A 212MPixels/s 4096×2160p Multiview Video Encoder Chip for 3D/Quad HDTV Applications

Li-Fu Ding, Wei-Yin Chen, Pei-Kuei Tsung, Tzu-Der Chuang, Pai-Heng Hsiao, Yu-Han Chen, Shao-Yi Chien, and Liang-Gee Chen
DSP/IC Design Lab, Graduate Institute of Electronics Engineering and Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
{lifu, wychen, iceworm, lgchen}@video.ee.ntu.edu.tw

ABSTRACT

A 4096×2160p multiview video encoder chip is implemented on a $3.95mm \times 2.90mm$ die with 90nm CMOS technology. A view-parallel macroblock-interleaved scheduling with 8-stage macroblock pipelined architecture achieves 212Mpixels/s throughput, which is 3.4× to 7.7× better than the state-of-the-art encoder chips. In addition, 94% on-chip SRAM area and 79% external system memory bandwidth are saved.

Categories and Subject Descriptors

C.3 [**Special-Purpose and Application-Based Systems**]: Signal processing systems; B.7.1 [**Integrated Circuits**]: Types and Design Styles–*Algorithms implemented in hardware*, *VLSI*

General Terms

Design

1. INTRODUCTION

To provide more vivid perception, TV resolution is getting higher and higher. In addition, 3D video becomes emerging because it can present immersive and complete scenes. Therefore, multiview video coding (MVC) is currently being developed as an extension of H.264/AVC [1]. H.264/AVC High Profile is adopted as the base layer. The most significant feature which differs from original H.264/AVC standard is inter-view prediction, which is also called disparity estimation (DE). DE can effectively exploit the inter-view redundancy and saves 20% to 30% of bit rates. Output bitstream of each views are assembled and then transmitted. The bitstream format is compatible with H.264/AVC, so a single-view H.264/AVC decoder can decode the the base layer. However, DE and motion estimation (ME) require ultra high computation and memory access. To encode a 3-view 1080p video, 82.4TOPS computing power and 54.6TB/s memory access are required with a full search algorithm. Moreover, view scalability is a critical functionality to deal with various coding structures of 3D video.

There are 3 challenges to design an efficient MVC encoder chip. 1) Encoding high-definition (HD) multiview video requires high processing capability. 2) Conventional macroblock (MB) pipelining and scheduling cannot deal with various MVC prediction structures. 3) With 3D and quad HDTV specifications, conventional ME architectures require 2.9Mb on-chip SRAM and 13.8GB/s external memory bandwidth, which is far beyond 6.4GB/s supported by DDR2800 at 100% utilization.

The proposed MVC encoder chip is characterized as follows: 1) View-parallel MB-interleaved (VPMBI) scheduling with 8-stage MB pipelining is introduced to overcome the first 2 challenges. With this technique, the processing capability is 212Mpixels/s, at least 3.4× better than the previous works [2, 3, 4]. In addition, view scalability is achieved and supports real-time processing from single-view 4096×2160p to 7-view 720p videos. 2) The cache-based prediction core with a search window (SW) prefetching scheme and a predictor-centered ME/DE algorithm effectively reduces 83% on-chip memory size and 39% external memory bandwidth compared with [4]. These techniques enable the design of H.264/AVC Multiview Extension and High Profile encoder. The MVC encoder chip with above techniques is finally realized on a $11.46mm^2$ die area, which contains 1732K gates using 90nm CMOS technology. The search range of ME/DE is 4× to 64× larger than the previous works [2, 3, 4] while only 20.1KB on-chip SRAM is used. This chip supports maximum throughput of 830kMB/s at 280MHz for 4096×2160p videos.

This paper is organized as follows. The proposed system architecture and scheduling are introduced in Section 2. Section 3 presents the architecture design of important modules. The chip design and verification flow are discussed in Section 4. The measured chip features and architectural comparison are shown in Section 5. Finally, Section 6 concludes the design and implementation of the MVC encoder chip.

2. SYSTEM ARCHITECTURE DESIGN

2.1 8-Stage MB Pipelined System Architecture

Fig. 1 shows the conventional 3- or 4-stage macroblock pipelined architecture [2, 3, 4]. The encoding task is split into integer ME, fractional ME, intra prediction, and entropy coding/deblocking. For MVC/quad HDTV encoding,

DAC 2009, July 26–31, 2009, San Francisco, California, USA.

