Low-Power and Power-Aware Video Codec Design: An Overview

Chung-Jr Lian, Po-Chih Tseng, Liang-Gee Chen
Graduate Institute of Electronics Engineering and Department of Electrical Engineering, National Taiwan University, Taipei 10617, Taiwan, China
Email: {cjlian, lgchen}@video.ee.ntu.edu.tw

ABSTRACT

Power consumption is the most critical design issue in a portable appliance. For a battery-operated mobile device to support high-complexity multimedia functions, low-power and power-aware techniques are the key to a successful design. This article reviews typical techniques for a low-power video IC design, and introduces emerging power-aware concepts toward a more power-efficient video codec.

Key words: low-power, power-aware, video codec

I. INTRODUCTION

We are now in a mobile multimedia era. Digital audio/video entertainment and services colorize our life. With the rapid advances of semiconductor and communication technologies, the mobile multimedia environment is more and more mature. We can now enjoy multimedia content not only at home but everywhere and anytime in our daily life. With 3G mobile phones, we can communicate to each other with not only voice but also real time video. With Portable Multimedia Player, we can enjoy music, photos, and movies on the road. Also, if you do not want to miss a special TV program, mobile phone TV broadcast over a DVB-H (Digital Video Broadcast - Handheld) network makes it come true.

Digitalized multimedia contents, especially the video contents, result in large digital data size. Data compression is required, and is a key technology to alleviate the limited storage and bandwidth problem. To provide better compression performance, researchers keep developing more powerful video compression algorithms. These powerful compression algorithms are inevitably more and more complicated [1]. Therefore, a video processing chip, no matter what architecture it is, has to execute more operations and run faster. Also, as there are a large amount of data to be processed, the compression engine itself faces the memory size and bus bandwidth problem while executing the compression algorithms.

Thanks to the fast progress of semiconductor technology, what is viewed as infeasible coding algorithm before is now implementable. For example, the Multiple Reference Frames (MRF) scheme [2] for Motion Estimation (ME) in video coding provides better coding performance but requires more memory spaces. It is not until now a tool in the latest H.264/AVC standard [2], although people have mentioned MRF before. The semiconductor technology makes progress continuously following the well-known Moore's Law. More and more transistors can be integrated on a chip. A dedicated video chip can provide enough power for the real time encoding and decoding of these highly complicated video compression algorithms.

The complexity issue seems solvable now, but here comes a more stringent power consumption problem for mobile multimedia design. According to the keynote speech of Pat Gelsinger from Intel Corporation in Design Automation Conference (DAC) 2004, the Moore's Law will be alive and well in the next two decades. In the future, it is expectable that there will be huge amount of transistors available on a single chip, and the System-On-a-Chip (SOC) with billion to trillion scales transistors will become the mainstream. Therefore, transistors are free, and power is the only real limiter, said Gelsinger.

In addition, according to the content of International Technology Roadmap for Semiconductor (ITRS) 2003 Edition, the power management is the Near-Term (through 2009) Grand Design Challenge, including power dissipation and power density issues. Based on the above two important references, one of the future trend of IC design is that the cost per transistor keeps going down, and the area issue is only with reasonable limitation, while the power becomes the major design constraint and the only real limiter. As a result, there

should be a shift in design methodology in order to cope with the future trend of IC design.

A battery-operated mobile device has only limited energy. A computationally powerful but high-power design is totally unacceptable. Therefore, for a mobile appliance to support high-computation video coding and decoding, the processing engine must be a low-power design. Low-power techniques should be applied in algorithm, architecture and circuit levels. Beyond that, further power-awareness can be added to the codec for more power optimization. The main idea is to have a configurable design that has multiple power consumption modes. The design can then adapt the configuration to a suitable power consumption status manually or automatically considering the video content variations, and the power and video quality tradeoff.

The rest of this article is organized as follows. In section II, we briefly introduce the video coding algorithms and standards, followed by the discussions of video codec architecture design. In section III, we discuss some low-power techniques for video codec design. The concept and feasible embodiment of a power-aware video codec are then described in section IV. Finally, a summary is given in section V.

