16.5 A 242mW 10mm² 1080p H.264/AVC High-Profile Encoder Chip

Yu-Kun Lin, De-Wei Li, Chia-Chun Lin, Tzu-Yun Kuo, Sian-Jin Wu, Wei-Cheng Tai, Wei-Cheng Chang, Tian-Shuean Chang

National Chiao-Tung University, Hsinchu, Taiwan

High-Profile H.264 has been adopted as the major coding standard in recently popular high definition video due to its excellent coding efficiency. Several implementations have been developed [1-3], but their performance is limited to baseline 720p [1,2] or SDTV [3]. The main stream 1080p high-profile application presents a series of new design challenges in throughput, cost, and power because of at least a 4x higher complexity than in the 720p baseline. Thus, a 0.13μm 1080p high-profile H.264 video encoder is presented with 10mm² core and 242mW power. Compared to a state-of-the-art 720p baseline design [2], this design achieves a 46.7% and 54% reduction in area and power, respectively. These savings are from parallelism enhanced throughput and a cross-stage sharing pipeline. The chip specification and features are summarized in Figure 16.5.1.

Figure 16.5.2 shows the system architecture of the proposed encoder with three macroblock pipelined stages. In the three pipelined stages, intra-prediction and fractional motion estimation (FME) are placed in the same stage to share the current block buffer and pipelined buffer. However, this could cause a timing conflict at the reconstruction of inter- and intra-prediction and thus reconstruction hardware has to be duplicated. The conflict is that the intra-predictor in the second stage needs the reconstructed boundary pixels of the previous block immediately for intra-mode decision and thus its reconstruction must finish in the second stage. However, the reconstruction after the final mode decision in different phase results in different reconstructed boundary data and thus will conflict with the intra- and inter-predictions which share the same hardware in different time slots. With the above reconstruction sharing technique, one extra reconstruction hardware unit and its associated power consumption can be eliminated.

The new 1080p high-profile coding tools, including intra 8x8 prediction and 8x8 integer DCT, increase the complexity by 37.5% and the throughput by 2.5x compared to the baseline profile. However, structural and data hazards will occur since new high-profile tools such as the intra 8x8 predictor needs extra reconstruction and different reconstructed boundary data and thus will conflict with the intra 4x4 modules. An independent boundary buffer is adopted for intra 8x8 prediction to eliminate the data hazard. To solve the high throughput demands and structural hazards, this design, except for ME and entropy coding, adopts eight-pixel parallelism, as shown in Fig. 16.5.3. Throughput is further improved by parallel processing intra 8x8 and intra 4x4/16x16, and interlaced scheduling is used to minimize the stall cycles introduced by data hazards. However, a direct implementation will incur high cost due to eight-pixel parallelism. To reduce cost, the reconstruction of intra 4x4/16x16 and intra 8x8 are merged into one and shared with the reconstruction of ME, as stated above. To further decrease the cost, intra 8x8/16x16 reconstruction is adopted so that the best mode result and its prediction value of intra 8x8/16x16 are not saved and recomputed if chosen. With this technique, memory can be reduced by 2560b. Using an eight-pixel architecture in the reconstruction and quantization phases can save the extra buffers used between different pixel-parallelism phases. In summary, these techniques can reduce gate count by 42% relative to a direct implementation.

Figure 16.5.4 shows the motion estimation (ME) algorithm and its architecture. To achieve 1080p resolution with bi-directional prediction, parallel single-step processing is adopted for ME. Thus, for integer motion estimation (IME), a paralleled subsampling algorithm known as, Parallel Multi-resolution ME (PMRME) [4] is used. It searches three subsampling levels of different search ranges in parallel so that all searches are done within 256 cycles in a single step. This provides higher throughput than both [2] and [3] using two and four steps in IME, respectively. With 256 cycles, the high-throughput IME can process bi-directional predictions sequentially while still meeting 1080p requirements. Thus, a single IME module is enough for both directional predictions. Besides, to support a search range (SR) of ±128 and to reduce the hardware cost within limited quality loss, level 1 and 2 provide different subsampling ratios according to the search range for large motion vectors, and the search centers of the two levels are at (0,0) for further data reuse. Furthermore, to compensate for quality loss due to subsampling to meet the 1080p requirement, the level 0 without subsampling is centered at the Motion Vector Predictor (MVP) to cover the most frequent motion vectors. After IME, we use Mode Filtering (MF) to select only the two best modes for FME refinement so that FME tests at most 18 motion vectors instead of 41 motion vectors in [1]. To further reduce the bandwidth of FME, we use the Non-Subsampling Reference Memory Sharing (NSRMS), another cross-stage technique, which uses three SRAM banks to enable sharing of level 0 reference memory of IME and that of FME. FME searches only six candidates in a single step so that only six processing units are needed, which eliminates 76% of the processing units used in [2].

Figure 16.5.5 shows the reduction of complexity and hardware cost of ME. The proposed algorithms can achieve a 98.7% complexity reduction and increased throughput compared to the full search [1] with only 0.1dB quality loss. In addition, the memory storage size can be reduced by 88.7% by subsampling and external memory access can be reduced by 46% through NSRMS, data reuse and subsampling. Finally, the proposed IME and FME modules can save 30% and 62.8% of the gate count, respectively, compared with [5] (an extension of [1]).

