An Improved Pipelined MSB-First Add-Compare Select Unit Structure for Viterbi Decoders

Keshab K. Parhi, Fellow, IEEE

Abstract—Convolutional codes are widely used in many communication systems due to their excellent error-control performance. High-speed Viterbi decoders for convolutional codes are of great interest for high-data-rate applications. In this paper, an improved most-significant-bit (MSB)-first bit-level pipelined add-compare select (ACS) unit structure is proposed. The ACS unit is the main bottleneck on the decoding speed of a Viterbi decoder. By balancing the settling time of different paths in the ACS unit, the length of the critical path is reduced as close as possible to the iteration bound in the ACS unit. With the proposed retimed structure, it is possible to decrease the critical path of the ACS unit by 12% to 15% compared with the conventional MSB-first structures. This reduction in critical path can reduce the level of parallelism (and area) required for a very high-speed Viterbi decoder.

Index Terms—Add compare select, MSB first, pipelining, redundant arithmetic, retiming, Viterbi decoder.

I. INTRODUCTION

TH E Viterbi algorithm (VA) was first introduced as a method for convolutional decoding in 1967 [1]. A Viterbi decoder is composed of three basic computation units as shown in Fig. 1. The branch metric unit (BMU) calculates the branch metrics $\lambda_{i,j}(n)$, which is the metric of the branch from state $s_i$ to state $s_j$ at time instance $n$. The branch metrics are fed into the add-compare select (ACS) unit to calculate the state metrics $\gamma_j(n+1)$ for state $s_j$. At time instance $n+1$, the state metrics can be computed recursively using

$$
\gamma_j(n+1) = \max_{\lambda_{i,j}(n)} \left\{ \gamma_i(n+1) \right\} = \max_{\lambda_{i,j}(n)} \left\{ \lambda_{i,j}(n) + \gamma_i(n) \right\}.
$$

(1)

A state metrics is the maximum of intermediate state metrics (ISM) $\gamma_{i,j}(n+1)$, which are the summations of the branch metrics $\lambda_{i,j}(n)$, and the state metrics $\gamma_i(n)$, at the previous time instance. The trellis representation of the computation is illustrated in Fig. 2(a). In this paper, the state metrics are the maximum ones among the intermediate state metrics. In some applications, they can be the minimum ones. The ACS units for maximum selection and minimum selection have similar structures. The improvement and conclusion on the maximum selection ACS unit in this paper can also be applied to the minimum selection ACS unit.

The survivor path memory unit (SMU) processes the decisions made in the BMU and ACS and outputs the decoded data. The BMU and the SMU are only composed of feedforward paths. It is relatively easy to shorten the critical path in these two units by utilizing pipelining and parallel processing techniques. However, the feedback loop in the ACS unit is the major bottleneck for the design of a high-speed Viterbi decoder. Although it is possible to implement the high-speed ACS unit with parallel processing techniques [2], it is still desirable to reduce the length of the critical path in the ACS unit. The reduction of the critical path can significantly lower the level of parallelism and the hardware complexity for a specified speed.

Several structures have been proposed to speed up the computation of ACS unit. In [2] and [3], the look-ahead technique was utilized to speed up the ACS unit. With the $M$-step look ahead, one iteration in the ACS unit is equivalent to $M$ iterations in the nonlook-ahead implementation. Thus, the speed requirements on the ACS units for a given decoding data rate are reduced by $M$ times. The trellis corresponding to 2-step look-ahead in a 4-state Viterbi decoder is shown in Fig. 2(b). Compared with Fig. 2(a), the number of branch metrics in the trellis increases exponentially as $M$ increases linearly. However, this exponential complexity can be reduced to linear hardware complexity.

Least significant bit (LSB) first computation is useful for accumulation operation, but most significant bit (MSB) first computation is more suitable for compare and selection operations in the ACS unit. An ACS structure combining MSB-first compare-select with carry-propagation-free addition was proposed based on redundant number representation in [4].