II. VIDEO COMPRESSION AND VIDEO CODEC DESIGN

2.1 Video coding algorithms and standards

The goal of video coding is to remove the data redundancy. There is temporal, spatial, statistical and perceptual redundancy in digital video data. Video coding is to represent the data in a more compact format. Most video coding algorithms adopt transform and motion compensation (MC) based hybrid video coding. Figure 1 shows a basic video coding framework. The temporal redundancy is effectively removed by motion compensation, which is realized by block-based motion estimation (ME) to search for a similar block in reference frame(s) for a current block. As for the spatial and statistical redundancy, Discrete Cosine Transform (DCT) and entropy coding such as Huffman coding or Arithmetic coding are adopted. Quantization is a scheme to control the compression ratio. The less important information is discarded by quantization scheme, which is basically based on the characteristics of human visual system.

The standardization of video coding algorithms is necessary to provide better interoperability. International experts groups such as ISO/IEC MPEG and ITU-T VCEG have dedicated to the standardization of video coding standards. A number of standards such as MPEG-1/2/4, H.261, H.263, H.263+, and H.264/AVC have been developed. Applications such as VCD, DVD, video conferencing, and video communications are powered by these standards. Currently, MPEG is working on the development of Scalable Video Coding (SVC), Multi-view Video Coding (MVC), and Reconfigurable Video Coding (RVC) for the video coding tool repository concept [3].

Fig. 1 Mobile Multimedia

Some other video codecs are also widely used, such as Microsoft's VC-1 and RealNetworks's RealVideo, etc. China is also developing his own video coding standard, AVS (Audio and Video Coding Standard), which adopts similar techniques as H.264/AVC. As for Motion JPEG and Motion JPEG 2000 based video compression, only intra-frame coding is used. That is, each frame of the video is coded as an independent image. Without inter-frame prediction, the coding performance is lower, but the complexity is much lower, too. Example applications such as surveillance system, and the live video of the messenger on the Internet adopt these intra-frame only coding methods.

2.2 Video codec design

Computational complexity is the issue a real-time video coder should face first. To provide better coding efficiency, complicated algorithms are adopted. A video codec has to complete the processing of very high data rate under real time constraint, say 30 frames per second video.

Video data are partitioned into regular coding units, such as 8 × 8 blocks, 16 × 16 macroblocks, and slices consisted of macroblocks. These video data to be processed come in a stream-like form. Part of the operations, such as transform, motion estimation and compensation, are inherently highly parallel processing operations, while part of the algorithm is highly sequential and consisted of bit-level processing, such as entropy coding and bitstream parsing and formation.

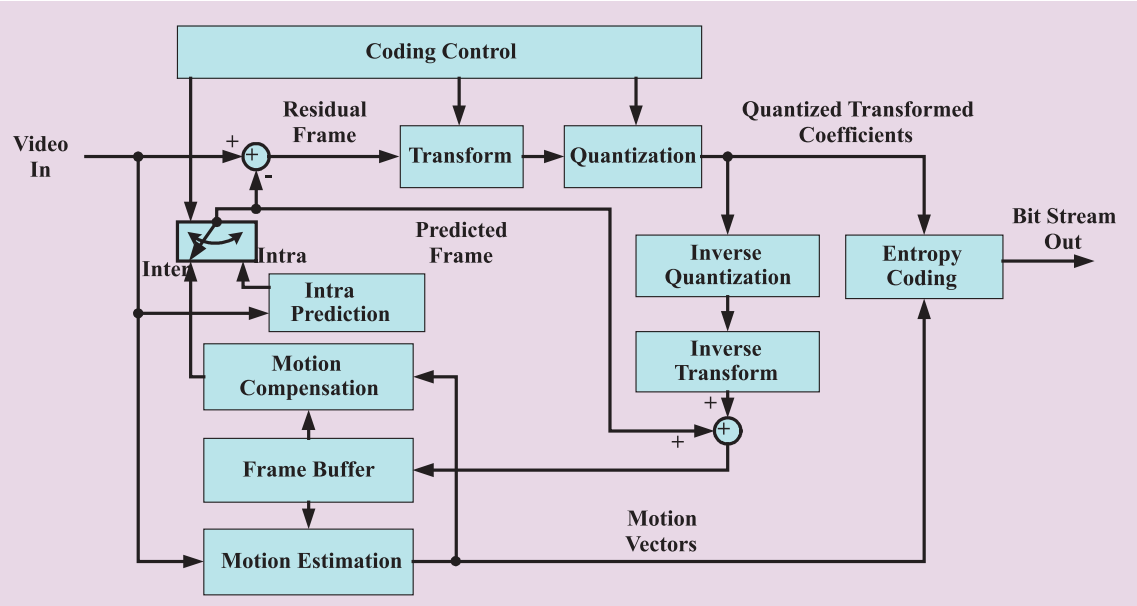

Fig. 2 Hybrid coding scheme of video compression

There should be different architecture mapping strategies for different styles of operations. The design goal is to process these video stream data smoothly and seamlessly [4][5][6].

