Chapter 3

A low-power portable H.264/AVC decoder using elastic pipeline

Yoshinori Sakata, Kentaro Kawakami, Hiroshi Kawaguchi, Masahiko Yoshimoto

Graduate School, Kobe University, Kobe, Hyogo, 657-8507 Japan
Email: {yoshi, kawakami, kawapy, yosimoto}@cs28.cs.kobe-u.ac.jp

Abstract. We propose an elastic pipeline architecture that can apply dynamic voltage scaling (DVS) to a dedicated hardware, and implement the elastic pipeline to a portable H.264/AVC decoder LSI with embedded frame buffer SRAM. A supply voltage and operating frequency are decreased by a feedback-type voltage/frequency control algorithm. In a portable H.264/AVC decoder, embedded SARM can be utilized as frame buffer since the frame buffer is not so large that an external DRAM is required. In the proposed pipeline architecture, the power in the embedded SRAM and even in a local bus connecting with the frame buffer SRAM can be controlled by DVS. We carried out simulation in the 320 × 180 pixels baseline profile and 320 × 240 pixels mail profile. The total power reduction in 320 × 180 pixels and 320 × 240 pixels are 30% and 31%, respectively.

Keywords. Dynamic voltage scaling, Elastic pipeline, Embedded sram, H.264/AVC, Memory bandwidth

3.1 Introduction

Dynamic voltage scaling (DVS) is a low-power technique that controls an operating frequency and a supply voltage on an LSI, according to an appli-

cation workload. DVS is well utilized on general-purpose processors to achieve both high peak performance and low average power [1, 2].

Figure 3.1 shows a relationship between an operating frequency and a power, with DVS and without DVS (= clock gating). In DVS, if a workload does not need a high operating frequency, we can choose a combination of a lower operating frequency and a lower supply voltage. Due to this power optimization, the power becomes lower than the conventional scheme without DVS, when a workload is low. If the maximum performance is instantaneously needed, the highest supply voltage and highest operating frequency are utilized so that DVS can accommodate the peak performance.

In a case of an application to hardwired logic circuits for real-time processing, there are a few problems; a dedicated hardware is often built with pipeline architecture for high performance. Considering the likely worst-case workload, the starting time of a pipeline process is segmented into the worst-case execution cycles (WCEC). Thus, the required operating frequency is uniquely fixed, and there is no room to apply DVS in the hardwired logic circuits.

To realize DVS in hardwired logic circuits, we propose an elastic pipeline architecture. Depending on characteristics of input data, this architecture can conserve process cycles in the pipeline operation. The slack time is exploited for DVS, which achieves lower power in hardwired logic circuits.

The rest of this paper is organized as follow. Section 3.2 mentions the conventional pipeline architecture. Section 3.3 describes the proposed elastic pipeline architecture for DVS in hardwired logic circuits. In Section 3.4, we exhibit the experimental results of the proposed architecture. Section 3.5 summarizes this paper.

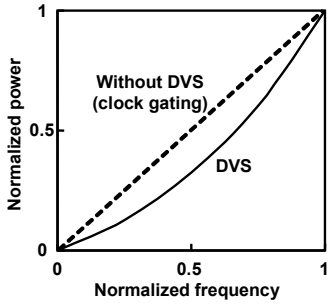

Fig. 3.1. Relationships between power and frequency in DVS and clock gating

3.2 Conventional pipeline architecture

Figure 3.2 illustrates a timing diagram of the conventional pipeline architecture. The WCEC is the maximum number of execution cycles required for one pipeline process. A gray area in the figure shows the number of processing cycles that a pipeline stage processes a datum. A hatching area means common idle cycles in a pipeline process after all pipeline stages were completed. Considering the worst-case workload, a starting time of each pipeline process is fixed to the WCEC in the conventional pipeline architecture. Hence, all the pipeline stages have to idle until the next starting time even if all the pipeline stages finish earlier than the WCEC.