Figure 16.5.6 presents a summary and comparison with other works. The present design can support 1080p resolution, high-profile encoding up to ±128 search range in only 256 cycles with only 2.5x overhead. The power consumption is only 116mW and 242mW for 720p and 1080p high-profile, and the operating frequency is 62.5MHz and 145MHz for 720p and 1080p, respectively. Compared to the state-of-the-art design [2] targeted at 720p baseline in low power mode, the power consumption and operating frequency are 53.4% and 13% lower than [2], due to our enhanced throughput and complexity reduction. The PSNR loss is only 0.1dB, which is better than [2] with its 0.6dB quality loss in low power mode. The core area is only 54% of [2]. For low resolution video, level 1 and level 2 of IME are turned off when the video size is below CIF, and level 2 is off for D1 video size. Therefore, the power consumption is 6.4mW, only 42.6% of [3] and comparable to the lower power mode in [2] for baseline CIF video. Figure 16.5.7 shows the chip micrograph.

Acknowledgements:
This work is supported by National Science Council and SeC Research Center of NCTU. We also thank UMC University Program for chip fabrication and Chip Implementation Center for chip testing facilities.

References:
**Name**  
H.264/AVC High Profile@ Level 4 Encoder

**Process**  
UMC 0.13 μm 1P8M Standard CMOS
1.2V core, 3.3V I/O

**Package**  
CQFP 208-pin

**Gate Count**  
593K

**Internal Memory**  
22KB

**Chip Size**  
3.76x3.76mm$^2$

**Core Size**  
3.17x3.17mm$^2$

**Maximum Processing Throughput**  
62.208 pixels/sec @ 145MHz

**Operating Frequency**  
145MHz@ 1080p/30fps
28.5MHz@ D1/30fps
7.2MHz CIF/30fps
1.8MHz QCIF/30fps

**Core Power consumption**

- Baseline Profile:  
  - 176.1mW@ 1080p/30fps/1.2V
  - 84.6mW@ 720p/30fps/1.2V
  - 23.61mW@ D1/30fps/1.2V
  - 6.74mW@ CIF/30fps/0.9V
  - 2.92mW@ QCIF/30fps/0.9V

- High Profile:  
  - 242.01mW@ 1080p/30fps/1.2V
  - 116.61mW@ 720p/30fps/1.2V
  - 28.5MHz@ CIF
  - 62.5MHz@ 720p

**Search Range Buffer Size**

- Full Search@ SR=±128  
  - 45.68
  - 0.31
  - 0.05

- RS:  
  - 45.68
  - 0.31
  - 0.05

**Overall Throughput (Mpixels/sec)**

- Full Search@ SR=±128  
  - 45.68
  - 0.31
  - 0.05

- RS:  
  - 45.68
  - 0.31
  - 0.05

**High Quality Loss (dB)**

- Full Search@ SR=±128  
  - 15.9mW@ CIF/1.3V
  - 13.5MHz@ CIF
  - V: -16~+15
  - H: -32~+31

- RS:  
  - 15.9mW@ CIF/1.3V
  - 13.5MHz@ CIF
  - V: -16~+15
  - H: -32~+31

**Throughput (MHz)**

- Full Search@ SR=±128  
  - 15.9mW@ CIF/1.3V
  - 13.5MHz@ CIF
  - V: -16~+15
  - H: -32~+31

- RS:  
  - 15.9mW@ CIF/1.3V
  - 13.5MHz@ CIF
  - V: -16~+15
  - H: -32~+31

**Work at 1.2V**

- Full Search@ SR=±128  
  - 15.9mW@ CIF/1.3V
  - 13.5MHz@ CIF
  - V: -16~+15
  - H: -32~+31

- RS:  
  - 15.9mW@ CIF/1.3V
  - 13.5MHz@ CIF
  - V: -16~+15
  - H: -32~+31

**Filtering+**

- Full Search@ SR=±128  
  - 15.9mW@ CIF/1.3V
  - 13.5MHz@ CIF
  - V: -16~+15
  - H: -32~+31

- RS:  
  - 15.9mW@ CIF/1.3V
  - 13.5MHz@ CIF
  - V: -16~+15
  - H: -32~+31

**Filtering-**

- Full Search@ SR=±128  
  - 15.9mW@ CIF/1.3V
  - 13.5MHz@ CIF
  - V: -16~+15
  - H: -32~+31

- RS:  
  - 15.9mW@ CIF/1.3V
  - 13.5MHz@ CIF
  - V: -16~+15
  - H: -32~+31

**Power Comparison**

- Full Search@ SR=±128  
  - 15.9mW@ CIF/1.3V
  - 13.5MHz@ CIF
  - V: -16~+15
  - H: -32~+31

- RS:  
  - 15.9mW@ CIF/1.3V
  - 13.5MHz@ CIF
  - V: -16~+15
  - H: -32~+31

Continued on Page 615
Figure 16.5.7: The chip micrograph.