Utilizing Power Management and Timing Slack for Low Power in High-Level Synthesis

Zong-Han Xie, Shih-Hsu Huang, Chun-Hua Cheng
Department of Electronic Engineering
Chung Yuan Christian University,
Taoyuan, Taiwan, R.O.C.

ABSTRACT

As the design size continues to increase, low power has become a very important concern. In the high-level synthesis stage, operation scheduling is critical for circuit performance. Previous algorithms do not consider power management and timing slack (operation delay selection) at the same time. Different from previous works, in this paper, we utilize power management and timing slack as possible to reduce the total power under the overall latency constraint. We propose an integer linear programming (ILP) approach to combine operation scheduling, power management, and timing slack selection in order to reduce the total power. Benchmark circuits show that our ILP approach has a significant improvement.

INTRODUCTION

In the high-level synthesis stage, the scheduling problem is to assign each operation in the control/data flow graph (CDFG) into a dedicated control step for starting its execution. As the design size continues to increase, low power is a very important concern for modern circuit design. Many operation scheduling algorithms [1-3] have been paid to deal with the trade-off between the number of control steps and the number of resources.

Monteiro et al. [4] show that large power savings can be derived by shutting down unused operations (i.e., power management). Huang et al. [5] propose an integer linear programming (ILP) to model the combination of operation scheduling and power management in the high level synthesis. However, these approaches [4,5] assume that the delay of each operation is a fixed constant value.

In fact, there exists a trade-off between operation delay and power consumption. Here we use the CDFG shown in Fig. 1 as an example. Suppose that the overall latency constraint is 4 control steps. However, if the execution time of each functional unit is 1 control step, the overall latency is only 3 control steps. Therefore, we can allow the multiplier be executed in lower voltage to utilize the time slack for power saving.

Yen et al. [6] mention that the trade-off between operation delay and power consumption can be derived in the high-level synthesis stage. However, their approach [6] does not consider the control constructs in the CDFG (i.e., they [6] do not deal with the power management technique shown in [4,5]). Furthermore, their approach [6] focus on the minimization of the cycle-by-cycle power differential.

Different from the previous work [6], in this paper, we propose a new ILP approach, which integrates operation scheduling, power management, and timing slack selection, together to minimize the total power (i.e., the average power) under a given overall latency constraint. The basic ideas are to shut down unused operations as possible (i.e., the power management technique) and utilize the timing slack of each operation as possible (i.e., the operation delay selection technique) in order to reduce the total power dissipation under the overall latency constraint. Benchmark circuits show that our approach has a significant improvement.

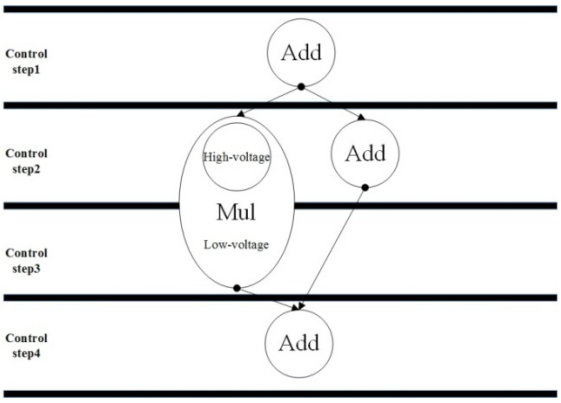

Fig. 1. The trade-off between the delay and the power in high-level synthesis.

PROPOSED APPROACH

We use the CDFG shown in Fig. 2 to explain our idea. By applying the operation delay selection technique [6], we can allow operation Mul3 be executed in two control steps (i.e., in lower voltage) to utilize the timing slack for power saving. Thus, we can have the CDFG as shown in Fig. 3. Furthermore, by applying the power management technique [4,5], we can make the following power savings: if control operation C7 is true, unused operations Add2 and Add5 can be shut down; on the other hand, if control operation C7 is false, unused operation Mul3 can be shut down. Thus, we can have the CDFG as shown in Fig. 4.

We propose an ILP approach, which combines operation scheduling, power management, and timing slack selection (operation delay selection), to minimize the total power under a given overall latency constraint. Note that our ILP approach is based on the ILP framework of the previous work [5] (i.e., the simultaneous operation scheduling and power management) and the ILP framework of the previous work [6] (i.e., the simultaneous operation scheduling and operation delay selection). We integrate the two ILP formulations into a

single ILP framework (i.e., the simultaneous operation scheduling, power management, and operation delay selection). Our optimization goal is to minimize the total power (i.e., the average power). The constraints used in the ILP framework include the resource constraints, the data dependency constraints, the control dependency constraints, and the operation delay selection constraints. Due to page limit, here we do not describe the detailed formulas.

