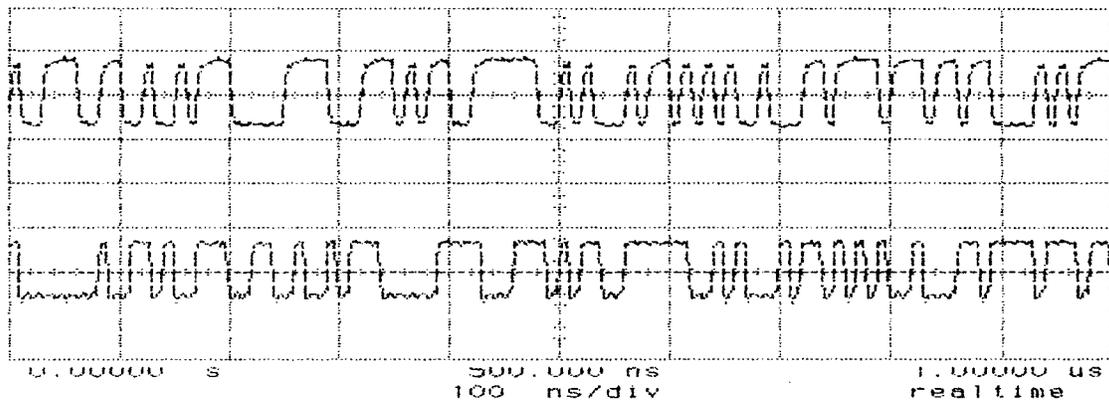
Fig. 23. A typical uncoded signal at 100 Mb/s (upper trace) and its decoded output (lower trace).

TABLE I
A SUMMARY OF THE CHIP SPECIFICATIONS AND TEST RESULTS

Chip	Analog Viterbi Decoder
Coding Scheme	Class-IV Partial-Response
Process	0.8 μm BiCMOS
Data Rate	200 MS/s
Power Consumption	30 mW @ 200 MS/s
Power Supply	3.3 V
Size	0.5 mm^2
Coding Gain at BER=10^{-6}	1.7 dB @ 200 MS/s, 2.4 dB @ 100 MS/s

memories into a single bit stream. A shared address line, externally available, controls the multiplexers. For class-IV operation, this line should be clocked by the master clock. By connecting the address line to either low or high, each individual dicode outputs its own decoded signal. This capability was extensively used during the tests. Fig. 20 depicts the de-interleaving circuit.

VII. EXPERIMENTAL RESULTS

Fig. 21 shows the layout of the chip, fabricated in a BiC-MOS process.[3] It contains two dicode decoders operating in a time-interleaved fashion. Each dicode consists of an analog processor core, a digital path memory, and a control signal generator. The small size of the processor demonstrates the efficiency of the proposed analog realization. The class-IV decoder was tested at up to 100 MS/s with the encoded signal contaminated by additive white Gaussian noise. However, since each individual dicode was also tested at this speed, the class-IV decoder should be capable of operating at 200 MS/s. Direct experiments at this speed were not possible due to the test equipment limitations which limited the rate of the partial-response signal to a maximum of 100 MS/s. The BER performance of the class-IV decoder was very similar to that of each individual dicode, as expected. Fig. 22 depicts the measurement results at two different speeds. The BER was

[3] The Northern Telecom BATMOS process, available through Canadian Microelectronic Corporation.

measured by counting the number of errors in a fixed period of time. The results are accurate, since due to the high-speed operation of the circuit thousands of errors could be counted in only few minutes, even at the lowest BER. The power of the generated noise could be accurately controlled in steps of 0.1 dB and the amplitude of the partial-response signal could be precisely adjusted. These capabilities allowed a fine control on the input SNR.

The results follow the Viterbi bound, with some expected degradation at high SNR. Recall that not all of this degradation is specific to the present analog realization. Also, it is believed that a part of the degradation at high speeds is due to the input test signal which could not be generated as reliably as it could be generated at low speeds. Fig. 22 shows that at an effective rate of 200 Mb/s and at a BER of 10^{-6}, a coding gain of 1.7 dB is achievable out of its 2.7 dB upper bound. This increases to 2.4 dB at 100 Mb/s.

In the present implementation, the path memory is truncated to a length of 12 b. The excess BER due to truncating the path memory is not negligible compared to the decoder BER at the high end of the SNR range of measurement. Tracing back a local-optimum sequence extends this SNR range toward lower values. Thus, any direct measurement would have been affected by the excess BER. To highlight the extremely low BER performance of the decoder, a different measurement technique was applied at high SNR's. The local trace back was performed on both states of the dicode decoder. From the resulting local optimum sequences, 2 b were detected. The detected bits were compared against the corresponding original bit and an error was flagged only when both of the detected bits were not correct. Having two opposite detected bits is an indication that two survivor sequences have not yet merged. These sequences could have merged if a deeper path memory had been used. Note that even if these sequences had merged, still an error could have occurred with a probability equal to the BER of the decoder. Ignoring these errors results in setting a BER target that is, in general, below the BER of the Viterbi decoder. Any measurement now should be compared to this fictitious target. However, simulations indicate that in the SNR and BER ranges of interest, and for the memory length of 12 b, this target is hardly distinguishable from the Viterbi bound and

the above technique can be used for low-BER measurements.

