Low-Energy Design Using Datapath Width Optimization for Embedded Processor-Based Systems

YUN CAO\textsuperscript{\dagger} and HIROTO YASUURA\textsuperscript{\dagger}

This paper presents a novel system-level technique that minimizes the energy consumption of embedded processor-based systems through datapath width optimization. It is based on the idea of minimizing energy consumed by redundant bits, which are unused during execution of programs by means of optimizing the datapath width of processors. To minimize the redundant bits of variables in a given application program, the effective size of each variable is determined by variable size analysis, and Valen-C language is used to preserve the precision of computation. Analysis results of variables show that there are average 39\% redundant bits in the C source program of MPEG2 video decoder. In our experiments for several real embedded applications, energy savings without performance penalty are reported and range from about 10.8\% to 48.3\%.

1. Introduction

Minimizing power consumption of embedded systems is a crucial task. Battery-operated portable systems demand tight constraints on energy consumption. Better low-power circuit design techniques and advances in battery technology have helped to increase battery lifetime. On the other hand, managing power dissipation at higher design levels can considerably reduce energy consumption, and thus increase battery lifetime. Energy consumption at all design levels should be considered to reduce power of the whole embedded system.

We have developed a design platform, which consists of a Valen-C retargetable compiler 2), soft-core processor (Bung-DLX) 1) 3) and a cycle-based simulator 4). We also have done some researches on reduction of area and cost for embedded core-based systems 5) 6). In this paper, we focus on minimizing energy dissipation and present a system-level technique for embedded processor-based systems, which minimizes energy consumption of the whole system while providing adequate performance level. In the initial design phase of our approach, we design a system with a soft-core processor, data RAMs, instruction ROMs and logic circuits. Then we analyze the effective bit width of each variable of a given application program. After that, using the results of analysis, we rewrite the application program in Valen-C language 2), in which we specify the word length of each variable satisfying accurate computation to reduce energy consumed by redundant bits in the application program. After verifying the functionality of the initial design, we modify several design parameters of the soft-core processor, including the datapath width, the number of registers and the instruction set. We can tune up the soft-core processor to minimize the energy consumption while satisfying the system performance constraints. To get first-cut estimates of energy consumption early in the design, a few component-based power estimation models are also developed, total energy is obtained by summing over all components of the system.

This paper is structured as follows: the next Section 2 gives an overview of related work. Section 3 describes our energy minimization technique by datapath width optimization. Section 4 presents our energy estimation models. Experiments and results are shown in section 5. Finally, Section 6 concludes our work.

2. Related Work

Hardware and software techniques to reduce energy consumption have become an essential part of current system designs. Extensive researches on power optimization from circuit level to system level have been conducted in these recent years. Such techniques have particularly targeted the memory system 7) 8) 9) due to the prevalent use of data-dominated signal and video applications. 14) focuses on exploiting cache to reduce power consumption. The work 11) presented an architecture-oriented power minimization approach. A power and performance simulation tool that can be used to do architecture-level optimizations has been in-

\textsuperscript{\dagger} Kyushu University

128
3. An Energy Minimization Approach

In the design of consumer electronic systems, designers have to manage rapid increase of complexity of a target system with requirements on high-performance and low-energy consumption under the tight constraint of short design time. Therefore, core-based solutions are proposed for embedded system design. Our approach gives designers a freedom to determine the datapath width of soft-core processor, because the datapath width of a processor has great impacts not only on power consumption and performance of the processor but also on those of memories. Optimizing datapath width for each given application is an effective approach to minimize energy consumption of the whole embedded systems.

3.1 Problem Description

We focus on reducing the energy consumption of an embedded system by tuning the datapath width of processors. We define an embedded system as a processor along with its data memory and instruction memory, all implemented on a single chip (SOC). The soft-core processor (Bung-DLX) that we used in this paper is a variable configuration RISC processor that can be synthesized to any bit-width and register file size. The instruction memory is implemented in ROM and the data memory is implemented in RAM that has a word length equal to the width of the processor datapath. The instruction word length remains invariant in this paper.

Valen-C retargetable compiler is produced by a multi-precision compiler generator that automatically emits the necessary multi-precision instructions, according to the processor datapath and the bit width specification of each variable.

The energy minimization problem is formulated as:

\[
\begin{align*}
\text{minimize} & \quad \text{Energy}(w) \\
\text{subject to} & \quad \text{Cycle}(w) \leq C_{\text{cst}} \\
& \quad \text{Area}(w) \leq A_{\text{cst}}
\end{align*}
\]

Where \( \text{Energy}(w) \), \( \text{Cycle}(w) \) and \( \text{Area}(w) \) are functions of the datapath width \( w \), \( C_{\text{cst}} \) and \( A_{\text{cst}} \) are the constraints on the execution cycle and area respectively. This is a nonlinear optimization problem. The overview of our energy minimization algorithm is described in Figure 2.

