Optimizing the Generation of Object-Oriented Real-Time Embedded Applications Based on the Real-Time Specification for Java

Marco Aurélio Wehrmeister¹, Carlos Eduardo Pereira¹,², Leandro Buss Becker³

¹Computer Science Institute, UFRGS, Brazil
²Electrical Engineering Department, UFRGS, Brazil
³Automation and Control Systems Department, UFSC, Brazil

{mawehrmeister@inf.ufrgs.br, cpereira@ece.ufrgs.br, lbecker@das.ufsc.br}

Abstract

The object-oriented paradigm has become popular over the last years due to its characteristics that help managing the complexity in computer systems design. This feature also attracted the embedded systems community, as today’s embedded systems need to cope with several complex functionalities as well as timing, power, and area restrictions. Such scenario has promoted the use of the Java language and its real-time extension (RTSJ) for embedded real-time systems design. Nevertheless, the RTSJ was not primarily designed to be used within the embedded domain. This paper presents an approach to optimize the use of the RTSJ for the development of embedded real-time systems. Firstly, it describes how to design real-time embedded applications using an API based on RTSJ. Secondly, it shows how the generated code is optimized to cope with the tight resources available, without interfering in the mandatory timing predictability of the generated system. Finally it discusses an approach to synthesize the applications on top of affordable FPGAs. The approach used to synthesize the embedded real-time system ensures a bounded timing behavior of the object-oriented aspects of the application, like the polymorphism mechanism and read/write access to object’s data fields.

1. Introduction

The interest in the use of object-orientation and Java in the embedded real-time systems (ERTS) domain has been increasing over the last years [13][14]. An important step was related to the definition of the Real-Time Specification for Java (RTSJ) [19], which specifies both an API to be used in development of real-time applications and also the expected behavior of a RTSJ-compatible Java Virtual Machine (JVM). Example of available RTSJ partially compliant virtual machines are found in [9][15][16]. However, such implementations are not primarily designed to be used within the embedded domain.

While object-orientation definitely increases the abstraction level, it also adds additional overhead. This overhead must be small on ERTS that have tight computational resources (e.g. memory and processing power).

To allow the use of the object-oriented (OO) paradigm by means of the Java language in ERTS design, this paper describes the use of a dedicated Java processor, named FemtoJava [1], which has versions from 8 to 32 bits with different organizations (e.g. multicycle, pipeline, VLIW).

The ERTS generation and synthesis (Java processor + application code) is made by a synthesis tool named SA-SHIMI (introduced in [1]), which originally allowed only the synthesis of static methods and attributes. In the current work this tool is extended to support the synthesis of OO code, adding capability to synthesize instance fields and methods, to make static object allocation, and to provide a predictable polymorphism mechanism to the method calls. The adopted approach in the ERTS synthesis ensures a bounded execution time of OO aspects of the application, such as the polymorphism mechanism and the object’s instance field access (even for the inherited fields).

The remainder of this paper is organized as follows. Section 2 describes related work dealing with the use of Java in the embedded systems market. The proposed synthesis tool is described in section 3. Section 4 presents our synthesis method and the modifications performed in the original synthesis tool. Some actual data related to the generated embedded system are presented in section 6. Finally, conclusions and future work are presented in section 7.

2. Related Work

According to the Java specification any application developed using the Java platform must be executed in a JVM [6]. A JVM can be implemented as software, which is a layer between the Java application and the real machine, or implemented as hw that executes Java bytecodes natively. Both approaches have pros and cons, and depending on the application requirements and constraints one choice is better than the other. The default JVM spec. can not be used in the real-time domain, specially due to undeterministic timing behavior introduced by garbage collector (GC) and also by inheritance mechanisms. In order to solve such problems, the RTSJ defines how a RTSJ-compatible JVM must behave.
JamaicaVM [8][22] is an example of JVM implemented as software layer and implements partially the RTSJ specification. JamaicaVM has two ways to execute embedded real-time applications: using a virtual machine layer or executing applications compiled into machine native code that incorporates also the JamaicaVM code. The JamaicaVM GC has a deterministic worst case execution time (WCET), it executes on each object allocation, and frees up to 32 bytes per execution. Another example of RTSJ-based JVM is jRate [9] which extends the GNU Compile for Java (GCJ) execution environment. This means that application code is compiled into target machine native code and therefore there is no JVM layer to execute the real-time Java application. In jRate the real-time threads implementation simply relies upon the underlying real-time operation system scheduler (which usually are preemptive and adopts a fixed priority mechanism). jRate implements only two kinds of memory, both from RTSJ: ScopedMemory and Immortal-Memory.