For a general-purpose processor, it is difficult to achieve real-time performance even clocked at a very high working frequency. That is because the general data path and sequential processing flow of a GPP are not efficient for video coding. As the multimedia market grows rapidly, designers are therefore considering a video application specific processor design, such as video or multimedia processors. In these architectures, either special instructions with parallel processing capability, or dedicated accelerators for the speed-up of some critical operations are added.

CIF (352 × 288) 30fps	MPEG-4 SP	H.264/AVC
Encoder	12 GOPS	80 GOPS
Decoder	200 MOPS	450 MOPS
SDTV (720 × 480) 30fps	MPEG-4 SP	H.264/AVC
Encoder	40 GOPS	272 GOPS
Decoder	680 MOPS	1500 MOPS

Table 1. Computational Complexity of MPEG-4 Simple Profile (SP) and H.264/AVC.

Fully dedicated video codec architecture is another design approach to achieve the optimized performance and power consumption. For every module in a video coding algorithm, the characteristics of the operations are carefully explored. Application Specific IC (ASIC) designers tend to provide a highly parallel architecture so that the required operational frequency is dramatically lowered and hence also lower the voltage and the power consumption. As the fast cost down of manufacturing process, the cost per transistor is lower. Therefore, highly parallel architecture is affordable.

The dedicated architecture design approach provides optimal performance but is lack of flexibility than a processor-based design. For applications requiring multi-standard supporting, platform-based and reconfigurable architectures are promising solutions to get an optimal point considering the flexibility and performance at the same time. A platform-based design is basically a processor with several function-specific accelerators. For example, an array processor for ME, a parallel architecture for DCT, and a bit-level processor for variable length coding and decoding are common accelerators. A reconfigurable architecture is an architecture that can support multiple algorithms, function parameters, etc. For example, a reconfigurable ME architecture may support full search and several other fast searching algorithms.

For a mobile device to provide multimedia functions, the power issue becomes the most important limiting factor. So far, the development of battery technology is still far beyond users' requirement, and the wireless energy transmission is still a question. With limited energy, mobile devices have to use the energy more efficiently. Low power video codec is a must for longer battery lifetime.

General low-power design techniques can be applied for a video codec IC design. In addition to those general ones, video specific features and considerations can provide more possibilities for low power design. Besides, a more advanced power-aware concept can be taken into consideration for further power optimization. In the following two sections, these power-oriented architecture

design issues are discussed.

III. LOW-POWER TECHNIQUES

There have been many researches about low power IC design. Generally, they can be classified into three different levels, the circuit, architecture, and algorithm level. In higher level, say, the algorithm level, a low power technique is usually with significant power saving gain, but the technique is usually more dedicated to a specific algorithm. That is, the technique may be only suitable for that algorithm. As for a low power technique at a lower level, it can usually be applied to VLSI circuits of different applications.

Examples of low power approaches in circuit level are low-power cell library, multiple supply voltages (multi-Vdd), multiple threshold voltages (multi-Vth), and Dynamic Voltage/Frequency Scaling (DVFS). The architecture designers can design their architecture according to the available circuit level techniques. Since designers have in-depth understanding of the algorithms, they can identify where the critical path/module is and do some partitions so that it is easier for EDA tools to apply, for example, higher voltage or lower Vth cells for some critical modules.

In the architecture level, IC designers focus more on the resource allocation and scheduling. Designers explore all possible parallelism in a coding algorithm, and map the algorithm to a high performance parallel architecture. One of the important guideline is to maximize the hardware utilization through resource allocation and data flow optimization. A design with higher hardware utilization usually means better performance over cost ratio and better power consumption, since fewer transistors are idled and wasted. When a low hardware utilization part does exist, using clock-gating technique or the data latch to avoid signal toggling should be considered to save the power.

Besides the area optimization of the logic part, memory optimization is more and more important as the area percentage of memories on a chip is increasing. Therefore, the effect of memory architecture optimization sometimes is more significant than logic optimization. Typical simple but effective memory optimization techniques are memory partition and hierarchical memory organization. When a memory is large, it is better to be partitioned into smaller memories so that when accessing data, only a smaller portion of memories is active. Power consumption is therefore reduced.

