

22.6 A 5mW MPEG4 SP Encoder with 2D Bandwidth-Sharing Motion Estimation for Mobile Applications

Chia-Ping Lin¹, Po-Chih Tseng^{1,2}, Yao-Ting Chiu¹, Siou-Shen Lin^{1,3}, Chih-Chi Cheng¹, Hung-Chi Fang¹, Wei-Min Chao^{1,4}, Liang-Gee Chen¹

¹National Taiwan University, Taipei, Taiwan

²NovaTek, Hsinchu, Taiwan

³MediaTek, Hsinchu, Taiwan

⁴Quanta Computer, TaoYuan, Taiwan

The MPEG-4 standard has been widely adopted for video compression on mobile devices. MPEG-4 codec designs [1-2] have been reported that address the low power requirements demanded by mobile devices. In this paper, a 5mW MPEG-4 SP encoder is presented with good encoding performance suitable for mobile applications.

Three sources consume most of the power in an MPEG-4 encoder. First, motion estimation (ME) consumes more than a half of the total power, in general, because of its high memory access requirements. Secondly, the discrete cosine transform/inverse discrete cosine transform (DCT/IDCT) consumes power because of complex computations. Thirdly, data buffering between motion estimation/motion compensation (ME/MC) and quantization/variable length code (Q/VLC) consumes power because of the SRAM accesses. Based on these three different power-consuming activities, different methodologies are applied to reduce the power consumption.

This chip supports MPEG-4 SP encoding and contains a rate control (RC) circuit. The core size is 7.7mm² using 0.18μm CMOS technology and contains 201k logic gates and a 4.56kB SRAM. The power consumption is 5mW at 1.3V and 9.5MHz for CIF 30fps or 18mW at 1.4V and 28.5MHz for VGA 30fps. The detailed chip features are shown in Fig. 22.6.1.

The system architecture is shown in Fig. 22.6.2. At the module level, the design focuses on ME and DCT designs to reduce power consumption. At the system level, the design reduces the amount of data buffering between Q/VLC.

Power analysis shows that memory accesses consume over 75% of ME total power. Therefore, minimizing the ME bandwidth is central to reducing the power.

For good bandwidth sharing and a reasonable number of search candidates, a predictive 16×16/16×8 adaptive moving search window method is proposed. The searching starts within the 16×4 search window around the predicted point. And then the next three 16×4 search windows are searched far from the initial point to complete a 16×16 search window. According to the search results of the first two 16×4 search windows, the next two 16×4 search windows may be skipped. Because of the 2D data-sharing scheme within each 16×4 search window and the adaptive search window, the bandwidth is only 0.65% compared with full search without data sharing. The bandwidth reduction is shown in Fig. 22.6.3. The encoding performance degrades less than 0.1dB, on average, as compared with a full search.

Figure 22.6.4 shows the proposed ME architecture. It supports 2D data sharing within a 16×4 search window. For integer motion estimation (IME), it loads 31×19 pixels from reference SRAMs to handle 16×4 candidates and reduces the bandwidth to 3.6% comparing to that without data sharing. It contains 16 processing units (PUs). Each PU contains 16 processing elements (PEs) followed by an adder tree to calculate a sum of absolute difference (SAD) of 16 pixels in a row. In one cycle, 16 PUs calculate 16 row SADs of successive candidates in the horizontal direction. This operation needs to read 31 pixels from SRAM and reduces the bandwidth to 12.1% because of data sharing in the horizontal direction. Furthermore, the vertical bandwidth sharing mechanism is achieved in the time domain. 31 pixels read from SRAM are kept for four cycles with registers. During this period, data from the current macroblock (MB)

changes. With data of different rows of the current MB, the same data reference is reused for calculating SADs of different candidates in the vertical direction. The clock associated with registers storing partial SADs is controlled to save power and multiplexer area. With advanced data sharing in the vertical direction, bandwidth requirements are reduced to 3.6% of the initial value.

For fractional motion estimation (FME), the same data sharing mechanism is also applied. The PUs and SAD registers are reused to calculate eight candidates of FME. A 3× data sharing of interpolated data in the horizontal direction is adopted. Data sharing in the vertical direction is achieved by the same mechanism as IME. With a row-based half-pixel interpolation module, data requirements of the reference for FME is only 18×18 pixels. (This is also the minimum requirement.)

The decimation mode for ME is also implemented for low-power. In decimation mode, MB size will be decimated to 16×8 for MB matching. With an interleave pattern in the search area, the bandwidth requirement and power consumption of IME becomes half of the normal mode. The quality of this mode degrades 0.3dB, on average, as compared with the normal mode.