Both Jamaica and jRate lead to an ERTS with higher memory usage. This is due to the fact that all method codes of application classes must be included into the application binaries, even the unused ones. Also all information present on class constant pool [6] of all application classes must be included into these binaries. Clearly, this issue can be optimized by removing the unused methods and the constant pool. Moreover, OO mechanisms such as object’s field access and polymorphism require allocation of additional memory.

On the other hand, the Java Optimized Processor (JOP) [10] is an example of a JVM implemented as hardware. JOP follows the Connected, Limited Device Configuration (CLDC) profile for the J2ME specification [11] and can be implemented using low-cost FPGAs. The GC is not implemented. aJile [12] is another example of JVM implemented in hardware, which also follows de CLDC profile for the J2ME. It implements a simple mark-and-sweep GC as a thread that uses Java synchronization mechanism. Applications can be compiled using a standard Java compiler, however before deploying it on the aJile processor the bytecodes must be adapted using the JEMBuilder. This builder generates RAM and ROM memory images and eliminates unused methods, fields, and constants, optimizing the footprint of the embedded real-time application. aJile has most of the functionality specified in the RTSJ. However, according to [17], compliance to RTSJ “will require business initiatives to pursue”. Both processors lack support to the RTSJ and also, for some applications, there is waste of FPGA area as it contains unused instructions.

3. The Synthesis Process

According to our approach, the embedded system design flow starts with the development of a Java application. This application can be implemented and tested into the host development environment. After the functional validation, the classfiles (which are generated by standard Java compiler) are used as input to the SASHIMI synthesis tool, which analyses them and makes additional optimizations into the code in order to generate an optimized and small footprint application code. The SASHIMI synthesis tool also generates a customized Femto-Java that implements only the instructions used by application. In other words, it analyses the application code and customizes the processor control unit in order to support only the used instructions, reducing the processor FPGA size. The synthesis flow can be observed in Fig. 1.

![Fig. 1 Synthesis Process Flow](image)

To cope with the synthesis process, designers must develop their Java application following a set of rules. For instance, only integer numbers and APIs provided by the environment are allowed. Moreover, in the original version of the tool designers could use only static methods and attributes, since there was no support for object allocation. This problem is now solved in the present work, as detailed in the next section.

Additionally our approach allows concurrent real-time programming by means of an operating system layer in charge of CPU scheduling. In this model each thread has its own stack, which must be initialized individually. Also, there are different scheduling algorithms which are supported by the environment. An evaluation of the impact of these algorithms in terms of footprint, power consumption and real-time performance are described in [5]. A drawback from the original environment is the lack a high-level API, enforcing designers to write programs “polluted” with low-level system calls to interact with the scheduler. Additionally, there was no mechanism to express clearly the tasks timing constraints. These issues are addressed by our RTSJ-based API. This API provides
classes to express concurrent real-time threads, release and schedule parameters, schedulers, timers, and also absolute and relative time values. Interested readers can refer [2] to obtain details on our RTSJ-based API.

4. Synthesizing Object-Oriented Code

As mentioned in the previous sections, the original version from our synthesis environment provided no support for object creation. Therefore, some adaptations are needed in order to provide full support for the proposed API in our FemtoJava platform. The following subsections describe how the code is synthesized.

4.1. Object allocation

One of the main motivations for working with Java is the related facilities for working with objects. Nevertheless, such observation contrasts with the restrictions from the previous version of our synthesis tool that had no support for objects, allowing designers to work only with static classes. Therefore this work presents an extension from the synthesis tool to support the synthesis of objects.