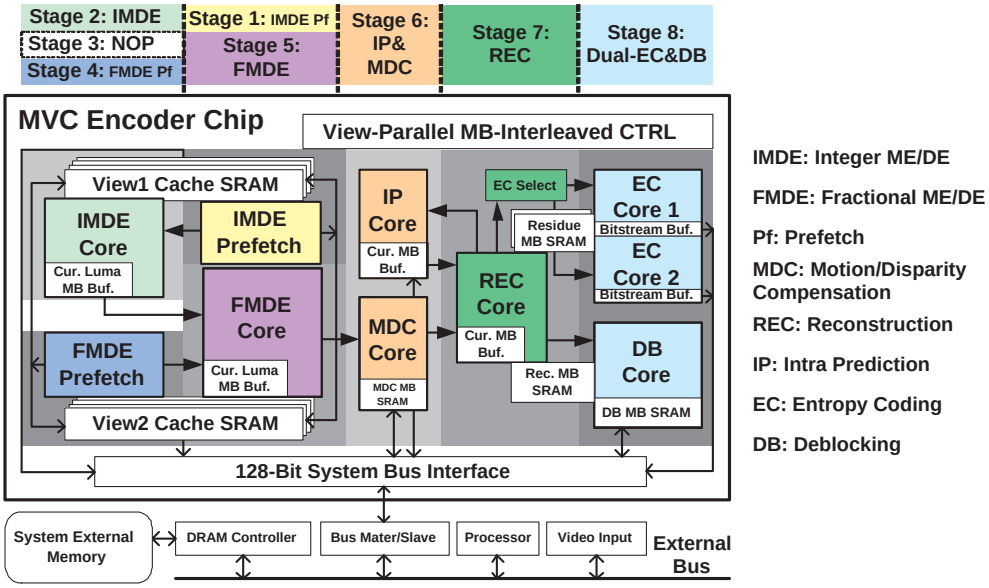

Figure 2: Block diagram of the proposed MVC encoder system.

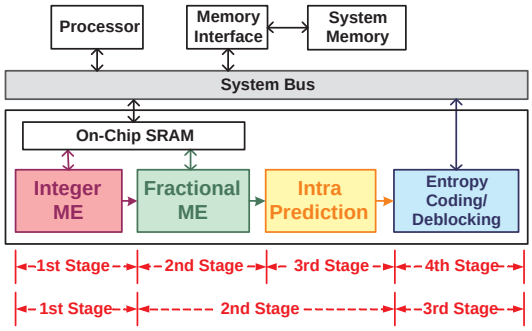

Figure 1: Conventional 3- or 4-stage macroblock pipelined architecture.

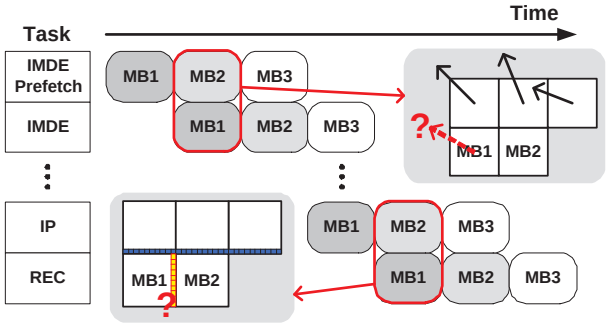

Figure 3: Conflict of data dependency occurs in the conventional MB pipelining.

there are only 350 cycles in an MB pipeline stage at the highest specification (4096×2160p/24fps/1 view@280MHz), where the conventional 3- or 4-stage MB pipelining [2, 3, 4] containing 600 to 1000 cycles in a pipeline stage is not feasible. In addition, if the conventional architectures directly scale up to support our target specification, a huge amount of on-chip SRAM area and external memory bandwidth are required.

