Towards Unified Design of Hardware and Software Components Using C++

Journal:	*Transactions on Computers*
Manuscript ID:	Draft
Manuscript Type:	Regular
Keywords:	C.3.d Real-time and embedded systems < C.3 Special-Purpose and Application-Based Systems < C Computer Systems Organization, D.2.3 Coding Tools and Techniques < D.2 Software Engineering < D Software/Software Engineering, B.0 General < B Hardware, C.0.e System architectures, integration and modeling < C.0 General < C Computer Systems Organization, D.2.10.g Object-oriented design methods < D.2.10 Design < D.2 Software Engineering < D Software/Software Engineering

Towards Unified Design of Hardware and Software Components Using C++

Tiago Rogério Mück, *Member, IEEE,* and Antônio Augusto Fröhlich, *Member, IEEE,*

Abstract—The increasing complexity of current embedded systems is pushing their design to higher levels of abstraction, leading to a convergence between hardware and software design methodologies. In this paper we aim at narrowing the gap between hardware and software design by introducing a strategy that handles both domains in a unified fashion. We leverage on *aspect-oriented programming* and *object-oriented programming* techniques in order to provide *unified* C++ descriptions of embedded system components. Such unified descriptions can be obtained through a careful design process focused on isolating aspects that are specific to hardware and software scenarios. Aspects that differ significantly in each domain, such as resource allocation and communication, were isolated in *aspect programs* that are applied to the unified descriptions before they are compiled to software binaries or synthesized to dedicated hardware using *high-level synthesis* tools. Our results show that our strategy leads to reusable and flexible components at the cost of an acceptable overhead when compared to software-only C/C++ and hardware-only C++ implementations.

Index Terms—System-level design, HW/SW co-design, High-level synthesis, Aspect-oriented system design.

1 Introduction

Embedded systems are becoming increasingly complex as the advances of the semiconductor industry enable the use of sophisticated computational resources in a wider range of applications. Depending on the application's requirements (e.g. performance, energy consumption, cost, etc.), the development of such systems may encompass an integrated hardware and software design that can be realized by several different architectures, ranging from simple 8-bit microcontrollers to complex *multiprocessor system-on-chips* (MPSoCs).

In order to deal with this complexity, embedded system designs are being pushed to the *system-level*. In this scenario, a convergence between hardware and software design methodologies is desirable, since a unified modeling approach would enable one to take decisions about hardware/software partitioning later in the design process, maybe even automatically. In the last few years, advances in *electronic design automation* (EDA) tools are allowing hardware synthesis from high-level, software-like descriptions. This process is known as *high-level synthesis* (HLS) and allows designers to describe hardware components using languages like C++, and higher-level techniques, such as *Object-Oriented Programming* (OOP). The focus of these tools [1], [2], [3], however, is hardware synthesis, and they do not provide a clear design methodology for developing components which could be implemented as both hardware and software.

Aiming to narrow this gap, in this paper we describe some design guidelines and mechanisms which support the implementation of both hardware and software components from a single, unified, C++ description. Our guidelines are built upon the *Application-driven Embedded System Design* (ADESD) methodology [4]. ADESD leverages on OOP and *aspect-oriented programming* (AOP) concepts, defining a domain engineering strategy which allows us to clearly separate the core behavior and the structure of a component from aspects that must be handled differently whether a component is implemented as hardware or software. Such specific characteristics are modeled in special constructs called *aspects* and are applied to the unified descriptions of components only after the hardware/software partitioning is defined, yielding the final implementation in the target domain (hardware or software). In order to generate descriptions that can be efficiently synthesized by HLS tools or compiled to a software binary, the implementation of both the components and the mechanisms which adapt them make extensive use of C++ *generative programming* [5] techniques such as *static metaprogramming*. To evaluate the efficiency of our approach, we present the implementation of a *Private Automatic Branch Exchange* (PABX) SoC of which some components were implemented using our unified design strategy.