As for memory hierarchy, a basic observation is that off-chip memory access consumes much more power than on-chip ones, since the board-level capacitance is much larger than chip-level ones. Therefore, it is better to reduce the frequent off-chip memory access by allocating on-chip data buffer, registers or on-chip memories. Data are read from off-chip memory for one time and write into the on-chip memory. Designers then try to finish all the computations that require these buffered data as operands. In this case, data do not have to be moved back and forth between the off-chip and on-chip memories, and hence reduce significant power dissipations.

In the implementation of a video encoder, motion estimation gives a good example showing the effectiveness and importance of memory optimization. The block matching process of ME is with large data access operations. For each macroblock, the search range data in reference frame(s) have to be accessed, and there is actually a large overlap among these memory accesses. There are, therefore, many data reuse scheme, such as level A to D data reuse [7], discussed in the literature for both better bandwidth and power optimization.

In the algorithm level, video data and algorithm characteristics can be carefully examined for possible low-power optimization. The ME, for example, is the most computation and power intensive module. It is good news that the ME searching algorithm is out of the scope of video coding standards. That is, the ME searching algorithm is not specified in the standard and not restricted to one single procedure. Therefore, there exists large room for algorithm optimization. A number of fast searching algorithms, and searching patterns are proposed in the literature [8]. Fast searching algorithms search fewer candidates than full search, so it requires fewer computations at the cost of some quality degradations. For full search ME, computation-saving schemes such as partial distortion elimination (PDE) and successive elimination algorithm (SEA) are still applicable to the reduction of unnecessary operations in advance.

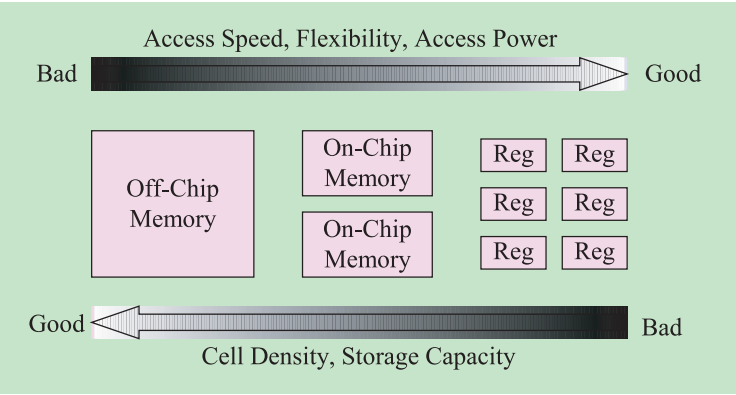

Fig. 3 Memory hierarchy: trade-offs and characteristics.

In addition to the power consumption from computations, data transfer also contributes large portion of the power consumption. Bus switching activity reduction is a low-power technique considering the correlation between signals. Embedded compression is another technique to compress the data inside the computing engine so that the memory size and bandwidth problem is lightened and the power consumption is lowered at the moderate cost of some coding and decoding overheads. To reduce this great source of power dissipation, the embedded compression of frame buffer access in a video coder is presented and adopted in some multimedia systems [9][10].

So far, more low power techniques focus on the reduction of dynamic power. As the leakage power phenomenon will be more serious in very deep sub-micron processes, how to design a power-efficient architecture when leakage power is not negligible and even dominant is a hot and challenging issue.

IV. POWER-AWARE VIDEO CODEC

A power-aware video codec is a design capable of scaling power dissipation in response to the changing operating conditions, including the time-varying nature of input, i.e. the signal content variations, the desired output quality, as well as the environmental conditions such as the battery condition. In modern mobile processors for notebook computers, power-aware schemes such as Intel's SpeedStep™ and AMD's PowerNow!™ technologies are implemented by dynamically adjusting their operating frequency and voltage to optimize the battery life. The idea is to provide just-enough performance on demand. When the application requires higher computing power, the voltage and frequency are scaled up to provide enough performance. While the application is less complicated, the voltage and frequency are scaled down to save the power.

In a function dedicated video codec IC, the same power-aware concept can be implemented. Actually, video codec designers can get a more optimal power-aware design by utilizing the high signal variations in video signals, and the trade-offs between video quality and power consumption [11].