The system architecture is shown in Fig. 2. The encoder contains 7 kinds of computation cores for integer ME/DE (IMDE) and fractional ME/DE (FMDE), intra prediction (IP), motion and disparity compensation (MDC), reconstruction (REC), entropy coding (EC), and deblocking filter (DB). 8-stage MB pipelining is proposed instead of simply raising the degree of parallelism. In the proposed system, the inter-prediction part is split into 5 MB pipeline stages, and the rest part is split into 3 ME pipeline stages. The cache-based prediction core is adopted as the inter-prediction part. The 2 prefetch stages for IMDE and FMDE not only reduce the burden of pipeline-cycle budget but also enhance the hardware utilization of IMDE and FMDE. The propose of NOP stage is introduced later. In the 6th MB pipeline stage, IP and MDC is performed and followed by the REC of an MB in the next pipeline stage. EC and DB are processed simultaneously in the 8th ME pipeline stage. To provide high symbol rate for detailed textured images, EC cores are doubled.

2.2 View-Parallel MB-Interleaved Scheduling

Directly increasing the number of MB pipeline stages causes conflict of data dependency and difficulties of resource sharing between computation cores. There are two critical issues, as shown in Fig. 3. Before beginning the prefetch stage, the initial guess of motion vectors (MVs) and disparity vectors (DVs) should be derived in advance. If the conventional MB pipelining is applied, IMDE for $MB1$ and IMDE prefetch for $MB2$ are performed simultaneously. Conflict of data dependency occurs because $MB2$ requires the MV predictors provided by $MB1$. Another data hazard occurs between the IP and REC pipeline stages. In H.264/AVC standard, if an MB is intra-coded, it is predicted by the reconstructed boundary pixels around each sub-block. Conflict of data de-

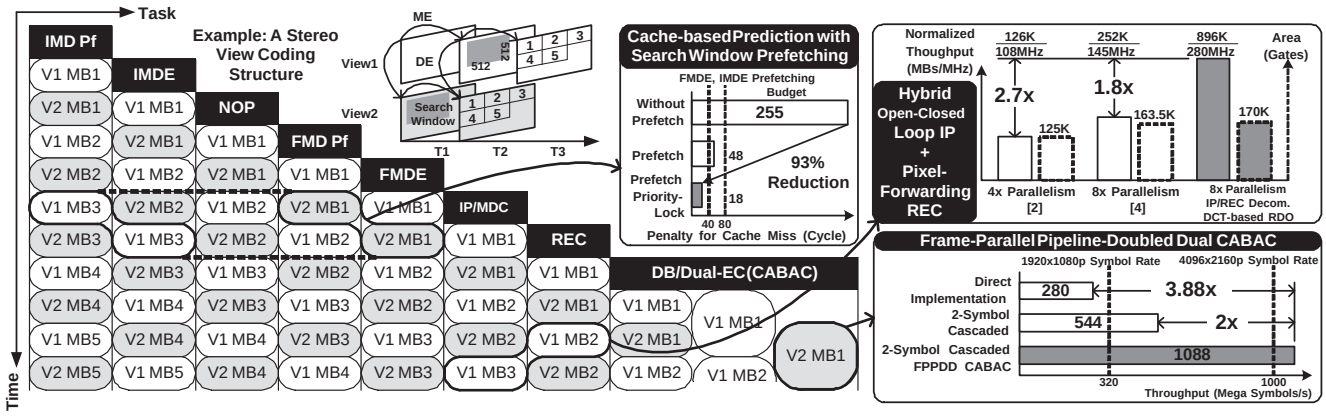

Figure 4: Features of view-parallel MB-interleaved scheduling.

pendency occurs when IP and REC are split into 2 pipeline stages. Therefore, view-parallel MB-interleaved (VPMBI) scheduling is proposed to overcome the above issues. With the VPMBI scheduling, the proposed system can process 9 MBs simultaneously without the above problems, so the throughput is enhanced to support 4096×2160p videos.