It is important to recall that it is not a goal of this work to discuss the quality of hardware implementation generated using HLS tools, neither the intrinsic differences between software and hardware that cannot yet be fully handled by such tools. We take in consideration the fact that these tools impose limitations to source descriptions (such as dynamic allocation and binding). As it will be demonstrated in the following sections, limitations of this kind can be circumvented by our unified approach.

The remaining of this paper is organized as follows: Section 2 presents a discussion about works related to the integration of the hardware and software design flows; Section 3 presents an overview of previous contributions

• *T. R. Mück and A. A. Fröhlich are with the Software/Hardware Integration Lab , Federal University of Santa Catarina, Florianópolis, Brazil. E-mail: {tiago,guto}@lisha.ufsc.br*

upon which this work is built; in Sections 4 and 5 we discuss the characteristics that are part of software and hardware scenarios and present our approach for their proper separation and implementation; Section 6 describes our case studies and shows our experimental results; and Section 7 closes the paper with our conclusions.

2 Related work

Several design methodologies and tools were proposed in order to provide more tightly coupled hardware and software design flows. Most of these methodologies are based on the concept of building a system by assembling pre-validated components. The *Ptolemy extension as a Codesign Environment* (PeaCE) [6], an extension of the Ptolemy project [7], is a synthesis framework for real-time multimedia applications. PeaCE takes models composed by synchronous data-flow graphs and extended *finite state machines* (FSMs) and either generates C code or maps the models to pre-existing IP cores and processors. ROSES [8] is a design flow which automatically generates hardware, software, and interfaces from an architectural model of the system and a library of pre-validated components. The *Metropolis* and its successor *Metro-II* [9] follow the *Platform-based design* (PBD) [10] methodology. They propose the use of a metamodel with formal semantics to capture designs. An environment to support simulation, formal analysis, and synthesis is also provided. Despite providing integration frameworks for the entire design flow, these methodologies do not define clear guidelines to design new reusable components. Also, mapping the application model to pre-existing components limits hardware/software partitioning.

In order to overcome these limitations, one must focus on closing the *design gap* of software and hardware components. State-of-the-art EDA tools (e.g. Calypto's CatapultC [1], Forte's Cynthesizer [2], Xilinx's AutoESL [3]) support hardware synthesis from high-level C++/C-based constructs. Indeed, several works have already demonstrated the applicability of HLS for implementing hardware components such as signal processing filters, cryptographic cores, and other computationally intensive components [11]. Our aims are different however. We want to describe components in a high level of abstraction, but those should be implementable in hardware as well as in software, and with performance close to software-only and hardware-only implementations.

On this track, the OSSS+R methodology [12] raises the level of *register transfer level* (RTL) SystemC by adding new language constructs to support synthesizable polymorphism and high-level communication. However, hardware/software partitioning must still be done in the beginning of the design process, and the inclusion of non-standard language constructs reduces the compatibility of the design with available compilers and hardware synthesis tools. The *Saturn* [13] design flow also contributes in this scenario, but follows a different approach. It aims to close the gap between UML-based modeling and the execution of the models for their verification. The authors have elaborated over *SysML*, an extension of UML for system-level design, and developed a tool which generates C++ for software and RTL SystemC for hardware. Although using a single language, software and hardware are still modeled separately. Additionally, it is not clear whether their tool generates code only for the interface and integration of components, or the behavior is also inferred from the SysML models.

SystemCoDesigner [14] is a tool which integrates a HLS flow with design space exploration. The design entry of SystemCoDesigner is an actor-based data flow model whose actors can be converted to synthesizable SystemC or to C++ for software compilation. However, as the authors themselves claim, SystemCoDesigner targets mostly data-flow-based applications, and they do not provide directions towards a more general deployment. The System-on-chip environment (SCE) [15] takes SpecC models as input and provides a refinement-based tool flow. Guided by the designer, the SCE automatically generates a set of *Transaction-level models* (TLM) that are further refined to pin- and cycle-accurate system implementation.