3.2 Our Approach

Figure 1 shows our proposed approach, which consists of the following phases:

- **Phase 1**: The source program of the target application, which was originally written in C or other language, is rewritten in Valen-C language, after the bit width of each variable is analyzed. For instance, if the variable \( x \) requires at most 11 bits, the programmer can write `int 11 x;` in the variable declaration of Valen-C program.
- **Phase 2**: Bung-DLX is customized to different soft-core processors by choosing different design parameters, such as the datapath width and the address size of the data memory.
- **Phase 3**: The Valen-C source program of an application is compiled for the customized...
In this paper, we explain our methods to analyze effective sizes of program variables in C programs. In order to optimize the datapath width, the effective size of each variable in an application needs to be analyzed. This section explains our approach to this problem, we adopted dynamic analysis. In dynamic analysis, we execute programs and monitor the value of each variable. Dynamic analysis is one kind of simulation-based method whose results depend on input data sets given to the programs. The other is static analysis.

For static analysis, when the maximum value of an unsigned integer variable \( x \) is \( n_{\text{max}} \), the effective size of \( x \), \( e(x) \), is given as follows:

\[
e(x) = \log_2(n_{\text{max}} + 1)
\]

(1)

For a signed integer \( x \) with a maximum value \( n_{\text{max}} \) and a minimum value \( n_{\text{min}} \), \( e(x) \) is defined as follows:

\[
e(x) = \lceil \log_2 N \rceil + 1
\]

(2)

where

\[
N = \max(|n_{\text{max}}| + 1, |n_{\text{min}}|)
\]

(3)

Static analysis is an efficient method to analyze the effective size of variables. However, in many cases when we can not predict the assigned value of a variable unless we execute the program, such as the case of unbounded loops, static analysis becomes insufficient. As a solution to this problem, we adopted dynamic analysis in our approach.

Figure 3 shows a part of the algorithm used for dynamic analysis.
3.4 Efficient Use of Data Memory

Since in many cases, high-level specifications are devoted to describe functionalities of target systems rather than implementation details, they often contain a lot of redundancies such as duplicated computations and never executed code. Therefore, the specifications must be optimized to remove the redundancies for energy-efficient design. Some redundancies are introduced in size of variables. For example, in C programs, a variable whose value is between 0 and 1000 is often declared as the int type, i.e., usually 16 or 32 bits depending on target processors, and then some upper bits make nonsense. This means that the memory has many unnecessary bits, which do not essentially contribute to the calculation of programs. Therefore redundant bits should be removed to reduce power consumption.

C language provides for three integer sizes, declared using the keywords short, int and long. The compiler designer determines the sizes of these integer types. In many processors, the size of short is 16 bits, int is 16 or 32 bits, and long is 32 bits. On the other hand, in Valen-C, programmers explicitly specify the required bit width of each integer data type. Thus it becomes possible to reduce the energy of the datapath and the data memory, which is dissipated by the redundant bits. For instance, if variables x, y, and z require 12, 20 and 24 bits respectively, the programmer can write “int12 x; int20 y; int24 z;” in the variable declaration of Valen-C program. If a processor with a datapath width of 20 bits is used in the system, the total memory size will be 80 bits. Moreover, the unused bits in the data memory will be 24 bits. On the other hand, if a processor of a datapath width of 12 bits is used, the total data memory size will become only 60 bits, and the unused memory size will decrease to 4 bits. As a result, specifying the word length required for each variable and changing the datapath width have a significant role in reducing the data memory size of a system. Therefore it also affects the power consumption of the system.

3.5 Datapath Width Optimization

System designers can tune the value of the datapath width in accordance with the characteristics of target system to deliver most suited processor. Designers can reduce the datapath width until the single precision point (SPP) without performance loss. SPP is the processor datapath width, which is equal to the bit width of the largest variable in a program. It is the smallest datapath width at which all instructions can remain single-precision. Designers may obtain better solutions, more power savings by shrinking the datapath less than SPP, under performance constraints. Figure 4 shows the overview of our datapath width optimization algorithm.
3.6 Power Versus Performance Trade-off

Minimizing power consumption is not simply an altruistic activity. A device consuming less power will accrue several desirable advantages such as longer battery life for wireless devices, but somewhat less obvious advantages, such as reliability and performance. The datapath width of a processor strongly affects the power consumption of the whole system including the processor, data memories and instruction memories, it also affects the execution cycles of a given task, i.e., narrowing the datapath width less than SPP will cause the increase of execution cycles because of multiple-precision operations. For example, that an addition of 20 bit data is executed by only one instruction on a 20 bit processor is assumed, If the datapath width becomes to 10 bits, two instructions including additions of lower 10 bits and high 10 bits with carry are required. So trade-offs exist between datapath width and execution cycles.

