

7.1 A 1.3TOPS H.264/AVC Single-Chip Encoder for HDTV Applications

Yu-Wen Huang¹, Tung-Chien Chen¹, Chen-Han Tsai¹, Ching-Yeh Chen¹, To-Wei Chen¹, Chi-Shi Chen², Chun-Fu Shen³, Shyh-Yih Ma², Tu-Chih Wang⁴, Bing-Yu Hsieh⁵, Hung-Chi Fang¹, Liang-Gee Chen¹

¹National Taiwan University, Taipei, Taiwan

²Chip Implementation Center, Hsin-Chu, Taiwan

³Vivotek, Taipei, Taiwan

⁴Chin Fong, Chang Hua, Tawain

⁵Mediatek Incorporation, Hsin-Chu, Taiwan

An H.264/AVC encoder [1] saves 25% to 45% and 50% to 70% of bit rates when compared with MPEG4 and MPEG2, respectively. New features include $\frac{1}{4}$ -pixel motion estimation (ME) with multiple reference frames (MRF) and variable block sizes (VBS), intra prediction, context-based adaptive variable length coding, deblocking, rate-distortion optimized mode decision, etc. Applications range from high definition digital video disc (HD-DVD) to digital video broadcasting for handheld terminals (DVB-H).

There are four critical issues to be addressed. First, the encoding algorithm is extraordinarily complex, making conventional macroblock (MB) pipelining impractical. Second, the integer ME requires ultra high memory access and computational power. Third, the fractional ME requires extremely complicated control and sequential processing for VBS. Fourth, the intra-prediction involves a great diversity of modes, making resource sharing very challenging. All these operations require high operating frequency and memory bandwidth. The reference software, JM7.3, requires computing power of 3.6 tera-operations/s (TOPS) and memory access of 5.6 tera-bytes/s (TB/s) on a general-purpose processor to encode HDTV720p videos (1280×720, 30frames/s) in real time.

Efficient techniques that enable H.264/AVC baseline profile coding for HDTV applications are presented in this paper. Figure 7.1.1 shows the detailed chip features. The core size of the chip is 31.72mm² using 0.18μm CMOS technology. It contains 922.8K logic gates and 34.72KB SRAMs. Power dissipation is 785mW at 1.8V and 108MHz for HDTV720p videos.

Figure 7.1.2 shows the system architecture. The encoder contains five engines for integer motion estimation (IME), fractional motion estimation (FME), intra prediction (IP), entropy coding (EC), and deblocking (DB). The bandwidth requirements of HDTV720p videos are 40MB/s and 240MB/s for the system bus and local bus, respectively. The system is characterized by the proposed four-stage MB pipelining. Conventional two-stage MB pipelining, composed of a prediction engine and a block engine (MC+DCT/Q/IQ/IDCT+VLC), constrains the throughput and utilization for H.264/AVC encoders. The IME has the most severe computation and memory requirements. FME is 100 times more complex than that of prior standards due to MRF, VBS, $\frac{1}{4}$ -pixel accuracy, and more precise distortion evaluation, and it cannot be parallelized with IME for the same MB. IP is also very time-consuming. Moreover, it is difficult to perform IME, FME, and IP on the same circuits. Hence, the prediction engine is partitioned into three stages, and EC/DB is placed at the 4th stage. MB data propagate through IME, FME, IP, and EC/DB stages. Four MBs are simultaneously processed, and the throughput is thus roughly doubled. IP must also integrate forward/inverse transform/quantization because reconstructed neighboring pixels are necessary for generating predictors. Cycles of the four stages are balanced to achieve high utilization, and local data transfer between stages is used to reduce bus traffic.

Figure 7.1.3 shows the parallel IME architecture comprising eight processing element (PE) arrays and sum of absolute differ-

ences (SAD) trees to dramatically reduce memory access, which is the most critical issue of IME. The full search pattern is adopted in which each PE array and its corresponding SAD tree compute the SAD of a search point. While prior systolic arrays require many extra partial SAD registers, the tree structure supports VBS without any overhead. The reference pixel array acts as caches between PE arrays and SRAMs to reuse search area data not only in the horizontal direction for the eight horizontally consecutive search points but also in the vertical direction for candidates on adjacent rows. With snake scan of search points, 41×8 SADs (41 blocks of seven sizes per search point × 8 search points) are continuously generated in each cycle. This parallel configuration reduces 81% (10.82GB/s) of SRAM access. Subsequent to the SAD trees are 41 comparator trees. Each comparator tree finds the smallest SAD among the eight search points and updates the best motion vector for one block. In addition, reuse of overlapped search areas between two horizontally adjacent MBs is adopted with on-chip padding to save 86% (0.98GB/s) of local bus bandwidth. Last but not least, pixel truncation, subsampling, and adaptive moving window are applied to reduce the encoder complexity (from 3.6TOPS to 1.3TOPS for HDTV720p videos) without noticeable quality loss. Figure 7.1.4 summarizes the bandwidth reduction techniques.