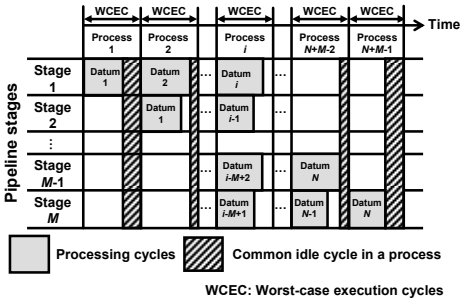

Fig. 3.2. Timing diagram of the conventional pipeline

3.3 Elastic pipeline architecture

3.3.1 Concept

We propose the elastic pipeline architecture as the solution to the issue of the conventional pipeline architecture [3]. Figure 3.3 (a) and (b) shows a concept and a timing diagram of the proposed elastic pipeline architecture. After each stage in the elastic pipeline was completed, it sends a completion signal to the pipeline controller. As soon as the pipeline controller collects all the completion signal from all the pipeline stages, each pipeline stage proceed to the next pipeline process with the start signal.

In the proposed architecture, the common idle cycles are built up in pipeline processes, and become a lump of time. As illustrated in Figure 3.3 (b), a pipeline process in the elastic pipeline requires less time than the conventional pipeline since the common idle cycles are put off. Thereby,

the elastic pipeline architecture produces the slack time, ΔH, compared to the conventional pipeline architecture.

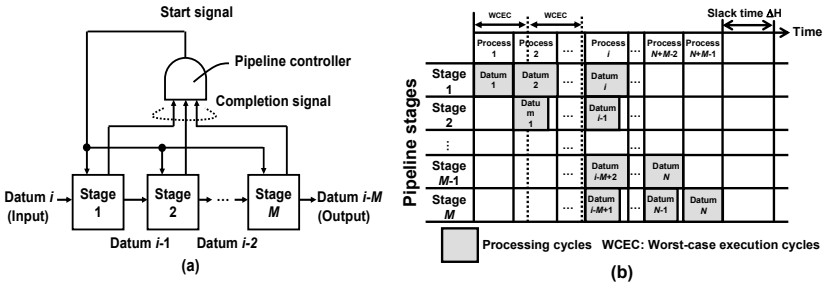

Fig. 3.3. Proposed elastic pipeline architecture: (a) concept and (b) timing diagram

3.3.2 Feedback-type voltage/frequency control algorithm

For DVS in the proposed elastic pipeline, a supply voltage and an operating frequency are changed by a feedback-type voltage/frequency control algorithm as illustrated in Figure 3.4 [4]. In an H.264 codec, data are processed in every single macro block (MB: 16 × 16 pixels). In this algorithm, a frame is divided into some slots; a set of MBs are assigned to a slot. The first and second slots are always processed with the maximum frequency (= f in Figure 3.4). However, these slots are potentially completed earlier since the elastic pipeline reduces the number of processing cycles. Now, pay attention to the third slots, where the slack time, ΔH, is left. Even considering a voltage/frequency transition time, T_{td}, the third slot has twice as long time as T_{slot} (a processing time for a slot), which allows the third slot to be processed at an half of f. Note that a real-time operation is guaranteed in this feedback algorithm. As described, we prepare the two kinds of operating frequencies, f and $f/2$ in this study [3, 4].

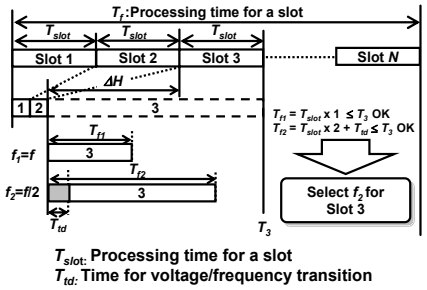

Fig. 3.4. Feedback-type voltage/frequency control algorithm

3.3.3 Architecture

To estimate the effect of the power reduction in the proposed pipeline architecture, we designed a H.264 decoder architecture as shown in Figure 3.5. SRAM utilized for reference images is embedded on a chip. The external DRAM is used as bit stream buffer, if the resolution is very large as HDTV. In this case, a supply voltage and operating frequency should be fixed since it is preferable that the DRAM interface operates at a fixed supply voltage and operating frequency for compatibility with other hardware cores. But in this study, we handle small resolution which is used for the portable product. So SRAM which utilized for bit stream buffer is embedded. We can adapt DVS for not only decoder core but SRAM and local bus connecting with the frame buffer SRAM.