Other works focus mainly on the interface between software and hardware. For instance, the ReconOS [16] and the BORPH operating system [17] provide a unified interface for both domains. In these works a task performed in hardware is seen with the same semantics as a software thread, and a system call interface is provided to hardware components. However, despite providing this unified interface, they do not cover the gap that still exists between the way the hardware and software threads themselves are implemented. The work of *Rincón et al* [18] is based on concepts from distributed object platforms such as CORBA and Java RMI and provides a tool which generates the communication glue necessary so that hardware and software components can communicate seamlessly.

3 Application-driven Embedded System Design

The *Application-driven Embedded System Design* (ADESD) methodology [4] elaborates on the *Object-Oriented Decomposition* [19] strategy to add the concept of *aspect* identification and separation at early stages of design. Variability analysis, as carried out in object-oriented decomposition, does not emphasizes the differentiation of variations that belong to the basic behavior of a component from those that emanate from the execution scenarios being considered for it. Components that incorporate environmental dependencies have a smaller chance of being reused in new scenarios, and, given that an application-driven system will be confronted with a new scenario virtually every time a new application is defined, allowing such dependencies could severely hamper the system.

Nevertheless, one can reduce such dependencies by applying the key concept of *aspect-oriented programming* (AOP) [20], i.e. aspect separation, to the decomposition process. Scenario dependencies can be encapsulated in special constructs called *aspects*. An aspect weaving semantic is then defined to describe when and how aspects are applied to components. By doing so, one can tell variations that shape new components from those that yield scenario aspects. For instance, instead of modeling a new component for a family of communication mechanisms that is able to operate in the presence of multiple threads, one could model multithreading as a scenario aspect that, when activated, would lock the communication mechanism (or some of its operations) in a critical section.

The aspect weaving semantics in ADESD is defined by constructs called *Scenario adapters* [21]. Scenario adapters were developed around the idea of components getting in and out of an execution scenario, allowing actions to be executed at these points. These actions may take place respectively before and after each of the component's operation in order to setup the conditions required by the scenario. As can be seen in the following sections, scenario adapters can be fully implemented in C++ using its OOP and template metaprogramming capabilities[1]. In contrast to traditional aspect weaving approaches, such as AspectC++ [24], scenario adapters do not require external tool support and can be compiled by any standard C++ compiler.

3.1 ADESD for HW/SW design

In ADESD, knowledge about implementation details should be driven to identify and isolate scenario aspects. In general, aspects such as identification, sharing, authentication, and encryption can be represented as scenario aspects. Such aspects are part of the *application domain*, since they describe behavior related to the application functional requirements. However, the design of new components in a HLS-capable system-level environment must also take into considerations the *platform domain*, therefore yielding at least two additional scenarios: a *software scenario*, in which the component is deployed as part of the software running in a processor; and a *hardware scenario*, in which the component is deployed as dedicated hardware.

An initial work [25] has already shown how the ADESD methodology was used to provide a C++ description of a resource scheduler suitable for high-level hardware synthesis. In this paper we further extend the previous case in order to show how ADESD's aspect separation mechanisms can be used to yield components susceptible to both hardware and software synthesis. We demonstrate how characteristics specific of a hardware or a software scenario can be isolated from the core behavior and the interface of a component. By following our design guidelines, the same component description can be tailored by the scenario adapter in order to feed either a software compilation flow or a hardware HLS flow.

1. this paper assumes prior knowledge about OOP, UML and C++. The reader may refer to [22], [5], and [23] for an indepth explanation of the relevant concepts

4 Unified Hardware/Software Design

HLS tools allow for hardware synthesis from algorithms described in languages such as C++. When aiming at this goal, the implementation of the algorithm itself is usually very similar to what is seen in software. However, at component level, several characteristics arise which distinguish descriptions aiming at hardware and software. In the following sections we first discuss these differences. Then, we present our strategy for the separation of hardware and software concerns and how we have achieved unified C++ descriptions.

4.1 Differences between hardware and software

Table 1 summarizes the most common component communication patterns. In the software domain, components may be objects which communicate using method invocation (considering an OOP-based approach), while in the hardware domain, components communicate using input/output signals and specific handshaking protocols. For communication through different domains, the software must provide appropriate *hardware abstraction layers* (HAL) and *interrupt service routines* (ISR), while the hardware must be aware that it is requesting a software operation.