Most DCT coefficients become zero after quantization, so the precision of these coefficients is less important. These can be calculated with less precision to save power, and ideally little drop in quality. A DCT design is adopted that depends on the content to decide the required precision. It consumes less power for lower-precision calculations reducing the total power consumption. For CIF 30fps at 1.8V, the power consumption of a P frame is 1.71mW (1.38mW) in high (low) bit rate, which is a 21% (36%) reduction.

The architecture of the DCT [3] is shown in Fig. 22.6.5. A classifier circuit decides the allocation of calculation resources. It is based on the value of the pixel-to-pixel amplitude (PPA) and the quantization parameter (QP). After classification, the number of calculation bits is decided. Both clock and combinational circuits are shut down for any unused additional bits. The quality degradation due to reduced precision is less than 0.1dB compared with a normal DCT.

A zero marker scheme is adopted to reduce the data access of the SRAM buffer between stages. The buffered data for VLC is quantized, and they are mostly zero. For every four entities stored in SRAM, a one bit register is used to record if they are all zeros. If this occurs, no reading and writing is required. This mechanism avoids most buffer accesses between the Q stage and VLC stage. It can save 86% of data buffering in low bit rate and 62% in high bit rate mode depending on the sequences.

Figure 22.6.6 shows the rate-distortion (R-D) curve of proposed MPEG-4 SP encoder in encoding the Stefan sequence in CIF resolution. On average, the quality drop is 0.05dB comparing with full search and lossless DCT. The power consumption can be reduced to 4.3mW in decimation mode with 0.3dB quality drop.

Figure 22.6.7 is the micrograph of the MPEG-4 SP encoder chip. The 2D-bandwidth-sharing ME and content-aware DCT enable this encoder to reduce the power consumption according to the content of the video source, while maintaining good encoding performance.

Acknowledgements:

The authors thank Prof. Shao-Yi Chien and all members of DSP/IC Design Lab. The authors also acknowledge Chip Implementation Center for supporting fabrication.

References:

- [1] H. Yamauchi et al., "An 81MHz, 1280 × 720pixels × 30frames/s MPEG-4 Video/Audio Codec Processor," *ISSCC Dig. Tech. Papers*, pp. 130-132, Feb. 2005.
- [2] T. Hagiya et al., "A Low-Power MPEG-4 Codec IP Macro for VGA-Video Applications," *Proc. COOL Chips VII (An International Symposium on Low-Power and High-Speed Chips)*, pp. 101-114, Apr. 2004.
- [3] C.-P. Lin et al., "Nearly Lossless Content-Dependent Low-Power DCT Design for Mobile Video Applications," *ICME Multimedia and Expo Dig.*, pp. 1238-1241, July 2005.

Technology	TSMC 0.18 μ m 1P6M CMOS
Supply Voltage	1.8V (Core) / 3.3V (I/O)
Core Area	1.78 x 1.77mm ²
Logic Gates	201K (2-input NAND gate)
SRAMs	4.56kB
Encoding Feature	MPEG-4 SP
Search Range	H[-16,+15.5] V[-16,+15.5]
Operating Frequency	9.5MHz for CIF 28.5MHz for VGA
Power Consumption	5mW (Encoding of CIF at 9.5MHz 1.3V) 18mW (Encoding of VGA at 28.5MHz 1.4V)

Figure 22.6.1: Chip features.

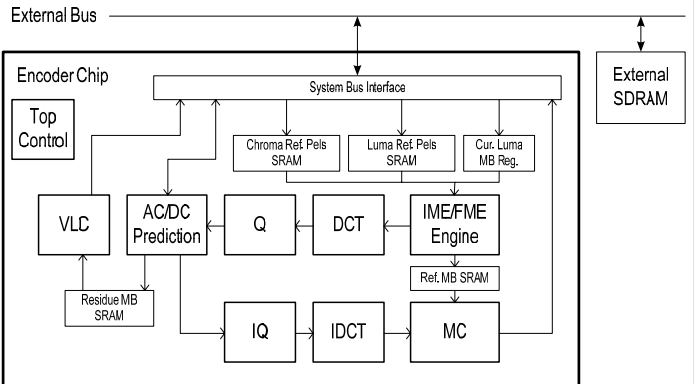


Figure 22.6.2: Block diagram of MPEG-4 SP encoder.

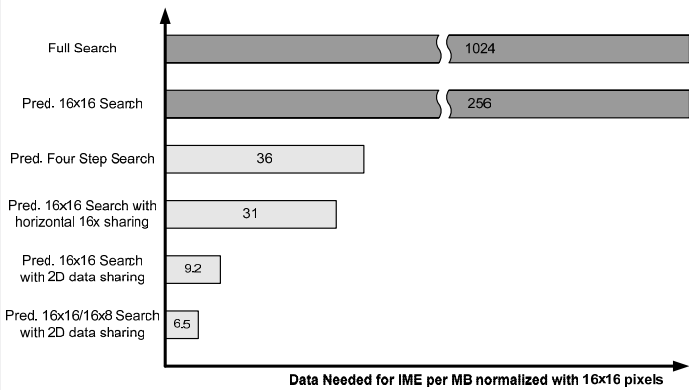


Figure 22.6.3: IME bandwidth reduction.

