2.8 to 67.2mW Low-Power and Power-Aware H.264 Encoder for Mobile Applications

Tung-Chien Chen, Yu-Han Chen, Chuan-Yung Tsai, Sung-Fang Tsai, Shao-Yi Chien and Liang-Gee Chen

Graduate Institute of Electronics Engineering and Department of Electrical Engineering,
National Taiwan University, Taipei, Taiwan

Abstract

A 2.8 to 67.2mW H.264 encoder is implemented on a 12.8mm^2 die with 0.18μm CMOS technology. The proposed parallel architectures along with fast algorithms and data reuse schemes enable 77.9% power savings. The power awareness is provided through a flexible system hierarchy that supports content-aware algorithms and module-wise gated clock.

Introduction

H.264 achieves 25-45% bit-rate savings over MPEG4. However, it consumes much more power due to the use of complex compression tools [1]. Low power consumption is required for portable devices in which the power is limited. In addition, power scalability is also important because it enables such devices to tradeoff compression performance with power consumption according to power levels and application requirements.

Efficient techniques that enable a low power and power scalable H.264 encoder for mobile applications are presented in this paper. There are three critical issues. First, motion estimation (ME) normally consumes about 85% of the encoder power [2]. To reduce power consumption, new data reuse (DR) schemes should be implemented in the parallel architectures for fast ME algorithms. Second, low power techniques have to be integrated across different design levels. This is not easy because fast ME algorithms are difficult to realize on parallel architectures due to their irregular and sequential natures. Furthermore, gated-clock techniques at the circuit level cannot be effectively supported without system-level considerations. Finally, to enable power scalability on an ASIC encoder, flexibility must be explored on the system and module architectures along with a computationally scalable algorithm.

Low-Power Motion Estimation

Figure 4 shows the proposed integer-pixel ME (IME) engine. To save memory power, reference pixels read from search window local memories (SWLMs) are stored and reused in systolic register array (SRA). Four-step search (FSS) fast algorithm reduces computational complexity but has an irregular searching flow. To combine fast algorithms and DR schemes, the SRA is designed with four configurations for four different searching directions—up-, down-, left- and right-shift. To cooperate with such SRA, the ladder-shaped data arrangement (LSDA) is proposed in SWLMs. The horizontally and vertically adjacent pixels such as "A2 to P2" and "B0 to B15" are both arranged in different memories and can thus be accessed in parallel. Then, the searching flow is designed to string up all candidates to be searched. For example in Fig. 1, reusable pixels are stored in the systolic array after Step1. During Step2, the shift direction of systolic array is successively configured as down-, right-, up- and left-shift configurations. The corresponding rows and columns of pixels are read from SWLMs, and then shifted and reused in SRA. 87.5% memory power of IME is thus saved.

Figure 2 shows the proposed fractional-pixel ME (FME) engine. Different from the sequential half-then-quarter refinement algorithm, the proposed one-pass algorithm searches 25 half-pixel/quarter-pixel candidates around the best integer-pixel candidate to facilitate parallel processing and DR. In the proposed parallel architecture, 25 processing units (PUs) process 25 half-pixel/quarter-pixel candidates in parallel. The half pixels interpolated by the 6-tap 2-D interpolation engine are shared by the nine half PUs, and 89% of memory accessing and 6-tap filtering power are saved for the half-pixel candidates. Then, because of the algorithmic linearity, the residues of half-pixel candidates in transform domain are reused by the bilinear filter array to generate the residues on quarter-pixel candidates in transformed domain for 16 quarter PUs. All memory access, 6-tap filter and Hadamard transform power are saved for the quarter-pixel candidates.

Flexible Encoder System and Power-Aware Algorithm

Based on the low-power ME engines, the proposed encoder further demonstrates the capability of power scalability with the flexible encoder system and power-aware algorithm. Figure 3 shows the system architecture. The encoder has 3 macroblock (MB) pipeline stages: coarse-prediction, fine-prediction and block-engine, and three MBs are pipelined and simultaneously processed with the power-aware algorithm shown in Fig. 4. A wide range of power scalability is achieved through this flexible system that adjusts the parameters of reconfigurable PEGs associated to all H.264 compression functionalities. Unlike the previous encoder system [1, 2], in which each pipeline controller is tightly coupled with the corresponding processing engines (PEGs) in each pipeline stage, our system hierarchy separates PEGs from pipeline controllers. PEGs can thus be flexibly reused between different pipeline stages.

