An MPEG-2 Video Encoder LSI with Scalability for HDTV based on Three-layer Cooperative Architecture

Mitsuo Ikeda, Toshio Kondo, Koyo Nitta, Kazuhito Suguri,
Takeshi Yoshitome, Toshihiro Minami, Jiro Naganuma, and Takeshi Ogura

NTT Human Interface Laboratories
1-1 Hikarinooka Yokosuka, 239-0847, Japan
Email: {ike, kond, koyo, suguri, tome, nami, jiro, ogura}@nttvdt.hil.ntt.co.jp

Abstract

This paper proposes a new architecture for a single-chip MPEG-2 video encoder with scalability for HDTV and demonstrates its flexibility and usefulness. The architecture based on three-layer cooperation provides flexible data-transfer that improves the encoder from the standpoints of versatility, scalability, and video quality. The LSI was successfully fabricated in the 0.25-μm four-metal CMOS process. Its small size and its low power consumption make it ideal for a wide range of applications, such as DVD recorders, PC-card encoders and HDTV encoders.

1 Introduction

Recent advances in video compression technology have enabled us to provide a much greater volume and range of digital image media. The MPEG-2 video standard [1], in particular, is currently seeing extensive worldwide use in digital satellite broadcasting, DVD storage media, and terrestrial digital broadcasting. MPEG-2 encoders that are compact and versatile, yet maintain high video-quality, are eagerly desired. Figure 1 shows the main target area of each profile in the MPEG-2 standard. The area covered by the profiles has expanded from the area of the initial standard so that it now includes profiles for high-quality use.

In 1996, we developed a two-chip SP@ML encoder [2][3]. The encoder satisfied the requirements of high video-quality and low encoding delay. It has been used in practical codec systems: in portable CODEC systems, for a PC-board encoder [4], and for encoder boards in digital CATV systems [5].

Several single-chip MPEG-2 video encoders have been developed mainly for use in digital storage media [6][7][8]. As far as real encoder applications go, however, no encoder LSIs adequately satisfy the requirements for both low cost and stable high video-quality. In addition, it is difficult to extend the function to a wide variety of encoding specifications, such as 4:2:0/4:2:2, CBR/VBR, and ML/HL[1]. Among the specifications, the need for an HL encoder has especially increased, since the HDTV (high-definition television) terrestrial digital broadcasting is close at hand.

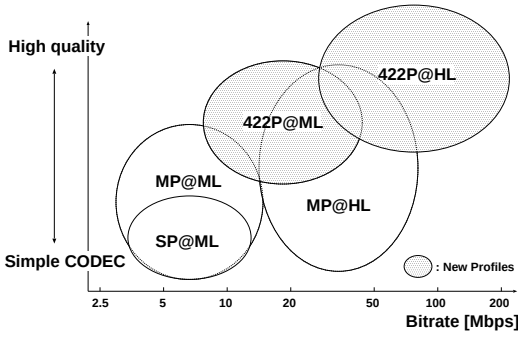

Figure 1: Target Field of the Profile

In response to the demands for versatility, compactness as well as video quality, we propose the concept of three-layer cooperative architecture for a single-chip MPEG-2 video encoder with scalability for HDTV. Flexible intra/inter-chip communication schemes on the architecture enable the encoder to provide a wide range of applications, scalability for HDTV, high video-quality, and compactness.

Section 2 describes the architecture of the encoder LSI and its design method. Section 3 covers implementation and design results. Section 4 is a brief conclusion.

