

## 18.2 81MS/s JPEG2000 Single-Chip Encoder with Rate-Distortion Optimization

Hung-Chi Fang, Chao-Tsung Huang, Yu-Wei Chang, Tu-Chih Wang, Po-Chih Tseng, Chung-Jr Lian, Liang-Gee Chen

National Taiwan University, Taipei, Taiwan

JPEG 2000 [1] is well known for its excellent coding efficiency as well as numerous useful features such as region of interest (ROI) coding and various types of scalability. Unlike JPEG, JPEG 2000 uses the Discrete Wavelet Transform (DWT) as the transformation algorithm and Embedded Block Coding with Optimized Truncation (EBCOT) as the entropy-coding algorithm. EBCOT produces finely embedded bit streams that enable post-compression Rate-Distortion (R-D) optimization.

There are three critical issues to design a high throughput JPEG 2000 encoder. First, the DWT requires high memory bandwidth and enormous computational power. Second, the EBCOT requires extremely complicated control and sequential processing. Third, R-D optimization requires a large memory for storing the lossless code-stream and R-D information. All of the above require high operating frequency, huge memory size, and high memory bandwidth for chip implementation.

In this paper, efficient techniques that enable JPEG 2000 compression at 81MS/s with R-D optimization are used to design an encoder chip. The processor is 5.5mm<sup>2</sup> using 0.25μm CMOS technology, and contains 163k gates and 7kb of SRAM. The processor consumes 348mW at 2.8V when operating at 81MHz. The detailed chip features are shown in Fig. 18.2.1.

The block diagram of the developed encoder chip is shown in Fig. 18.2.2. The encoder consists of a main controller, a 2-level DWT module, a pre-compression R-D optimization controller, a parallel EBCOT module, and a dedicated Bit Stream Formatter (BSF). The input format can be either original or color-transformed raw pixel data and the output is the JPEG 2000 code-stream. Two 24kB off-chip SRAMs are required for tile-level pipelining.

Memory bandwidth analysis indicates that the minimization of external memory access is the most critical problem. A line-based architecture, as opposed to the block-based approach in [2], is used to prevent data retransmission from external memory. The line buffers can be classified into two categories: the data buffer and the temporal buffer. The data buffer, which requires 1.5 lines of pixel data, stores the intermediate decomposition coefficients after the 1-D row DWT. The coefficients are then read by the 1-D column DWT to produce the 2-D results. The temporal buffer stores the intermediate data for the 1-D column DWT module, which requires 2 lines of pixel data for the (5,3) filter. Two 1-level 2-D line-based DWT modules are cascaded to implement the 2-level 2-D DWT decomposition and achieve a throughput of two pixels per cycle. Figure 18.2.3 shows the block diagram of this 2-level 2-D DWT module. The 1-D row and column DWT modules are implemented using a lifting scheme. 8512b of on-chip memory, implemented by registers and on-chip SRAM, is required to accommodate the 128x128 tile size.

The most critical challenge to increase the throughput of a JPEG 2000 design is the EBCOT, which requires a lot of sequential operations and complicated controls. The state-of-the-art of EBCOT design is the two-bit plane parallel architecture proposed in [2]. In this work an EBCOT architecture capable of processing a DWT coefficient in parallel, regardless of bit-width, is achieved by use of three techniques. First, a parallel context

modeling approach is taken instead of a traditional bit plane-by-bit plane, to increase the processing speed. Second, a reconfigurable FIFO architecture that reduces bubble cycles is obtained by exploiting the features of the EBCOT and the DWT. Third, a folded Arithmetic Encoder (AE) architecture is devised to reduce the area.

Figure 18.2.4 shows the block diagram of the parallel EBCOT architecture. The proposed architecture is capable of processing one 11b DWT coefficient per cycle. This architecture processes 28 passes in parallel, and therefore operates at 1/28th the frequency of a traditional architecture. The state variables for context formation are calculated on the fly for each bit plane of each coefficient so the 16kb state variable memory is eliminated. The folding technique reduces the hardware cost of the AE by 99k gates.

There are two fatal drawbacks of the recommended post-compression R-D optimization in the reference software. First, the computational power and the processing time are wasted since the source image must be losslessly coded regardless of the target bit rate. Second, a large temporary memory is required to buffer the bit stream and side information for rate control. A pre-compression R-D optimization algorithm is proposed to solve these problems.

The flowchart of the proposed pre-compression R-D optimization algorithm is shown in Fig. 18.2.5. It is comprised of two stages: accumulation and decision. During accumulation the distortion and bit-count are calculated and accumulated. In the decision stage the truncation points are determined according to the normalized distortion and estimated bit rate. The proposed algorithm allows the truncation point of a code-block to be determined before EBCOT encoding. Hence, only required coding passes are processed reducing the computational power as the compression ratio increases (e.g., compression ratio of 8 requires 8 times less EBCOT computation). In addition, the memory for lossless code-stream and R-D information is eliminated. Figure 18.2.6 shows that the performance of the proposed pre-compression algorithm degrades by only 0.3dB on average compared with the post-compression algorithm.

