A Unified Parallel Radix-4 Turbo Decoder for Mobile WiMAX and 3GPP-LTE

Ji-Hoon Kim[1,2] and In-Cheol Park[2]
[1]Samsung Electronics Co., Ltd., Republic of Korea
[2]School of EECS, KAIST, Republic of Korea

***Abstract*- This paper describes the energy-efficient implementation of a high performance parallel radix-4 turbo decoder, which is designed to support multiple fourth-generation (4G) wireless communication standards such as Mobile WiMAX and 3GPP-LTE. We propose a new hardware architecture that can share hardware resources for the two standards. It mainly consists of eight retimed radix-4 soft-input soft-output (SISO) decoders to achieve high throughput and a dual-mode parallel hardware interleaver to support both almost regular permutation (ARP) and quadratic polynomial permutation (QPP) interleavers defined in the two standards. A prototype chip supporting both Mobile WiMAX and 3GPP-LTE standards is fabricated in a 0.13µm CMOS technology with eight metal layers. The decoder core occupies 10.7mm^2 and can exhibit a decoding rate of more than 100Mb/s with eight iterations while achieving an energy efficiency of 0.31nJ/bit/iter.**

I. INTRODUCTION

As the demand for supporting multiple standards in a single handheld device increases, the efficient implementation of the advanced channel decoders, which is the most area-consuming and computationally intensive block in baseband modem, becomes more important. Accordingly, the unified decoder architecture which can support multiple standards becomes necessary since the separate implementation for different standards requires much hardware resources leading to huge silicon area occupation.

Recently, turbo codes have been adopted in the 4G mobile communication systems such as Mobile WiMAX (IEEE 802.16e) and 3GPP-LTE in the form of the double-binary and the single-binary, respectively [1][2]. In this paper, we propose a multi-standard parallel radix-4 turbo decoder architecture to support both Mobile WiMAX and 3GPP-LTE. First of all, a unified radix-4 SISO decoding structure is presented for both single-binary and double-binary turbo decoding, where the retiming technique is exploited to reduce critical path delay. Additionally, since the memory takes significant area in the implementation, how to share the memory in implementing the two different decoding modes is presented considering the bit-level extrinsic information exchange technique proposed in [3]. Finally, a dual-mode hardware interleaver is presented to support both the ARP interleaver and the QPP interleaver defined in Mobile WiMAX and 3GPP-LTE, respectively. By generating the interleaved addresses *on the fly*, we can avoid the RAM-based interleaver that is usually associated with huge area overhead. To verify the proposed architecture, a unified radix-4 turbo decoder is implemented for Mobile WiMAX and 3GPP-LTE with eight SISO decoders.

TABLE I
MAX-LOG-MAP ALGORITHM COMPARISON

	Double-Binary	Radix-4 Single-Binary
State Metric Calculation	4-operand MAX	4-operand MAX
a priori info. in Branch Metric	$L_{e,IN}(u_k = z)$	$L_{e,IN}(v_k) + L_{e,IN}(v_{k+1})$
# of Exchanged Extrinsic Info.	3	2

II. SISO DECODING ALGORITHMS FOR TWO STANDARDS

As the Max-log-MAP algorithm has low computational complexity, it is widely used for SISO decoding in the turbo decoder implementation. Table I indicates the comparison of the Max-log-MAP algorithm that can process two received bits per clock cycle for double-binary SISO decoding and for radix-4 single-binary SISO decoding, doubling the throughput for a given clock rate over the radix-2 architecture [4]. For the forward/backward metric calculation, both SISO decoding require 4-operand MAX operation for the Add-Compare-Select (ACS) operation.

In the double-binary turbo codes, the branch metric, γ, which indicates the transition probability, can be defined as follows;

$$\gamma_k = x_k^{s_1} y_k^{s_1} + x_k^{s_2} y_k^{s_2} + x_k^{p_1} y_k^{p_1} + x_k^{p_2} y_k^{p_2} + L_{e,IN}(u_k = z) \quad (1)$$

where z belongs to $\phi = \{00, 01, 10, 11\}$. Also, u_k is the input symbol at time k and $P(u_k)$ is *a priori* probability of u_k, $\mathbf{x}_k$ and $\mathbf{y}_k$ are transmitted and received codewords associated with u_k, respectively. The superscripts p and s denote the parity bits and systematic bits, respectively. In (1), $L_{e,IN}(u_k)$ is the *a priori* information received from the other SISO decoder. In the double-binary Max-log-MAP algorithm, three extrinsic information values should be exchanged as indicated in Table I, since $L_{e,IN}(u_k = 00)$ is defined to zero.