2 LSI architecture

2.1 Requirements for the LSI

Figure 2 is a diagram of the functions of an MPEG-2 video encoder. As the whole process, the encoder

[1] 4:2:0/4:2:2: chroma-formats (Appendix A.2). CBR/VBR: Constant/Variable Bit-Rate. ML/HL: Main-level/High-level, that indicate the performance of encoders and decoders. [1]

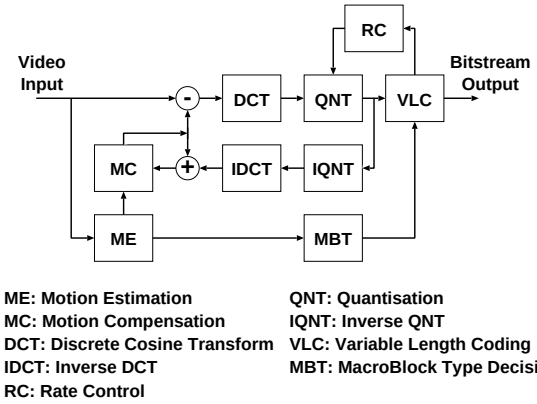

Figure 2: Function Model of MPEG-2 Encoding

Table 1: Function Modes for Applications

	Communicat.	Storage	Broadcast†
CBR	*A*		*A*
VBR		*A*	*B*
IPB	*B*	*A*	*B*
IPP	*A*		*B*
I I I		*B*	*B*
LD	*A*		*B*
4:2:2		*B*	*A*

A: absolute necessity, *B*: necessity
†: including the distribution, the relay and the source handling.
IPB,...: GOP structures (Appendix A.1)
LD: Low-delay Coding Mode
4:2:2: 4:2:2 chroma-format (Appendix A.2)

receives video input, and outputs a bitstream. Initially, motion estimation (*ME*) and motion compensation (*MC*) are performed on the input in order to remove temporal redundancy. Motion estimation involves finding out where a macroblock[2] has moved to when the current picture is compared with reference pictures. The result is a motion vector that is included in the output as the macroblock header information (*MBT*). The video data is transformed by discrete cosine transform (*DCT*) and quantisation (*QNT*). It is compressed by means of variable-length coding (*VLC*). For the motion compensation on the following picture, the compressed data is decoded by means of inverse quantisation (*IQNT*) and inverse discrete cosine transform (*IDCT*). The rate-control (*RC*) function adjusts the output to the proper bit rate. This involves counting the number of encoded bits and calculating the quantiser scale for feedback control.

MPEG-2 video encoding is a combination of various types of video processing [2]. It requires both high-performance/high-throughput of video signal processing (*ME*, *DCT* etc.) and flexibility/programmability in the field (*RC*, *MBT*, *MC mode decision* etc.). In addition to the necessary conditions for the implementation of MPEG-2 encoding, the following items are required for the encoder LSI.

Req. I: *Versatility for a wide range of applications*
For the applications in the various media, the encoder has to provide the function modes shown in Table 1. For example, VBR technology is important for storage media like DVDs and authoring tools, while CBR is necessary in many real-time applications.

Req. II: *Scalability for HDTV*
MPEG-2 MP@HL encoding with high video-quality requires parallel processing by more than one encoder LSI. In this case, it is difficult for multiple LSIs to share data, especially in the case of reference pictures for motion compensation. Furthermore, the encoder has to have scalability for different resolutions for all HDTV specifications.

Req. III: *Video quality with high-performance ME/MC*
ME/MC is one of the key technologies of MPEG-2 video encoding, and it has a great influence on the video quality.

Req. IV: *short time-to-market*
Needless to say, the requirements for short time-to-market is essential.

Req. V: *compactness*
Chip size, chip package, power consumption, chip cost including peripherals, are also important issues for such practical encoders as recordable DVDs, portable CODECs and PC-card encoders.

2.2 Three-layer cooperative architecture

To meet the requirements for the encoder LSI, the functions are partitioned onto software/hardware modules, and the architecture with three-layer construction is introduced.