A Performance Index (PI), defined as throughput per unit area at 1MHz is used to make a fair comparison to existing work. The PI of this work is 0.182 (<sup>81</sup>/<sub>5.5x81</sub>). The area of the JPEG 2000 core in [2] is estimated as 25mm<sup>2</sup> and the PI will be 0.030 (<sup>20.7</sup>/<sub>25x27.4</sub>). Hence, the developed chip is 6 times better than [2] using this metric. The improvement is mainly due to the proposed parallel EBCOT architecture. Furthermore, the developed chip provides an R-D optimized code-stream, an important feature of JPEG 2000.

Figure 18.2.7 is the micrograph of the 81MS/s JPEG 2000 single chip encoder. Parallel and pipeline techniques reduce the required operating frequency while reconfigurable and folding techniques reduce the silicon area. Pre-compression rate-distortion optimization algorithm reduces computational power and memory requirements.

### Acknowledgements:

The authors thank Prof. Chien-Mo James Li and members of DSP/IC Lab. for contributions and discussion. The multi-project chip support from the National Science Council of Taiwan/Chip Implementation Center is also acknowledged.

### References:

- [1] ISO/IEC IS 15444-1, "Information Technology—JPEG 2000 Image Coding System—Part 1:Core Coding System," ISO/IEC JTC1/SC29/WG1 (Dec. 2000), AMENDMENT 1: Codestream restrictions (Mar. 2002).
- [2] H. Yamauchi, et al., "Image Processor Capable of Block-Noise-Free JPEG2000 Compression with 30frames/sec for Digital Camera Applications," *ISSCC Dig. Tech. Papers*, Feb. 2003.

Technology	TSMC 0.25- $\mu$ m 1P5M CMOS
Supply Voltage	2.8 V
Core Area	2.73 $\times$ 2.02 mm <sup>2</sup>
Logic Gates	162.5 K (2-input NAND gate)
SRAM	7 K bits
Operating Frequency	81 MHz
Power	348 mW
Package	PGA 256
Image Size	Up to 32K $\times$ 32K
Processing Rate	81 M samples/sec
DWT	(5,3) filter, 2-level decomposition
Tile Size	128 $\times$ 128
Code-block Size	64 $\times$ 64

Figure 18.2.1: Chip features.

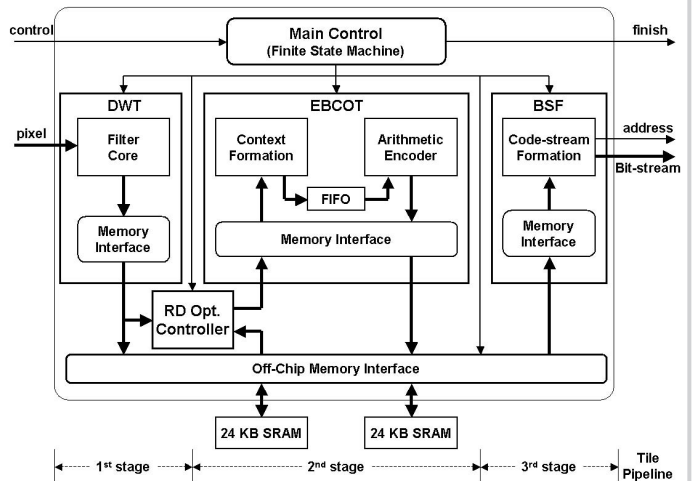


Figure 18.2.2: Block diagram of the JPEG2000 encoder.

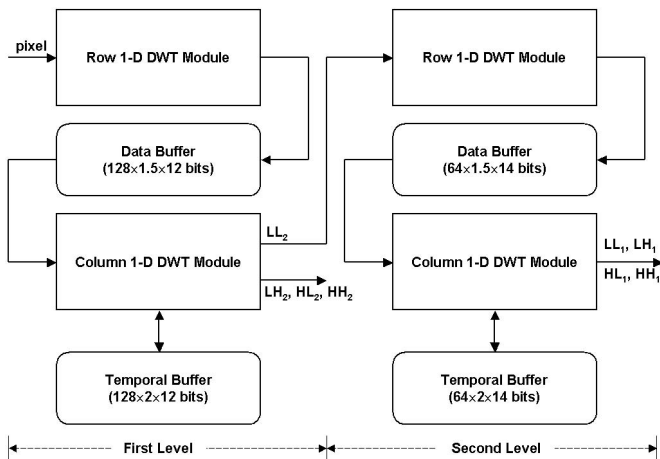


Figure 18.2.3: Block diagram of DWT module.