Fig. 4 shows the operation and features of the VPMBI scheduling. A stereo view video coding structure is taken for an example. In this case, 2 views are processed in parallel, and MBs are processed in an interleaving manner. Each capsule unit represents the cycle budget for an MB pipeline. VPMBI is characterized as follows: 1) Cache-based prediction with SW prefetching, which is composed of 5 pipeline stages, is proposed. SW prediction and prefetching are to lower cache miss rate. The purpose of inserting the NOP stage is to prevent IMDE and FMDE from fighting for the same cache reading/writing port. 2) Hybrid open-closed loop IP and pixel-forwarding REC are decomposed into 2 pipeline stages without any conflict of data hazard. Reconstructed pixels in neighboring MB boundaries are forwarded to IP and adopted as intra predictors, while intra predictors inside the current MB use original pixels instead of reconstructed pixels. DCT-based rate-distortion optimization (RDO) is also adopted to avoid quality degradation. The throughput of the proposed architecture is 1.8× and 2.7× better than previous works with similar silicon area [2][4]. 3) To achieve the symbol rate of 4096×2160p resolution, EC cores are doubled to perform frame-parallel pipeline-doubled dual (FPPDD) Context-Based Adaptive Binary Arithmetic Coding (CABAC). Each EC core encodes 2 symbols per cycle. The cycle budget of this pipeline stage is doubled, and 2 EC cores operate in a ping-pong manner to connect with REC stage. FPPDD CABAC provides 3.88 times of symbol rates over direct implementation so that it can meet the real-time requirement for encoding 4096×2160p resolution. The detailed architectures of these main modules are introduced in the next section.

3. MODULE ARCHITECTURE DESIGN

3.1 Cache-Based Temporal/Inter-view Prediction

The cache architecture for reference frames shown in Fig. 5 replaces traditional SW buffer. For better locality, the in-

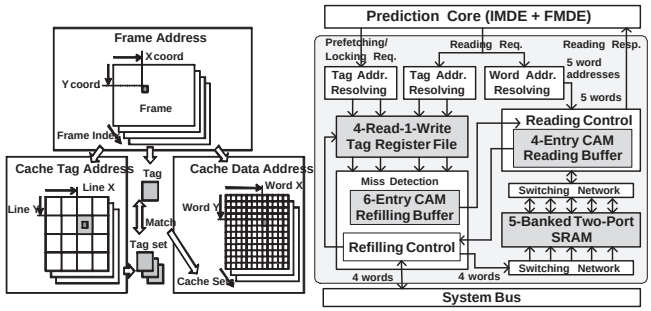

Figure 5: Address-resolving flow and hardware architecture of the cache controller.

ternal addressing in the cache keeps the intrinsic 2D nature of frames. The 3-tuple vector (x, y, frame-index) is translated to the tag address and the tag. A tag-set is located by the tag address, and the tag is compared to that set. Upon cache-hit, the word address locates the word in a 5-banked on-chip SRAM. The bank assignment is determined by the 3 constraints shown in the figure. In this 4-way non-blocking architecture, the control logic supports reading after up to 6 misses and concurrent reading and prefetching/locking. To meet the throughput of the prediction core, the proposed architecture supports sustained rate of matching 4 cache lines, reading 5 words, and refilling 4 words per cycle without cache line split penalty. With these techniques, the length of pipeline stage is shorter than 350 cycles. In the prediction core, the predictor-centered fast ME/DE algorithm is used. First, several predictors are classified into intra-frame and inter-frame predictors, including the 16×16 MVs of the left, top-left, top, and top-right MBs. They are from highly correlated sources of MVs like neighboring and the best matching MBs. These MV predictors are set as the refining centers and evaluated by sum of absolute difference (SAD) cost. Then a ±16×±16 searching range is used around the best predictor. The computation of the proposed algorithm is 3 orders lower than that of full search and 95% less than that of hierarchical search.