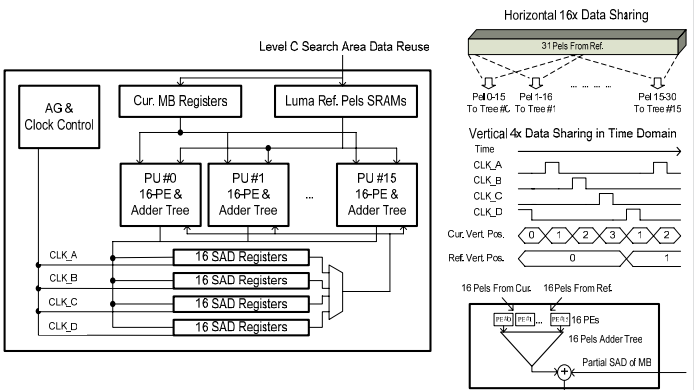


Figure 22.6.4: ME architecture.

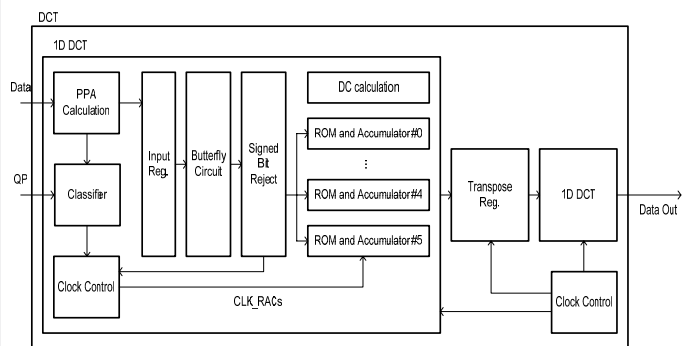


Figure 22.6.5: DCT architecture.

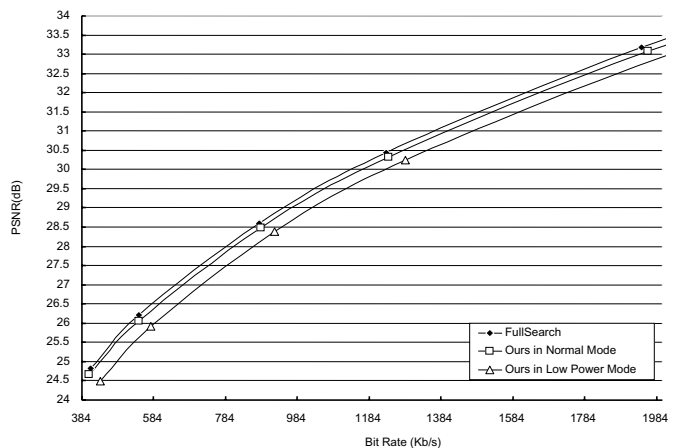


Figure 22.6.6: Performance in Stefan CIF 30fps encoding.

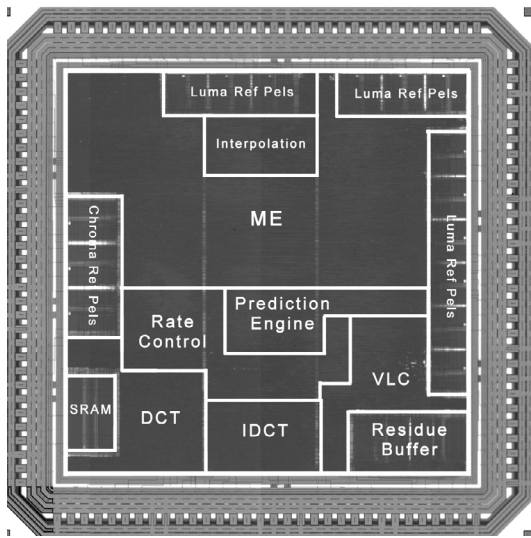


Figure 22.6.7: Die micrograph.

Technology	TSMC 0.18 μ m 1P6M CMOS
Supply Voltage	1.8V (Core) / 3.3V (I/O)
Core Area	1.78 x 1.77mm ²
Logic Gates	201K (2-input NAND gate)
SRAMs	4.56kB
Encoding Feature	MPEG-4 SP
Search Range	H[-16,+15.5] V[-16,+15.5]
Operating Frequency	9.5MHz for CIF 28.5MHz for VGA
Power Consumption	5mW (Encoding of CIF at 9.5MHz 1.3V) 18mW (Encoding of VGA at 28.5MHz 1.4V)

Figure 22.6.1: Chip features.