To reduce the complexity in maximum selection operation, a code converter (CC) was introduced for removing the coding redundancy for the digit “1” in the carry-save digit $\{0,1,2\}$ [5]. The circuit of CC and its truth table are summarized in Fig. 3. In Fig. 3, $\gamma_{C_b}$ and $\gamma_{C_c}$ are the sum and carry bits with the same weight, respectively. One redundant representation of digital “1” ($\gamma_{C_b} = 0$ and $\gamma_{C_c} = 1$) is removed at the output of the CC. After the CC, the maximum selection operation reduces to a bit-wise OR operation.

It was also discovered that comparing the local intermediate metric with the maximum one of all other ISMs is more area efficient than pairwise comparison [6], [7].

This paper is organized as follows. In Section II, more precise estimates of the critical path and iteration bound of the ACS unit are obtained. Based on the critical path analysis, it is feasible to achieve even shorter critical path by moving the retiming cutset to balance the settling time of different paths. The improvement on the critical path by applying a novel retiming design is discussed in Section III. With the proposed approach, for a 4-state

Manuscript received June 6, 2003; revised October 29, 2003. This paper was recommended by Associate Editor Y. Wang.

The author is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: parhi@ece.umn.edu).

Digital Object Identifier 10.1109/TCSI.2004.823657
II. CRITICAL PATH OF BIT-LEVEL PARALLEL ACS UNIT

An MSB-first ACS unit can be further divided into 3 basic function blocks as shown in Fig. 4. The three basic function blocks are adders, CCs, and maximum selection (MS) units. The adders compute the summation of the state metrics and the branch metrics, i.e., the intermediate state metrics. The CCs remove a redundant state in the sum-carry representations of the ISM, as illustrated in Fig. 3. In the MS units, the local ISM is compared with the maximum one of all other ISMs for the same state. The comparison scheme can achieve higher locality compared with pairwise comparisons [6], [7]. The results of the comparisons are fed back to the adders.

For an 8-state Viterbi decoder, it is possible to increase the speed by about 13%.

After applying the retiming cutsets every 2 bits as shown in Fig. 6 [5], [6], the length of the critical path is reduced to

$$T_{\text{crit}} = T_{FA} + T_{CC} + W \cdot T_{MS}$$

where $T_{FA}$, $T_{CC}$ and $T_{MS}$ are the computational latencies of a full adder, a CC, and a maximum selection (MS) unit, respectively. The critical path of the circuit $T_{\text{crit}}$ is shown in Fig. 5 as a dotted line

$$T_{\text{crit}} = 2T_{FA} + 2T_{CC} + W \cdot 2T_{MS}.$$
Although the numbers of full adders and CCs in the critical path are doubled, the number of the MS units is significantly reduced. At the same time, the length of the critical path is independent of the word length used to compute the state metrics.

To reduce the critical path by novel retiming, the detailed structures of the MS units in the bit-level ACS structure must be explored. The structure of the MS unit is shown in Fig. 7, which compares the local ISM with the maximum one among all other ISMs of the same state [6]. It can be found that the MS unit can be further divided into three blocks, from left to right: the maximum value block (m-block), the decision block (d-block), and the partial maximum block (pm-block). The pm-block uses the signals from other bit-level ACS structure for the same state with different ISMs. The output of the pm-block is the maximum one of all other ISMs. With the signals from the pm-block (s^\text{pm}\text{max} and c^\text{pm}\text{max}), the CC (s^d and c^d) and the d-block in the last more significant bit (d^\text{d}\text{last} and d^\text{d}\text{last}), m-block computes the maximum value among all of the intermediate state metrics and feeds it back to the full adder in the ACS unit as the state metric γ^0(n+1). The state metric at the output of m-block is expressed in the form of sum and carry, (γ^m(n+1), c^m(n+1)). The inputs of the d-block include the signals from the pm-block, the CC, and the d-block in the circuits of last more significant bit. The outputs of the d-block are sent to the MS unit of the next LSB.