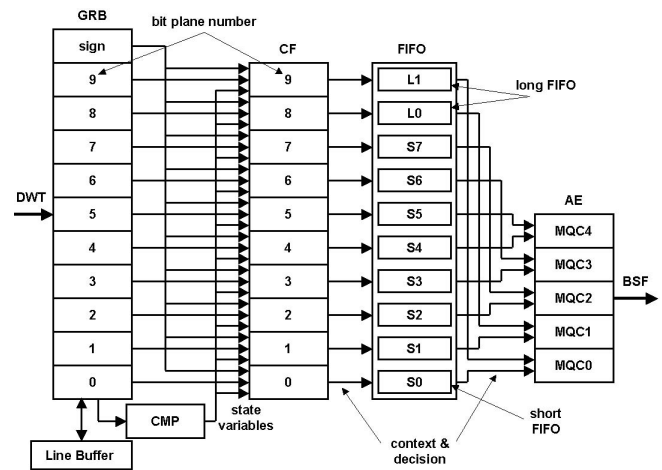


Figure 18.2.4: Word-parallel EBCOT architecture.

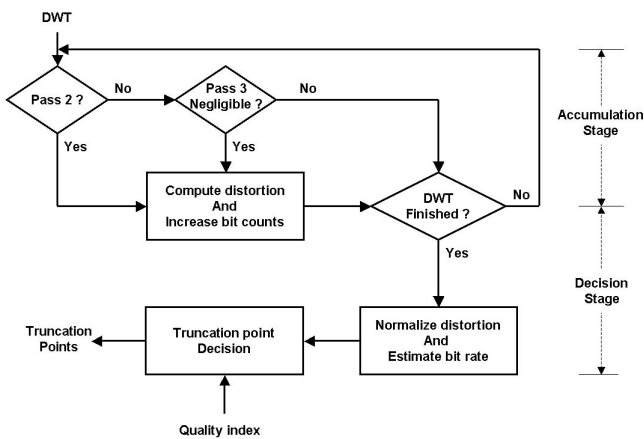


Figure 18.2.5: Flowchart of the pre-compression rate-distortion optimization algorithm.

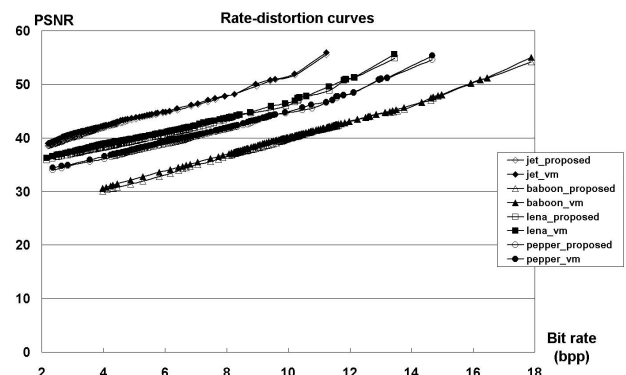


Figure 18.2.6: Performance comparison with the reference software.

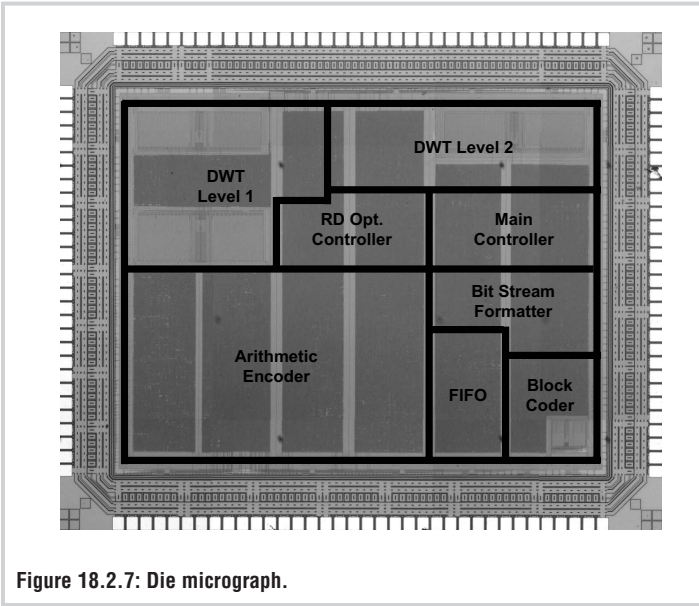


Figure 18.2.7: Die micrograph.

Technology	TSMC 0.25- $\mu\text{m}$ 1P5M CMOS
Supply Voltage	2.8 V
Core Area	2.73 $\times$ 2.02 mm <sup>2</sup>
Logic Gates	162.5 K (2-input NAND gate)
SRAM	7 K bits
Operating Frequency	81 MHz
Power	348 mW
Package	PGA 256
Image Size	Up to 32K $\times$ 32K
Processing Rate	81 M samples/sec
DWT	(5,3) filter, 2-level decomposition
Tile Size	128 $\times$ 128
Code-block Size	64 $\times$ 64

Figure 18.2.1: Chip features.