Figure 3 illustrates the concept of video-data buffering, which is not presented in the figure of the function model (Fig. 2). Video data at the picture level

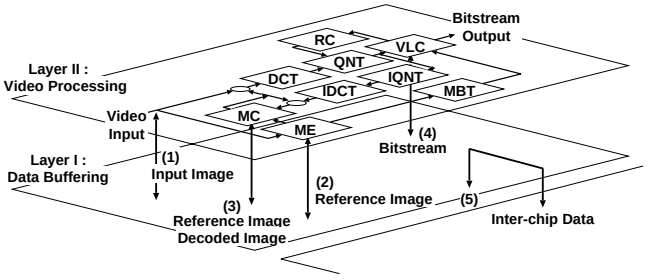

Figure 3: Function Layers

[2] A basic unit in MPEG-2 standard. Composed of 16×16 pixels.

are stored in external memory (Layer I). To reduce the area of the LSI, its video-data memory on the LSI is restricted to macroblock-level data (Layer II). The bandwidth of the external memory and throughput of the interface module are estimated in this phase.

Next, there are certain functions which require flexibility and programmability in particular. Rate-control (RC) should be partitioned onto a software module, because the improvement of RC is indispensable for high video-quality in the field. Macroblock-type processing (MBT) is also performed by the software, since some optimization is necessary for each application. RC and MBT, however, don't require video-data transfer at Layer I. These functions are partitioned in the third layer (Layer III), where the control of whole encoding sequence is performed by an embedded RISC processor. The three layers are composed as shown in Fig. 4.

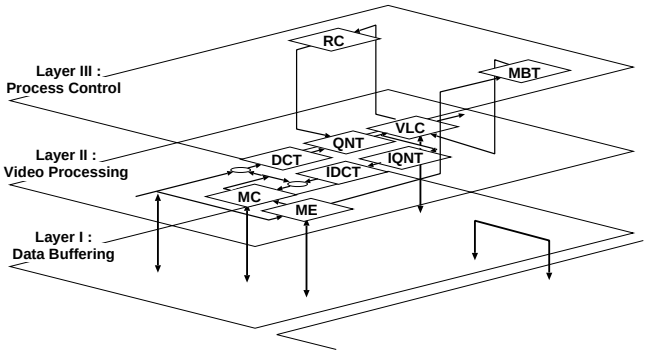

Figure 4: Function Layers (2)

There is another function that requires an appropriate level of programmability. To achieve high video-quality (Req. III), the ME/MC module has be able to optimize motion compensation adaptively as well as perform a powerful motion search. Hence, we combined a search engine (SE) and a SIMD processor (SIMD) for the ME/MC. The SE executes a wide area search with rough precision for motion vectors. The SIMD performs a fine search and the MC. The SE consists of hard-wired logics to obtain enough performance for the ME, and the SIMD has an embedded software to provide the adaptive MC.

Figure 5 is a block diagram of the encoder chip. Its three-layer construction corresponds to the three function layers. The encoding process is controlled by three software modules: the encoding system by RISC on Layer III, ME/MC by SIMD on Layer II, and data-flow control by the memory interface module(SDIF) on Layer I. The encoding functions are partitioned onto hardware/software modules, as shown in Table 2.

In this architecture, the independency of each module is one of the important features. Each hardware

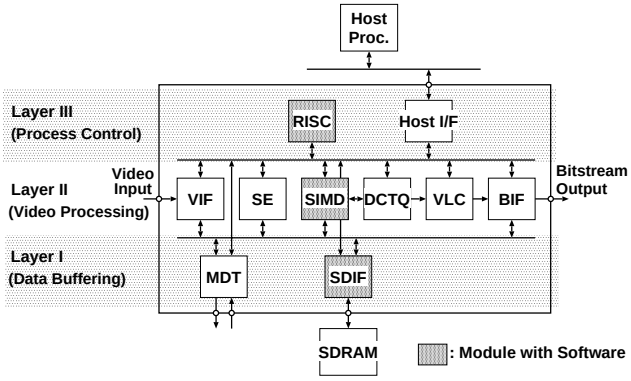

Figure 5: Block Diagram

Table 2: Modules on the LSI

Module	Main Function	Layer	SW	I.P.
HIF	Host I/F	III	–	B
RISC	System-ctrl, RC, MBT	III	SW	A
VIF	Video I/F	II	–	B
SE	1st ME	II	–	C
SIMD	2nd ME / MC	II	SW	C
DCTQ	$DCT/QNT/IQNT/IDCT$	II	–	A
VLC	VLC	II	–	A
BIF	Bitstream I/F	II	–	C
SDIF	SDRAM I/F, Buffer-ctrl	I	SW	C
MDT	Multi-chip Data Transfer	I	–	C