The MS unit, according to Fig. 7, can be logically divided into two subunits (the m-subunit and the d-subunit, as shown
Fig. 7. Circuit of an MS unit for a 2-step look-ahead 4-state Viterbi decoder.

Fig. 8. Block diagram of an MS unit. The m-subunit includes pm-block and m-block. The d-subunit includes pm-block and d-block. The pm-block, with inputs from other states, is shared by both m-subunit and d-subunit. 

in Fig. 8. The m-subunit, which includes the pm-block and the m-block, computes the state metrics. The d-subunit, which includes the pm-block and the d-block, computes the preliminary decision information for the next less significant bit. Thus, the settling time of the MS unit is equal to the maximum one of the two subunits. The circuits in the d-subunit are more complex than the m-subunit. Thus,

\[ T_{\text{max}} = \max(T_d, T_m) = T_d \]  

where \( T_d \) and \( T_m \) are the settling times for the d-subunit and the m-subunit, respectively.

If we redraw the bit-level structure of the ACS unit, as shown in Fig. 9(a), it is clear that the critical path is actually shorter than the one determined by either (4) or (5). Although the critical path passes through the MS unit twice, it passes the m-subunit and the d-subunit separately, i.e., the critical path can be more precisely computed as

\[ T_{\text{crit, actual}} = 2T_{FA} + 2T_{CC} + T_m + T_d. \]  

Similarly, the iteration bound of the ACS unit can be calculated as

\[ T_{\text{crit, iter}} = T_{FA} + T_{CC} + T_m + T_d. \]  

III. NOVEL PIPELINED AND RETIMED ACS UNIT

A. Improvement of Retiming Cutsets

In Section II, the precise estimate of the critical path of the ACS unit is obtained by exploring the structure of the MS unit. Based on the fact that the d-subunit is more complex than m-subunit and requires longer propagation delay, if the register at the output of the d-subunit can be moved backward by redesigning the retiming cutset which balances the settling time of the m-subunit and the d-subunit, the critical path in the bit-level ACS structures can be further reduced. Fig. 9(b) illustrates a possible retiming cutset design. The possible candidates of critical path are the dotted lines marked with 1, 2, and 3. Path 1 is a shortened version of the critical path in Fig. 9(a), while paths 2 and 3 are new candidates due to the retiming cutset applied. The complexity and the settling time of the m-subunit and the d-subunit depend on the number of the ISMs for a state in the Viterbi decoder. With a large number of states (such as a 3-step look-ahead 8-state Viterbi decoder).
decoder), the settling time of d-subunit may be greater than the sum of the settling times of a full adder and a coder converter. The critical path is the maximum one among paths 1, 2 and 3, i.e., as shown in (8) at the bottom of the page, where \( T_{(1)} \) is the settling time of the path 1, which includes two full adders, two CCs and two m-subunits, \( T_{(2)} \) is the settling time of the path 2, which includes one full adder, one CC, two m-subunits, and one d-subunit and \( T_{(3)} \) is the settling time of the path 3, which includes two d-subunits and one m-subunit. Noticing that the m-subunit and d-subunit share the same pm-block and the decision logic in the d-subunit is composed of three levels of multiplexers as shown in Fig. 10, we conclude that the condition \( T_d < T_{FA} + T_{CC} + T_m \) in (8) always holds true, i.e., \( T_{(3)} < T_{(2)} \). Thus, the critical path of the retimed structure shown in Fig. 9(b) is as shown in (9) at the bottom of the page. The critical path \( \hat{T}_{crit} \) is smaller than the critical path \( T_{crit} \) obtained in Section II, if any one of the following conditions holds:

\[
\begin{align*}
T_d &< T_{FA} + T_{CC} \\
T_d &\geq T_{FA} + T_{CC} \text{ and } T_m < T_{FA} + T_{CC}.
\end{align*}
\] (10a)

\[
\hat{T}_{crit} = \max(T_{(1)}, T_{(2)}, T_{(3)})
\]