Table 1
Usual component communication patterns. The *Caller* requests operations from the *Callee*.

Direction	Type of communication	
	Caller	Callee
SW→HW	HAL sends commands to the HW using a communication infrastructure (e.g. a bus or a NoC).	Communication interface receives commands and triggers the operations.
HW→SW	HW interrupts SW and waits for the operation to finish. It may transfer data to/from the main memory (e.g. DMA).	An ISR calls the requested operation and signalizes HW when finished.
SW→SW	Function call interface	
HW→HW	Signals/Handshaking	

Another concern is related to the composition of hardware/software components in an object-oriented model and the mapping of high-level components to physical implementations. Figure 1 shows a simple system modeled using such strategy and illustrates two different mappings. In OOP, the original structure may be "disassembled" in the final physical implementation if different objects in the same class hierarchy represent

components that are to be implemented in different domains. For example, *C2* is "inside" *C1* in the OO model in Figure 1, but, in a possible final implementation, *C2* could be implemented as a hardware component while *C1* could run as software in a processor (first mapping in Figure 1).

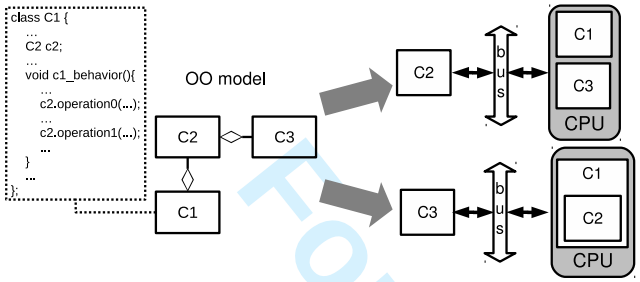


Figure 1. Possible mappings of an object-oriented model to physical implementations

When focusing on an OOP-based methodology, one must provide ways to make these different mapping transparent to the designer. Furthermore, the different communication patterns between hardware and software must also be properly abstracted. For instance, in the second mapping shown in Figure 1, *C1* can call *C2*'s methods directly, while in the first mapping a hardware/software communication mechanism must be defined. In this sense, the method call interface must also be properly adapted when a component is to be implemented in hardware, so it can be compliant with HLS tools requirements. In Calypto's CatapultC [1] and Xilinx's AutoESL [3], for instance, the top-level interface of the resulting hardware block (port directions and sizes) is inferred from the signature of a *single function*[2].

Another important characteristic that distinguishes hardware from software is resource allocation. The hardware is frozen and common software features, such as dynamic resource allocation, are not easily available. Therefore, in code suitable for hardware synthesis, all data structures must reside in statically allocated memory. Some works focus on this problem by relying on dynamic reconfiguration technologies of FPGAs to support hardware components instantiation at run-time[26]. However, this is not the focus of our current work. We aim at providing more general guidelines for describing components.

4.2 Defining C++ unified descriptions

C++ code unified and suitable for automatic implementation in both hardware and software must follow a careful design process so it will not contain characteristics specific of hardware or software. In this sense, the goal of this section is to demonstrate how the application-oriented design process proposed by ADESD yields designs that allow such concerns to be handled externally and implemented as aspect programs. To illustrate our design process, we describe in more details one of our case studies: the design and unified implementation of the EPOS's thread scheduler. The *Embedded Parallel Operating System* (EPOS) [27] is a case study on which the design artifacts proposed by ADESD were implemented and validated. The complex behavior of the scheduler motivated its choice as our main case study. Task scheduling involves both synchronous (e.g. creating a new thread) and asynchronous (e.g. preempting the execution of another component) operations. Furthermore, enabling a hardware/software scheduler allows a system with less jitter and better support for real-time applications [28]. Figure 2 shows the entities identified in the domain of real-time scheduling.