SW: Software on the module
I.P.: Reuse of Intellectual Property
(A: Description-level, B: Architecture-level, C: Brand new)

module performs its processing independently except for an interaction that occurs once in a macroblock sequence, and a flexible pipeline can be built with a combination of the processing. Figure 6 indicates the basic protocol of data transfer. The protocol of "Request"

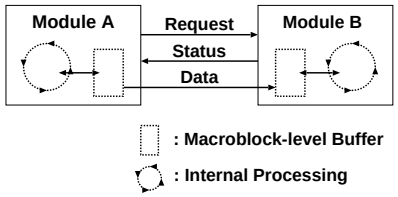

Figure 6: Basic Protocol of Data Transfer

and "Status" provides a simple handshake between the modules. Each module has macroblock-level buffers that enable the module to perform its own processing and data transfer independently at any time during a macroblock sequence. This loosens the time restrictions and gives flexibility to the macroblock pipeline. Even though the area for the macroblock-buffers is increased, the overhead is not so great, as will be discussed later.

Module independency is also effective in implementation. Each module can be designed with the optimal module architecture for such properties as parallelism, complexity, throughput, and flexibility of function. The concept of optimal module architecture effects the optimization of area and power consumption in each module as well as gives versatility of function in future designs. It also leads to efficient development (Req. IV), because of reusability of modules as intellectual property, ease of simulation that results from the software simulator of the functions [10], and local modifications that are closed within the module.

2.3 Flexible communication scheme

Flexible intra-chip and inter-chip communication schemes on the architecture are essential to satisfy the requirements of versatility (Req. I), multi-chip encoding with inter-chip data transfers (Req. II), and integration of the memory-interface module for the reduction of peripherals (Req. V). The main features of the schemes are as follows:

1. The memory interface modules are integrated and the protocol of the data transfer is unified for all modules. The video-data transfers are controlled by the task control on the interface module.
2. A program sequencer is embedded in the interface module so that the order of the transfers and their input/output are determined flexibly by its software. In other words, the sequence of data transfers in/between the chip are spatially and temporally programmable.

Utilizing these features, flexible intra/inter-chip data transfers are available.

Intra-chip Data Transfer: Figure 7 shows the intra-chip communication model. The arrows in Fig. 7 indicate that the memory interface module can change the input/output modules of the data by its software. It can also relay the data without accessing the external memory. The timing of the data transfer is controllable by the software so that the buffering interval is maintained. For example, the SE and the SIMD operate independently (Fig. 5). Therefore, the motion vectors from the SE can be stored in the external memory until fine search on SIMD is performed after some time interval has elapsed. This enables the encoder to perform the fine search and the processes that follow, such as *DCT*, by utilizing the motion information of "future" pictures.

Inter-chip Data Transfer: Figure 8 shows the inter-chip communication model. The inter-chip data transfer through the Multi-chip Data Transfer module(MDT) enables the encoder chips to use the data

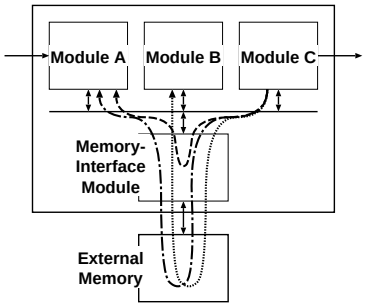

Figure 7: Intra-chip Communication Model

on other encoder chips. This scheme effects the scalable configuration of multi-chip encoding for HDTV.