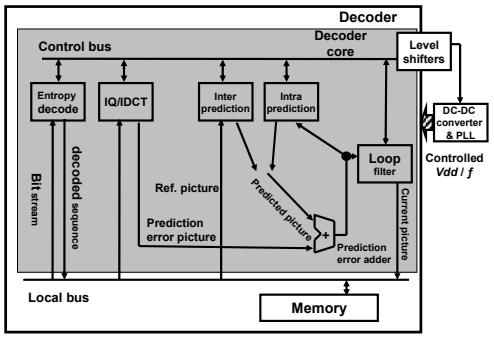

Fig. 3.5. H.264 decoder block diagram

3.4 Experimental results

3.4.1 Test sequence

Assuming portable H.264 image sequences, we handle two kinds of resolution (320 × 180 pixels and 320 × 240 pixels). As test sequences, we chose six sequences: Bus (BUS1), Cheer (CHER), Flower (FLOW), Foreman (FORE2), Football (FTBL), Girl (GIRL). Then, we encoded the six test sequences under the configurations in Table 3.1 (a) to prepare the test sequences: The baseline profile with 320 × 180 pixels complies with the Japanese portable television standard, and the main profile with 320 × 240 is adopted by Sony PSP [5].

Table 3.1. (a) Encoding configuration, (b) Simulation parameter

(a)

Frame size (Resolution)	320 × 180 pixels	320 × 240 pixels
# of test sequence	6 sequences	6 sequences
Profile	Baseline profile	Main profile
Frame rate	15	30
# of reference frames	1	2
Reference software	JM9.6[6]	JM9.6[6]

(b)

Frame size (Resolution)	320 × 180	320 × 240
# of slots	75	30
Operating frequency (MHz)	6.75/3.38	15.84/7.92
Supply voltage (V)	0.62/0.6	0.8/0.63
SRAM (Mbits)	2	4
WCEC	1760	1760
# of logic gates	600329	600329

Table 3.1 (b) illustrates the simulation parameters. The respective supply voltage and operating frequency are prepared for the two kinds of resolutions. The capacities of the embedded SRAMs are 2 M bits and 4 M bits, respectively. The SRAM capacity of the 320 × 240 pixel main profile is twice as large since the number of reference frames is two. However, note that the WCECs and the numbers of logic gates are equal between the two kinds of resolutions. In other words, the operating frequencies are different, but the sizes of the decoder cores are the same.

3.4.2 The optimum number of slots per frame

Since the elastic pipeline architecture reduces the processing cycle, we can apply DVS to the decoder LSI. In this subsection, the optimum number of slots is discussed.

Figure 3.6 illustrates the simulation result of the relationship between the power of the decoder core and the number of slots, using the BUS1 sequence. The power reduction factor depends on the number of slots. In this simulation, the transition time is assumed to be 50 μs [1]. The baseline profile with 320 × 180 pixels has the power minimum when the number of slices is 75. On the other hand, the optimum number of slots is 30 in the main profile with 320 × 240 pixels.

If the number of slots is smaller than the optimum point, the power reduction drastically becomes worse. For instance, if there are merely several slots, there are few chances to make the operating frequency and the

supply voltage lower, which causes high power consumption. Alternatively, if there are many slots, there are many chances to change the operating frequency and supply voltage. The voltage/frequency transition time, however, becomes longer. The power reduction gradually becomes worse with the increase of the slice number.

3.4.3 Power saving

As well as the decoder core, we can apply DVS to the embedded frame buffer SRAM and the local bus connecting to the SRAM. Figure 3.7 (a) shows the respective power reduction factors in the local bus, the embedded SRAM, and the decoder core. In the case of the frame size of 320 × 180 pixels, the proposed elastic pipeline reduces the powers by 7%, 25%, and 41%, respectively, in the local bus, the embedded SRAM, and the decoder core. In the case of the frame size of 320 × 240 pixels, the respective factors are 38%, 27%, and 33% in the resolution of 320 × 180 pixels.