\[
\hat{T}_{crit} = \begin{cases} 
2T_{FA} + 2T_{CC} + 2T_m, & \text{if } T_d < T_{FA} + T_{CC} \\
2T_{FA} + 2T_{CC} + 2T_m + T_d, & \text{if } T_d < T_{FA} + T_{CC} + T_m \\
2T_{FA} + 2T_{CC} + 2T_m + 2T_d + T_m, & \text{if } T_d \geq T_{FA} + T_{CC} + T_m
\end{cases}
\] (8)

\[
\hat{T}_{crit} = \begin{cases} 
T_{(1)} = 2T_{FA} + 2T_{CC} + 2T_m, & \text{if } T_d < T_{FA} + T_{CC} \\
T_{(2)} = T_{FA} + T_{CC} + 2T_m + T_d, & \text{if } T_d < T_{FA} + T_{CC} + T_m \\
T_{(3)} = 2T_d + T_m, & \text{if } T_d \geq T_{FA} + T_{CC} + T_m
\end{cases}
\] (9)

Fig. 9. Bit-level ACS unit structure by considering the MS unit as two logically separated subunits: m-subunit and d-subunit. Due to the similarity of the ACS structure, only a portion of the bit-level structure is shown in the figure to illustrate the cutset and the candidates of the critical path. (a) Detailed circuits of the bit-wise ACS unit by exploring the structure of the MS block. Only the circuits for 2 bits are shown. (b) Retimed bit-level ACS unit by considering the MS unit as two separated subunits.
It is worth noticing that the d-subunit contains the same pm-block as the m-subunit. To achieve the critical path $T_{\text{crit}}$, the pm-block must be physically duplicated for both the m-subunit and the d-subunit. The complexity of the decoder will significantly increase and the efficiency of the hardware will decrease.

Instead of using the retiming cutset which achieves a critical path $T_{\text{crit}}$, a more efficient retiming cutset can be applied inside the d-subunit, which is illustrated by the dashed line in Fig. 11. With this retiming design, the pm-block can be physically shared by both d-subunit and m-subunit. The d-subunit is divided by the cutset into two parts. The path $\bar{2}$ only passes through d-block instead of the whole d-subunit. The settling times on the paths $\bar{1}$ and $\bar{2}$ are more balanced. Without modifying the settling time on the path $\bar{1}$, the critical path is

$$T_{\text{crit}} = \max(T_{\bar{1}}, T_{\bar{2}}) = \max(2T_{\text{FA}} + 2T_{\text{CC}} + 2T_{\text{m}}T_{\text{FA}} + T_{\text{CC}} + 2T_{\text{m}} + T_{\text{db}})$$

(11)

where $T_{\text{db}}$ is the settling time of the d-block, which is shorter than $T_{\text{db}}$, the settling time of the d-subunit.

By exploring the structure of the decision logic, the settling time of the path $\bar{2}$ can be further reduced by moving the delays at the output of the d-block in Fig. 9(a) inside the block. Fig. 12 illustrates the decision logic after applying the new retiming cutset. The dashed line in Fig. 12 is the retiming cutset, which moves the registers from the output of the circuit to the inputs and some wires inside the logic. By applying the retiming cutset inside the decision logic, the settling time on the path $\bar{2}$ is further reduced to

$$T_{\bar{2}} = T_{\text{FA}} + T_{\text{CC}} + 2T_{\text{m}} + T_{\text{db}}$$

where $T_{\text{db}}$ is the delay of the multiplexers inside the retiming cutset in the decision logic. As shown in Fig. 12, $T_{\text{db}}$ equals the sum of the settling time of two multiplexers. It is independent of the numbers of states and ISMs in the Viterbi decoder. Thus, $T_{\text{db}}$ is always smaller than the sum of the settling time of a full adder and a coder converter. It is worth noticing that the settling time of the remaining part of the d-subunit is just equal to the settling time of the pm-block which is always smaller than the settling time of the m-subunit.

