Exploiting Narrow Bitwidth Operations for Low Power Embedded Software Design

Yamaguchi, Seiichiro
Department of Computer Science and Communication Engineering, Graduate School of Information Science and Electrical Engineering, Kyushu University

Muroyama, Masanori
System LSI Research Center, Kyushu University

Ishihara, Tohru
System LSI Research Center, Kyushu University

Yasuura, Hiroto
Department of Computer Science and Communication Engineering, Graduate School of Information Science and Electrical Engineering, Kyushu University | System LSI Research Center, Kyushu University

https://doi.org/10.15017/6320

Exploiting Narrow Bitwidth Operations for Low Power Embedded Software Design

Seiichiro Yamaguchi† Masanori Muroyama‡ Tohru Ishihara‡ Hiroto Yasuura‡

† Department of Computer Science and Communication Engineering, Graduate School of Information Science and Electrical Engineering, Kyushu University
6-1 Kasuga-koen, Kasuga-shi, Fukuoka 816-8580 JAPAN
{seiichiro, yasuura}@c.csce.kyushu-u.ac.jp

‡ System LSI Research Center, Kyushu University
3-8-33 Momochihama, Sawara-ku, Fukuoka-shi, Fukuoka 814-0001 JAPAN
{muroyama, ishihara}@slrc.kyushu-u.ac.jp

Abstract - This paper proposes a low power software design technique for processor-based embedded systems. A basic idea is to reduce switching activities in sign extension bits of instruction operands through shifting the operands. To the best of our knowledge, this is the first software-level power reduction technique which exploits narrow bitwidth operations. Our technique, called shifted operation technique in this paper, consists of following three steps: 1) shift operands from LSB side toward MSB side by optimal shift amount, 2) execute instructions with the shifted operands, and 3) shift back computational results to original positions. Experimental results exploiting shifted operation technique show about 10.7% energy reduction of datapath, and about 5.5% energy reduction of overall processor.

I. Introduction

Short time to market, low cost and low power are important requirements in embedded system design. Especially, low power is the most important requirement in portable systems such as cellular phones and PDAs. In embedded system design environment, degrees of freedom in hardware are often very limited, whereas much more freedom is available in software. In addition to the benefit, software-oriented optimization is applicable to general purpose processors. In this paper, we propose a low power software design technique for processor-based embedded systems.

Processors are implemented by digital CMOS circuits. There are three major sources of power consumption in digital CMOS circuits. These sources are switching power, short-circuit power and leakage power [1]. In particular, the switching power for charging and discharging of load capacitance is the most major source of the power consumption. The switching power is shown as follows:

\[ P_{sw} = \alpha \cdot C_L \cdot V_{dd}^2 \cdot f_{clk} \]  \( \cdots \)(1)

where \( \alpha \) denotes the switching activity factor, \( C_L \) represents the load capacitance, \( V_{dd} \) is the supply voltage and \( f_{clk} \) is the clock frequency. According to (1), several approaches can reduce the switching power. In this paper, we focus on reducing the switching activity factor. Since the short-circuit power is consumed at the time of transistor switching, the short-circuit power can be reduced through reducing the switching activity factor as well. The switching power and the short-circuit power are called dynamic power, which is our target to reduce.

To design a low power processor-based embedded system, we propose a novel low power design technique exploiting narrow bitwidth operations at software-level. This is the first software-level power reduction technique. Since software-level power reduction techniques do not require any hardware modifications, the techniques can be easily applied to existing processor-based embedded systems. Therefore, our technique satisfies the requirements of short time to market and low cost. Two’s complement representation is typically chosen to represent numbers since arithmetic operations are easy to perform in processor systems. One of problems with two’s complement representation is sign extension. Due to sign extension, an arithmetic operation for narrow bitwidth operands requires the dynamic power throughout entire word length. We reduce the dynamic power consumption due to sign extension through shifting the narrow bitwidth operands. Our technique, called shifted operation technique in this paper, consists of following three steps: 1) shift operands from LSB side toward MSB side by optimal shift amount, 2) execute instructions with the shifted operands, and 3) shift back computational results to original positions. Shifted operation technique is effective to overall datapath of processor, because sign extension affects the dynamic power consumption of not only arithmetic circuits, but also buses, registers and logical circuits.

In this paper, we present a mechanism of the power reduction exploiting narrow bitwidth operations. To find optimal shift amount for each variable of a target application source code, we formulate shift amount optimization problem. Shifted operation technique generates low energy source codes through inserting shift operations to the target application source code according to the optimal shift amount for each variable. This paper is organized as follows: Section II discusses related work and presents our approach. In Section III, we formulate shift amount optimization problem. Experimental results are presented in Section IV. Finally, Section V concludes this paper.