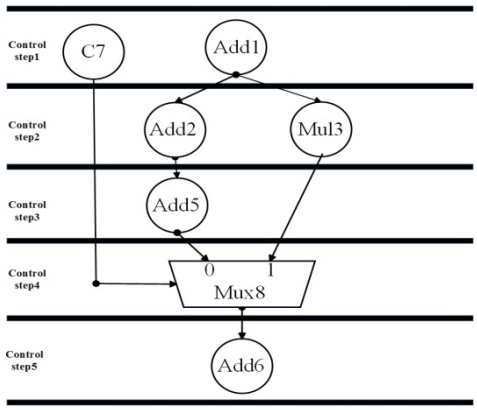

Fig. 2. Motivational example.

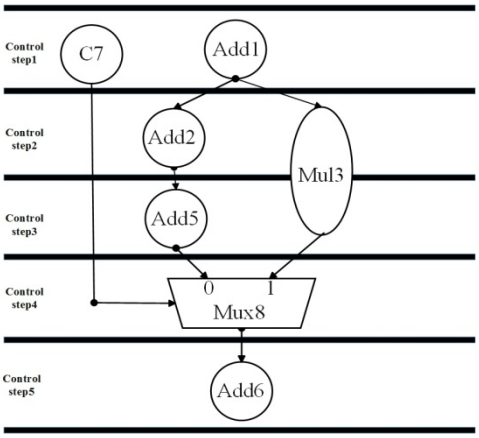

Fig. 3. Operation delay selection is applied.

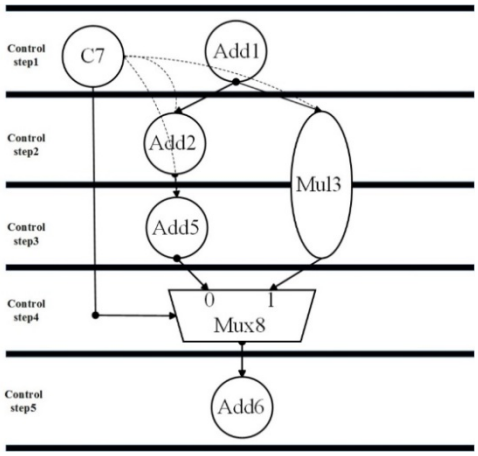

Fig. 4. Power management technique is applied.

EXPERIMENT RESULTS

In the experiments, we use two benchmark circuits to test the effectiveness of the proposed approach. We also implement the previous work [5] for comparisons. If the operation delay is 1 control step, we adopt the same power consumption model as the previous work [5]. We say the power consumption for completing an operation in one control step is the original power consumption. If the timing slack is utilized, an operation can be executed in more than one control step and the power consumption can be reduced. Here we assume the power consumption is reduced to be p/n, where p is the original power consumption and n is the number of control steps.

Table I gives the experimental results. The column Steps denote the overall latency constraint. The column Power Saving denotes the power saving. Since the proposed approach utilizes the timing slack (i.e., the operation delay selection technique), the power saving can be greatly increased.

Table I: Experimental results on benchmark circuits,

Circuit	Steps	Power Saving		
		[5]	Ours	Imp(%)
G2	8	108.5	123.5	14%
R1	11	79.0	159.5	102%

CONCLUSIONS

In this paper, we utilize the timing slack as possible to reduce the total power consumption under a given overall latency constraint. We present an ILP approach to combine operation scheduling, power management, and timing slack selection in order to reduce the total power. Benchmark circuits show that our ILP approach works well in practice.

It is noteworthy to mention that, although the proposed approach can reduce the total power consumption in the high-level synthesis stage, it requires different voltages for different functional units in the physical layout stage. As a result, the implementation of multi-voltage design will introduce some overheads in the physical layout stage. Our future work is to take these overheads into account in the high-level synthesis stage.

ACKNOWLEDGMENTS

This work was supported in part by Ministry of Science and Technology, Taiwan, R.O.C., under grant number 106-2221-E-033-071-MY3.

REFERENCES

[1] R. Camposano, “Path-Based Scheduling for Synthesis”, IEEE Trans on CAD, pp. 85-93, 1991.

[2] K. Wakabayashi and H. Tanaka, “Global Scheduling Independent of Control Dependencies Based on Condition Vectors”, Proc. of IEEE DAC, pp. 112—115, 1992.

[3] R.A. Walker and S. Chaudhuri, “Introduction to the Scheduling Problem”, IEEE Design and Test of Computers, pp. 60-69, 1995.

[4] J. Monterio, et al., “Scheduling Techniques to Enable Power Management”, Proc. of DAC, pp. 349-352, 1996.

[5] S.H. Huang, et al., “An ILP Approach to the Simultaneous Application of Operation Scheduling and Power Management”, Proc. of APCCAS, pp.868-871, 2006.

[6] W.T. Yen, et al., “Simultaneous Operation Scheduling and Operation Delay Selection to Minimize Cycle-by-Cycle Power Differential”, Proc. of International Conference on Embedded and Ubiquitous Computing, pp. 638-647, 2007.