Although a processor with narrower datapath width dissipates lower power per clock cycle, the total energy for the task is not reduced always by narrowing the datapath width. Thus, for a given target system, trading off the power consumption and performance is an important work.

4. Energy Estimation Models

This section describes energy consumption models. The total energy consumption, $E$, is the summation of energy consumed by the processor ($E_{\text{proc}}$) and memories ($E_{\text{mem}}$).

$$E = E_{\text{proc}} + E_{\text{mem}}$$  \hspace{1cm} (4)

We estimated $E_{\text{proc}}$ and $E_{\text{mem}}$ separately, and got the energy consumption model of our soft-core processor generated by HITACHI 0.5um CMOS technology and the energy consumption models of memory generated by Alliance CAD System Ver.3.0 with 0.5um double metal CMOS technology.

$E_{\text{proc}}$ is given by

$$E_{\text{proc}} = \sum_{i \in I} e_i \times \text{Cycle}_i$$  \hspace{1cm} (5)

where

- $e_i$ : Average energy of instruction $i$
- $\text{Cycle}_i$ : The number of execution of instruction $i$
- $I$ : Instruction set of Bung-DLX

$e_i$ is obtained by performing post-layout simulation of switch-level. After several simulations, we obtained the empirical energy model at several datapath widths in Figure 5, where power savings are got by comparing to the power consumption of 32bits Bung-DLX. The power dissipation in static CMOS can be divided into static, dynamic and short-circuit power. Because static power and short-circuit power are far less than dynamic power, we just focus on dynamic power, which consists of Cell Internal Power($P_c$) and Net Switching power($P_s$).

$$e_i = \frac{1}{2} \times V_{dd}^2 \sum_{\text{net}} \left( C_j \times S_j + E_{ck} \times S_k \right)$$  \hspace{1cm} (6)

where

- $V_{dd}$: Supply voltage
- $C_j$: Load capacitance of net j
- $S_j$: The average number of switching of net j per clock cycle
- $E_{ck}$: Internal power of cell k
- $S_k$: The average number of switching of cell k per clock cycle

$E_{\text{mem}}$ is estimated as follows:

$$E_{\text{mem}} = E_{\text{ROM}} + E_{\text{SRAM}}$$  \hspace{1cm} (7)

$$E_{\text{ROM}} = \epsilon_{\text{ROM}} \times \sum_{i \in I} C_i$$

$$E_{\text{SRAM}} = \epsilon_{\text{SRAM}} \times C_{\text{load}} + \epsilon_{\text{Sw}} \times C_{\text{store}}$$  \hspace{1cm} (8)
The access energy of memories ($e$) is obtained from the SPICE simulation of the memories.

$e_{ROM}$, $e_{SR}$, $e_{SW}$ is obtained from the SPICE simulation of several memories with the different configurations. As the result, we have obtained the estimation models as follows:

$$e_{ROM} = 50.97 \cdot b \cdot \sqrt{N_w} + 1.4[pJ/C] \quad (9)$$

$$e_{SR} = 24.9 \cdot b \cdot \sqrt{N_w} + 56[pJ/C] \quad (10)$$

$$e_{SW} = 197 \cdot b \cdot \sqrt{N_w} + 369[pJ/C] \quad (11)$$

Where $b$ is the word width of the memory and $N_w$ is the number of words. $pJ/C$ means one $pJ$ per cycle.

## 5. Experiments and Results

In this section we present experiments and results based on several real applications to evaluate our proposed approach. We mainly illustrate how we use our approach to minimize energy consumption of MPEG2 video decoder, a relatively large program.

In the experiments, we assumed the target system, a SOC chip, which consists a Bung-DLX processor, a ROM and a SRAM. Bung-DLX is a non-pipelined, simple RISC processor, which has several design parameters including the datapath width and the number of registers. All instructions are executed within a single machine cycle. The ROM and the SRAM are used as instruction memory and data memory respectively. These memories are generated by Alliance CAD System Ver. 2.0 with 0.5µm double metal CMOS technology. For simplicity, we assumed that no other core is integrated in the SOC chip.