In the video specification level, there are several parameters that are good candidates for power-aware considerations. For example, one can adjust the resolution, frame rate, ME search range, or bit-rate to have different coding complexity. The SVC standard, for example, can inherently provide this algorithm complexity and power scalability in both the encoder and decoder side easily. Especially, the decoder can decode a scalable bitstream into a video with just enough resolution, frame rate, and quality according to the available decoding capability, energy status, or user preferences.

In the module level of a video codec, there are many factors that can be further considered for power-awareness. A function-specific module can have multiple operating modes by a reconfigurable architecture. For example, a reconfigurable ME engine can execute several different searching algorithms on a function-specific hardware, and a content dependent DCT/IDCT module [12][13] can adjust its computing precisions and skip some operations by detecting the content to be processed.

In [14], a multi-mode content-dependent ME algorithm is proposed for power-aware video coding. Based on the predictions and judgments of the motion complexity, the ME execution is switched among one of the four searching modes: Full Search (FS) mode, adaptive search range mode, adaptive enhanced four-step search (E4SS) and FS mode, and adaptive E4SS and three-step search mode. Those modes with lower search range or adaptive fast algorithms are of lower computational complexity, while the image quality in terms of PSNR (Peak Signal to Noise Ratio) is dropped a little accordingly. A reconfigurable architecture for this multi-mode ME is implemented. Post-layout gate-level power dissipations under 1.8V and 49 MHz are 51.27mW, 24.88mW, 13.76mW, and 8.46mW for each mode respectively, and with only 0.0036 dB, 0.01 dB, and 0.16 dB average output quality drop in the later three modes. This design shows the power-awareness of a ME processor.

In a conventional video coding system, rate-distortion optimization is the main consideration. For mobile devices, the power issue becomes very critical so that the original R-D optimization problem should be extended to be a R-D-Complexity or R-D-Power optimization problem [15]. That is, a power-aware device cares about not only R-D performance but also how much energy has to be paid to boost the video quality. When the power budget is limited, devices have to adjust the sconfiguration automatically and intelligently to mimic human being's thinking of priority.

V. SUMMARY

Low-power and power-aware video codec design is a key to successful mobile multimedia applications. It is always a challenging task providing a more powerful processing engine but at the same time asking for less power consumption. In this article, we, from a video codec IC designer's perspective, provide an overview of the low-power and power-aware techniques. These

techniques demonstrate good results in extending the battery lifetime. By embedding the power-aware scheme, a mobile device will be more power-efficient and intelligent.

REFERENCES

[1] T. Sikora,"Trends and perspectives in image and video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 6(17, Jan. 2005.

[2] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, "Video coding with H.264/AVC: tools, performance, and complexity," *IEEE Circuits and Systems Magazine*, pp.7 (28, 1st quarter 2004.

[3] The MPEG Homepage, ISO/IEC Moving Picture Experts Group (MPEG) [online] http://www.chiariglione.org/mpeg/

[4] P. Pirsch, N. Demassieux, and W. Gehrke, "VLSI architectures for video compression-a survey," *Proceedings of the IEEE*, vol. 83, No. 2, pp. 220(246, Feb. 1995.

[5] P.-C. Tseng, Y. -C. Chang, Y.-W. Huang, H.-C. Fang, C.-T. Huang, and L.-G. Chen, "Advances in hardware architectures for image and video coding-a survey," *Proceedings of the IEEE*, vol. 93, No. 1, pp. 184(197, Jan. 2005.

[6] S.-Y. Chien, Y.-W. Huang, C.-Y. Chen, H.H. Chen, and L.-G. Chen, "Hardware architecture design of video compression for multimedia communication Systems," *IEEE Communications Magazine*, vol. 43, no. 8, pp. 123 (131, Aug. 2005.

[7] J.-C. Tuan, T.-S. Chang, and C.-W. Jen, "On the data reuse and memory bandwidth analysis for full-search block-matching VLSI architecture," *IEEE Trans*. CSVT, vol. 12, no. 1, pp. 61(72, Jan. 2002.

[8] Y.-W. Huang, C.-Y. Chen, C.-H. Tsai, C.-F. Shen and L.-G. Chen, "Survey on block matching motion estimation algorithms and architectures with new results," *Journal of VLSI Signal Processing Systems*, vol. 42, no. 3, pp.297(320, March 2006.

[9] H. Shim, N. Chang, and M. Pedram, "A compressed frame buffer to reduce display power consumption in mobile systems," in *Proc. of ACM/IEEE Asia South Pacific Design Automation Conference*, 2004, pp.819(824.