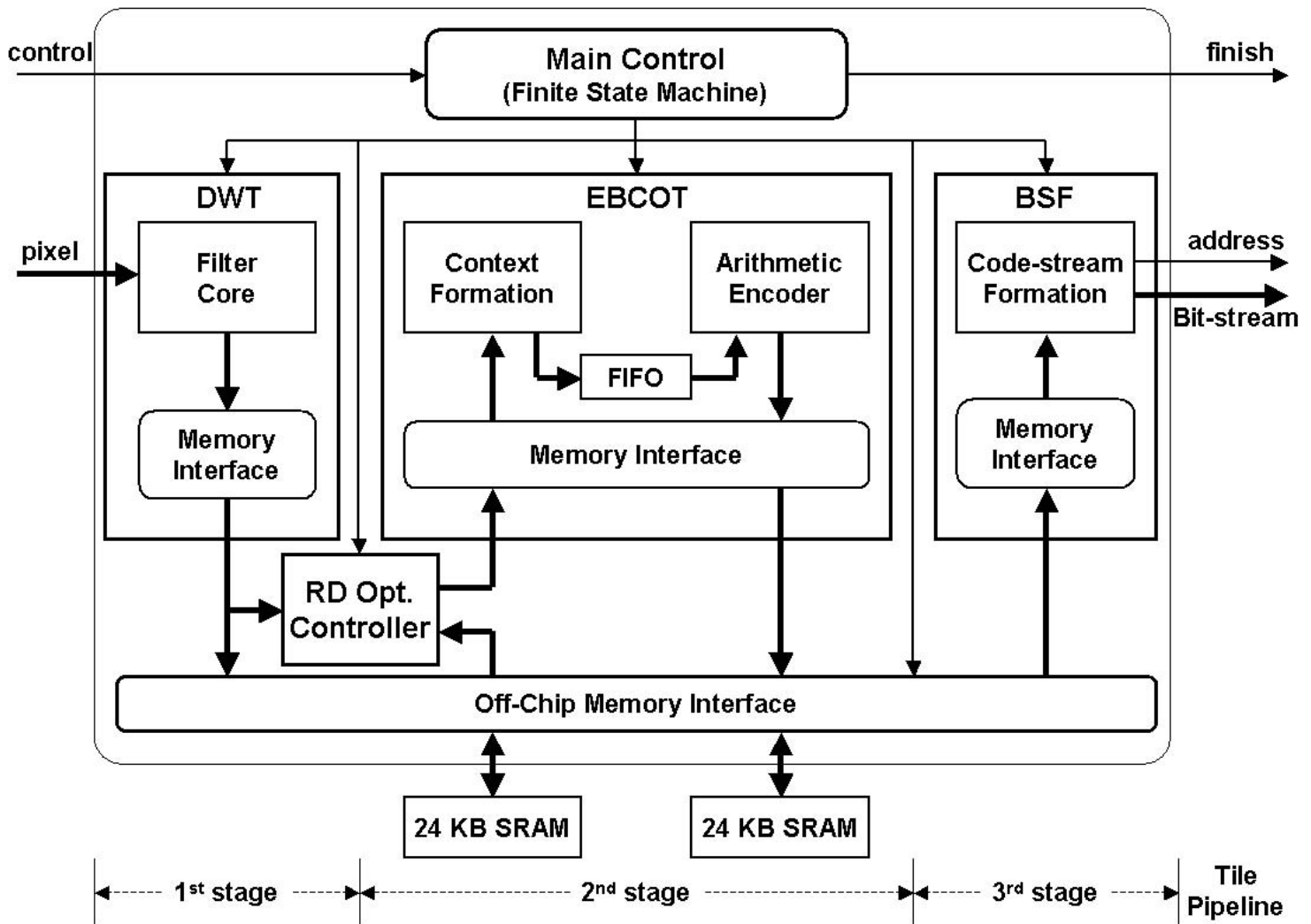


Figure 18.2.2: Block diagram of the JPEG2000 encoder.