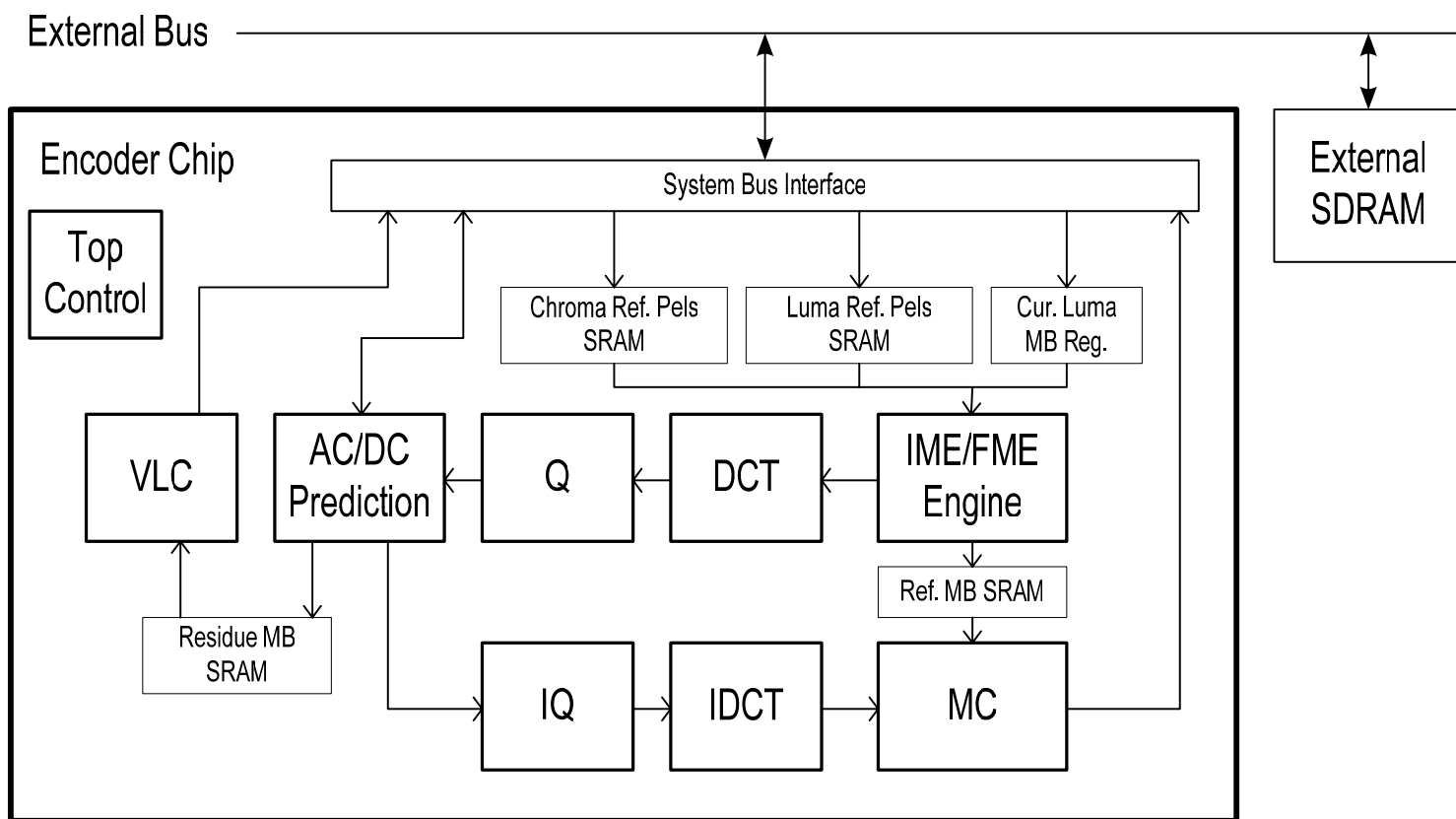


Figure 22.6.2: Block diagram of MPEG-4 SP encoder.