3.2 Hybrid Open-Closed Loop IP and Pixel-Forwarding REC

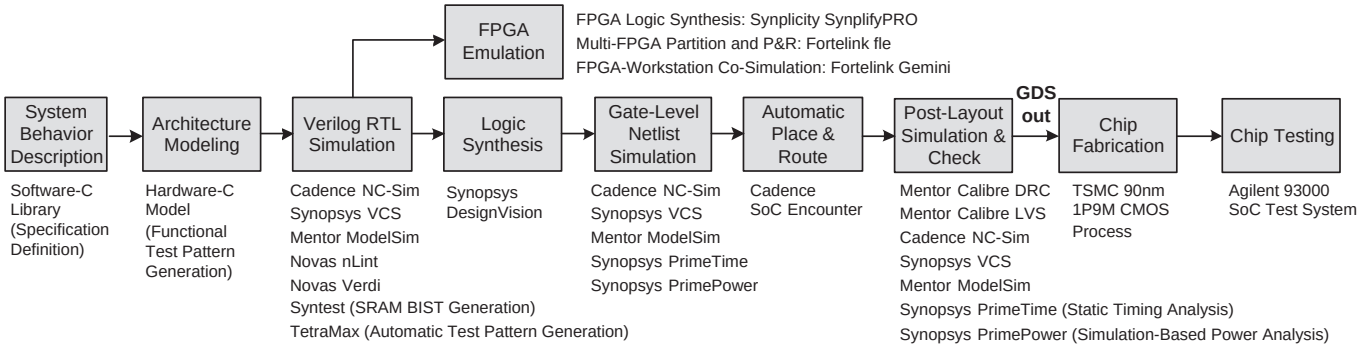

Figure 7: Design flow and EDA tools adopted the MVC encoder design process.

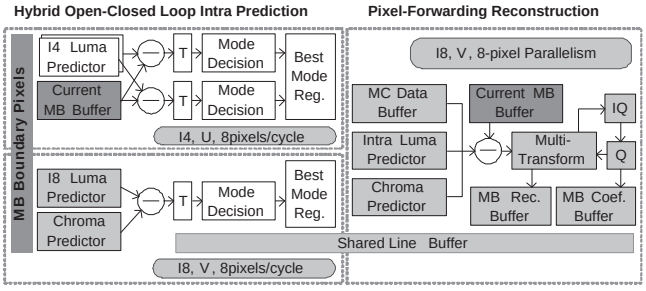

Figure 6: Architecture of IP and REC computation cores.

In order to improve the processing parallelism limited by data dependency in Sec. 2.2, the proposed hybrid open-closed loop IP use original pixels instead of reconstructed pixels for most intra predictors. This is because that original pixels are close to reconstructed pixels in our target high definition application which PSNR is always greater than 35dB. The architecture of hybrid open-closed loop intra prediction and pixel-forwarding reconstruction is shown in Fig. 6. In order to be consistent with the throughput of 8×8 DCT in Intra_8×8 prediction, the parallelism of our architecture is set to be 8-pixel parallel. Since the Intra_8×8 prediction mode is similar to Intra_4×4 prediction, a reconfigurable intra luma predictor generator is proposed that can generate 8 predictors for Intra_8×8 mode, or 8 predictors for two 4×4 sub blocks for Intra_4×4 mode. Besides, the multi-transform can be configured as two 4×4 Hadamard/DCT/IDCT or one 8×8 DCT/IDCT transform for cost estimation and reconstruction. The proposed hardware architecture improves processing capability with high area efficiency by using these reconfigurable 8-pixel parallel PEs.

3.3 Frame-Parallel Pipeline-Doubled Dual (FPPDD) CABAC

To achieve the symbol rate of 1000Msymbols/s for 4096×2160p videos, the binary arithmetic coder in CABAC is cascaded to become a 2-symbol architecture. Applying 2-symbol CABAC architecture can enhance the throughput to be doubled. However, for some textured MBs, 2-symbol CABAC architecture still does not meet the throughput requirement. Therefore, FPPDD CABAC is proposed. Dual CABAC computation cores are adopted, and each CABAC core has doubled pipeline-cycle budget of 700 cycles. Dual CABAC computation cores process in an interleaved manner to be compatible with the proposed VPMBI scheduling, so the MB scheduling is preformed smoothly without being stalled by the final CABAC stage. The throughput of the FPPDD CABAC architecture is 3.88× and 2× better than direct implementation and 2-symbol cascaded architectures, respectively, as shown in Fig. 4.

4. CHIP DESIGN FLOW AND VERIFICATION STRATEGIES

4.1 Design Flow

The design flow and the corresponding adopted EDA tools are illustrated in Fig. 7. The design flow covers from algorithm level, system level, RTL level, and the physical level. In the algorithm-level phase, the encoder specification is firstly defined, and some hardware-oriented algorithms are simulated and proposed with the self-developed software-C library.

