A LOW-POWER DSP CORE ARCHITECTURE FOR LOW BITRATE SPEECH CODEC

Hiroyuki OKUHATA, Morgan H. MIKI, Takao ONOYE, and Isao SHIRAKAWA

Department of Information Systems Engineering, Osaka University
2-1 Yamada-Oka, Suita, Osaka, 565 Japan
Phone: +81-6-879-7808, Fax: +81-6-875-5902
E-mail: {okuhata, miki, onoe, sirakawa}@ise.eng.osaka-u.ac.jp

ABSTRACT

A VLSI implementation of a low-power DSP is described, which is dedicated to the G.723.1 low bitrate speech codec. A number of sophisticated DSP microarchitectures are devised mainly on dual multiply accumulators, rounding and saturation mechanisms, and two-banked on-chip memory. The proposed DSP architecture has been integrated in the total area of 7.75 mm^2 by using a 0.35μm CMOS technology, which can operate at 10MHz with the dissipation of 45mW from a single 3V supply.

1. INTRODUCTION

In order to realize a variety of low bitrate multimedia communication facilities, such as videophone and TV conference over GSTN (General Switched Telephone Network), there arises a strong demand of the realtime speech codec at a very low bitrate. The ITU-T G.723.1[1], which is an audio part of the ITU-T H.324[2] audiovisual communication standard, specifies a speech codec at a very low bitrate. Fig. 1 illustrates a block diagram of H.324, in which the shaded part is related to G.723.1. This paper describes a DSP core architecture dedicated to G.723.1, which aims at the single chip implementation of a realtime H.324 codec.

Since G.723.1 is intended for the speech codec, it can not represent music and other audio signals so faithfully as speech signals. However, the compression ratio for the speech is by far superior to those for the others; actually, the digital voice is to be coded at 6.3/5.3 kbps with the ratio of 20:1/24:1. On the other hand, this G.723.1 codec incurs an incredible amount of computation, for which any general purpose DSP should operate at an extremely high frequency. Moreover, considering that this speech codec has to be integrated in conjunction with a moving picture codec[6], a radical innovation should be pursued for the reduction of not only the area occupancy but also the power dissipation. Thus this paper puts the main focus on the low power implementation of G.723.1 realtime codec.

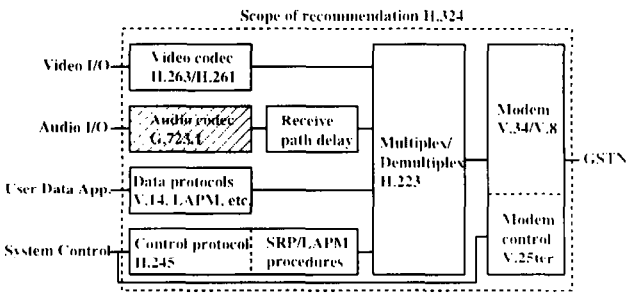

Fig. 1 Multimedia communication standard H.324.

In what follows, first the analysis of the computational labor of G.723.1 codec is evaluated by software simulation, and then on the basis of this analysis, DSP microarchitectures are devised for a number of operation units and memory facilities. Finally, a part of VLSI implementation results are shown, which are attained with the use of an ASIC design system *COMPASS Design Tools*.

2. COMPUTATIONAL COST OF G.723.1

Table 1 indicates the specifications of G.723.1, where two bitrates of 5.3kbps and 6.3kbps are specified. The Algebraic Code Excited Linear Prediction (ACELP) is used as the excitation signal for the lower rate (5.3kbps) coder, and the Multipulse Maximum Likelihood Quantization (MP-MLQ) for the higher rate (6.3kbps) coder. The input speech signal is sampled by the 16 bit linear PCM at 8kHz, and a sequence of 240 samples constitutes a unit of *frame*, which corresponds to 30 msec.

Fig. 2 shows an outline of G.723.1 codec process. A digital voice is encoded through means of parameters: gain, pitch, prediction coefficient, and non-periodic component. A gain and a set of non-periodic components are generated by ACELP/MP-MLQ, a pitch by Pitch Estimator and Pitch Predictor, and a set of predic-

Table 1 Specifications of G.723.1

Bitrate	5.3kbps/6.3kbps
Coding method	ACELP/MP-MLQ
Input/output PCM	16bit linear PCM
Sampling rate	8kHz
Frame size	240 samples