Fig. 23 shows a typical pseudorandom binary signal (uncoded) and the decoded output at 100 Mb/s for one dicode. The decoded output shows a delay slightly more than the expected 13 b (12 b due to the length of the path memory plus 1 b processing time). This extra delay is because of the latency introduced by the propagation time and was not observed at low speeds.

Table I summarizes the specifications of the chip as well as some of the experimental results.

VIII. CONCLUSIONS

Analog Viterbi decoders result in significant savings in power and size, while operating at higher speeds, compared to their conventional digital counterparts. This paper described a successful attempt toward realizing a class-IV partial-response Viterbi decoder in the analog domain. It was demonstrated that such a decoder can be efficiently realized using a few simple building blocks. This goal was achieved by examining the difference-metric algorithm from an analog implementation point-of-view. The outcome was the new input-interleaved algorithm. It was demonstrated that the complexity of the decoder is comparable to that of a 2-b A/D. This illustrates a substantial decrease compared to that of the typical 6-b prestage A/D required in digital realizations. Furthermore, the decoder was shown to be faster and more robust than other reported analog decoders. The fast operation of the decoder was illustrated in practice, whereas extensive simulations were appealed to confirm the robustness of the structure to various analog imperfections which vary from one implementation to the other.

The decoder was fabricated in a 0.8-μm BiCMOS process, tested, and achieved a speed of 100 MS/s per dicode, corresponding to 200 MS/s for the class-IV operation. Direct experiments on the class-IV decoder were limited to 100 MS/s due to the test equipment limitations. The power consumption of the chip was only 30 mW from a 3.3-V single power supply. The core area is 0.5 mm^2, from which only 25% is dedicated to the analog circuitry. These features make the analog detector an extremely attractive alternative in commercial products for demanding high-speed, low-power, and small-size applications such as magnetic disk-drive storage systems.

REFERENCES

[1] P. Kabal and S. Pasupathy, "Partial-response signaling," *IEEE Trans. Commun.*, vol. 23, pp. 921–934, Sept. 1975.

[2] A. Lender, "The duobinary technique for high-speed data transmission," *IEEE Trans. Commun., Electron.*, vol. 82, pp. 214–218, May 1963.

[3] E. R. Kretzmer, "Generalization of a technique for binary data communication," *IEEE Trans. Commun. Technol.*, vol. 14, pp. 67–68, Feb. 1966.

[4] M. J. Ferguson, "Optimal reception for binary partial response channels," *Bell Syst. Tech. J.*, vol. 51, no. 2, pp. 493–505, Feb. 1972.

[5] H. Kobayashi and D. T. Tang, "Application of partial-response channel coding to magnetic recording systems," *IBM J. Res. Develop.*, vol. 14, pp. 368–375, July 1970.

[6] H. Kobayashi, "Correlative level coding and maximum-likelihood decoding," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 586–594, Sept. 1971.

[7] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363–378, May 1972.

[8] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 260–269, Apr. 1967.

[9] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, Mar. 1973.

[10] T. W. Matthews and R. R. Spencer, "An integrated analog CMOS Viterbi detector for digital magnetic recording," *IEEE J. Solid-State Circuits*, vol. 28, pp. 1294–1302, Dec. 1993.

[11] M. H. Shakiba, D. A. Johns, and K. W. Martin, "Analog implementation of class-IV partial response Viterbi detector," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 4, pp. 91–94, 1994.

[12] R. G. Yamasaki *et al.*, "A 72 Mb/S PRML disk-drive channel chip with an analog sampled-data signal processor," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 1994, pp. 278–279.

[13] M. H. Shakiba, D. A. Johns, and K. W. Martin, "A 200 MHz 3.3 V BiCMOS class-IV partial response analog Viterbi decoder," in *Proc. IEEE Custom Integrated-Circuits Conf.*, 1995, pp. 567–570.

[14] R. W. Wood and D. A. Petersen, "Viterbi detection of class IV partial response on a magnetic recording channel," *IEEE Trans. Commun.*, vol. COMM-34, pp. 454–461, May 1986.

[15] T. W. Matthews and R. R. Spencer, "An analog CMOS Viterbi detector for digital magnetic recording," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 1993, pp. 214–215.

[16] R. Wood, "Magnetic megabits," *IEEE Spectrum*, vol. 27, no. 5, pp. 32–38, May 1990.