In the system-level phase, the 8-stage MB-pipelining scheme with the corresponding processing flow is designed. To verify the proposed system architecture and scheduling, another hardware-C model are established. In addition, the functional test pattern is generated in this step for further RTL, gate-level, and post gate-level simulation.

In the RTL-level phase, the Verilog hardware description language (HDL) is adopted to represent our hardware design in the register transfer level (RTL). Novas nLint is used for HDL syntax checking. Cadence NC-Sim, Synopsys VCS, and Menter ModelSim are used to simulate the hardware behavior represented in RTL, and Novas Verdi served as debugging tool is used to observe the waveform extracted these simulation tools. After that, FPGA emulation can be started earlier for the functional verification of huge amount of test pattern.

In the physical-level phase, the logic synthesis is done by Synopsys DesignVision. The generated netlist are simulated again by the same tools adopted in the RTL level. In addition, Synopsys PrimeTime and PrimePower are used for static timing analysis and power simulation, respectively. After the gate-level HDL codes are fully verified, the physical layout is generated by Cadence SoC Encounter. The timing-driven placement and routing (P&R) are applied to improve

Technology	TSMC 90nm 1P9M CMOS
Supply Voltage	Core 1.2V, I/O 3.3V
Temperature	25℃
Core Area	3.95x2.90mm ^{2}
Logic Gate Count	1732K (2-input NAND gate)
On-chip SRAM	20.1KB
Encoding Features	H.264/AVC Multiview Extension/High Profile@Level5.1
View Scalability	4096x2160p for 1 view
	1920x1080p for up to 3 views
	1280x 720p for up to 7 views
Maximum Throughput	212Mpixels/s, 830kMB/s@280MHz
ME/DE Search Range H/V	[-256,+255]/[-256,+255]
Operating Frequency &	522mW@280MHz for 4096x2160p/24fps/Single view
Power Consumption	366mW@166MHz for 1920x1080p/30fps/Stereo views
	317mW@144MHz for 1280x 720p/30fps/Quad views
	148mW@181MHz for 1920x1080p/30fps/Single view
	58mW@ 36MHz for 1280x 720p/30fps/Single view

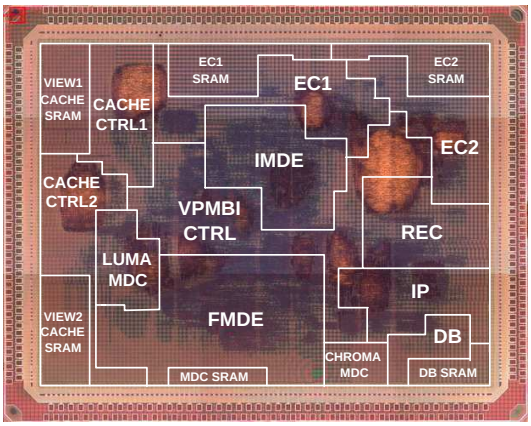

Figure 8: Measured chip features and the chip micrograph.

the P&R performance under the timing constraint. Clock tree synthesis and optimization are performed to minimized the clock skew for the highest target operating frequency, 280MHz. The GDS file and the corresponding post-layout gate-level HDL coded are then extracted. The chip layout is verified as follows. First, Mentor Calibre is used in verification of design rule checking (DRC) and layout versus schematics (LVS). Then, we use the same EDA simulation tools applied in the RTL and gate levels to perform post-layout gate-level simulation to confirm the timing after clock tree synthesis, gate sizing, and buffer insertion done by SoC Encounter. The chip is finally fabricated with TSMC 90nm 1P9M process.

To sum up, 18 EDA tools are utilized in the MVC encoder design process for simulation, verification, layout generation and so on. This chip is verified to ensure the validity of the functionality and the target operating frequency.

4.2 Testing Considerations

Design for testability (DfT) is critical for the enhancement of the testability and observability of fabricated chips, especially for a large-scale VLSI system design. 3 DfT techniques are utilized in the MVC encoder design. 1) SRAM build-in self-test (BIST) is implemented with March C algorithm and Syntest tool for 28 embedded SRAM modules. 2) Scan chain is inserted and combined with automatic test pattern generation (ATPG). The ATPG pattern is obtained by TetraMax tool. 3) An ad-hoc testing scheme is adopted to connect the I/O ports of main computation cores with the chip I/O. The memory mapped registers and other control registers of each cores can be directly written, and then the corresponding result can be monitored via the chip I/O. Therefore, the observability of 9 main computation cores, including IMDE, FMDE, IP, and so on, is enhanced.

