A Hardware/Software Partitioning Algorithm for Pipelined Instruction Set Processor

Nguyen Ngoc Binh, Masaharu Imai,
Akichika Shiomi, and Nobuyuki Hikichi*

Department of Information and Computer Sciences
Toyota University of Technology
Toyota, 441 Japan
* Software Research Associates, Inc.
Tokyo, 170 Japan

Abstract
This paper proposes a new method to design an optimal instruction set for pipelined ASIP development using a formal HW/SW co-design methodology. The co-design task addressed in this paper is to find a set of HW implemented operations to achieve the highest performance of a pipelined ASIP under a given gate count and power consumption constraint. The method enables to estimate the performance and pipeline hazards of the designed ASIP very accurately. The experimental results show that the proposed method is effective and quite efficient.

1 Introduction
Pipelining is a simple yet effective technique for increasing parallelism and the utilization of Functional Units (FUs) [1]. It is frequently used for increasing the performance of an instruction set processor by overlapping the execution of instructions. The total execution cycles needed for a series of instructions are reduced in proportion to the amount of overlap in their processing (i.e., the number of stages in the pipeline).

In designing the pipelined instruction set processor it is necessary to deal with all types of pipeline hazards, especially when multi cycle and pipelined FUs are used. In the traditional Application Specific Integrated Processor (ASIP) design methodology, system architects decide which operations will be implemented in hardware (HW) or in software (SW). In order to produce an efficient design in a reasonable turn-around-time (TAT), an efficient HW/SW co-design partitioning method should be used. In practice, design process still depends on the designer’s skill, and the best possible design is usually achieved after investigating a number of design candidates.

This paper focuses on the optimization of a pipelined instruction set processor using a formal HW/SW co-design methodology. In the case of area and power consumption constrained design, some of the operations may be implemented in SW using an ALU only. During the execution of these SW operations, other ALU operations must be stalled. However, FUs may operate simultaneously as long as no more than one instruction can be issued at each clock cycle and there are no two instructions which complete their execution at the same time. That is, the method must handle all types of data hazards and structural hazards.

The rest of the paper is organized as follows. Section 2 reviews related work and describes an architecture model to be considered in this paper. Section 3 gives definitions and notations. Section 4 describes a new problem formalization and the proposed algorithm. The effectiveness and efficiency of the proposed algorithm are shown in Section 5 through design samples. Section 6 gives the conclusion and future work.

2 Related Work
A HW/SW co-design system PEAS-I (Practical Environment for ASIP development - type I) [2] has been developed to synthesize an optimal instruction set processor by solving Instruction set implementation Method Selection Problems (IMSP) type 1 and 2. IMSP-1 [3] is set up assuming no interaction among the operations, and each operation was to be implemented using a separate HW module. However, IMSP-2 [4] is an extension of IMSP-1 by taking resource sharing into account.

The target CPU to be generated by PEAS-I belongs to a class of Harvard architecture with separate data bus and instruction bus. The PEAS-I CPU core architecture is shown in Figure 1, where ‘Kernel’ consists of an ALU, a one-bit shifter, and a register file. The CPU core may include other FUs such as multiplier, divider, and so on. The pipelined architecture synthesized by PEAS-I consists of four stages: IF (Instruction Fetch and decode), EX (EXecution), MEM (MEMoery access) and WR (Write back to Register file), respectively. While each of IF, MEM, and WR stages takes only one cycle, EX stage takes one or more cycles. The PEAS-I CPU has a RISC type load/store (register-register) architecture and each control step...
Corresponds to one clock cycle. While the CPU may contain the Kernel and different types of FUs, it is assumed that there are no identical FUs. The architecture has a register bypass to forward the computation results to Kernel or FUs. Each FU can be multi cycle and pipelined. However, in the IMSP-1 and 2 formalizations, the pipeline was not taken into account, in particular, pipeline hazards were not addressed and left for future work [2]. Therefore, we need to develop a method to estimate the effect of pipeline hazards as accurately as possible because they affect the performance of the CPU. Moreover, the method should estimate the execution cycles of the application program, i.e., the performance of the ASIP, as accurately as possible.

![Kernel Diagram](image)

Figure 1: PEAS-I CPU core architecture.