Figure 7.1.5 shows the parallel FME architecture comprising nine 4×4-block processing units (PUs) per reference frame to thoroughly parallelize the rate-constrained mode decision with high utilization, which is the toughest challenge of FME. Sum of absolute transformed differences (SATD) and all side information are considered in the matching criterion. Therefore, sequential processing of different block modes is inevitable. After detailed analysis, the loops of MRF, 3×3 search points, and two 1-D Hadamard transforms are unrolled. Each PU contains two parallel 1-D transforms and transpose registers. The parallel 2-D transform has four times the throughput of the traditional sequential design but requires a similar gate count. The nine PUs calculate the SATDs of 3×3 search points in parallel. This configuration benefits from extensive sharing of interpolation, and 764K logic gates are saved. For larger blocks, a folding technique is applied to iteratively utilize the interpolation circuits and PUs. An efficient schedule is also arranged to reuse interpolated pixels of previous 4×4-blocks in the vertical direction, and 26% of the cycles are saved.

A reconfigurable intra predictor generator [2] is designed for resource sharing of all intra prediction modes, which is the most important issue of IP. Furthermore, partial distortion elimination terminates the intra mode decision prematurely when the partial intra cost is already larger than the minimum inter cost.

Figure 7.1.6 shows that the encoded video quality of this chip is competitive with that of JM7.3 using full search. With improved Lagrange multipliers, ours is even better at high bit rates.

Figure 7.1.7 shows the die micrograph of the H.264/AVC encoder. Parallel and pipeline techniques reduce the frequency and increase the utilization, while folding and reconfigurable techniques reduce the area. Full search quality is achieved with 1,200 times of speed-up in comparison with JM7.3 on a PC with a Pentium IV 3GHz CPU.

Acknowledgements:

The authors would like to thank Professors Shao-Yi Chien, Homer Chen, members of DSP/IC Design Lab, and Chip Implementation Center.

Reference:

[1] Joint Video Team of ISO/IEC MPEG & ITU-T VCEG, "Draft ITU-T Recommendation H.264 and Final Draft International Standard 14496-10 Advanced Video Coding," ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6, May, 2003.

[2] Y.-W. Huang, B.-Y. Hsieh, T.-C. Chen, and L.-G. Chen, "Analysis, Fast Algorithm, and VLSI Architecture Design for H.264/AVC Intra-Frame Coder," to appear, *IEEE Trans. Circuits Syst. Video Technol.*

Technology	UMC 0.18µm CMOS 1P6M
Supply Voltage	1.8V
Core Area	7.68x4.13mm ²
Logic Gates	922.8K (2-input NAND gate)
SRAMs	34.72KB
Encoding Features	All Baseline Profile Compression Tools
Max. Number of Ref. Frames	4
Max. Search Range (Ref. 0)	H[-64,+63] V[-32,+31]
Max. Search Range (Ref. 1-3)	H[-32,+31] V[-16,+15]
Operating Frequency	81MHz for D1 (720x480, 30Frames/s, 4 Ref. Frames, Max. Search Range) 108MHz for HDTV720p (1280x720, 30Frames/s, 1 Ref. Frame, Max. Search Range)
Power Consumption	581mW for D1 785mW for HDTV720p

Figure 7.1.1: Chip features.

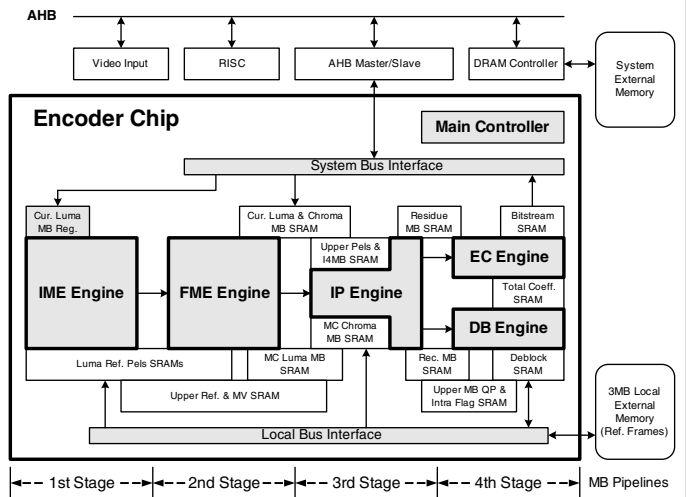


Figure 7.1.2: Block diagram of the H.264/AVC encoding system.