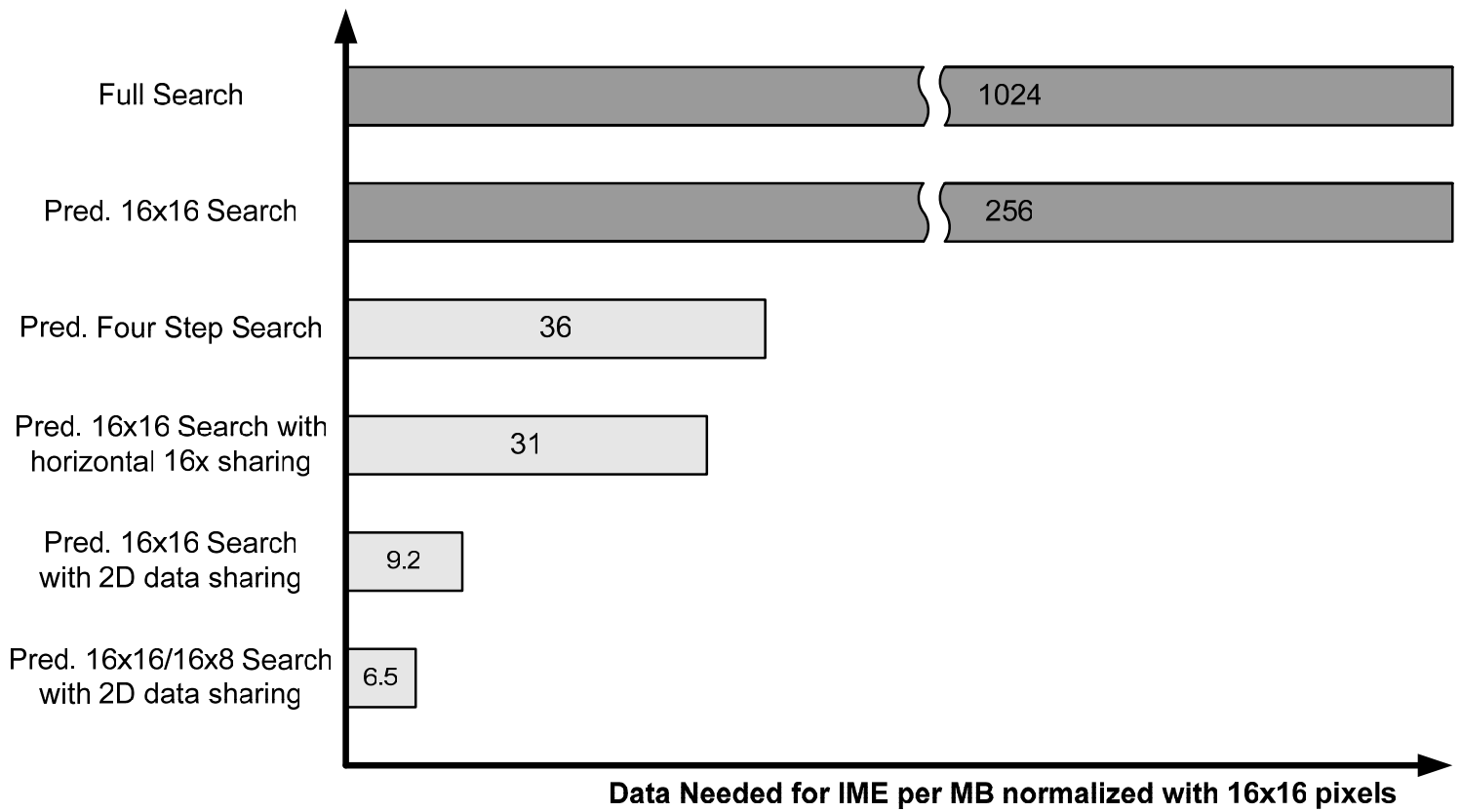


Figure 22.6.3: IME bandwidth reduction.