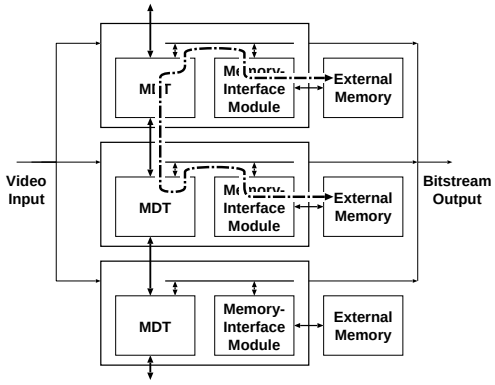

Figure 8: Inter-chip Communication Model

3 Implementation and evaluation

Concerning implementation, the way the software / hardware modules cooperate in terms of their processing on the architecture is a key issue for driving the whole encoding smoothly.

3.1 Hardware/software cooperation

Figure 9 shows a pipeline schedule for several macroblocks on the architecture. The combination of three software modules on three function layers gives enough programmability to the pipeline to satisfy the versatility requirement. In other words, it enables the pipeline schedule to alter its timing on demand for a specific application. Each hardware module interacts with neighboring modules once in a macroblock. The macroblock-level buffers described before give flexibility to the macroblock pipeline. For example, even if a process in the upper level, which is shown as a shaded block on Layer III in Fig. 9, is appended onto the RISC software, the other modules are able to perform their processing without loss, and the sequence on the macroblock pipeline continues smoothly, maintaining the required performance of encoding.

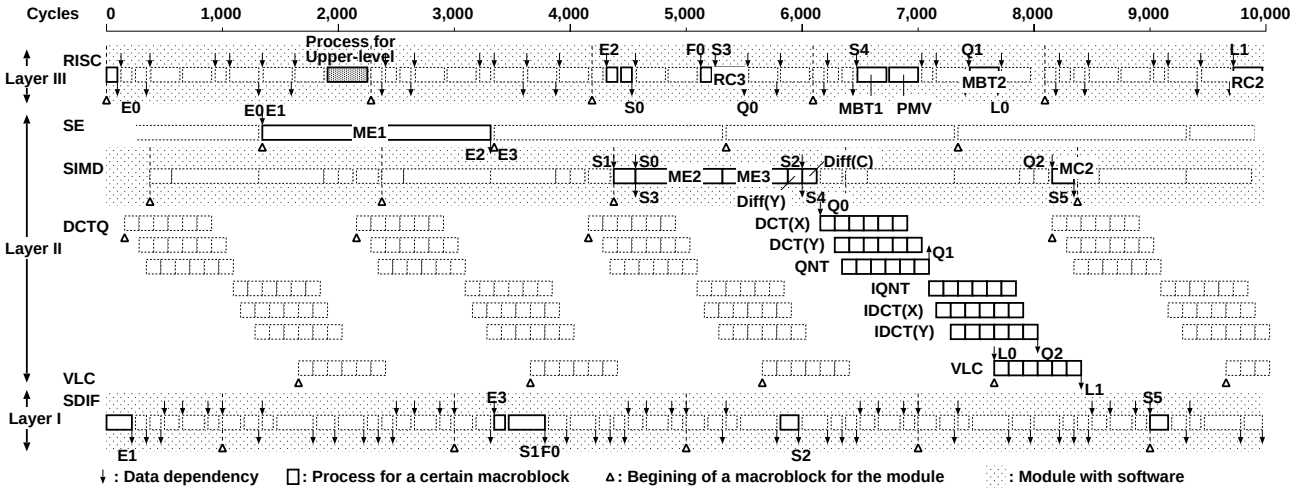

Figure 9: Macroblock-Pipeline Schedule

As another example of flexible macroblock pipeline, the schedule can be transformed so as to encode with the 4:2:2 chroma format. Because the chrominance data for the 4:2:2 format is twice that of the 4:2:0 format (Appendix A.2), the process times for a macroblock on the SIMD and *DCT/QNT* are increased. In this case, the SIMD software can alter the timing of the encoding process on Layer II. The influence of the altered timing of Layer II on other layers is absorbed by the macroblock-level buffer on Layer II. Therefore, the encoding sequence for the 4:2:2 chroma format can be established simply by changing the software configuration on the SIMD.