Huang and Despain [5] propose a systematic approach to synthesize an instruction set that the given application SW can be efficiently mapped to a parameterized, pipelined microarchitecture. While their work is similar to our work in terms of the inputs and part of outputs, it is different from our method in terms of approach and efficiency. First, Huang and Despain assume that the designers are required to specify the number of HW resources (in particular, the number of each type of FU), which will take several iterations to find the best allocation; whereas in our method, an optimal pipelined architecture can be selected and generated automatically. Second, Huang and Despain synthesize an instruction set from the assembly code by grouping some instructions into a new one whereas we generate the optimal instruction set from a super set regarding the optimal HW/SW partitioning.

3 Definitions and Notations

The architecture of an ASIP synthesized by the PEAS-I system is based on the GNU C Compiler (GCC) abstract machine model [6]. The GCC Register-Transfer Language (RTL) operations are divided into primitive and basic operations. The primitive operations contain the minimum operations that can be included in the ASIP chip so that it can execute any C program. The primitive operations should be implemented in HW as the Kernel. The basic operations contain other C operators that are not included in the primitive operations. A basic operation can be implemented using some HW choices (such as fast or slow HW modules) or using a SW subroutine (runtime routine) that uses primitive operations and some other basic operations.

The HW/SW partitioning problem in the current version of PEAS-I is defined as follows [4]:

For a whole set of all candidate instructions representing a given application domain, select a set of implementation methods which maximizes the performance of the CPU under the constraints of chip area and power consumption, taking into account the functional module sharing relation among instructions.

In order to formalize the IMSP-2P the following definitions and notations are used in the remainder of this paper.

(1) “n” denotes the total number of basic operations to be considered.

(2) “f_i” denotes the execution frequency count of basic operation #i in the given set of application programs, where 1 ≤ i ≤ n. We denote frequency count of all primitive operations as f_0.

(3) “a_i” denotes an implementation method that realizes operation #i, where x_i may be HW choice or SW, 0 ≤ i ≤ n. Then X = (x_a, x_1, ..., x_n) is a combination of implementation methods to be considered.

(4) “t_i(x_j)” denotes the execution cycles of operation #i when implemented by method x_j, where 0 ≤ i ≤ n.

(5) “u(x_i)” and “p(x_i)” denote the area and power consumption required for implementation method x_i respectively, where 0 ≤ i ≤ n.

(6) “A_{max}” and “P_{max}” denote the available chip area and the maximum power consumption allowable for the computing module in the ASIP chip.

(7) “N” denotes the total number of basic blocks in the application program’s GCC RTL code.

(8) “t(B_j, X)” denotes the execution cycles needed to execute basic block B_j using a combination of implementation methods X, where 1 ≤ j ≤ N.

(9) “F_j” denotes the execution frequency count of basic block B_j in the given set of application programs, where 1 ≤ j ≤ N.

(10) “c_j” denotes clock cycles needed to define control (e.g., branch delay) from block B_j to another one, where 1 ≤ j ≤ N. Here, it is assumed that all branches are taken and delay slot scheduling is not performed.

(11) “b” denotes execution cycles reduced by un-taken branches in execution of the given application program.
Note that $f_i$, $F_j$, $c_j$, and $b$ are computed from the application program and associated input data by using the application program analyzer (APA) of the PEAS-I system.

4 Proposed Method
The problem addressed in this paper for designing an optimal pipelined instruction set processor in PEAS-I is called IMSP-2P (P stands for Pipeline) and can be considered as an extension of IMSP-2 toward the pipelined architecture. Our goal is to detect and resolve pipeline hazards and to increase the performance of pipelined ASIP to be designed as much as possible in the new problem formalization.

4.1 IMSP-2P Formalization
Find a solution vector

$$X = (x_0, x_1, \ldots, x_n)$$

which minimizes the objective function:

$$T(X) = \sum_{j=1}^{N} (F_j \times (t(B_j, X) + c_j)) - b,$$  \hspace{1cm} (1)

subject to the constraints:

$$\sum_{x_i \in S} a(x_i) \leq A_{\text{max}},$$  \hspace{1cm} (2)

$$\sum_{x_i \in S} p(x_i) \leq P_{\text{max}},$$  \hspace{1cm} (3)

where

$$S = \bigcup_{i=0}^{n} \{x_i\}.$$  \hspace{1cm} (4)

4.2 Consideration
We have developed a HW/SW partitioning-oriented pipeline scheduling algorithm [7] to estimate $t(B_j, X)$ for basic block $B_j$ for given HW resources $X$. The pipeline control hazards are addressed in introducing the coefficients $c_j$. Note that the number of clock cycles due to control hazards is equal to $\sum_{j=1}^{N} (F_j \times c_j) - b$. The pipeline scheduling algorithm detects and resolves all types of data hazards and structural hazards by ensuring that no more than one instruction can be issued or completed at each control step. Therefore, the IMSP-2P solver estimates the pipeline execution cycles accurately if the code optimization is performed by the same strategy taken in the scheduling algorithm.