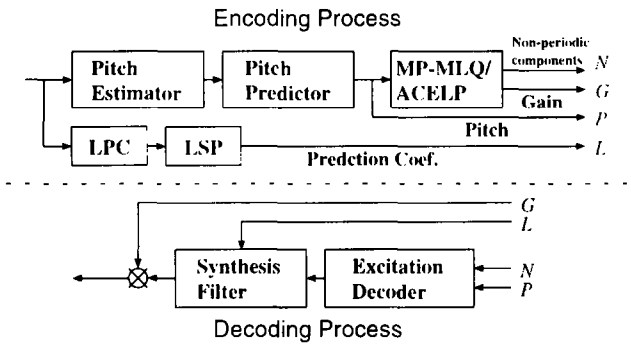

Fig. 2 Outline of G.723.1 codec process.

tion coefficients by LPC (Linear Predictive Coding) and LSP (Line Spectrum Pair). In the decoding process, the prediction coefficients are used to construct a synthesis filter by which the pitch and non-periodic components are filtered, and the gain is multiplied by the output signal of the filter so as to restore the original speech signal.

To analyze the computational labor of this G.723.1 codec, the number of clock cycles necessary for executing each codec process is calculated with the use of a general purpose DSP. Since G.723.1 input/output is the data sampled by the 16 bit linear PCM, most of the operations are done with the 16 bit precision. Therefore we choose a 16 bit fixed point DSP as a measuring device. The number of clock cycles for each operation code is defined as indicated in Table 2. As a result of our G.723.1 software simulation, the number of clock cycles per frame of each encoding/decoding process is shown in Table 3.

As can be readily seen from Table 3, the encoding process needs 1.0 and 1.4 million cycles per frame at 5.3 and 6.3 kbps, respectively, and the decoding process needs 80 thousand cycles per frame at each of these bitrates. Since 1 frame is composed of 240 samples with the sampling rate of 8kHz, the whole 1 frame codec process has to be completed within 30 msec so that a decoding speech signal should not be broken. Consequently, in order to realize a realtime codec with the use of this conventional 16 bit DSP, a clock frequency needs

$$(1.407,255 + 82.518)/(30/1000) = 49.66MHz \quad (1)$$

Table 2 Latency of each operation code.

cycle	operation
1	add sub abs_s shl shr mult L_mult negate extract_h extract_l
2	L_add L_sub L_negate mult_r L_shl L_shr L_deposit_h L_deposit_l
3	L_abs
4	L_mac L_msu round
15	norm_s
18	div_s
20	div_l
30	norm_l

Table 3 Number of clock cycles at each process (cycle/frame).

process		bitrate 5.3kbps	6.3kbps
Encod.	LPC	53,280	53,263
	LSP	11,264	11,278
	Pitch Est.	145,600	135,063
	Pitch Pred.	505,006	124,748
	MP-MLQ		599,255
	ACELP	139,449	
	Others	149,597	153,648
	Total	1,034,197	1,407,255
Decod.	Unpack	10,884	10,913
	Syn. filter	57,232	57,457
	Others	14,402	14,399
	Total	82,518	82,769

in case of the bitrate of 6.3kbps.

Seeing that the computational labor of Pitch Predictor and MP-MLQ amounts to almost 70% of the total of the encoding process as shown in Table 3, the detailed analysis is executed mainly for these processes. As a result, it turns out that the multiply accumulation (MAC) operation frequently appears in the deepest loop part of these two processes. Moreover, we find that there are two kinds of MAC's which can be classified according to whether or not there needs a shift operation after the multiplication. Generally, audio data are treated as fixed point decimal data, and if operands of MAC are of the fixed point type, the result of multiplication should be shifted by 1 bit to left. In contrast with this, if operands are of the integer type, there is no shift operation after the multiplication. Thus in the G.723.1 codec process, there are only these two kinds of MAC operation. Paying attention to this feature, henceforth let us devise an architecture of DSP.

3. ARCHITECTURE OF THE DSP

Since the MAC operations, decimal or integer, appear frequently both in Pitch Predictor and in MP-MLQ, our primary subject is how to execute these operations in a limited number of clock cycles. A general DSP has an MAC unit which is composed of a multiply block and an accumulate block. In order to improve the performance of such a MAC unit, the following two methods can exist:

A: to insert pipeline registers to improve the clock frequency[4].

B: to construct a number of MAC units to be run in parallel[5].

Method **A** improves the processing ability, but increases the chip size because of the added pipeline registers. On the other hand, Method **B** enlarges the chip size, but not the clock frequency. It should be added that Method **A** requires a higher clock frequency than Method **B**, and that Method **B** increases the chip area much more than Method **A**. For our purpose of the low-power consumption, it is a major priority to reduce the clock frequency, and hence Method **B** is employed here.