In addition to the DfT techniques which aim to increase the observability of the chip, "near-pad logic" is implemented for making the chip fit in with the highest target operating frequency. Agilent 93000 mixed-signal SoC test system shown in Fig. 9 is used for chip testing. Due to the large

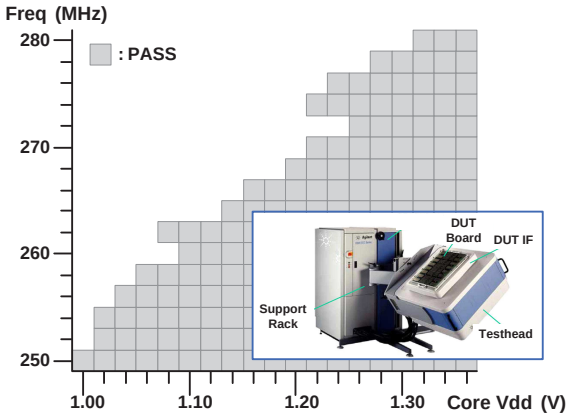

Figure 9: SHMOO plot generated by Agilent 9300 mixed-signal SoC test system.

output loading capacitance (∼40pF) of the DUT board in the tester, we place registers close to each output pad. The SHMOO plot which describes the functional correctness in terms of core supply voltage and operating frequency can be derived from the tester as illustrated in Fig. 9. The maximum measured frequency is 280MHz. As a result, the average available MB pipeline cycles is 354 cycles, which is sufficient for processing an MB in the highest specification.

5. EXPERIMENTAL RESULTS

5.1 Measured Chip Features

The measured chip features, specifications, and the chip macrograph are shown in Fig. 8. The core size of the chip is 11.46mm^2 (3.95mm×2.90mm), which contains 1732K gates using 90nm CMOS technology. This chip supports both H.264/AVC Multivew Extension and High Profile at Level 5.1. In addition, view scalability, which depends on the frame resolution, is supported for 1 to 7 views. This chip supports maximum throughput of 212Mpixels/s and 830kMB/s at 280MHz for 4096×2160p videos.

	This Work	ISSCC'05 [2]	ISSCC'07 [3]	ISSCC'08 [4]
Maximum Resolution	4096x2160@24fps	1280x720@30fps	1280x720@30fps	1920x1080@30fps
Maximum Throughput	212Mpixels/s	27.6Mpixels/s	27.6Mpixels/s	62.2Mpixels/s
H.264 Profile	Multiview/High@Level5.1	Baseline	Baseline	High@Level4
Search Range H/V	[-256,+255]/[-256,+255]	[-64,+63]/[-32,+31]	[-32,+31]/[-32,+31]	[-128,+127]/[-128,+127]
Quality Loss *	0.03 to 0.08dB	~0dB	<0.6dB	0.1dB
Technology	TSMC 90nm	UMC 0.18µm	TSMC 0.13µm	UMC 0.13µm
Core Size	3.95x2.90mm^2	7.68x4.13mm^2	4.30x4.30mm^2	3.17x3.17mm^2
Gate Count	1732K	923K	470K	593K
On-chip SRAM	20.1KB	34.7KB	13.3KB	22.0KB
Power Consumption	522mW@280MHz	785mW@108MHz	183mW@108MHz	242mW@145MHz