There are three main advantages. First, it can provide flexibility in system scheduling. Compared with [2], our system has a more compact and balanced schedule, which reduces pipeline stages and subsequently saves power for data pipelining. Second, it provides the flexibility for algorithm development. For example, to enable the pre-skip algorithm as shown in Fig. 4, the FME, which normally operates in the fine-prediction stage, can also calculate the matching cost associated with the motion vector (MV) predictor at the beginning of the coarse-prediction stage. Because best inter-prediction modes of a significant number of MBs are the skip modes, our design monitors this event in the early stage to skip the corresponding Integer ME (IME) and FME operations. 20-40% of encoder power is saved with this approach. Third, based on our system hierarchy in Fig. 3, a fine-grained module-wise gated clock circuit is

implemented to precisely turn off the clocks of static register files (RFs) and inactive PEGs. The gated clock circuits in our system save 18.8-33.2% power for the encoder.

Implementation Results

Figure 5 shows the chip features. The encoder containing 452.8k logic gates and 16.95kB SRAMs is implemented on a 12.84mm^2 die with 0.18µm CMOS technology. Power dissipation is 2.8-67.2mW. Figure 6 shows the power-distortion curves of the proposed H.264 encoder and the previous arts. Our encoder consumes 22.1% power with similar compression performance compared to [2] and makes a 1.96dB quality improvement with about 5mW extra power compared to [3].

References

[1] T. Wiegand, et. al., "Overview of the H.264/AVC video coding standard," IEEE Transactions on CSVT, vol.13, no.7, pp.560–576, July 2003.

[2] Y.-W. Huang, et. al., "A 1.3TOPS H.264/AVC Single-Chip Encoder for HDTV Applications," ISSCC Dig. Tech. Paper, Feb., 2005.

[3] C.-P. Lin, et. al., "A 5mW MPEG4 SP Encoder with 2D Bandwidth-Sharing Motion Estimation for Mobile Applications," ISSCC Dig. Tech. Paper, Feb., 2006.

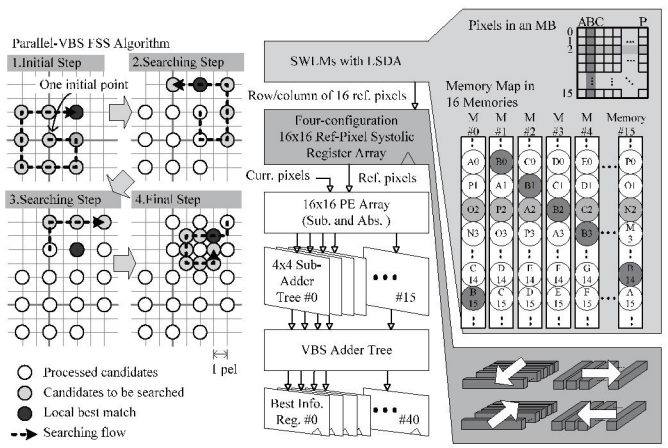

Fig. 1. Parallel VBS FSS algorithm and corresponding parallel architecture of IME.

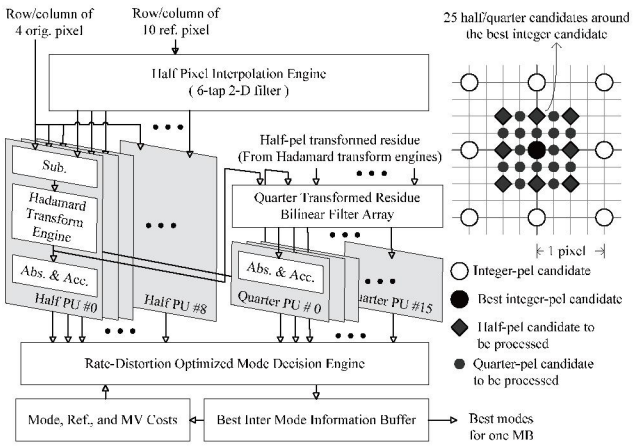

Fig. 2. One-pass FME algorithm and corresponding parallel architecture..