In the radix-4 single-binary Max-log-MAP algorithm, the SISO decoder can process two bits per each cycle by merging two trellis sections [4]. Similar to the above equations, the metrics can be defined as

$$\gamma_k = x_k^s y_k^s + x_k^p y_k^p + L_{e,IN}(v_k) + x_{k+1}^s y_{k+1}^s + x_{k+1}^p y_{k+1}^p + L_{e,IN}(v_{k+1}) \quad (2)$$

where v_k is the input bit. Since $L_{e,IN}(v_k = 0) / L_{e,IN}(v_{k+1} = 0)$ is defined to zero, two extrinsic information should be exchanged during the decoding. Although the number of *a priori* information is different as shown in (1)-(2), the overall calculation is almost the same as indicated in Table I, which

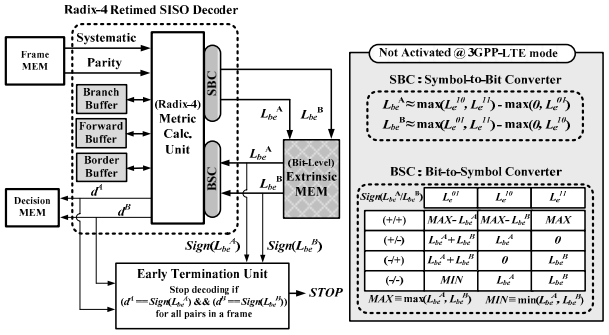

Fig. 1. Overall unified turbo decoder architecture with time-multiplexing

enables a unified SISO decoder that can be used for both decodings.

III. Proposed Chip Architecture

Fig. 1 shows the architecture of the proposed unified turbo decoder, where one SISO decoder is used in a time-multiplex manner to calculate the two decoding steps of an iteration. In addition, the decoder contains two low-complexity converters required for bit-level extrinsic information exchange which can reduce the number of extrinsic information values from three to two in the case of double-binary turbo decoding [3]. Extrinsic information values are exchanged by storing and loading the values in an extrinsic information memory. The memory is accessed in a sequential order for the first decoding, and in an interleaved order for the second decoding. The SISO decoder calculates the extrinsic information values iteratively, and then stops decoding when the stopping criterion for radix-4 processing [5] is satisfied or the iteration number reaches the pre-determined limit. At the last SISO decoding, hard-decision values are calculated and stored in a decision memory. This Section explains in detail the proposed chip architecture developed to share hardware resources for Mobile WiMAX and 3GPP-LTE.

A. Parallel Turbo Decoding

To support high-speed data transmission of the 4G mobile communication systems, eight SISO decoders are adopted in the proposed chip, and part of or all the SISO decoders operate in parallel. The overall architecture of the proposed parallel turbo decoder is depicted in Fig. 2. The number of SISO decoders to be involved in the decoding, M, is scalable

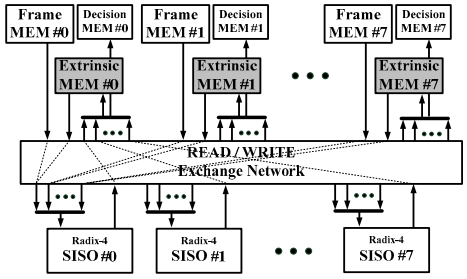

Fig. 2. The proposed chip architecture with eight SISO decoders

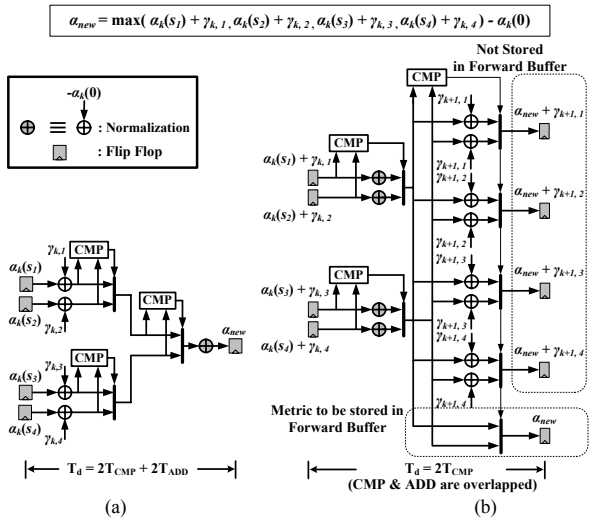

Fig. 3. (a) Conventional and (b) retimed ACS block with four operands.

according to the frame size, N. Each SISO decoder decodes a sub-frame of $L = N/M$. If M is less than eight, some SISO decoders stay idle. For the deactivated SISO decoders, their clocks are gated to reduce the power consumed in such decoders. Each SISO decoder accesses a specific memory by using the address computed in the interleaver. The access passing through the exchange networks does not induce any *collision* due to the *collision-free* property of the ARP/QPP interleavers defined in Mobile WiMAX/3GPP-LTE standards.