* Compared with full search algorithm

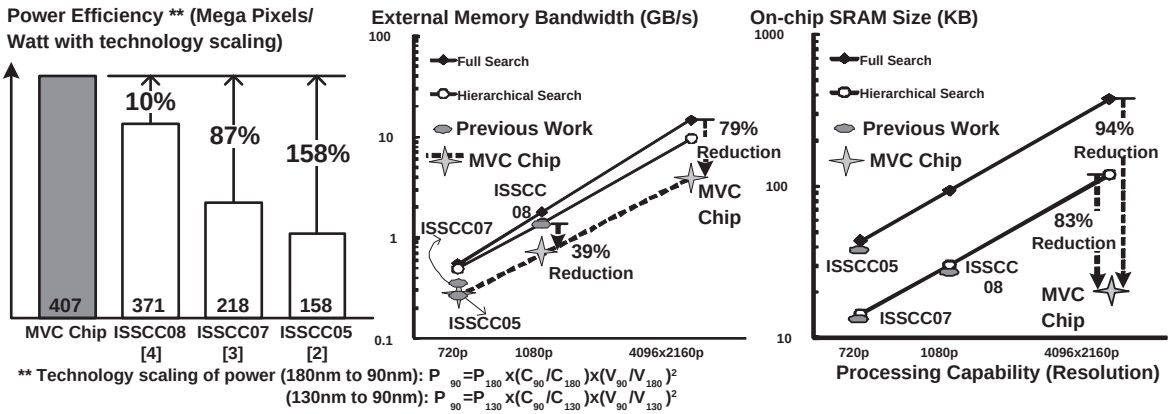

Figure 10: Comparison with the state-of-the-art encoder chips.

5.2 Architectural Comparison

Fig. 10 summarizes the the performance evaluation of the MVC encoder chip with the state-of-the-art encoder chips [2, 3, 4]. With the VPMBI scheduling and the 8-stage MB pipelining, our work provides 3.4× to 7.7× throughput better than the previous works and supports the maximum frame resolution. The search range of ME/DE is 4 to 64 times larger than the previous works while only 20.1KB on-chip SRAM is used with the penalty of only 0.1dB quality degradation. The power efficiency defined as mega pixels per Watt is compared. Note that the technology is scaled from 0.18μm and 0.13μm process to 90nm process. The MVC encoder chip provides the power efficiency 10% to 153% better than the previous works. In addition, external memory bandwidth and on-chip SRAM size among these works are evaluated. The external memory bandwidth and on-chip SRAM requirement for full search and hierarchical search algorithm are also illustrated. In the 3 kinds of HD resolution, the MVC chip requires the least external memory bandwidth and on-chip SRAM area. The proposed predictor-centered ME/DE algorithm is most suitable for the hardware implementation. Compared with [4], the proposed cache-based prediction core along with SW prefetching scheme reduces 39% external memory bandwidth. Also, 83% to 94% on-chip SRAM size is saved compared with the previous works scaled up to 4096×2160p resolution.

6. CONCLUSION AND FUTURE WORK

The proposed MVC single-chip encoder supports view scalability for encoding 1-view 4096×2160p, 3-view 1080p, and 7-view 720p videos for future 3DTV and quad HDTV applications. The 212Mpixels/s throughput is 3.4× to 7.7× higher than the state-of-the-art encoder chips. In addition, the highest power efficiency of 407Mpixels/Watt is achieved with the VPMBI scheduling, economic access of external system memory bandwidth, and highly parallelism to reduce internal memory access. 79% system memory bandwidth and 94% on-chip SRAM are saved with cache-based prediction core. The chip design is accomplished with the solid design flow and the robust testing and verification strategies.

Furthermore, the VPMBI scheduling can be regarded as a design methodology for ASIC-based HD video encoder. By allocating more view-cache SRAM in the design, the parallel-processing capability is enhanced. It also enables more efficient access of external system memory bandwidth for complex MVC prediction structures. They are challenging research topics and also belong to our future work.

7. REFERENCES

[1] Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, "Joint draft 7.0 on multiview video coding," Apr. 2008.

[2] Y.-W. Huang et al., "A 1.3TOPS H.264/AVC single-chip encoder for HDTV applications," in *ISSCC Dig. Tech. Papers*, Feb. 2005, pp. 128–129.

[3] H.-C. Chang et al., "A 7mW to 183mW dynamic quality-scalable H.264 video encoder chip," in *ISSCC Dig. Tech. Papers*, Feb. 2007, pp. 280–281.

[4] Y.-K. Lin et al., "A 242mW 10mm^2 1080p H.264/AVC High Profile encoder chip," in *ISSCC Dig. Tech. Papers*, Feb. 2008, pp. 314–315.