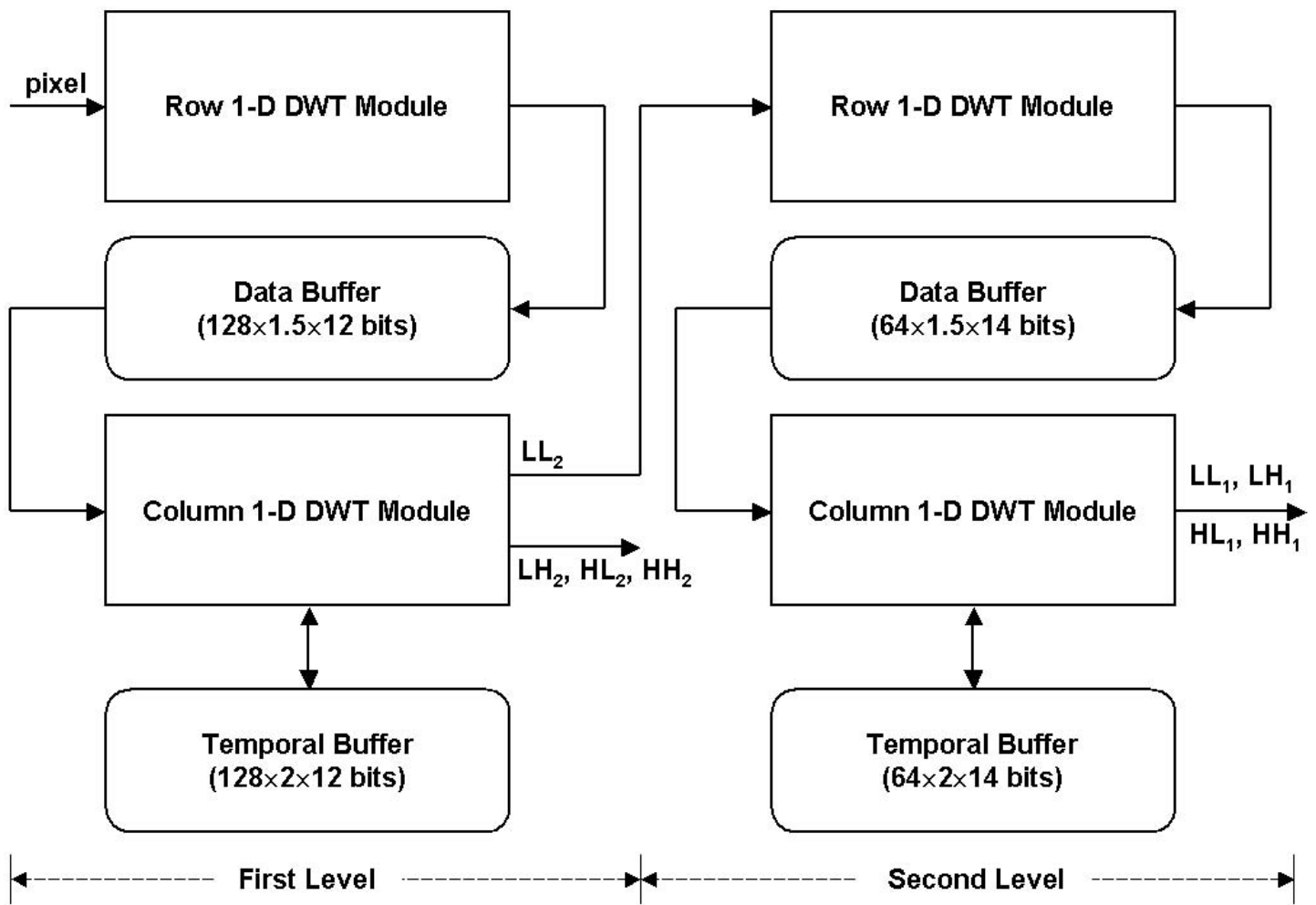


Figure 18.2.3: Block diagram of DWT module.