Fig. 3 shows a structure of our MAC scheme in the proposed DSP. Two parallel MAC units are used to reduce the the total number of instruction cycles. The required clock frequency is not so high for two MAC units to be pipelined, and hence these two units are connected in parallel so that two MAC operations can be executed within 1 clock cycle. The multiplexer can select the 1 bit left shift according to whether or not MAC operations are fixed decimal, and hence this mechanism can achieve two MAC operations in 1 clock cycle.

The proposed DSP has a 16 bit ALU and a barrel shifter unit in addition to these MAC units. The 16 bit ALU can execute add/sub, logic operation, negate, and absolute operation in 1 clock cycle. Each of 32 bit precision operations is executed in two clock cycles. The barrel shifter can shift the 16 bit data ranging from 1 to 16 bits in both of the left and right directions. Each of the ALU and the shifter has a facility of saturation against overflow, which can reduce the number of comparison and branch instructions

The on-chip memory is composed of two-banked RAM for 1 frame codec work memory, IROM (Instruction ROM) for instructions, and TROM (Table ROM) for filter coefficient table. According to our G.723.1 software simulation, it is decided that the sizes of each RAM and TROM are 2k and 5k words, respectively. Since the 1 frame codec process can be executed without any access to the external memory, not only the total processing speed is improved, but also the band width to the external memory is reduced.

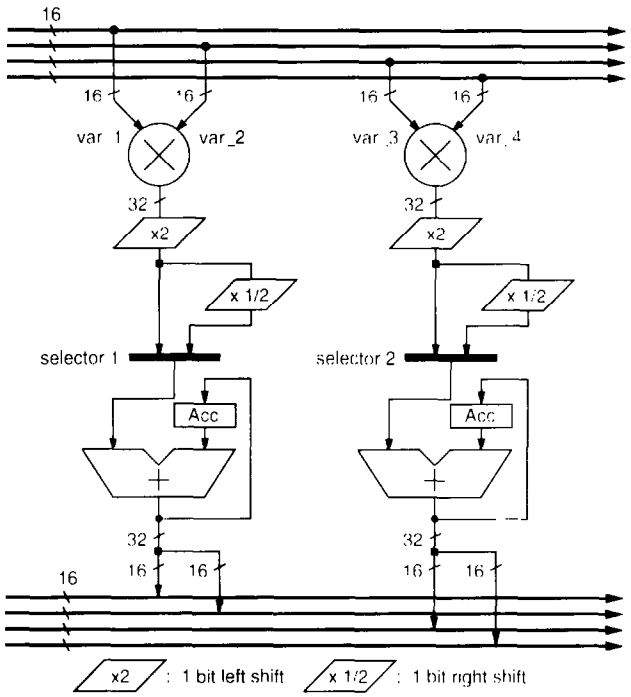

Fig. 3 Block diagram of MAC.

Fig. 4 shows the whole structure of the proposed DSP. In order to operate two MAC units in parallel, a 64 bit internal data bus is employed. This internal bus is divided into two 32 bit buses, say, bus A and bus B. The MAC units are connected to both of these buses, the ALU and the shifter are connected to bus A. The switching multiplexer satisfies the connection requests between RAMs and buses.

A 13 bit address port and a 16 bit data port are provided as Input/Output ports. The I/O port band width is 128kbps for the output of generated PCM, and 6.3 or 5.3kbps for the input of a bitstream. Accordingly, the total band width requires about 130kbps.

In order to realize the G.723.1 encoding process with the proposed DSP, it needs 57,702 cycles in Pitch Predictor, 67,251 cycles in MP-MLQ, and 250,295 cycles in the total encoding process. On the other hand, the decoding process needs 26,327 cycles. Consequently, in order to realize a realtime G.723.1 codec with the proposed DSP, the clock frequency is given by

$$(250,995 + 26,327)/(30/1000) = 9.23MHz. \quad (2)$$

4. IMPLEMENTATION RESULTS

The proposed DSP core has been integrated by using a top-down ASIC design system, *COMPASS Design Tools*. Table 4 summarizes the main features of the proposed DSP core, and Fig. 5 shows the layout pat-