II. Related Work and Our Approach

A. Previous Work

There are several approaches for reducing the switching activities in sign extension bits. Using other number representation is one of the approaches. Sign-magnitude representation in which only one bit is allocated for the sign and the rest for the magnitude [2]. Significance compression is also effective through appending two or three extra bits to represent significant bytes [3]. Reduced two's-complement...
representation generates a representation dynamically according to a magnitude of number [4]. In addition, low power adder operating which considers narrow bitwidth operands dynamically is proposed in [5], and several bus coding techniques are discussed in [6], [7], [8]. Brooks et al. showed that over half of integer operation executions require 16-bits or less across SPECint95 benchmarks [9]. We also exploit this fact for low power embedded software design.

These techniques mentioned above need some hardware modifications. On the other hand, shifted operation technique does not need any hardware modifications. Therefore, shifted operation technique can be applied to existing processor-based embedded systems.

B. Motivational Example

To illustrate a key point of shifted operation technique, we introduce a motivational example by using 32-bits Ripple Carry Adder (RCA). In Fig. 1, two input operands, \( x(t) \) and \( y(t) \), represented by two’s complement are added where \( x(t-1) \) and \( y(t-1) \) are previous input operands. Since the dynamic power of RCA originates from the switching activities of two input operands changing from \( x(t-1) \) and \( y(t-1) \) to \( x(t) \) and \( y(t) \), four input operands have to be considered. Significant bits of all the operands are only 8-bits. Rectangles indicate significant bits of the operands. The MSB in significant bits is the sign bit. Remaining upper 24-bits are sign extension bits for conventional operation. Meanwhile, lower 24-bits are all ‘0’ for shifted operation because the operands are shifted from LSB side toward MSB side firstly. A large number of the switching activities due to sign extension are generated when conventional operation is executed. On the other hand, there is no switching activity due to sign extension when shifted operation is applied.

We compared the dynamic power consumptions at gate-level by using Cadence NC-Verilog to count the switching activities of all nets, and using Synopsys Power Compiler to estimate the dynamic power consumptions. A process technology we used is Hitachi CMOS 0.18μm standard cell library. Supply voltage is 2.5V and clock frequency to RCA is 10MHz. In this case, the dynamic power consumptions of RCA are 61.3\( \mu \)W for conventional operation and 12.0\( \mu \)W for shifted operation. The dynamic power consumption is reduced 80.4% through applying shift operation technique. Note that we ignored overheads to shift operations of each instruction. On the other hand, we shift each variable of a target processor-based embedded system.

overheads. It is clear that the overheads dramatically increase if we insert shift operations before and after every instruction. To keep down the overheads, we shift each variable of a target application source code only once according to the operational bitwidth. Since the dynamic power consumption strongly depends on the operational bitwidth and the overheads, it is necessary to find optimal shift amount for each variable. We should consider additional time/power overheads due to shift operations. Shifted operation technique reduces energy consumption which is a product of average power consumption and execution time. Fig. 2 shows the overview of our approach, and details of the overview are as follows:

**Given data [abbr.]:**
- C source code of target application [Target-C]
- Chip or HDL code of target processor [IP]
- Target process technology [Process]
- Training data for instruction trace [Data1]
- Training data for energy characterization [Data2]

**Given tools [abbr.]:**
- Compiler for the target processor [Compiler]
- Instruction set simulator for the target processor [ISS]
- Energy estimation tools [Tool]

**Procedure prototype [abbr.] (in1, ...)- out1 [abbr.], ...:**
- Compile [Compile] (Target-C, Compiler)
- Assembler source code of the target application [Asm]
- Object code of the target application [Obj]
- Variable size analysis [VSA] (Target-C)
- Effective bitwidth for each variable [EB]
- Operand size analysis [OSA] (Asm, EB)
- Minimum operational bitwidth for each instruction [MOB]
- Instruction trace [Trace] (Obj, Data1, ISS)
- Trace data [T-Data]
- Energy characterization [EC] (IP, Process, Data2, Tool)
- Energy consumption model for each circuit module [EM]
- Shift amount analysis for each variable [SAA] (EB, MOB, T-Data, EM)
- Code Optimization [Optimization] (Target-C/Asm, OSA)
- Optimized C/assembly source code [Optimized-C/Asm]
Four parameters are needed to find the optimal shift amount for each variable: effective bitwidth for each variable, minimum operational bitwidth for each instruction, trace data and energy consumption model for each circuit module.