[17] C. M. Rader, "Memory management in a Viterbi algorithm," *IEEE Trans. Commun.*, vol. COMM-29, pp. 1399–1401, Sept. 1981.

[18] R. R. Spencer, "Simulated performance of analog Viterbi detectors," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 277–288, Jan. 1992.

[19] F. Hemmati and D. J. Costello, Jr., "Truncation error probability in Viterbi decoding," *IEEE Trans. Commun.*, vol. COMM-25, pp. 530–532, May 1977.

[20] D. A. Johns and K. W. Martin, *Analog Integrated Circuit Design*. New York: Wiley, 1997.

[21] B. Razavi and B. A. Wooley, "Design techniques for high-speed, high-resolution comparators," *IEEE J. Solid-State Circuits*, vol. 27, pp. 1916–1926, Dec. 1992.

[22] L. Singer, "Analog circuit techniques utilizing BiCMOS technology," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 3, pp. 1979–1982, 1990.

[23] P. J. Lim and B. A. Wooley, "An 8-bit 200-MHz BiCMOS comparator," *IEEE J. Solid-State Circuits*, vol. 25, pp. 192–199, Feb. 1990.

Mohammad Hossein Shakiba (S'97) was born in Hamedan, Iran, in 1960. He received the B.Sc. and M.Sc. degrees from the Isfahan University of Technology, Iran, in 1985 and 1988 and the Ph.D. degree from the University of Toronto, Canada, in 1997, all in electrical engineering.

From 1986–1991, he worked as a Research and Teaching Member in the Department of Electrical and Computer Engineering, Isfahan University of Technology. Along the course of the Ph.D., he has held the position of Research Fellow in the Department of Electrical and Computer Engineering, University of Toronto. During this time, his research has focused on the area of high-speed data communication systems. His doctoral work was related to the analog implementation of Viterbi detectors with emphasis on the partial-response systems. He is currently interested in analog integrated circuit design for wired and wireless data communication systems.

David A. Johns (S'82–M'88–SM'96) received the B.A.Sc., M.A.Sc., and Ph.D. degrees from the University of Toronto, Canada, in 1980, 1983, and 1989, respectively.

From 1980–1981, he worked at Mitel Corp., Ottawa, Canada, while from 1983–1985 he was an Analog IC Designer at Pacific Microcircuits Ltd., Vancouver, Canada. His doctoral work focused on analog and digital adaptive filters including the development of an orthonormal structure for analog filters. In 1988, he was hired at the University of Toronto where he is currently an Associate Professor. He has ongoing research programs in the areas of analog integrated circuits, oversampling, and digital communication circuits resulting in more than 40 publications and one textbook. He has been involved in numerous industrial short courses and has just completed a one-year research leave with Brooktree Corp., San Diego, CA, in the area of high-speed digital communications.

Dr. Johns is a past and present Associate Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS.

Kenneth W. Martin (S'75–M'80–SM'89–F'91) received the B.A.Sc., the M.A.Sc., and the Ph.D. degrees from the University of Toronto, Canada, in 1975, 1977, and 1980, respectively.

From 1977–1978 he was a Member of the Scientific Research Staff at Bell Northern Research, Ottawa, Canada, where he did some of the early research in integrated, switched-capacitor networks. Between 1980 and 1992, he was consecutively an Assistant, Associate, and Full Professor at the University of California at Los Angeles. In 1992, he accepted the endowed "Stanley Ho Professorship in Microelectronics" at the University of Toronto. He has also been a consultant to many high technology companies including Xerox Corp., Hughes Aircraft Co., Intel Corp. and Brooktree Corp. in the areas of high-speed analog and digital integrated circuit design. He has ongoing research programs in the areas of integrated circuits and systems. Recently, most of his research has focused on data-communication systems, both wired and wireless, as well as CAD for analog IC design. Recently, he completed, along with co-author D. Johns, a textbook entitled *Analog Integrated Circuit Design* (Wiley, 1997).

Dr. Martin was appointed as the Circuits and Systems IEEE Press representative (1985–1986). He was selected by the Circuits and Systems Society for the Outstanding Young Engineer Award that was presented at the IEEE Centennial Keys to the Future Program in 1984. He was elected by the Circuits and Systems Society members to their Administrative Committee (ADCOM 1985–1987). He served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS from 1985 to 1987 and has served on the technical committee for many International Symposia on Circuits and Systems. He is currently an Associate Editor for the PROCEEDINGS OF THE IEEE, Chairman of the IEEE CAS Guillemin-Cauer Best Paper Awards Committee, and a member of the IEEE Circuits and Systems Board of Governors (1995–1997). He was awarded a National Science Foundation Presidential Young Investigator Award that ran from 1985–1990. He was a co-recipient of the Beatrice Winner Award at the 1993 ISSCC.