3.2 Simulation tools

For the hardware/software concurrent design, some tools were used in each design phase. In the architecture-level design, a software encoder, where the MPEG-2 core algorithm was described in C language, was utilized for the analysis of MPEG-2 encoding. The method of the analysis for the previous encoder chipset [2] was extended for this LSI. In the function-level design, a high-speed software-based platform was developed to shorten the turn-around-time of the software/hardware concurrent development [10].

Because of the wide range of the simulation cycles to be executed, it is necessary to use various types of simulators, which have enough performance and functionality suitable for the purpose of simulation and the design phase. Figure 10 presents the simulation time for each simulator/emulator in our computer environment. The line segment indicates the run-time of simulation executed by each simulator/emulator practically. For example, it takes about seven hours for our compile-based RTL simulator to simulate an encoding process of one picture. In order to keep short turn-around-times, the RTL simulators are used to simulate the process below the picture layer, and the FPGA emulator is adopted to emulate the whole encoding-sequences.

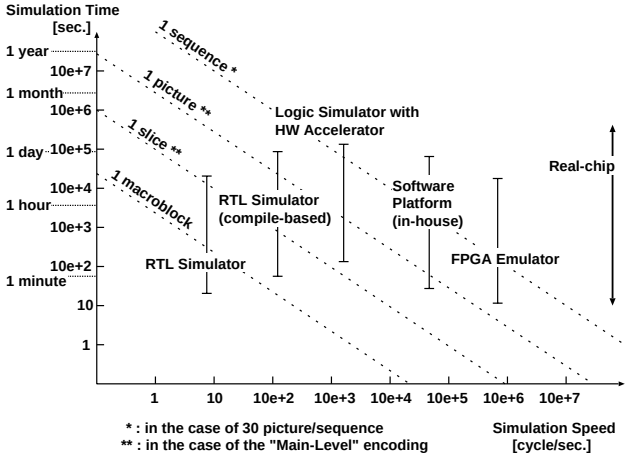

Figure 10: Simulation Time

3.3 Evaluation

Table 3 shows the LSI characteristics. The functional features of the LSI are shown in Table 4.

Table 5 shows the picture-level data stored in the external memory, SDRAM. Since the external memory and its usage are configurable to the application, Table 5 shows the values for a typical case. Compared with the usage of the external memory, the capacity of each internal memory in Layer II is at most 128K bits. The architecture makes it possible to restrict the internal memories to the minimum size necessary to

Table 3: Characteristics of LSI

Die Size	9.8×9.8 mm^2
Technology	0.25-μm four-level metal CMOS
Supply Voltage	2.5 V/3.3 V
Transistors	5.0 M Tr.
Clock Frequency	81 MHz
Power Consumption	1.5 W
Package	208-lead PQFP

Table 4: Functional Features of LSI

Profile and Level	MP@ML, 4:2:2Profile@ML
(multi-chip)	*MP@HL, 4:2:2Profile@HL*
Search Range	−113.5/+99.5 (H), ±57.5 (V)
(wide-area mode)	*±211.5 (H), ±113.5 (V)*
Encoding Delay	Min. 85 ms at M=1
External Memory	16-Mbit SDRAM×2 *or*
	64-Mbit SDRAM×1

(H): Horizontal, (V): Vertical

drive the macroblock pipeline. The macroblock-level buffers also enable the SDIF to keep its bandwidth usage above 80% because the timing of data transmission on the bus is allowed to be flexible in a macroblock sequence.