B. Unified Radix-4 SISO Decoder with Retiming

As denoted in Section II, the metrics required in the SISO decoding of both standards can be calculated with almost the same hardware because their mathematical expressions are very similar to each other, where only several multiplexers are additionally required since the trellis structures for ACS are slightly different to each other. For the branch metrics, the branch metric recovery scheme proposed in [5] can be applied to (6) in order to reduce the branch memory size. As shown in Fig. 3(a), a max operation with four operands is required for the calculation of the forward and backward metrics. However, the cascaded 2-input ACS shown in Fig. 3(a) leads to a long computation time and thus becomes the critical path limiting the operating frequency. To reduce the critical path delay without affecting the input/output characteristics, the locations of delay elements are retimed. Applying the *retiming* [6] and migrating the common operators can reduce the critical path delay significantly at the cost of a little increase of hardware resources as indicated in Fig. 3(b). The reduced critical path delay enables the turbo decoder to operate at the higher operating frequency and to provide the higher throughput. If the desired throughput is less than the peak one, the supply voltage can be lowered to reduce the power consumption by decreasing the operating frequency.

C. Memory-Sharing with Bit-level Extrinsic Information

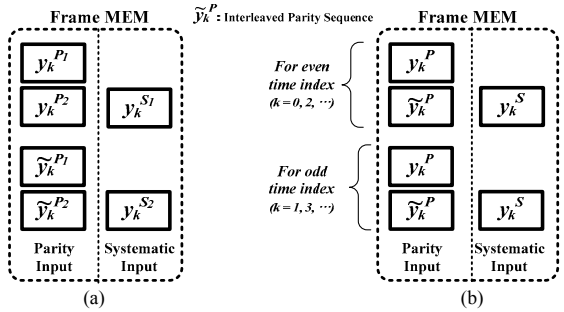

Fig. 4. Input frame memory configurations for (a) double-binary turbo decoding mode and (b) radix-4 single-binary turbo decoding mode.

In the turbo decoder implementation, it is well known that significant area is taken by the input frame memory and the extrinsic information memory [5]. The extrinsic information memory for the double-binary turbo decoder can be reduced by applying the bit-level extrinsic information exchange proposed in [3]. The exchange is achieved by using two converters, symbol-to-bit and bit-to-symbol converters shown in Fig. 1. In this case, the number of extrinsic information values to be exchanged between two SISO decodings is reduced to two, though the double-binary turbo code requires three extrinsic information values. Therefore, the extrinsic information memory can be shared with the radix-4 single-binary turbo decoding, as it exchanges two extrinsic information values, too.

As shown in Fig. 4(a), six memory banks are required to store the input frame in double-binary turbo decoding mode (two systematic inputs and four parity inputs). For the input memory sharing, the *parity property* of the QPP interleaver in 3GPP-LTE is exploited, which means that even (odd) positions in the input are mapped to even (odd) positions in the output. Even in the second SISO decoding that reads the values in an interleaved order, the memory configuration shown in Fig. 4(b) enables two values accessed in each cycle to be resident in different memory banks because of the parity property. Therefore, the proposed memory configuration partitioned according to the parity of the time index is also effective in the radix-4 processing.

D. Dual-Mode Hardware Interleaver

Mobile WiMAX and 3GPP-LTE employ different interleaving procedures – the ARP interleaver and the QPP interleaver, respectively [1][2]. To support both standards with a single hardware, a dual-mode hardware interleaver is proposed to avoid the huge area overhead caused by separate RAM-based interleavers [4].

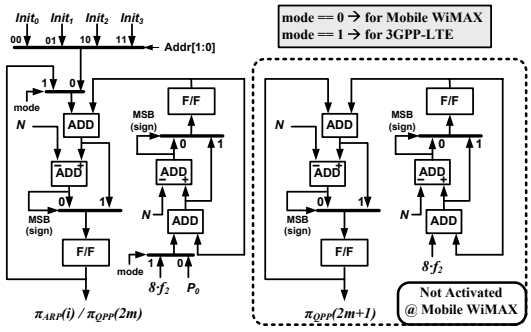

Fig. 5. Dual-mode dedicated hardware interleaver.