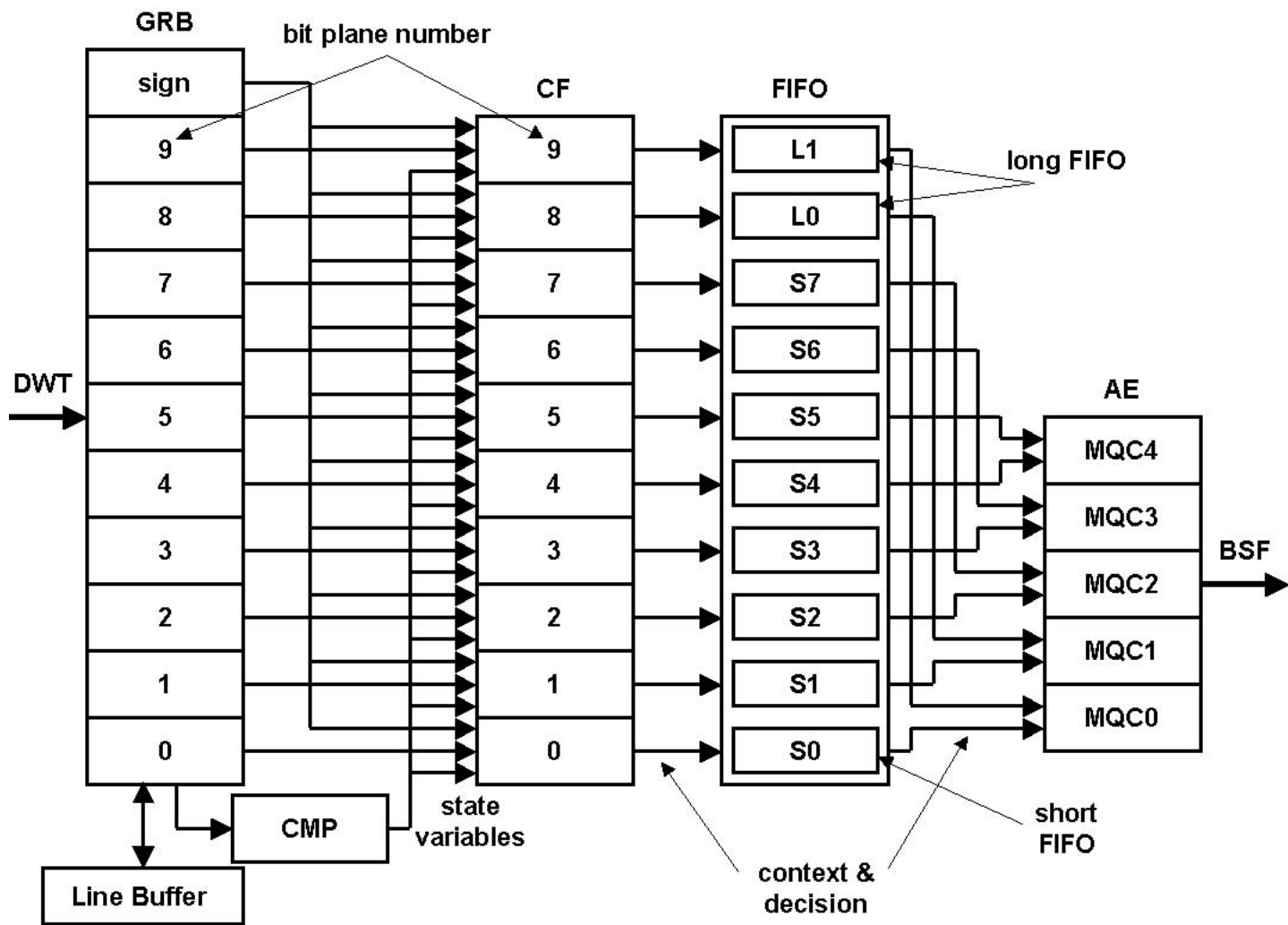


Figure 18.2.4: Word-parallel EBCOT architecture.