[10] E. G. T. Jaspers and P. H. N. de With, "Embedded compression for memory resource reduction in MPEG systems," in *Proc. of IEEE Benelux Signal Processing Symposium (SPS)*, 2002, pp.17-20.

[11] P. Jain, A. Laffely, W. Burleson, R. Tessier, and D. Goeckel, "Dynamically parameterized algorithms and architectures to exploit signal variations", *Journal of VLSI Signal Processing Systems*, vol. 36, no. 1, pages 27-40, January 2004.

[12] T. Xanthopoulos, A. P. Chandrakasan, "A low-power IDCT macrocell for MPEG-2 MP@ML exploiting data distribution properties for minimal activity," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 5, pp. 693-703, May 1999.

[13] T. Xanthopoulos, A. P. Chandrakasan, "A low-power DCT core using adaptive bitwidth and arithmetic activity exploiting signal correlations and quantization," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 5, pp. 740-750, May 2000.

[14] S. -S. Lin, P. -C. Tseng, C. -P. Lin, and L. -G. Chen, "Multi-mode content-aware motion estimation algorithm for power-aware video coding systems," in Proc. *IEEE Workshop on Signal Processing Systems (SIPS 2004)* pp. 239-244, 2004

[15] Z. He, Y. Liang, L. Chen, I. Ahmad, and D. Wu, "Power-rate-distortion analysis for wireless video communication under energy constraints," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 15, no. 5, pp. 645-658, May 2005.

BIOGRAPHIES

Chung-Jr Lian *received the B. S., M.S. and Ph.D. degrees in electrical engineering from National Taiwan University, Taipei, Taiwan, China in 1997, 1999 and 2003, respectively. He is currently a postdoctoral research fellow in Graduate Institute of Electronics Engineering National Taiwan University, Taipei, Taiwan. His major research interests include image and video coding (JPEG, JPEG 2000, MPEG, and H.264/AVC), image and video codec VLSI architecture design.*

Po-Chih Tseng *was born in Tao-Yuan, Taiwan, China, in 1977. He received the B.S. degree in electrical and control engineering from National Chiao Tung University, Hsinchu, Taiwan, China, in 1999 and the M.S. and Ph.D. degrees in electrical engineering from National Taiwan University, in 2001 and*

2005, respectively. He is currently with Novatek Microelectronics Corp., Ltd., Hsinchu, Taiwan. His research interests include energy-efficient reconfigurable computing for multimedia systems and power-aware image and video coding systems.

***Liang-Gee** Chen received the B.S., M.S., and Ph.D. degrees in electrical engineering from National Cheng Kung University, Taiwan, China. in 1979, 1981, and 1986, respectively. In 1988, he joined the Department of Electrical Engineering, National Taiwan University, During 1993-1994, he was a Visiting Consultant in the DSP Research Department, AT&T Bell Labs, Murray Hill, NJ. In 1997, he was a Visiting Scholar of the Department of Electrical Engineering, University of Washington, Sea-ttle. Currently, he is Professor at National Taiwan University. His current research interests are DSP architecture design, video processor design, and video coding systems.*

Dr. Chen is an IEEE Fellow. He has served as an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY since 1996, as Associate Editor of the IEEE TRANSACTIONS ON VLSI SYSTEMS since 1999, and as Associate Editor of IEEE TRANSACTIONS CIRCUITS AND SYSTEMS II since 2000. He has been the Associate Editor of the Journal of Circuits, Systems, and Signal Processing since 1999, and a Guest Editor for the Journal of Video Signal Processing Systems. He was also the Associate Editor of the PROCEEDINGS OF THE IEEE. He was the General Chairman of the 7th VLSI Design/CAD Symposium in 1995 and of the 1999 IEEE Workshop on Signal Processing Systems: Design and Implementation. He is the Past-Chair of Taipei Chapter of IEEE Circuits and Systems (CAS) Society, and is a member of IEEE CAS Technical Committee of VLSI Systems and Applications, the Technical Committee of Visual Signal Processing and Communications, and the IEEE Signal Processing Technical Committee of Design and Implementation of SP Systems. He is the Chair-Elect of the IEEE CAS Technical Committee on Multimedia Systems and Applications. During 2001-2002, he served as a Distinguished Lecturer of IEEE CAS Society.