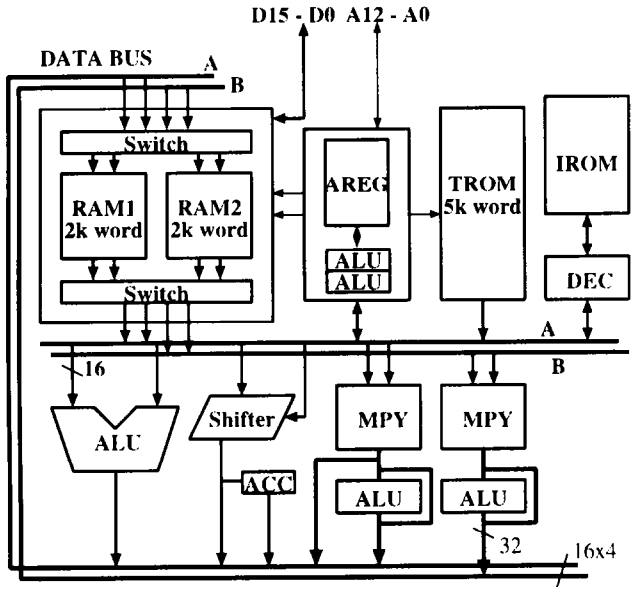

Fig. 4 Block diagram of DSP.

terns obtained by employing 0.35μm CMOS standard-cell and datapath libraries. The maximum clock rate is 58MHz, and hence the realtime codec of G.723.1 can be realized. Since on-chip memories totally occupy the greater part of the chip area, the increase for an additional MAC unit is only 2.76% of total chip size.

Table 4 Main features of DSP core.

Technology	0.35μm CMOS 3LM
Chip size	7.75 mm^2 (3.05 mm × 2.54 mm)
Maximum clock rate	57.73MHz
Power dissipation	44.91 mW (3 V, 10MHz)

5. CONCLUSION

A VLSI implementation of a low-power DSP is described, which is dedicated to the G.723.1 low bitrate speech codec. A number of sophisticated DSP microarchitectures for the G.723.1 codec have been devised, such as the dual multiply accumulators, the rounding and saturation mechanism, two-banked on-chip memory, etc. The proposed DSP has been integrated in the total area of 7.75 mm^2 by using a 0.35 μm CMOS technology, which can operate at 10MHz with the dissipation of 45mW from a single 3V supply. Development is continuing further on the refinement of VLSI implementation of the G.723.1 audio coding.

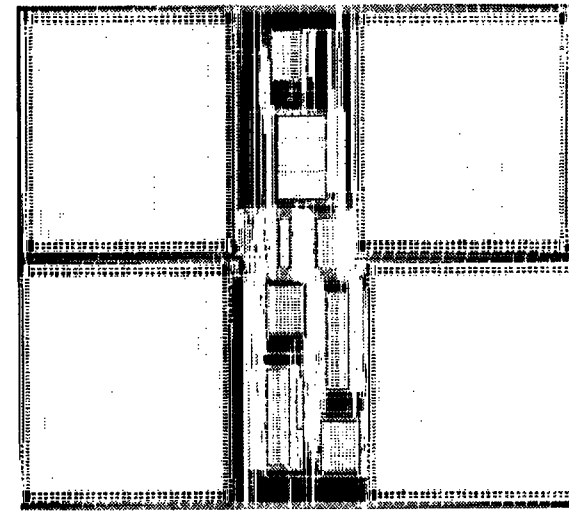

Fig. 5 Layout patterns.

6. REFERENCES

[1] ITU-T Rec. G.723.1: "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s." International Standard, Mar. 1996.

[2] ITU-T Rec. H.324: "Terminal for low bitrate multimedia communication." Draft International Standard, Nov. 1995.

[3] ITU-T Rec. H.263: "Video coding for low bitrate communication." Draft International Standard, May 1996.

[4] H. Kabuo, M. Okamoto, I. Tanaka, H. Yasoshima, S. Marui, M. Yamasaki, T. Sugimura, K. Ueda, T. Ishikawa, H. Suzuki, and R. Asahi: "An 80-MOPS-peak high-speed and low-power-consumption 16-b digital signal processor." *IEEE Journal of Solid-State Circuits*, vol. 31, no. 4, pp. 494-503, Apr. 1996.

[5] T. Shiraishi, K. Kawamoto, K. Ishikawa, H. Sato, F. Asai, E. Teraoka, T. Kengaku, H. Takata, T. Tokuda, K. Nishida, and K. Saitoh: "A 1.8V 36mW DSP for the half-rate speech CODEC," in *Proc. IEEE Custom Integrated Circuits Conf.*, pp. 371-374, May 1996.

[6] M. H. Miki, G. Fujita, T. Kobayashi, T. Onoye, and I. Shirakawa: "A low-power H.263 video codec core dedicated to mobile computing," in *Proc. IFIP Int'l Conf. on VLSI*, Aug. 1997.