For a frame size N, the ARP interleaver is defined as

$$\pi_{ARP}(i) = \left(P_0 \cdot i + O + d(i)\right)_{\mathrm{mod}\, N} \quad (3)$$

where $0 \le i \le N\text{-}1$ is the sequential index of the symbol positions after interleaving, $\pi(i)$ is the symbol index before interleaving, P_0 and O are constant values defined in the Mobile WiMAX standard, and $d(i)$ is also a constant value determined by two lowest bits of i. Based on the incremental calculation, an efficient implementation of the ARP interleaver was proposed in [5], as shown in Fig. 5 where P_0 is accumulated and added to an initial value selected by two LSBs.

Similarly, the QPP interleaver defined below can be rearranged to share the hardware resources with ARP interleaver.

$$\pi_{QPP}(i) = \left(f_1 \cdot i + f_2 \cdot i^2\right)_{\mathrm{mod}\, N} \quad (4)$$

where f_1 and f_2 are the coefficients defined in the 3GPP-LTE standard. The QPP interleaver should generate two interleaved addresses per clock cycle, $\pi(2m)$ and $\pi(2m+1)$ where $0 \le m \le N/2\text{-}1$, to support the radix-4 single-binary turbo decoding. Transforming (4) to a recursive form, we can obtain the following relation.

$$\pi_{QPP}\left(2(m+1)\right) = \left(\pi_{QPP}(2m) + f_1 \cdot 2 + f_2 \cdot 8m + f_2 \cdot 4\right)_{\mathrm{mod}\, N} \quad (5)$$

$$\pi_{QPP}\left(2(m+1)+1\right) = \left(\pi_{QPP}(2m+1) + f_1 \cdot 2 + f_2 \cdot 8m + f_2 \cdot 8\right)_{\mathrm{mod}\, N} \quad (6)$$

As illustrated in Fig. 5, (5)-(6) can be implemented using two accumulator-based circuits with the appropriate initial flip-flop values where one of which is disabled in the Mobile WiMAX mode. With employing small look-up tables keeping the initial constant values, the proposed dual-mode hardware interleaver can calculate the interleaving patterns for all the frame sizes specified in ARP and QPP interleaving procedures.

IV. Unified Turbo Decoder Implementation Results

To verify the proposed architecture, a unified radix-4 turbo decoder is implemented for Mobile WiMAX and 3GPP-LTE with eight SISO decoders as shown in Fig. 2. To avoid unnecessary iterations at high signal-to-noise ratio (SNR), as indicated in Fig. 1, the stopping criterion presented for the double-binary turbo decoding [5] is also adopted for the radix-4 single-binary turbo decoding. Compared to the HDA criterion [4][7], it does not require the additional memory needed to store the decision bits and has no error floor at high SNR. Figure 6 shows the frame-error rate (FER) performance obtained by applying the proposed stopping criterion. Table II compares the characteristics of the developed turbo decoder with previous turbo decoders. A prototype chip containing eight SISO decoders, hardware interleavers, and on-chip dual-port SRAMs is fabricated in a 0.13μm CMOS process with 8 metal layers. The decoder core occupying 10.7mm^2 operates at a maximum frequency of 250MHz, and such a high operating frequency is mainly resulted from the retiming actively

TABLE II
COMPARISON OF DECODER IMPLEMENTATION

Publication	Proposed	[5]	[8]	[4]	[7]
Standard Compliant	Mobile WiMAX/ 3GPP-LTE	WiMAX	HSDPA	HSDPA	UMTS/ CDMA2000
Decoding Algorithm	Double-Binary/ Radix-4 Single-Binary Max-log-MAP	Double-Binary MAX-log-MAP	Radix-2 Single-Binary Max-log-MAP	Radix-4 Single-Binary Log-MAP	Radix-2 Single-Binary Log-MAP
CMOS [µm]	0.13	0.13	0.13	0.18	0.25
Supply Voltage [V]	1.2 (0.9[1])	1.2	1.2	1.8	2.5
# of SISO Decoders	8	1	1	1	1
Gate Count	800K (including 300K for Buffers)	64.2K	44.1K	410K	34.4K
Core Area [mm^2]	10.7	2.2	1.2	14.5 (7.3[2])	8.9 (2.2[2])
Max. Throughput [Mb/s] with 8 iterations	187.5/186.0 @ 1.2V, 250MHz 105[1]/104[1] @ 0.9V, 140MHz	24.3	14.0	18.0 (24.9[2])	4.1 (7.9[2])
Energy Efficiency [nJ/bit/iteration]	0.61 @ 1.2V, 250MHz 0.34[1] @ 0.9V, 140MHz	0.63	0.7	10.0 (2.7[2])	6.9 (0.9[2])