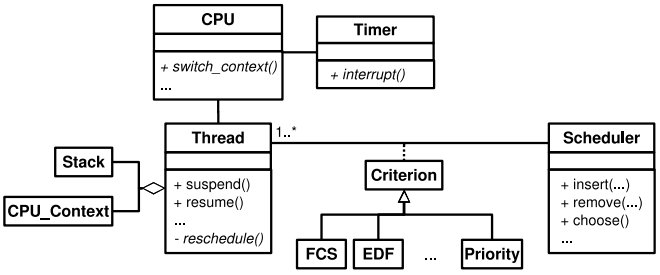


Figure 2. Simplified UML class diagram of scheduling-related classes in EPOS.

In this domain decomposition, a task is represented by the class *Thread* which defines the execution flow of the task, implementing the traditional functionality of such kind of abstraction (e.g. timing interrupt handling and context switching). Concerns such as timing management are performed through OS abstractions for CPU-dependent hardware (e.g. the *Timer* class).

The classes *Scheduler* and *Criterion* define the structure that realizes the task scheduling itself. Traditional implementations of scheduling algorithms are usually done by a hierarchy of specialized classes of an abstract *Scheduler* class, which can be further specialized to bring new scheduling policies to the system. In order to reduce the complexity of code maintenance (generally present in such hierarchy of specialized classes), as well as to promote its reuse, our design detaches the scheduling policy from its mechanisms (lists implementations) and also detaches the scheduling policy from the thread it represents. The separation of the mechanism from the scheduling policy is what in fact allows the deployment of the scheduler as both hardware and software. The unified description of the *Scheduler* component implements only the mechanisms that realize the ordering of the tasks, based on the selected policy. In this sense, the same component can realize distinct policies, as the definition of the policy is confined in the *Criterion* component, which isolates the comparison algorithm between the elements of the scheduler queue.

2. both tools also allow the specification of the entry point as a SystemC module with signal ports defining the IO protocol

4.2.1 Scheduler implementation

According to such decomposition, the scheduler must be implemented in a flexible way to enable the decoupling of the policy and the resource from the scheduling mechanisms. To achieve such flexibility without imposing performance penalties, the implementation of the scheduler itself makes extensive use of *static metaprogramming techniques* [5]. Figure 3 shows a more detailed definition of the scheduler. The component is actually defined through the class template *Scheduler<T>*, which receives the type of the resource being scheduled as template parameter (in this case, an object of the type `Thread`).

The *List*⊲⊢*Ordered_List* class hierarchy in Figure 3 defines a linked list used to implement the scheduling queues. An important aspect in this implementation lies in the fact that the linking and storage concerns are factored and implemented separately, while static metaprogramming is used to integrate these concerns at compile-time. That is, the list classes are not responsible for the allocation of memory for the internal nodes and the objects it stores. They implement only the list algorithms and deal with references to such elements. The motivation for this approach is to avoid the excessive allocation/deallocation of nodes when the same object is repeatedly inserted and then removed from the list, which tends to happen frequently in a thread scheduling environment. As can be seen in Section 4.3, this separation of concerns also allows us to handle memory allocation as a separated aspect according to the target implementation scenario.