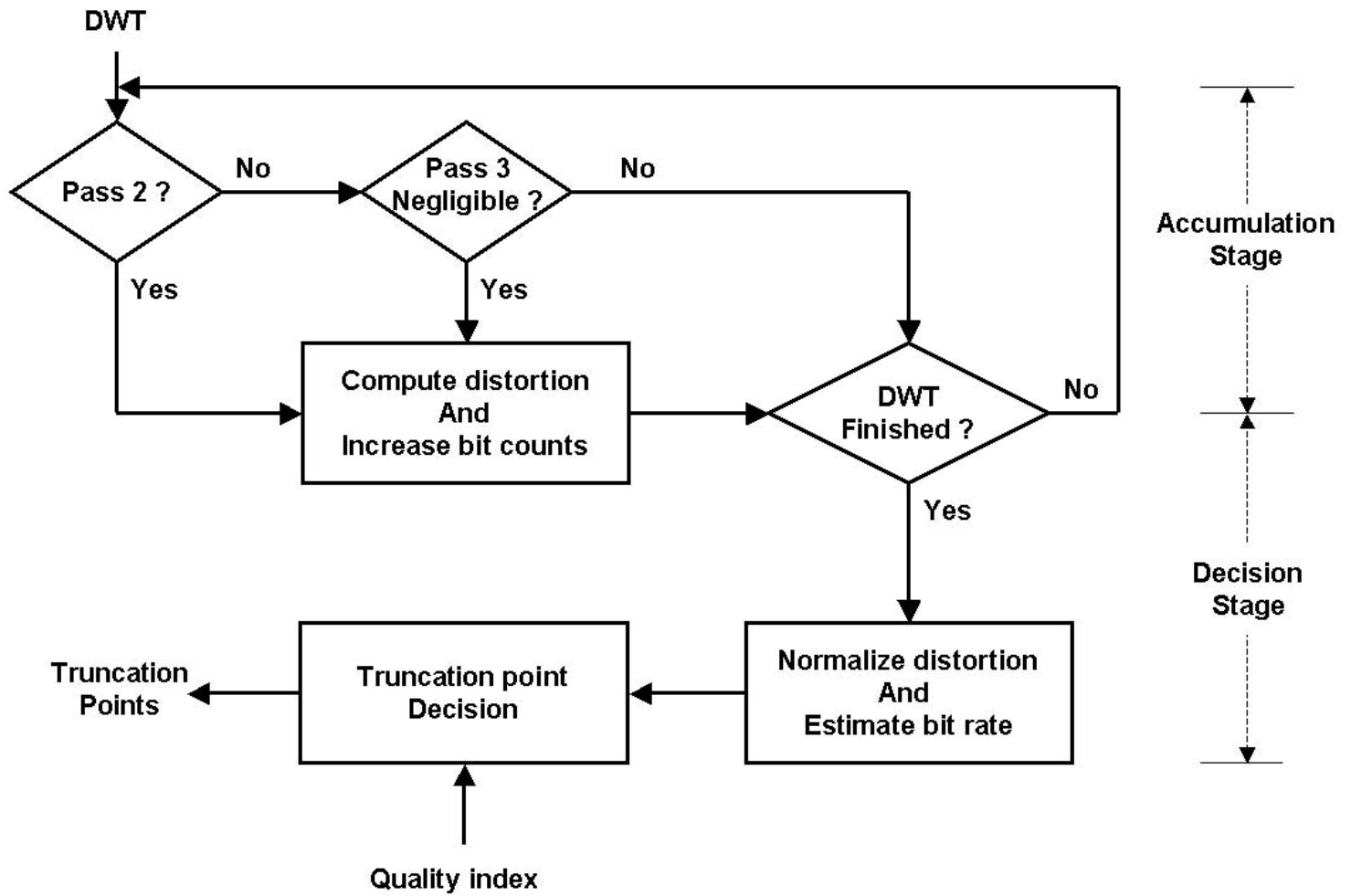


Figure 18.2.5: Flowchart of the pre-compression rate-distortion optimization algorithm.