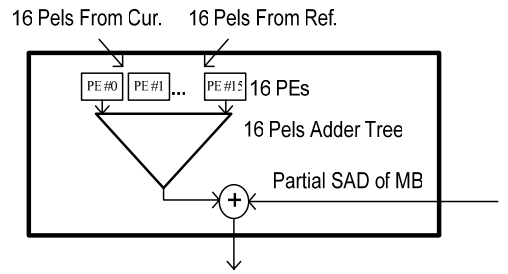
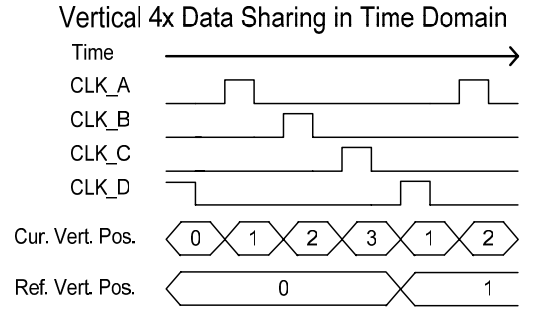
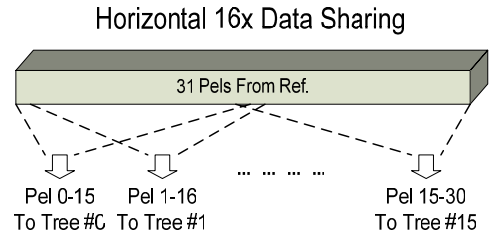
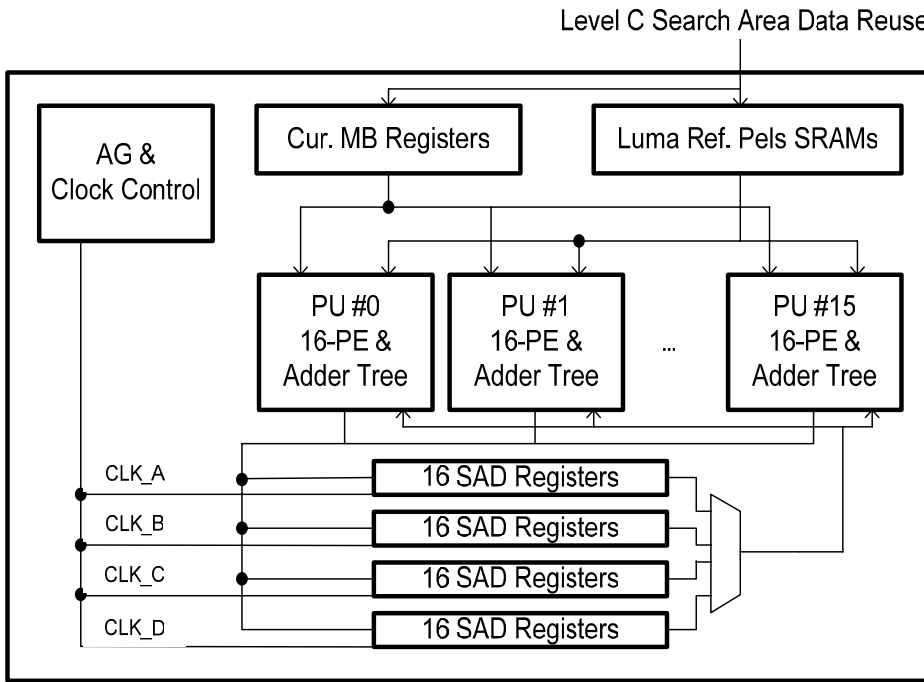


Figure 22.6.4: ME architecture.