The adopted scheme followed the proposal from Pushner and Wellings in [18], which defines that all objects in the system must be defined a priori, avoiding overheads related to dynamic memory allocation and management. Such practice avoids the use of the GC, which is the main source of indeterminism in Java applications. As a result, in our approach all application objects are allocated statically at synthesis time. For each application object the SASHIMI tool analyses the class hierarchy in order to identify all fields that an instance of the class must contain. After all, the size and the structure of the object are known at synthesis time. The object allocation follows the model presented in Fig. 2.

The first field of the allocated object is used to hold the class unique identifier. Each application class has its own number, which is used by the polymorphism mechanism and also by the class related opcodes (e.g. “instanceof” and “checkcast”) in order to distinguish instances of different classes. The memory for all inherited fields are allocated after the class identifier field. The highest class fields (i.e. the fields of the most abstract class) are allocated first, followed by its subclasses fields. Finally, the class fields are allocated after the inherited fields. The order of field allocation obeyed by the class hierarchy is as follows: the fields from the topmost class (in hierarchy) are allocated first, followed by the fields from its child class, and so on.

An example of this statically allocation of objects is shown in Fig. 3 containing a fragment of the VHDL-compatible file that was generated using our tool and which represents the FemtoJava RAM memory. In this figure the allocated object represents an instance of \textit{RTTask1Class} class which extends the \textit{RealtimeThread} class, thereby represents a real-time task in the generated embedded system. The class unique identifier can be observed at line 63h. The following nine fields (lines 64h to 6Ch) are the instance fields inherited from the \textit{RealtimeThread} class, and the remaining fields are instance fields from the class.

```
... 63 : 0b;-- ----> ID for RTTask1Class
64 : 00;-- m_relParam[LReleaseParameters]
65 : 00;-- m_schedParam[LschedulingParameters]
66 : 00;-- m_isStarted [Z.1]
67 : 00;-- m_isFinalized [Z.1]
68 : 00;-- m_isRunning [Z.1]
69 : 00;-- m_isBlocked [Z.1]
6a : 00;-- m_BaseStackPointer [I.1]   
6b : 00;-- m_stackPointer [I.1]
6c : 00;-- m.ResumeTime [LAbsoluteTime.1]
6d : 00;-- m_aux [I.1]
6e : 00;-- m_loop [I.1]
6f : 00;-- m_schedCount [I.1]
...
```

Fig. 3 Object statically allocated into memory

Another required adaptation is changing the bytecode of the Java class files in order to allow objects static allocation. Therefore the reference for the allocated object (RAM memory address of the first attribute in the object) is inserted into the bytecode in substitution of the “new” opcode. All bytecode occurrences related to the Java command “new” are replaced by this fixed memory address. It is important to emphasize that the object is never released from memory, once there is no GC running. For example, in \textit{RTTask1Class} object allocation the instruction “new” is changed to “sipush 63h” (push object reference in the top of the stack).

To allow accessing the object attributes, the “getfield” and “putfield” opcodes [6] also need to be supported in the synthesis. The proposed allocation approach eliminates the need of a class constant pool [6] because the class structure (i.e. fields from class and its superclasses) is analyzed at synthesis time, so it is removed from the synthesized classes. As consequence, the semantic of “getfield” and “putfield” opcodes must be modified. The reference to the constant pool was replaced by the memory offset of the field related to the initial address of the object. In summary, to access an object field its memory address should be calculated based on the object initial address and the respective field offset. For example, to access the value of \textit{RTTask1Class}'s m_loop field that is stored on the 6eh memory address (see Fig. 3), the “getfield ConstatPoolReference” instruction is changed to “getfield 0Bh”. Therefore when this instruction is executed, the value 63h (reference of the \textit{RTTask1Class} object) is popped from top of stack and be added to 0Bh offset in
order to obtain the RAM address to access m_Loop field value (i.e access the RAM address 6Eh). The “putfield” instruction is changed similarly to the “getfield” one, however the “putfield” writes a value into the calculated RAM address instead of reading the memory address.