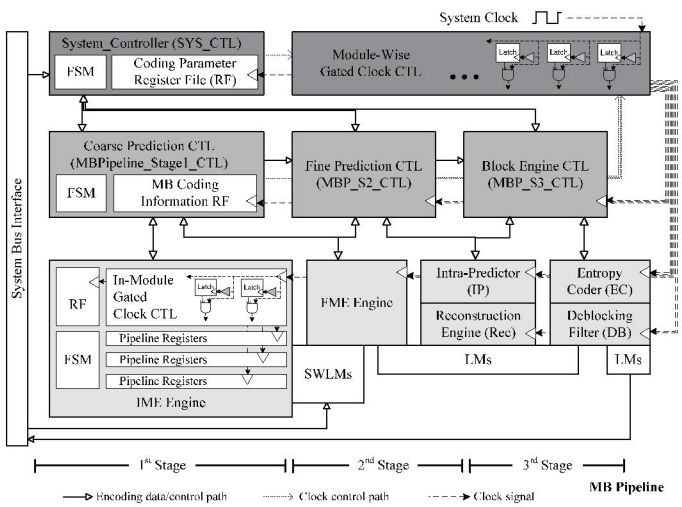

Fig. 3. Block diagram of the proposed H.264 encoding system..

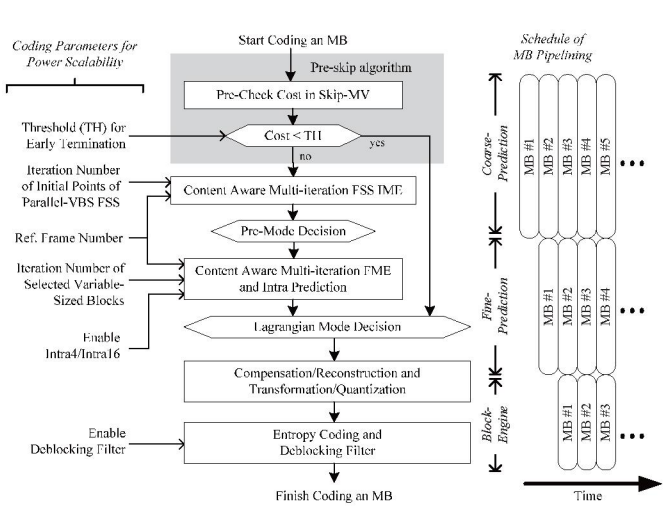

Fig. 4. MB pipelining schedule and power scalable algorithm.

Technology	TSMC 0.18 μm 1P6M CMOS
Pad/Core Voltage	3.3(Core)/1.8(I/O) V
Core Area	3.47×3.70 mm^2
Logic Gates	452.8 K (2-input NAND gate)
SRAM	16.95 KByte
Max. Ref. Frame	2
Max. Horizontal SR	[-32, +31]
Max. Vertical SR	[-16, +15]
Power Consumption (Measured Results)	67.2-43.5 mW for SDTV, 1 ref @ 54MHz, 1.8V 40.3 mW for CIF, 2 ref @ 27MHz, 1.8V 15.9-9.8 mW for CIF, 1 ref @ 13.5MHz, 1.3V 8.7 mW for QCIF, 2 ref @ 6.25MHz, 1.3V 4.3-2.8 mW for QCIF, 1 ref @ 3.125MHz, 1.3V
Power Consumption (Simulated Results with TSMC 0.13 μm process)	16.3-9.1 mW for SDTV, 1 ref @ 54MHz, 1.3V 12.9 mW for CIF, 2 ref @ 27MHz, 1.3V 8.2-5.1 mW for CIF, 1 ref @ 13.5MHz, 1.3V 4.5 mW for QCIF, 2 ref @ 6.25MHz, 1.3V 2.2-1.5 mW for QCIF, 1 ref @ 3.125MHz, 1.3V

Fig. 5. Chip features.

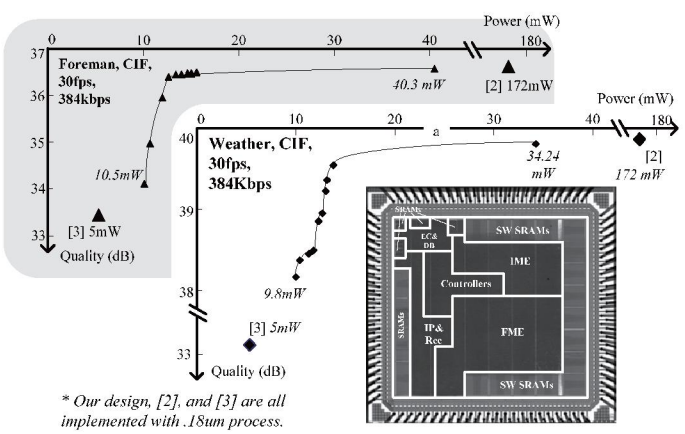

Fig. 6. Power-distortion curves and die micrograph.