Table 5: Picture-level Data in SDRAM

Data	Module	Capacity
Input Image	VIF	12,800 Kb
Decoded Image	SIMD	8,400 Kb
Subsampled Image	SE,SIMD	4,500 Kb
Image after Filtering	VIF,SIMD	4,500 Kb
Bitstream	BIF	3,800 Kb
Inter-chip Data	MDT	—
Instructions of RISC	RISC	500 Kb

Table 6 shows the software characteristics. By utilizing the intra-chip communication scheme, software on the RISC processor can use the external memory of 16K words as a second instruction memory. A portion of the software can be swapped onto the internal instruction memory on demand. This reduces the size of the internal memory. The capacity of the external instruction memory that is used for the software can be changed according to the system configuration.

The LSI was successfully fabricated using a 0.25-μm four-level metal CMOS technology [9]. It took only one week to confirm its high-quality encoding on evaluation boards after the chips were packaged.

Table 6: Types of Software on the LSI

Module	Language	Size of Program Memory
RISC	C	32 bit × 4 K word †
SIMD	Assembler	32 bit × 4 K word
SDIF	Mnemonic	20 bit × 64 word

†: the first instruction memory on the LSI

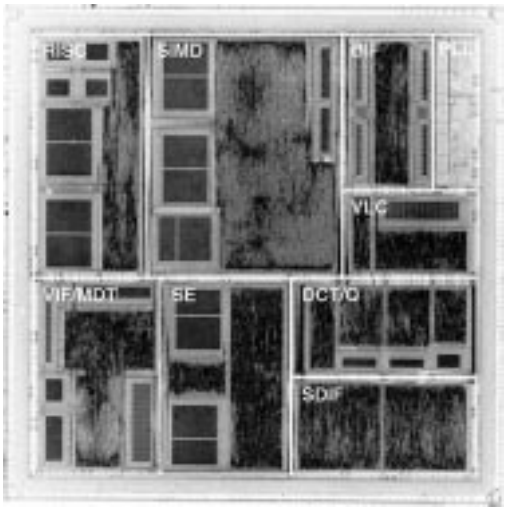

Figure 11: Photograph of the LSI

4 Conclusion

A new architecture and design method for a single-chip MPEG-2 video encoder was developed. The main features of the encoder are versatility for a wide range of applications, scalability for HDTV, high video-quality supported by high performance of motion compensation, and compactness. To achieve these features, the authors introduced the concept of three-layer cooperative architecture, which is suitable for the hybrid video-processing LSIs such as an MPEG-2 encoder.

The encoder chip was successfully fabricated. It can be used as an MP@ML encoder as well as a 4:2:2P@ML encoder and several can be used together as an MP@HL / 4:2:2P@HL encoder for HDTV. Its compactness and versatility make the chip applicable to a wide range of encoder systems, including DVD recorders, portable CODECs, PC-card encoders, and portable HDTV encoders.

Acknowledgements

The authors would like to thank Dr. Susumu Ichinose of the NTT Human Interface Laboratories and Dr. Ryota Kasai of the NTT System Electronics Laboratories for supporting this work. Thanks are also due to the members of the NTT Visual Communication Laboratory for their helpful suggestions.

References

[1] "ISO/IEC 13818-2 International Standard (Video)," November 11, 1994.

[2] M. Ikeda et al., "A Hardware/Software Concurrent Design for a Real-Time SP@ML MPEG2 Video-Encoder Chip Set," Proc. of the ED&TC, pp. 320-326, March 1996.

[3] T. Kondo et al., "Two-Chip MPEG-2 Video Encoder," IEEE Micro, pp. 51 - 58, April 1996.

[4] Y. Tashiro et al. "MPEG2 Video and Audio CODEC Board Set for a Personal Computer," IEEE Global Telecommunications Conference, Vol. 1, pp. 483-487, 1995.

[5] N. Terada et al. "An MPEG2-Based Digital CATV and VOD System using ATM-PON Architecture," Proc. of the International Conference on Multimedia Computing and Systems, 1996.