The main advantage of this approach is that an object field can be accessed using only two instructions (i.e. object reference push instruction and access field instruction) in contrast to execute a lot of instructions to access constant pool information to know where field’s data are stored in memory. A disadvantage is that application class hierarchy must be evaluated at synthesis time, forbidding the instantiation of classes that were not previously defined.

4.2. Predictable Polymorphism Mechanism

Polymorphism is another desirable characteristic from object-orientation that was also not supported in the original version of the SASHIMI synthesis tool. Our goal here is to define a solution to make the polymorphic method invocation predictable. Therefore our mechanism is neither implemented using tables to store the class hierarchy nor overrides method tables as specified in [6].

The proposed mechanism uses the class hierarchy information extracted from application code that was collected during analysis phase. For each method that is overridden, an extra piece of code is inserted before the users code. This code represents the polymorphism mechanism shown in Fig. 4.

This algorithm induces the conclusion that the proposed polymorphism mechanism introduces indeterminism into ERTS. Particularly, if the application classes are loaded dynamically, the above conclusion is true because this algorithm is highly dependent on the input (i.e. using dynamic load of classes, the class hierarchy level is evaluated at runtime what introduces indeterminism in polymorphism mechanism). On the other hand, if class hierarchy is known at synthesis time, this algorithm can be unrolled and transformed on an algorithm that follows the single path paradigm [20].

According to Pushner and Burns [20], in the single path paradigm the behavior of algorithms is fully temporally predictable because they execute on the single possible execution path, independently from the input data. Therefore, knowing the class hierarchy the algorithm presented in Fig. 4 can be unrolled into several nested “if” instructions, one for each class that overwrites methods from its superior classes (in hierarchy). The use of single path paradigm fulfills the two identified central facts to determine the WCET of a program: (i) the possible sequence of actions and (ii) the time needed for each action of this sequence [21]. In summary, the WCET for polymorphic method invocation can be determined at synthesis time, using the following formula:

\[ f_{WCET}(meth.call) = t_1 + \sum_{i=1}^{n_i} t_{2i} + \sum_{i=1}^{n_i} t_{3i} \]

Where:

- \( n \) is the number of subclasses that overwrite the original method
- \( t_1 \) is the time spend by “has overridden method” condition plus “Read class of first overridden method” command;
- \( t_{2i} \) is the time needed by “Self if instance of current child class” condition;
- \( t_{3i} \) is the time needed by “has overridden methods to evaluate” condition plus “Read class of next overridden method”.

Other issue supported by the polymorphism mechanism is the “instanceof” and “checkcast” instructions. The first one is used to test if an object is an instance of a determined class, the later evaluates an object typecast. As in our approach the constant pool was removed from classfiles, the semantics of these instructions had to be changed.

The “instanceof” instruction had its reference to the constant pool replaced by the class identifier relative to the class that is originally referenced by the constant pool entry used by this instruction. So, the object class identifier is compared with this unique identifier and, if the object is an instance of the tested class, a “true” value is pushed into the stack. The “checkcast” instruction follows the same approach. Its reference to constant pool is replaced by a class identifier. So, this instruction compares the class identifier field of the object which is on the top of stack with the class identifier of “checkcast” instruction. If they are equal, the stack remains unchanged; otherwise an exception is generated.

The proposed modification in these two instructions allows some inheritance evaluation. It avoids spending time to evaluate application class hierarchy at runtime, in
In order to discover the type of a class instance. Evaluation of class instances can be made using only two instructions (e.g. push object reference instruction and “instanceof” or “checkcast” instruction). However, this approach has a drawback: the type of a class instance comparison always will be “the class of objectX is equal to classY?” opposite to original comparison “objectX belongs to classY or its superclasses?”. This limitation is imposed by the constant pool removal.

4.3. Threads Stack Initialization

In our scheme each real-time thread has its own stack (i.e. the global stack is divided in parts, one for each thread). This thread stack must be initialized with all method frames needed to execute a real-time thread (thread frame). In the previous version of SASHIMI tool this initialization was done by the programmer. This means that at each compilation, different code addresses needed to be pushed in the thread stack. This could lead to initialization errors and a lot of extra-work to fix this address code changes. To solve this problem, an automatic thread stack initialization routine was proposed.