First of all, we compile C source code of a target application program. For example, if an integer type variable is set to 1 if instruction \(i\) is executed on circuit \(m\) immediately after instruction \(i'\). Note that the \(E_{m,i,i',w,w'}\) can be modeled by following parameters:

- Type of circuit module \(m\) on which instruction \(i\) and \(i'\) are executed.
- Operational bitwidth of instruction \(i\) (w-bits).
- Operational bitwidth of instruction \(i'\) (w'-bits).
- Effective bitwidths for each operand of instruction \(i\).
- Effective bitwidths for each operand of instruction \(i'\).

\(X_{m,i}\) : Number of executions for instruction \(i\) executed on circuit module \(m\) immediately after instruction \(i'\).

\(B_i\) : Minimum operational bitwidth which guarantees instruction \(i\) to be executed without sacrificing any computational precisions.

\(v_{m,i,i',w,w'}\) : 0-1 integer variable to be determined. The variable is set to 1 if instruction \(i\) is executed on circuit module \(m\) with \(w\)-bits and instruction \(i'\) is executed with \(w'\)-bits immediately before instruction \(i\). Otherwise it is set to 0.

B. ILP Formulation

**Shift amount optimization problem** is formulated as follows:

**Objective function to be minimized:**

\[
\sum_{m=1}^{M} \sum_{i=1}^{N_m} \sum_{i'} \sum_{w=1}^{W} \sum_{w'=1}^{W} E_{m,i,i',w,w'} X_{m,i,i',w,w'} v_{m,i,i',w,w'} \cdots (2)
\]
For each \(i\) and \(i'\), \(\sum_{w=1}^{W} \sum_{w'=1}^{W} v_{m,i,i',w,w'} = 1 \cdots (3)\)

where \(i'\) denotes an instruction executed on circuit module \(m\) immediately before instruction \(i\). Note that a number of \(i's\) can be more than one.

For each \(i\) and \(i'\), \(\sum_{w=1}^{W} \sum_{w'=1}^{W} w \cdot v_{m,i,i',w,w'} \geq B_i \cdots (4)\)

where \(i'\) denotes an instruction executed on circuit module \(m\) immediately before instruction \(i\). Note that a number of \(i's\) can be more than one.

For each \(i\) and \(i'\), \(\sum_{w=1}^{W} \sum_{w'=1}^{W} w' \cdot v_{m,i,i',w,w'} = \sum_{z=1}^{W} \sum_{z'=1}^{W} z \cdot v_{m,j,j',z,z'} \cdots (5)\)

where \(i'\) denotes an instruction executed on circuit module \(m\) immediately before instruction \(i\). Note that a number of \(i's\) can be more than one.

For each \(i\) and \(j\), \(\sum_{w=1}^{W} \sum_{w'=1}^{W} w \cdot v_{m,i,i',w,w'} = \sum_{z=1}^{W} \sum_{z'=1}^{W} z \cdot v_{m,j,j',z,z'} \cdots (6)\)

where \(j\) denotes an instruction which shares at least one variable with instruction \(i\). Note that a number of \(j's\) can be more than one.

The first constraint (3) indicates that only one operational bitwidth should be assigned to each instruction. The second constraint (4) means that operational bitwidth for instruction \(i\) must be greater than or equal to the minimum operational bitwidth which guarantees instruction \(i\) to be executed without sacrificing any computational precisions. The third constraint (5) indicates that operational bitwidth \(w'\) determined by \(v_{m,i,i',w,w'}\) and that determined by \(v_{m,i,i',w,w'}\) should be consistent. The last one (6) means that operational bitwidths of instructions which use a same variable as its operand should be equal to each other. Shift amount optimization problem can be formally defined as follows:

“For given \(P_{m,b}, E_{m,i,i',w,w'}, X_{m,i,i',w,w'}\) and \(B_p\), find a set of \(v_{m,i,i',w,w'}\) which minimizes the objective function (2) under constraints of (3), (4), (5) and (6).”

Clearly shift amount optimization problem defined above is an integer linear programming (ILP) problem. Therefore, complexity of the problem depends on a polynomial of the number of variables to be determined \((M \times N^2 \times W^2)\) and the number of constraints \((M \times N^2)\). Here \(N\) denotes the number of instructions in the object code. Shift amount optimization problem cannot be solved within a practical time if \(M, N\) and \(W\) are large numbers. Our idea for reducing the complexity of the problem is to reduce \(N\) and \(W\). Operational bitwidth for each instruction can be any integer values between 0 to \(W\). However, if we limit the shift amount into predefined values, the complexity can be reduced. For example, if we limit the shift amount into multiples of 4, the problem size can be quartered. Since \(N\) is the number of instructions in the object code, \(N\) can be decreased by reducing the target code size. For example, we can reduce \(N\) by applying our technique to frequently executed functions only. If the complexity is still high, then we need to come up with new heuristic algorithm.