[6] M. Mizuno et al., "A 1.5W Single-Chip MPEG-2 MP@ML Video Encoder,.," IEEE J. Solid-State Circuits, vol. 32, pp. 1807-1816, November 1997.

[7] E. Miyagoshi et al., "A 100mm^2 0.95W Single-Chip MPEG2 MP@ML Video Encoder,.," Proc. of the International Solid-State Circuits Conference, pp. 30 - 31, February 1998.

[8] E. Ogura et al., "A 1.2W Single-Chip MPEG2 MP@ML Video Encoder LSI,.," Proc. of the International Solid-State Circuits Conference, pp. 32 - 33, February 1998.

[9] T. Minami et al. "A Single-chip MPEG2 MP@ML Video Encoder LSI with Multi-chip Configuration for a Single-board MP@HL Encoder," Symposium HOT Chips X, August 1998.

[10] K. Ochiai et al. "High-speed Software-based Platform for Embedded Software of a Single-chip MPEG-2 Video Encoder LSI with HDTV Scalability," Proc. of the Design, Automation and Test in Europe (DATE99), March 1999.

A Appendix

A.1 Picture coding type

Figure 12 shows the layer of MPEG-2 video sequence. A moving picture is a sequence of pictures. These pictures are separated into many groups of pictures (GOP). Each picture consists of some slices, and each slice is made up of some macroblocks. A macroblock is composed of 16 $\times$ 16 pixels. In MPEG-2 technology, such as *DCT* and *QNT*, a macroblock is treated as a basic unit of encoding.

Figure 13 shows the structure of the GOP and the picture type. There are three picture types. Intra-coded pictures (I-Pictures) are coded without reference to other pictures. Predictive-coded pictures (P-Pictures) are coded using motion compensated prediction from a previously coded picture. Bidirectionally-predictive coded pictures (B-Pictures) have both past and future reference pictures for motion compensation.

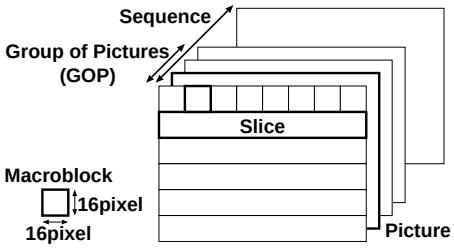

Figure 12: Layer of MPEG-2 Video

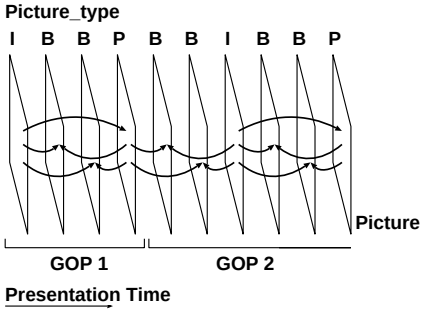

Figure 13: GOP structure and Picture Type

A.2 4:2:2 chroma mode

Figure 14 shows the difference in chroma information between the 4:2:0 and 4:2:2 formats. For the "Main Profile" in MPEG-2 video standard, the chroma information is subsampled to 4:2:0. At the higher profile, it can be encoded directly as 4:2:2. The "4:2:2 profile @ HL" is under process of amendment and will be announced as the standard very soon.

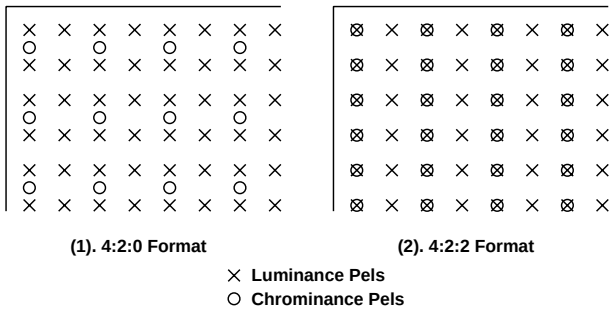

Figure 14: 4:2:0/4:2:2 Format