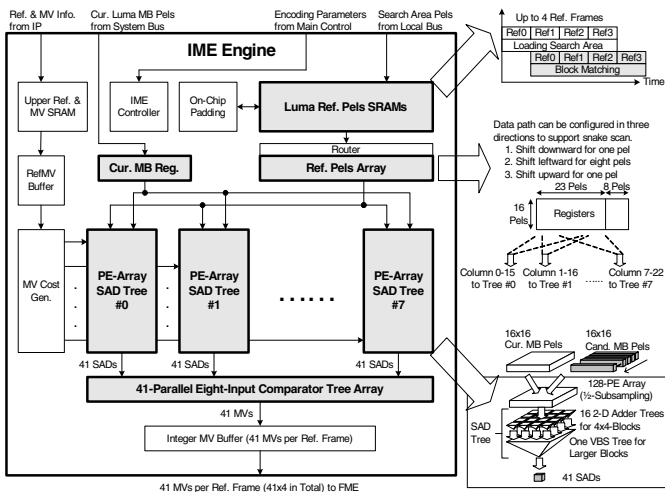


Figure 7.1.3: Block diagram of the IME engine.

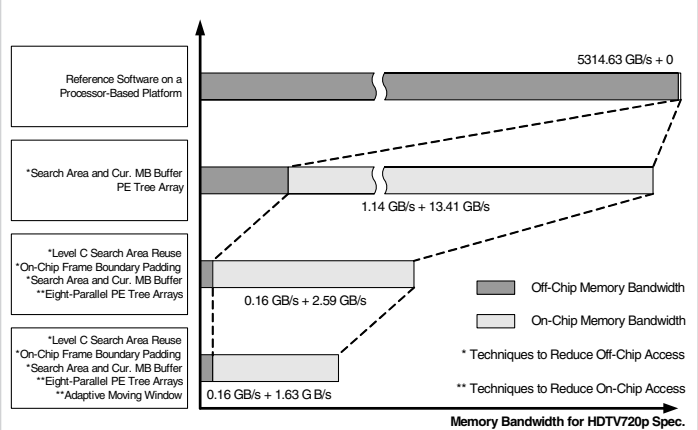


Figure 7.1.4: Bandwidth reduction techniques in the IME engine.

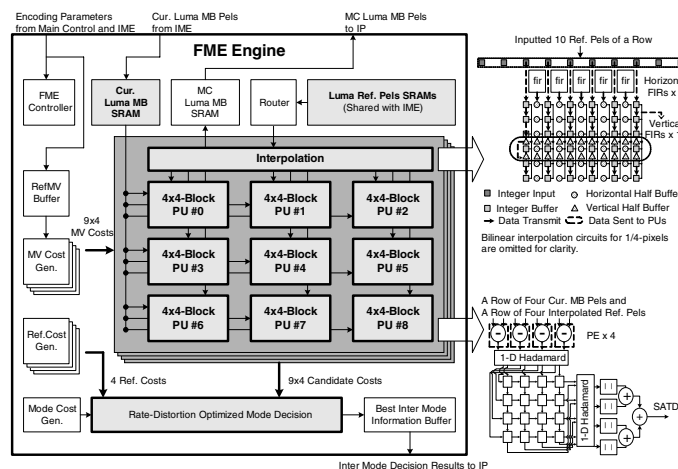


Figure 7.1.5: Block diagram of the FME engine.

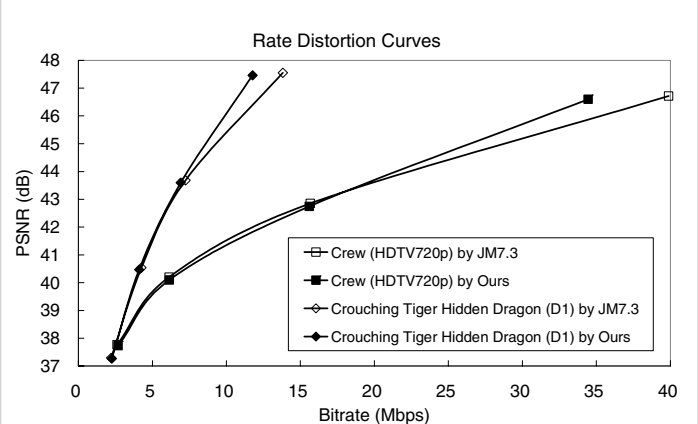


Figure 7.1.6: Comparison of coding quality with the reference software JM7.3.

Continued on Page 588

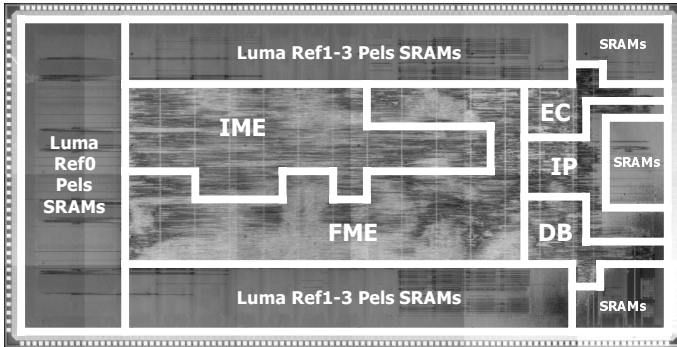


Figure 7.1.7: Die micrograph of the H.264/AVC encoder.