[1] Optimistic estimation based on $t_{pd} \sim v_{dd} / (v_{dd}-v_{th})^2$, Energy ~ v_{dd}^2
[2] Optimistic technology scaling to 0.13µm assumed [8] : A ~ $1/s^2$, $t_{pd} \sim 1/s$, $P_{dyn} \sim 1/s^3$

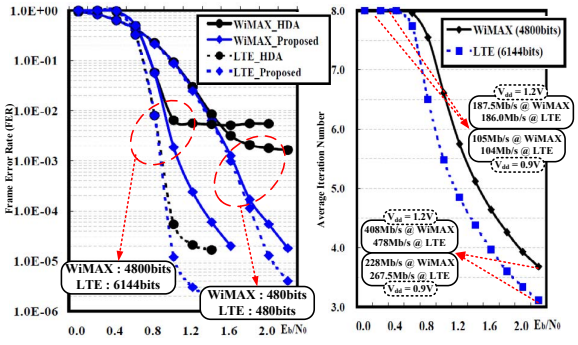

Fig. 6. FER performance and average iteration number with early termination in an AWGN channel.

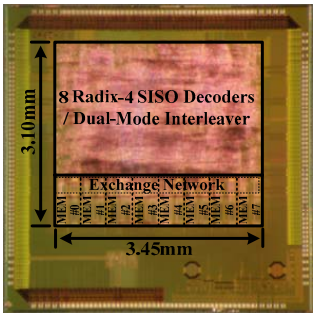

Fig. 7. Micrograph of the chip.

applied to the complex ACS operation as shown in Fig. 3. As denoted in Table II, the proposed decoder achieves an energy efficiency of 0.34nJ/bit/iteration while achieving more than 100Mb/s with fixed eight iterations. The supply voltage is scaled for this measurement since the peak operating frequency is relatively high due to the retiming. A die-photo is shown in Fig. 7 where we can see that the memory is partitioned into eight groups to support the parallel accesses of eight SISO decoders. The complexity of the exchange network connecting the memories and SISO decoders is significantly lowered due to the reduced number of extrinsic information values. To avoid the routing congestion, the exchange network is routed over the memory macros.

V. CONCLUSION

We have presented a unified parallel radix-4 turbo decoder that can support both Mobile WiMAX and 3GPP-LTE standards. New SISO decoding structure, memory configuration, and dual-mode interleaver structure have been proposed to allow hardware sharing for the two standards. The proposed decoder containing eight radix-4 SISO decoders and a low-complexity dual-mode interleaver occupies 10.7mm^2 in 0.13µm CMOS process. The proposed decoder can exhibit more than 100Mbps in case of eight iterations and its throughput can be increased further at high SNR by adopting the stopping criterion. In addition, it achieves an energy efficiency of 0.31nJ/bit/iter.

REFERENCES

[1] *Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems—Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands*, IEEE Std 802.16e/D5-2004, Nov. 2004.

[2] 3GPP TS 36.212 v8.0.0 (2007-09): "Multiplexing and Channel Coding (FDD) (Release 8)".

[3] J. H. Kim and I. C. Park, "Bit-Level Extrinsic Information Exchange Method for Double-Binary Turbo Codes," *IEEE Trans. Circuits Syst. II, Express Briefs*, vol. 56, no. 1, pp. 81–85, Jan. 2009.

[4] M. Bickerstaff, L. Davis, C. Thomas, D. Garret, and C. Nicol, "A 24Mb/s radix-4 LogMAP Turbo decoder for 3GPP-HSDPA mobile wireless," in *ISSCC Dig. Tech. papers*, 2003, pp. 150–151.

[5] J. H. Kim and I. C. Park, "A 50Mbps Double-Binary Turbo Decoder for WiMAX Based on Bit-level Extrinsic Information Exchange," *IEEE Asian Solid-State Circuit Conference,* 2008, pp. 305–308.

[6] K. Parhi, *VLSI Digital Signal Processing Systems: Design and Implementation*. New York: Wiley, 1999.

[7] M. C. Shin and I. C. Park, "SIMD Processor-Based Turbo Decoder Supporting Multiple Third-Generation Wireless Standards," in *ISSCC Dig. Tech. Papers*, 2003, pp. 154–155.

[8] C. Benkeser, A. Burg, T. Cupaiuolo, and Q. Huang, "A 58mW 1.2mm^2 HSDPA Turbo Decoder ASIC in 0.13µm CMOS," in *ISSCC Dig. Tech. Papers*, 2008, pp. 264–265.