4.3 IMSP-2P Solver
4.3.1 Input and Output
The input to the IMSP-2P solver includes the following items:

1. the GCC’s RTL code of the given application program,
2. $F_j$’s for $j = 1, \ldots, N$,
3. $b$ (# clock cycles reduced by un-taken branches),
4. area and power consumption constraints $A_{\text{max}}$ and $P_{\text{max}}$, and
5. the module information database, which includes execution cycle count, latency, area, and power consumption of each implementation method of all operations.

The output of the IMSP-2P solver includes the optimum implementation method of each basic operation and pipelined schedules of basic blocks. The instruction set of the designed ASIP will include the primitive operations as default and those basic operations that are selected to be implemented in HW. The algorithm tries to automatically integrate the functional modules, which share basic operations, into one HW module whenever possible.

4.3.2 Algorithm
The IMSP-2P can also be solved using the branch-bound method as IMSP-2can. The key to solving problems efficiently by this method is to find a tight lower-bound function to prune as many non-optimum solutions as early as possible. The lower-bound function used in the IMSP-2P solver is as follows:

$$\text{Lower bound} = (f_0 + \text{Stall}_{\text{fast}})$$

$$+ \sum_{i=1}^{d-1} (f_i \times u_i(x_i)) + \sum_{i=d}^{n} f_i,$$  \hspace{1cm} (5)

where

$$u_i(x_i) = \begin{cases} 1, & \text{if } x_i \text{ is a HW implementation } \\ \xi_i(x_i), & \text{if } x_i \text{ is a SW implementation } \end{cases},$$

where the parameter $d$ represents the depth of the node under consideration. The first term in Eq.(5) represents the value independent of $X$, where $\text{Stall}_{\text{fast}}$ is the number of stalls in executing the given application program using the hypothetical FUs of one cycle denoted by $X_{\text{fast}}$ and is computed as follows:

$$\text{Stall}_{\text{fast}} = T(X_{\text{fast}}) - \sum_{i=0}^{n} f_i,$$  \hspace{1cm} (6)

where $T(X_{\text{fast}})$ is computed by using Eq.(1).

The second term in the lower-bound in Eq.(5) represents the value less than determined cost of the already searched path, while the third term represents the minimum cost of the remaining path. Note that instructions are executed in a pipeline manner and can be overlapped. Therefore, in the best case, instructions may be overlapped maximally and there are no idle clock cycles. In this case, instructions will be executed as if they were executed in one cycle each. However, when operations are implemented in SW, they cannot be overlapped since only the Kernel is in
5 Experiments and Results

The IMSP-2P algorithm has been implemented in C and examined on a workstation. A set of sample programs has been performed to evaluate the effectiveness of the algorithm.

5.1 Module Library

We use a module library with both non-pipelined FUs and pipelined FUs such as multipliers and dividers generated using a high-level synthesis system called PARTHENON [9] and cell library VSC470lib (0.8umCMOS) from VLSI Technology, Inc. A 16 MHz clock was assumed in the design of HW modules. The database contains 14 basic operations, each of them has different implementation methods ranging from 2 to 11. The number of leaf nodes in the search tree is of $11^2 \times 7^2 \times 2^8 = 74,373,376$. The whole search tree ranges from $8.2 \times 10^{9}$ to $1.5 \times 10^{8}$ nodes depending on the order of variables ($x_i$) to be examined. Therefore, it is necessary to have an efficient strategy to explore the search space to get an optimal solution in a reasonable time.