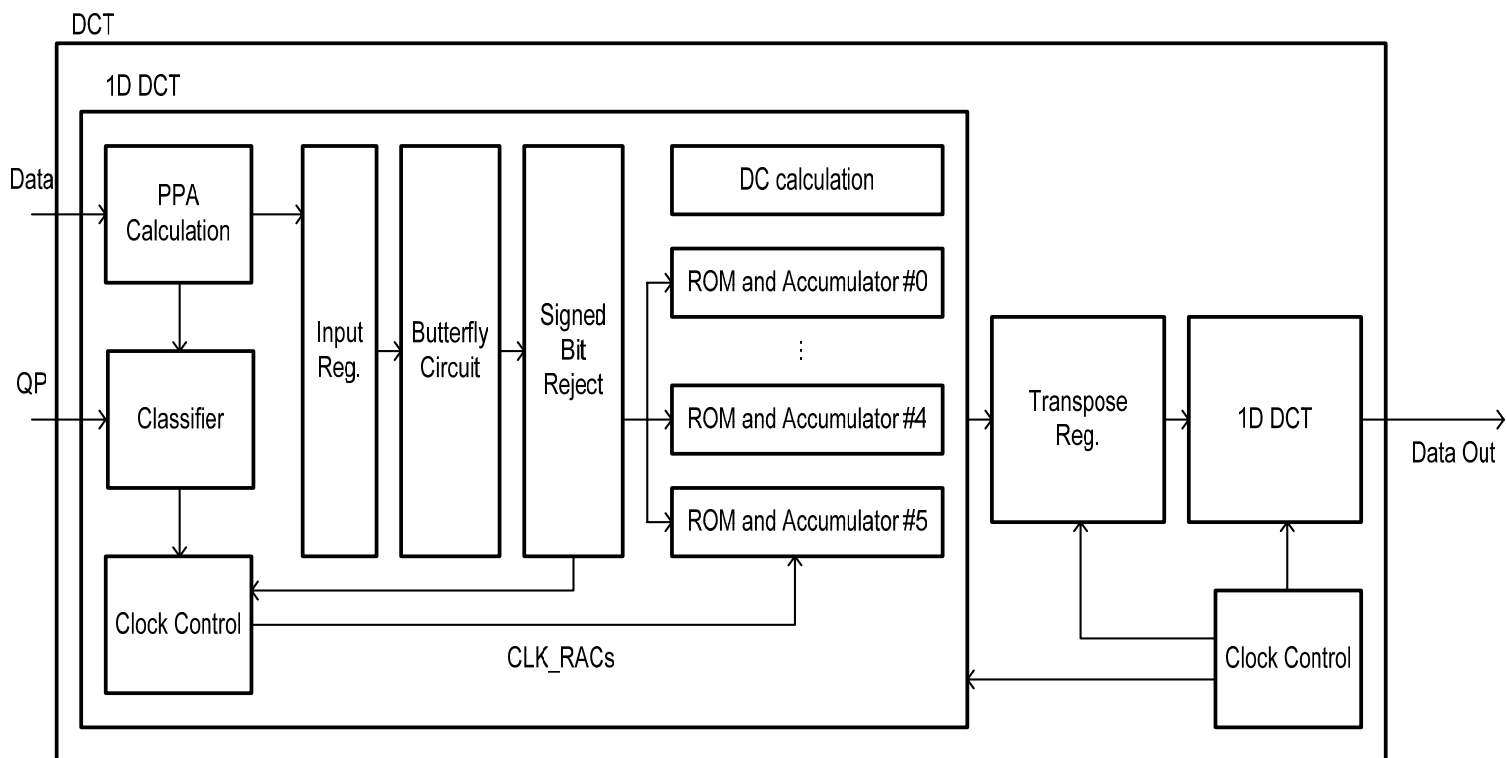


Figure 22.6.5: DCT architecture.