The code for the thread frame allocation in the global stack is inserted automatically by the synthesis tool during synthesis of the ERTS. The maximum number of tasks allowed in the embedded system generated by synthesis tool is direct related with the amount of available RAM memory. The method frames are constructed according the FemtoJava frame organization [1] and is popped from stack by the method return instructions (e.g. “return”, “ireturn”, “areturn”, etc.). The thread stack is composed by three method frames. Firstly, the method frame of the scheduling algorithm method is pushed. Afterwards the mainTask() method frame is pushed on the stack. The last pushed frame is a frame that is used only for the first task context switch, which means that this frame is consumed on the first task activation.

5. Results on ERTS Generation

To illustrate the proposed ERTS synthesis method, this section presents results from two case studies. The first one is a wheelchair movement control, on each users can control the wheelchair’s velocity and direction using a joystick. This system has four periodic tasks with different periods: one to read the joystick sensor; other to read movement sensors (speed and angle); a control task; and a movement mode controller task. The second case study is a crane control system [7], which is a known benchmark for ERTS. It moves a load through its track to a desired position. The implemented system has also four tasks: one to read the angle sensor; other to read the position sensor; a control task; and a diagnosis task.

After having generated the RTSJ-based Java code for each application, this code is then compiled and linked with the RTSJ-based API using a standard Java compiler. The resulting class files are used as input to the synthesis tool and are analyzed in order to generate the VHDL files that represent the customized FemtoJava processor. In the final step the resulting VHDL files are synthesized with Altera Quartus II 4.0 Web Edition Full using the Cyclone EP1C20F400C6 FPGA device. The synthesized ERTS characteristics are shown in Table 1.

<table>
<thead>
<tr>
<th>Table 1: Size of the generated systems</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Compiled code</td>
</tr>
<tr>
<td>Application size</td>
</tr>
<tr>
<td>17536 bytes</td>
</tr>
<tr>
<td>Used API classes size / %</td>
</tr>
<tr>
<td>Total size (application+API)</td>
</tr>
<tr>
<td>Synthesized code</td>
</tr>
<tr>
<td>Application size (ROM)</td>
</tr>
<tr>
<td>Used API classes size (ROM) / %</td>
</tr>
<tr>
<td>Total ROM size (application+API)</td>
</tr>
<tr>
<td>RAM size</td>
</tr>
<tr>
<td>Application size (ROM+RAM)</td>
</tr>
<tr>
<td>Code size reduction / %</td>
</tr>
<tr>
<td>Total of different used instructions</td>
</tr>
<tr>
<td>FemtoJava Size</td>
</tr>
</tbody>
</table>

The compiled code represents the size of all class files generated by the standard Java compiler. After synthesis, the total application code size was reduced from approximately 43,4 kb to 5,4 kb (87,5%) in the wheelchair system and from 60,8 kb to 9,2 kb (84,9%) in the crane system. This reduction is even higher if we consider the size of a RTSJ-compatible JVM that must be deployed together with the application code. For example, Jamaica VM needs a minimum of 128 kb [22]. Considering this information the reduction of total application size is, respectively, 96,9% and 95,2%.

Considering the additional code related to the RTSJ-based API, which main benefit is to increase the code readability, in the wheelchair experiment it represents 53% of total application code, and 33% in the crane experiment. This difference is due to the more complex control mechanisms in the crane experiment. This increased code shows the reduction of designer work that is provided by high-level languages. Similarly, the total number of FemtoJava instructions is higher in the crane experiment. As stated before, the size of the control unit of the Java processor is directly proportional to the total number of instructions used in the application. Therefore the total number of logic cells needed in the crane experiment is higher than in the wheelchair one.

As stated before, the approach used to synthesize the

---

1 A method frame contains all information (local variables and return context) for current executing method
Additional GC with bounded execution time (it must have a known worst case execution time). Also other RTSJ constructions should be supported, as for example Asynchronous constructs should be supported, as for example AsynchronousEventHandler and Timed.

7. References