### 5.1 Variable Size Analysis for MPEG2

Our program is based on Mpeg2decode program from the MPEG Software Simulation Group. It is a player for MPEG-1 and MPEG2 video bitstreams. Mpeg2decode is an implementation of an ISO/IEC DIS 13818-2 decoder, whose emphasis is on correct implementation of the MPEG standard and comprehensive code structure. We rewrote it in Valen-C with about 6650 lines. The MPEG2 core consists of several function blocks such as a soft-core processor, IDCT blocks, a couple of motion estimation blocks, a motion compensation block, variable length encoding, decoding blocks and so on.

We analyzed the C source program of MPEG2 video decoder and got some analysis results. The number and types of the variables in MPEG2 decoder are described in Table 2.
Table 1. The results of static variable analysis are depicted in Table 2 (E.Size means effective size of variable; N.of Variables means the number of variables), and that of dynamic analysis are shown in Table 3 (V.name means variable name). From Table 2 and Table 3, we can see that there are many redundant bits in the variables of MPEG2 decoder C source program. We got 34% reduction of bits from the static analysis and 52% from the dynamic analysis.

To verify our analysis results of variable size, we used the following model.

\[
PSNR = 10 \times \log_{10} \frac{1}{E} \times 255^2 [dB] \quad (12)
\]

where, \( PSNR \) : Ratio of pick signal to noise
\( E \) : Mean-square error

Our experimental results of \( PSNR \) are infinite, so it shows that the variables, which are assumed according to the analysis results can work exactly as that of the source program of MPEG2 video decoder. Therefore, our analysis results are verified.

5.2 Power and Performance Estimation

This section reports some experimental data concerning the use of our approach to reduce energy consumption. The cycle count is obtained by using our instruction-level simulator. The input of the simulator is the assembly code, which is generated by the retargetable Valen-C compiler. Results of energy consumption \( E_t \) (shown in Figure 6) include energy of a soft-core processor \( (E_p) \), a data RAM \( (E_d) \) and an instruction ROM \( (E_r) \), where \( D.W \) is datapath width. We use the energy consumption models in section 4. Apparently, the energy consumption changes nonlinearly.

Figure 7 shows the energy consumption, execution cycles and area (gates) of MPEG2 video decoder, and we got the optimal datapath width, 28bits for MPEG2 video decoder. Figure 8 describes the energy savings of our benchmarks, such as Lempel-Ziv algorithm, ADPCM encoder, and MPEG2 AAC decoder and so on. No Opt. means the original datapath width of Bung-DLX (32bits). Opt. is the datapath width where the whole system has the minimization energy consumption without performance loss. For Lempel-Ziv algorithm, we got energy savings of 48.3% at datapath width of 15bits, for ADPCM encoder, energy savings is 22.8% at datapath width of 19bits and for MPEG2 video decoder, the energy savings is 10.8% at datapath width of 28bits. For different application, the number of variables is different and the effective size of variables is also different, therefore the optimal datapath width of minimal energy is different. For a given application, our approach just tries to take advantage of the characteristics of the application to reduce the energy consumption.

6. Conclusions

In this paper, we proposed a system-level energy minimization technique through datapath width optimization, which can suit the complexity of embedded systems and stringent time-to-market constraints. We also presented a set of algorithms that minimize energy consumption in system-level. We illustrated issues and tradeoffs involved in the design. Our experimental results show that for a given application we can reduce significantly the energy consumption by datapath width optimization. We have demonstrated energy savings without performance penalty range from about 10.8% to 48.3%, which based on a number of real embedded applications. Extending parameter-tuning for low power to DSPs is our future work.
Acknowledgments  This research was partly supported by the Grant-in Aid for Scientific Research (B) (2) 12558029 and VCDS project of STARC.

References

(Received September 12, 2001)  
(Accepted January 25, 2002)

Yun Cao is a PhD candidate in Department of Computer Science and Communication Engineering, Kyushu University, Japan. She received her B.E. degree of Electronics and Information Engineering from Huazhong University of Science and Technology, China, in 1989. Her research interests are hardware/software co-design, low power/low energy system design and system design methodology. She is a student member of IPSJ, IEEE and ACM.

Hiroto Yasuura is a professor of Department of Computer Science and Communication Engineering, Graduate School of Information Science and Electrical Engineering, and the director of System LSI Center of Kyushu University, Fukuoka, Japan. He received the B.E., M.E., and PhD degrees in computer science from Kyoto University. His current interests include parallel computer architectures, hardware algorithms for VLSI, VLSI CAD, and system design methodology. He is a member of IPSJ, ACM, and IEEE.