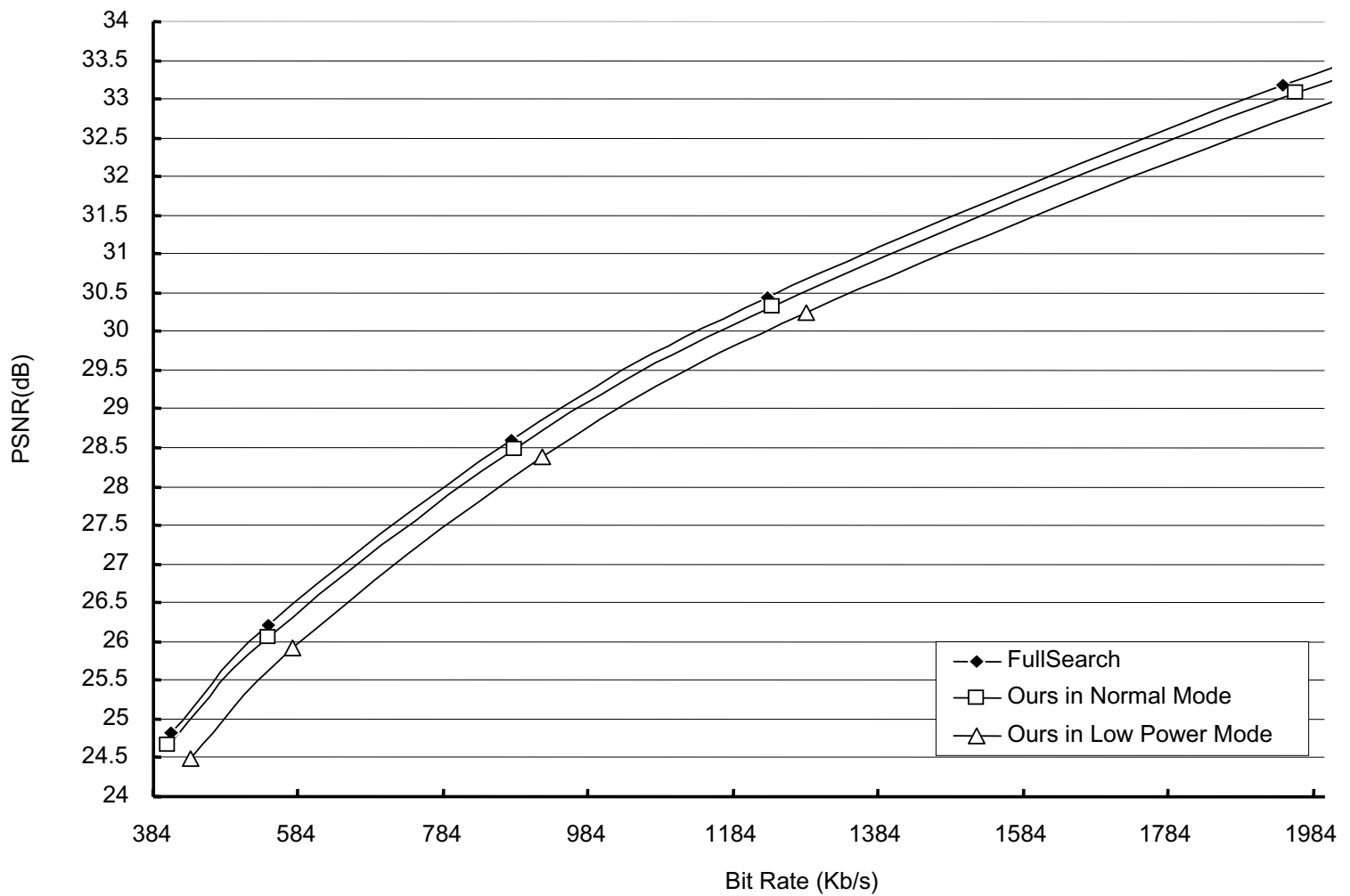


Figure 22.6.6: Performance in Stefan CIF 30fps encoding.

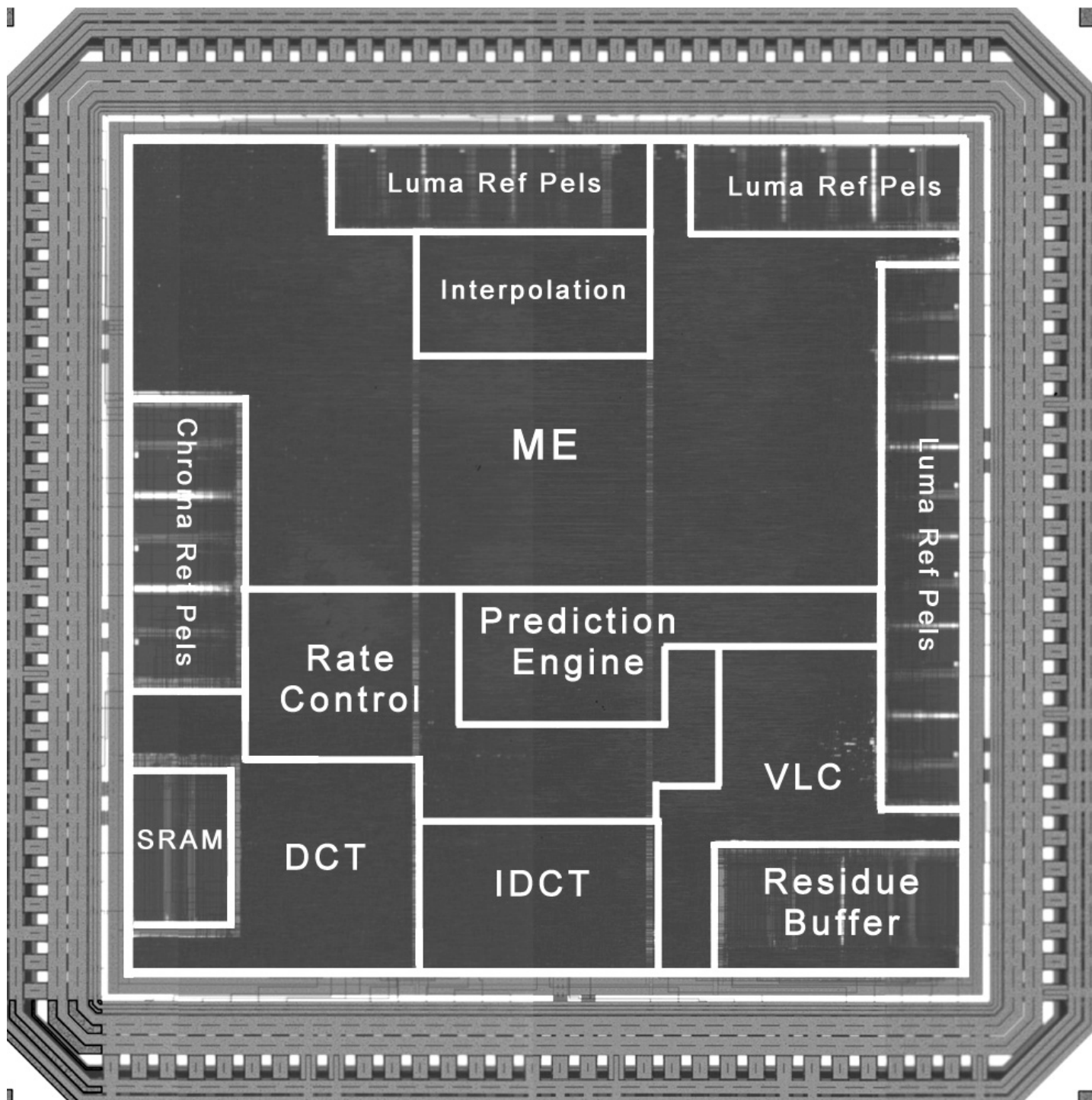


Figure 22.6.7: Die micrograph.