Table 1: Part of Module Database with Pipelined Multipliers and Dividers

<table>
<thead>
<tr>
<th>Module</th>
<th>Name</th>
<th>Gate Count</th>
<th>Power*</th>
<th>L</th>
<th>D</th>
<th>Implied Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>kernel</td>
<td></td>
<td>14018</td>
<td>1002.9</td>
<td>1</td>
<td>1</td>
<td>(primitive)</td>
</tr>
<tr>
<td>s_adf1</td>
<td></td>
<td>466</td>
<td>876.3</td>
<td>1</td>
<td>1</td>
<td>ashr, ashr, ashr</td>
</tr>
<tr>
<td>e_adf1</td>
<td></td>
<td>127</td>
<td>172.8</td>
<td>1</td>
<td>1</td>
<td>extendh, extendd, extendq, z_extendh, z_extendd</td>
</tr>
<tr>
<td></td>
<td>mul_p</td>
<td>7744</td>
<td>1106.9</td>
<td>1</td>
<td>1</td>
<td>mul, umul</td>
</tr>
<tr>
<td></td>
<td>mul_q</td>
<td>6118</td>
<td>3008.6</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mul_qp</td>
<td>3367</td>
<td>3042.9</td>
<td>11</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mul_sq</td>
<td>2567</td>
<td>1771.1</td>
<td>32</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mul_sqp</td>
<td>14560</td>
<td>17138.4</td>
<td>4</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mul_sqq</td>
<td>7866</td>
<td>2869.1</td>
<td>8</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>div_sp</td>
<td>4062</td>
<td>2897.1</td>
<td>16</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>div_sqp</td>
<td>19562</td>
<td>25103.2</td>
<td>2</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>div_sqq</td>
<td>10149</td>
<td>13209.9</td>
<td>4</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>div_spq</td>
<td>6800</td>
<td>6407.6</td>
<td>8</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

Part of the module information database used in the experiments is shown in Table 1. In this table, ‘kernel’ represents the minimal HW components. The module ‘s_adf1’ denoted a barrel shifter that performs both arithmetic and logical shift operations such as ashr, ashr, lshl, lshr. The module ‘extend’ performs extension operations such as extendh, extendd, z_extendh, and z_extendd. The modules ‘mul_pco’, ‘mul_3cl’ ‘mul_bpr’, and ‘mul_seq’ denote multipliers that execute a 32-bit × 32-bit multiplication such as mul or umul in 1, 3, 17, and 32 clock cycles, respectively. The modules ‘div_2seq’ and ‘div_seq’ represent dividers that execute 32-bit division operations such as div, udiv, mod, and umod in 19 and 35 clock cycles, respectively. These modules are non-pipelined, therefore, each delay and latency are the same. The pipelined multipliers are denoted as ‘mul_seq_p17’, ‘mul_bpr_p5’, and dividers as ‘div_seq_p17’, ‘div_2seq_p9’.

Table 2: Expected execution cycles of SW implemented operations

<table>
<thead>
<tr>
<th>Basic Operation</th>
<th>#Cycles</th>
<th>Basic Operation</th>
<th>#Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>div</td>
<td>216</td>
<td>udiv</td>
<td>96</td>
</tr>
<tr>
<td>udiv</td>
<td>202</td>
<td>mod</td>
<td>214</td>
</tr>
<tr>
<td>mod</td>
<td>121</td>
<td>umod</td>
<td>201</td>
</tr>
<tr>
<td>umod</td>
<td>31</td>
<td>extendh</td>
<td>10</td>
</tr>
<tr>
<td>ashr</td>
<td>31</td>
<td>extendd</td>
<td>9</td>
</tr>
<tr>
<td>lshl</td>
<td>31</td>
<td>z_extendh</td>
<td>2</td>
</tr>
<tr>
<td>lshr</td>
<td>31</td>
<td>z_extendd</td>
<td>1</td>
</tr>
</tbody>
</table>

Another part of the module database describes the expected execution cycles for each basic operation to be implemented in SW by using the kernel only. This part is shown in Table 2, where trunc支部 and truncq represent the truncation operations.

5.2 Sample Programs

The sample programs used in the experiments are as follows:

(1) ESS : Equation System Solver program, which solves a system of two linear equations using Cramer’s rule.
(2) IMC : Inverse Matrix Calculator program that computes the inverse of a non-singular 3 × 3 matrix using Cramer’s rule.
(3) diffq : A program for solving a second order differential equation from Ref. [8].

These sample programs were fed to APA of the PEAS-I system. The code optimization was performed by the GNU C Compiler [6].