With the cutset inside the decision logic, the critical path becomes

$$T'_{\text{crit}} = 2T_{\text{FA}} + 2T_{\text{CC}} + 2T_{\text{m}}.$$  

(12)

It is concluded that

$$T'_{\text{crit}} \leq T_{\text{crit}}.$$  

(13)

The equality in (13) holds true if and only if $T_{\text{FA}} + T_{\text{CC}} \geq T_{\text{d}}$. If the number of the ISMs in the Viterbi decoder for a single state is large, such as in a 3-step look-ahead 8-state Viterbi decoder, due to complexity of the pm-block, $T_{\text{d}}$ will be always larger than $T_{\text{FA}} + T_{\text{CC}}$. The critical path $T'_{\text{crit}}$ is always smaller than the critical path $T_{\text{crit}}$ obtained in Section II, independent of the number of the ISMs.

B. Critical Path Comparison

The analysis of the critical path and the improved retiming cutset are not limited to 4-state Viterbi decoders. They can be extended to Viterbi decoder with arbitrary number of states. At the same time, they can also be applied to decoders involving look-ahead [3] or sliding block [9].

The assumed settling times of the basic units on the critical paths are shown in Table I. If the number of the ISMs for a single state in the trellis is large, the settling time of the pm-block increases due to the increase of the complexity of the logic. Thus, the settling time of the m-subunit and the d-subunit also increases. The critical paths for different pipelining cases are listed in Table II. The critical path of a 4-state Viterbi decoder is reduced from 2.2 ns as in the conventional approach to 1.9 ns in the novel proposed approach. For an 8-state decoder, the critical path reduces from 3.1 ns to 2.7 ns. Therefore, the critical path reduction is about 14% in a 4-state and about 13% in an 8-state Viterbi decoder. The values of $T'_{\text{crit}}$ are only 15% and 5.9% larger than the iteration bound, $T_{\text{crit}}$, for a 4-state Viterbi decoder and an 8-state Viterbi decoder, respectively.

With the improved retiming cutset, it is possible to implement a high-speed Viterbi decoder with relatively lower hardware complexity. For instance, to implement a 8-state 10 Gb/s Viterbi decoder, a parallel Viterbi decoder based on either look-ahead or sliding block needs to be implemented. For a design using conventional retiming design with critical path $T_{\text{crit}}$, requires a clock period of at least 3.4 ns, if an additional 0.3-ns clock setup/hold time is counted in. However, for a 32-parallel design, the maximum allowed clock period is 3.2 ns to achieve a decoding speed of 10 Gb/s. Thus, either a 64-parallel design in the look-ahead Viterbi decoder or a 48-parallel design in sliding block Viterbi decoder is mandatory. With the improved retiming cutsets, the critical path $T'_{\text{crit}}$ is 2.7 ns. After considering the clock setup/hold time, the clock period required can be as small as 3.0 ns, which is shorter than 3.2 ns. Thus, it is feasible to implement 10 Gb/s Viterbi decoder with only 32-parallel design. The hardware complexity of the hardware is significantly reduced compared with a 48-parallel or 64-parallel design.

IV. Conclusion

In this paper, after exploring the detailed circuits in the bit-level ACS structure in the Viterbi decoder, a more precise estimate of the critical path of the bit-level parallel ACS

1For a look-ahead Viterbi decoder, the level of parallelism is constrained to be a power of 2.

2For a sliding-block Viterbi decoder, the level of parallelism is assumed to be a multiple of eight.
structure is obtained. Several improvements on the retiming cutset are proposed. Among them, the one with the cutset inside the decision logic achieves the smallest settling time, which leads to a reduction of the critical path by about 12%–15%. The shortening of the critical path makes it feasible to implement the high-speed Viterbi decoders with a relatively low level of parallelism. The reduction of the level of parallelism significantly decreases the complexity of the hardware, independent of whether the parallel decoder is implemented using look-ahead or sliding-block technique.
ACKNOWLEDGMENT

The author would like to thank A. Abnous for useful discussions during the course of this work and J. Tang for his help in preparation of the paper. This work was carried out while the author was at Broadcom Corporation, Irvine, CA, while on leave from the University of Minnesota.

REFERENCES