IV. Experiments and Results

A. Experimental Framework

To evaluate the effectiveness of shifted operation technique, we experiment under following conditions:

Given data:
- Target-C: C source code of 16-points moving average filter
- IP: HDL code of M32R-II processor core
  - (32-bits RISC processor of Renesas Technology)
- Process: Hitachi CMOS 0.18μm standard cell library
- Data1: Sign wave (# of samples: 4096, 1 sample: 16-bits)
- Data2: Random data

Given tools:
- Compiler: m32r-linux-gcc
- ISS: m32r-linux-run
- Tool: Cadence NC-Verilog, Synopsys Design Compiler and Synopsys Power Compiler

Other conditions, supply voltage is 2.5V and clock frequency to M32R-II processor core is 10MHz. Fig. 3 shows the architecture of 16-points moving average filter we used as Target-C.

We use an optimization option “-O3” when we compile the C source codes. Energy characterization estimates the average dynamic power consumptions for each circuit module with synthesized HDL code of processor core by using Synopsys Design Compiler to synthesize the HDL code, and using Cadence NC-Verilog to count the switching activities of all net, and using Synopsys Power Compiler to estimate the average power consumptions. We insert shift operations to original source codes by hand according to the results of shift amount optimization problem which is the optimal shift amount for each variable.

After inserting shift operations, we compile the optimized C source code with an optimization option “-O3”, and assemble the optimized assembler source code with no options. In the first case, there is a possibility that the original object code and the optimized one are different. In the latter case, those object codes are always same except shift operations inserted by shifted operation technique. This means that the shifted operation technique at assembler source-level has execution cycle overheads absolutely. These three object codes are simulated on the synthesized HDL code of processor core to estimate the average dynamic power consumptions, and to obtain the execution times.

B. Experimental Results

The result of variable size analysis for 16-points moving average filter is tabulated in TABLE I. Since the bitwidth of
input data is 16-bits, the effective bitwidth of all the variables except \( \text{sum}[n] \) are 16-bits. However, we cannot shift the variables by 16-bits, because the minimum operational bitwidth is defined as the largest effective bitwidth of instruction operands.

Fig. 4 shows power estimation results of 32-bits RCA for shifted operation technique while changing the operational bitwidth at \( t \) and \( t-1 \), respectively. Assume that the effective bitwidths of input operands are equal to each other, and equal to the operational bitwidth. Therefore, if the operational bitwidth is 8-bits, then lower 24-bits of the operands are all ‘0’. Fig. 3 shows that the average dynamic power consumption is roughly proportional to the larger operational bitwidth either at \( t \) or \( t-1 \).

We find that the optimal shift amount for all the operands are 12-bits through solving shift amount optimization problem. Additional time overheads due to shift operations are shown in TABLE II. The number of execution cycles increases only 1.2% in assembler source-level optimization. On the other hand, the number of execution cycles decrease 3.4% in C source-level optimization. We have to discuss about reasons of the results. This is our future work.

TABLE III shows both cases can reduce power/energy consumption of datapath even if assembler source-level optimization increased the number of execution cycles. The results show about 7.6%/4.9% power reduction and about 10.7%/3.7% energy reduction for shifted operation technique at C/Assembler source-level optimization. In addition to the effectiveness, TABLE IV shows about 2.2%/2.7% power reduction and about 5.5%/1.5% energy reduction of overall processor.
V. Conclusions

We have proposed a novel low power software design technique, called *shifted operation* technique, exploiting narrow bitwidth operations for processor-based embedded systems. *Shifted operation* technique consists of following three steps: 1) shift operands from LSB side toward MSB side by optimal shift amount, 2) execute instructions with the shifted operands, and 3) shift back computational results to original positions. We also have presented *shift amount optimization problem* as a 0-1 integer linear programming problem. Experimental results exploiting *shifted operation* technique have shown about 10.7% energy reduction of datapath and about 5.5% energy reduction of overall processor. Our future work are as follows:

- To define an algorithm to solve *shift amount optimization problem*
- To discuss reasons of the energy reduction
- To experiment by using other benchmarks and other processors

Acknowledgements

This work has been supported in part by the Grant-in-Aid for Creative Scientific Research No. 14GS0218 and the grant of Fukuoka project in the Cooperative Link of Unique Science and Technology for Economy Revitalization (CLUSTER) of the Ministry of Education, Culture, Sports, Science and Technology (MEXT). This work is supported by Core Research for Evolutional Science and Technology (CREST) of Japan Science and Technology Corporation (JST). This work is supported by VLSI Design and Education Center (VDEC), The University of Tokyo with the collaboration of Renesas Technology, Hitachi, Ltd, Cadence Design Systems, Inc. and Synopsys, Inc. We are grateful for their support.

References