The overall power reduction is 30% and 31% on average, in the two kinds of resolutions, as shown in Figure 3.7 (b).

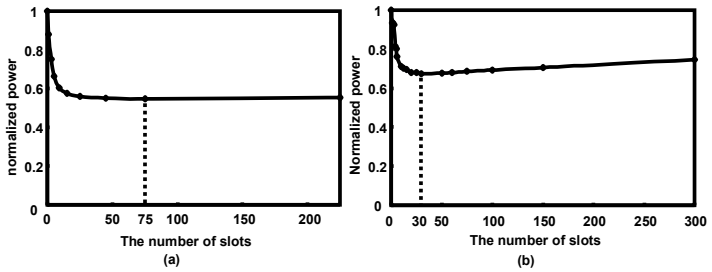

Fig. 3.6. The number of slots in a frame vs. power using the test sequence "Bus": (a) 320 x 180 pixels and (b) 320 x 240 pixels

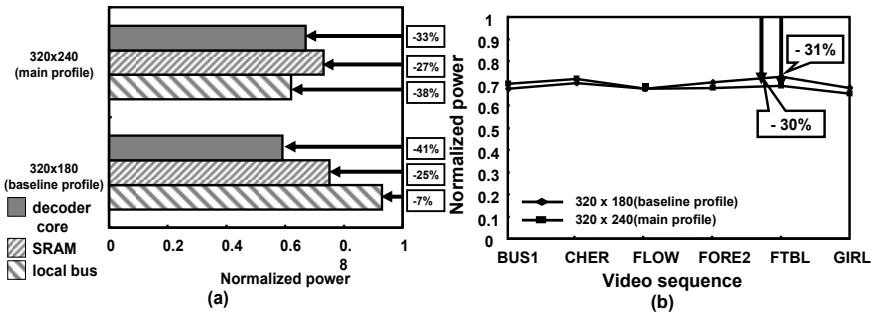

Fig. 3.7. Power reduction ratio: (a) decode core, embedded SRAM, local bus, (b) overall decoder

3.5 Summary

We proposed the elastic pipeline architecture that can apply DVS to a hardwired circuit. We implemented a H.264 decoder LSI, and controlled the embedded frame buffer SRAM and the local bus connecting to the embedded SRAM with DVS, as well as the decoder core.

We verified that the proposed elastic pipeline reduces the power on the H.264 decoder LSI by 7% in the local bus, by 25% in the frame buffer SRAM, and by 41% in a decoder core in a 320 × 180 pixel baseline profile. In a case that 320 × 240 pixel main profile, the power is reduced by 38%, 27%, and 33% in the local bus, the frame buffer SRAM, and the decoder core, respectively. The total power reductions in the baseline profile and the mail profile pixels are 30% and 31%, respectively.

References

1. Nowka KJ, Carpenter GD, MacDonald EW, Ngo HC, Brock BC, Ishii KI, Nguyen TY, Burns JL (2002) A 32-bit PowerPC system-on-a-chip with support for dynamic voltage scaling and dynamic frequency scaling. IEEE J. Solid-State Circuits 37(11):1441-1447
2. Kawakami K, Kanamori M, Morita Y, Takemura J, Miyama M, Yoshimoto M (2005) Power-minimum frequency/voltage cooperative management method for VLSI processor in leakage-dominant technology era. IEICE Trans. Fundamentals E88-A(12):3290-3297
3. Kawakami K, Kuroda M, Kawaguchi H, Yoshimoto M (2007) Power and memory bandwidth reduction of an H.264/AVC HDTV decoder LSI with elastic pipeline architecture. In: Proceeding of Asia and South Pacific Design Automation Conference (ASP-DAC)
4. Kawaguchi H, Shin Y, and Sakurai T (2005) μITRON-LP: power-conscious real-time OS based on cooperative voltage scaling for multimedia applications. IEEE Trans. Multimedia 7(1):67-74
5. http://manuals.playstation.net/document/jp/psp/current/video/filetypes.html
6. Joint Video Team (JVT) of ISO/IEC MPEG&ITU-T VCEG (2003) ISO/IEC 14496-10.