5.3 Algorithm Efficiency

Experimental results show that the proposed algorithm with the lower bound function in Eq.(5) is very efficient. Selecting the optimum architecture out of $7.44 \times 10^9$ combinations is not an easy task, even to be solved by using the integer linear programming (ILP) approach.

An analyzer has been developed to reduce the search space and to get necessary information for the proposed algorithm, such as GCC RTL code, execution frequency ($F_j$) of each basic block (BB), data dependencies between instructions in BB, and the num-
ber of basic operation types in each BB. A basic block is said to be dependent on $X$ if it contains instruction(s) with the basic operation(s), otherwise it is called independent of $X$. Note that a combination of implementation methods $X$ is determined when a leaf node in the search tree has been reached. Then, using the pipeline scheduler [7] the IMSP-2P algorithm computes the values $t(B_j, X)$ for BBs only when these BBs are dependent on $X$, for other BBs (i.e., independent of $X$) they are computed once before exploring the search tree.

It has been found that the number of BBs dependent on $X$ (denoted as $N'$) is usually much smaller than the total number of BBs of the given application program. The number of basic operation types is six for these sample programs. The reduced search space contained 4935 nodes. The analyzed results are summarized in Table 3. The number of performed experiments is shown in Table 3 (#Cases) for each sample program. Note that while the total number of nodes in each reduced search tree is about 5000, the average numbers of visited nodes are 75, 114, and 112 for the ESS, IMC and diff eq examples, respectively. That means Eq.(5) employs a good lower bound function. Throughout the experiment, the algorithm gave the optimum solution within a few seconds on a SPARC station 10 workstation, including the time for analyzing the input data to get necessary information such as data flow graph, the number of basic operations, reduced search space, and so on. For a given area constraint, the IMSP-2P solver has taken CPU time of 2.6s, 2.3s, and 2.1s in average for IMC, ESS, and diff eq, respectively.

Table 3: Analyzed results and statistics

<table>
<thead>
<tr>
<th>Specification</th>
<th>ESS</th>
<th>IMC</th>
<th>diff eq</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>66</td>
<td>76</td>
<td>50</td>
</tr>
<tr>
<td>$N'$</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td># Basic Operation Types</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td># Nodes of Reduced Tree</td>
<td>4055</td>
<td>4035</td>
<td>4035</td>
</tr>
<tr>
<td># Cases</td>
<td>26</td>
<td>31</td>
<td>29</td>
</tr>
<tr>
<td># Visited Nodes in Average</td>
<td>16</td>
<td>114</td>
<td>112</td>
</tr>
</tbody>
</table>

5.4 Algorithm Effectiveness

The effectiveness of the IMSP-2P algorithm has been evaluated by using it to select the implementation methods of the basic operations. The IMSP-2P selected the optimum partitioning for different values of area constraint. The power consumption was ignored to simplify the experimental cases.

FUs needed to implement basic operations for these sample programs are a multiplier, a divider, a barrel-shifter, an extender and so on, where the multiplier and the divider can be pipelined or non-pipelined.

Figures 2 and 3 show estimation errors by IMSP-2 and IMSP-2P, respectively. Note that IMSP-2P can estimate the execution cycles much more accurately than IMSP-2. That is, estimation errors are below 1.3% for designs with gate count constraints exceeding 22 Kgates, where all operations are implemented in HW. On the other hand, the estimation errors by IMSP-2 range from 5% to 20% because of not taking into account the pipeline hazards.

Figure 2: Estimation errors by IMSP-2

Figure 3: Estimation errors by IMSP-2P

Please note that in the scheduling process it was assumed that each SW implemented basic operation is executed in the expected fixed number of clock cycles as shown in Table 2. In practice, these SW implemented basic operations will take different execution cycles depending on the input data. In the measurement of the execution cycles, the actual behavior of these SW implemented basic operations were simulated. As a result, the execution cycle estimation reported by IMSP-2P will contain some error as in the case of IMSP-2 due to the execution cycles dependent on input data (up to 25% and 32% as shown in Figures 2 and 3, respectively.)