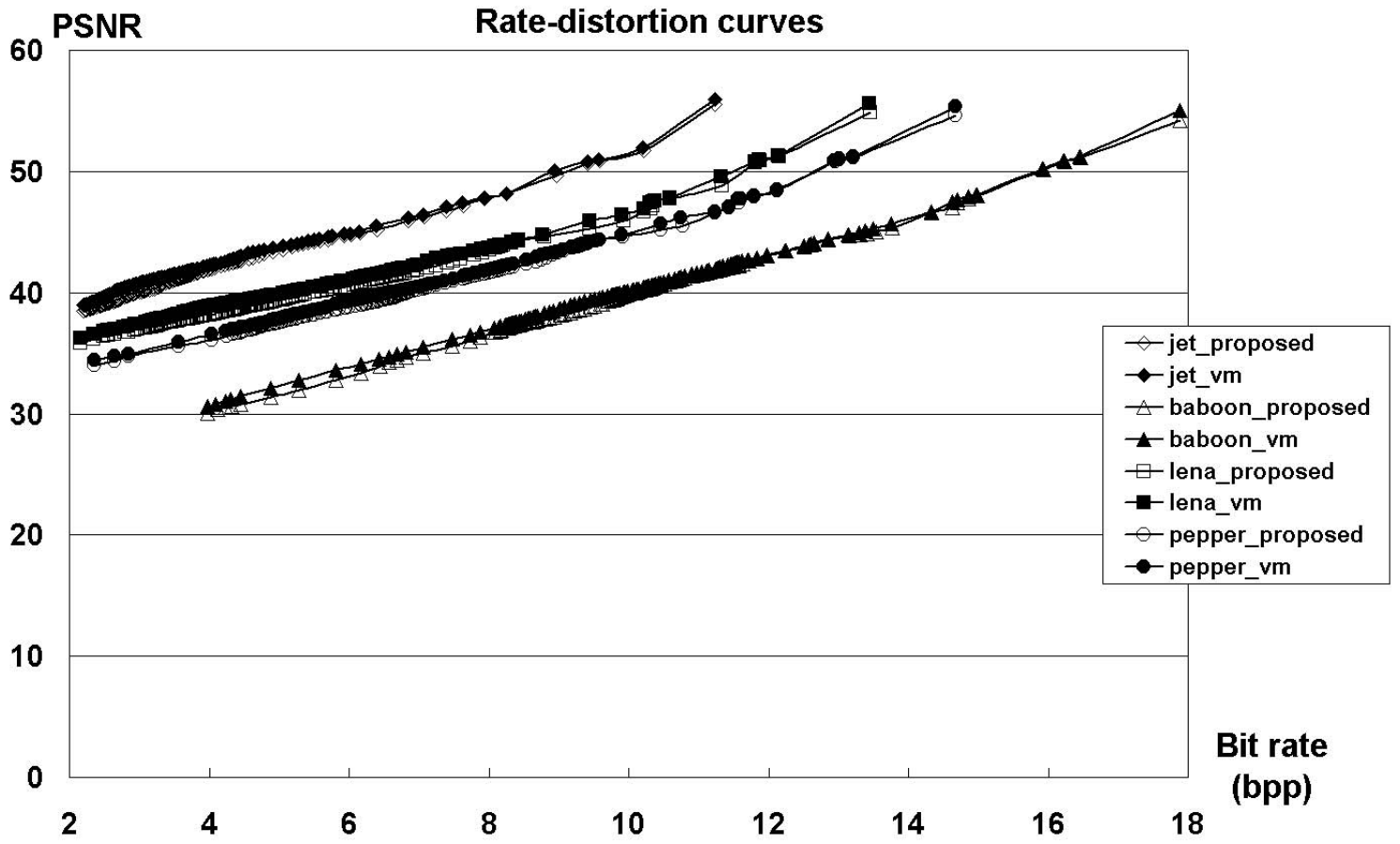


Figure 18.2.6: Performance comparison with the reference software.

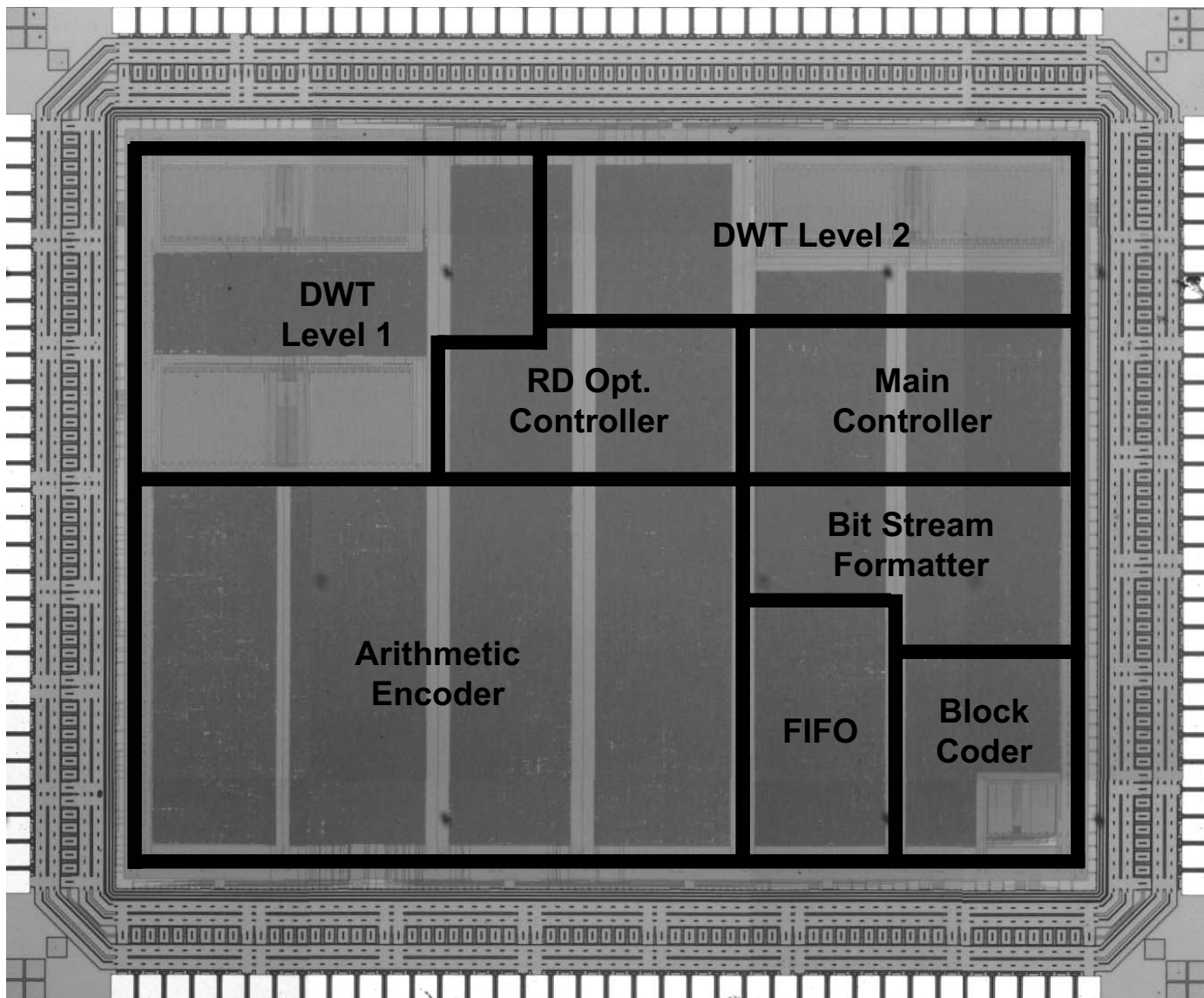


Figure 18.2.7: Die micrograph.