Some of the design results for IMC are shown in Table 4. In this table, the second column represents area constraints $A_{max}$ in Kgates. The third column represents the estimated execution cycles $T(X)$ by the IMSP-2P solver. The fourth column shows the execution cycle errors (Err) measured by using the simulator generated by the PEAS-1 system, assuming that the HW interlock is used. Note that estimation errors are almost 0% for designs with all HW implemented operations. The last column shows the HW modules
that implement basic operations to be implemented by HW. The optimum instruction set is then defined on the selected HW modules. For example, in design #7 with $A_{max} = 20$ K Gates only the modules 'b_adds' and 'div seq' have been chosen together with the kernel, therefore the instruction set of the designed ASIP contains the primitive operations and those basic operations belonging to 'b_adds' and 'div seq', i.e. ashr, ashl, bshr, bslh, div, udiv, mod, and umod. The remaining basic operations such as mul, extendhi, and so on are implemented in SW using the Kernel. For any given area constraint, the shown partitioning represents the optimum one. Other combinations were not selected by the algorithm because of their inferiority. The design results for the ESS and diffseq programs are not shown here due to lack of space.

Table 4 : Estimated execution cycles, estimation errors and selected HW modules for IMC by IMSP-2P

<table>
<thead>
<tr>
<th>#</th>
<th>$A_{max}$ K Gates</th>
<th>T(Cycles)</th>
<th>Bwr (%)</th>
<th>Selected HW Modules (with kernel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>56</td>
<td>60689</td>
<td>0.5</td>
<td>b_adds, extends, mult, div, seq</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>60688</td>
<td>0.1</td>
<td>b_adds, extends, mult, div, seq</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>61817</td>
<td>0.1</td>
<td>b_adds, extends, mult, div, seq</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>71400</td>
<td>0.0</td>
<td>b_adds, extends, div, seq, mult, lca</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>84329</td>
<td>0.0</td>
<td>b_adds, extends, div, seq, mult, lca</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
<td>123396</td>
<td>1.4</td>
<td>b_adds, extends, div, seq, mult, lca</td>
</tr>
<tr>
<td>7</td>
<td>20</td>
<td>164396</td>
<td>27.9</td>
<td>b_adds, div, seq, mult, lca</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
<td>260096</td>
<td>31.6</td>
<td>extends, div, seq, mult, lca</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>344448</td>
<td>9.1</td>
<td>b_adds, extends, mult, lca</td>
</tr>
<tr>
<td>10</td>
<td>16</td>
<td>380888</td>
<td>8.4</td>
<td>b_adds, extends, mult, lca</td>
</tr>
<tr>
<td>11</td>
<td>15</td>
<td>406228</td>
<td>4.1</td>
<td>(kernel only)</td>
</tr>
</tbody>
</table>

Another major feature of IMSP-2P is that it can reduce the pipeline execution cycles by optimally selecting the pipelined FUs in comparing to the IMSP-2 solver. In experiments, the performance (via the execution cycles of the application program) of the designed ASIP's was measured by using the PEAS-I simulator. It is found that up to 4.2%, 8.4%, and 6.7% execution cycle reduction rates due to selecting the pipelined FUs by IMSP-2P compared to IMSP-2 were achieved for the ESS, IMC, and diffseq examples, respectively.

6 Conclusion and Future Work

We have proposed an efficient method to design an optimal pipelined instruction set processor in the PEAS-I system. The method uses the pipeline scheduler which is capable of detecting and resolving pipeline hazards and handling multicycle operations simultaneously. The method with IMSP-2P is introduced as a HW/SW partitioning problem and selects the implementation method of the operations that implement a pipelined ASIP instruction set so that the performance is maximized under the given design constraints. The effectiveness of the IMSP-2P algorithm was demonstrated through design examples. The method estimates the performance of a designed pipelined ASIP accurately for most designs. Considerable execution cycle reduction rates due to selecting pipelined FUs have been obtained. The algorithm is so efficient that all the optimal solutions performed in the experiments in partitioning process were obtained within a few seconds on a conventional workstation.

The estimation error reduction for SW implemented operations is left for future work. Our future work also includes the development of a HW/SW partitioning algorithm for pipelined ASIP design with the least gate count under a given power consumption and execution cycle constraints. Moreover, the design with the lowest power consumption under gate count and execution cycle constraints is also planned.

Acknowledgments

Authors would like to express their thanks to NTT Communication Science Laboratories, VLSI Technology, Inc., Science Create, Co., Ltd., Japan, for their kind help and assistance. Part of this research is supported in part by a research grant from the Ministry of Education, Science and Culture of Japan under contracts No's. 07558038 and 07608053.

References