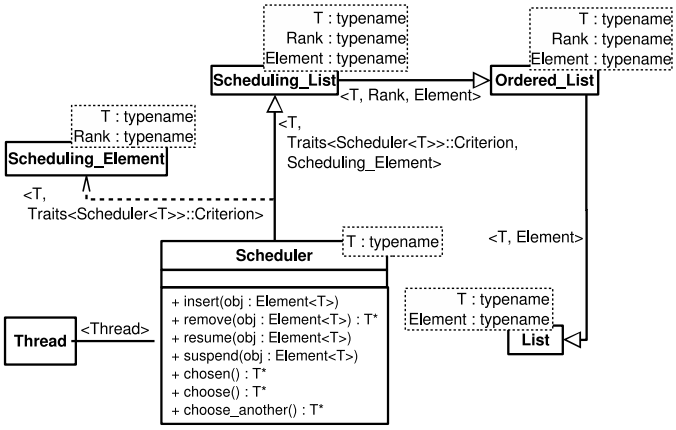


Figure 3. Definition of the Scheduler component

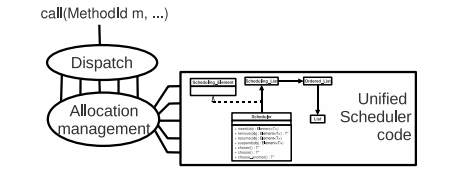
As can be seen in Figure 3, a node for the list is defined by the class template *Scheduling_Element<T,Rank>*. The *Rank* parameter defines the ordering criterion and maps directly to one of the criteria shown on Figure 2 (*Rank* is also a parameter for the class *Scheduling_List*). *Scheduler* then inherits from *Scheduling_List*, defining *Scheduling_Element* as the type for list nodes and *Traits<Scheduler<T> >::Criterion* as the ordering criterion. The criterion depends on static configurations defined in template classes called *Traits* [29]. In Figure 3, when the scheduler is used with the class *Thread*, *Traits<Scheduler<T> >* maps to *Traits<Scheduler<Thread> >*, which can be defined as follows:

```
template <> struct Traits<Scheduler<Thread> > {
  typedef Scheduling_Criteria:: Priority  Criterion ;
};
```

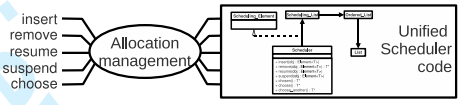
Trait classes provide a convenient way to associate related types, values, and functions with a template parameter type without requiring their definition as members of the type [29].

4.2.2 Scheduler implementation for HW/SW

Current HLS tools support ANSI C++ and allow the use of advanced language constructs, such as pointers, classes, arrays, and template metaprogramming. Therefore, the core scheduling algorithm and the class hierarchy shown previously required only minor modifications so it could serve as input for both hardware and software implementation flows. However, as mentioned in section 4.1, the *communication interface* and *resource allocation* must be handled differently whether a component is implemented as hardware or software. Figure 4 gives an overview of the resulting final implementations that incorporate these concerns.



(a) Hardware component



(b) Software component

Figure 4. Scheduler implemented in both hardware and software

As mentioned in section 4.2.1, the allocation of list nodes is not part of the core implementation of the scheduler. However, to implement the scheduler as a self-contained component in hardware, an *allocation management* block was defined separately and associated to the scheduler in order to keep the linked list node in the same "synthesis scope" of the final component. This was necessary since pointers to, for instance, memory locations allocated in software would not make sense in the scope of a self-contained hardware block. The interface of the hardware component was also redefined so that HLS tools could infer the top-level interface of the hardware block. The *call* function provides a single entry-point for the hardware scheduler. It receives a *MethodId* that defines which method of the scheduler should be called. This call is performed by a dispatching mechanism that wraps the original method call interface of the scheduler.

In summary, only resource allocation and the allocation interface were handled differently in hardware/software.

The case study presented in this section shows that a careful domain engineering and system design has favored a scheduler implementation in which specific hardware/-software characteristics were handled externally without affecting the core algorithmic implementation. In Figure 4, this core implementation is called *unified*, since exactly the same code is used for both hardware and software implementations. In the remainder of this section, we summarize some additional considerations must also be taken into account in order to produce such unified code. Then, in section 4.3, we present the generalization of these specific hardware/software concerns into aspect programs that can be weaved with the unified code only during the final stages of the design process

4.2.3 Handling pointers, polymorphism and bitfields

An extra care must be taken in order to produce code that can be synthesized to hardware and run efficiently in software. This is not due to specific hardware or software characteristics, but due to some characteristics of high-level synthesis [25]. One of these characteristics is related to the use of C++ *pointers*. Pointers have no intrinsic meaning in hardware. They are mapped by HLS tools to indices of the storage structures to which they point or to objects that can be statically determined. Thus, *no null or otherwise invalid pointers* are allowed in the source code. However, it is a common coding style to use null pointers to report failures. To avoid this, one can change the code to utilize *option types*. An option type is a container for a generic value, and has an internal state which represents the presence or absence of this value. We implemented an option type in the C++ class template *SafePointer<T>*, which has the following constructors:

```
template<typename T> class SafePointer {
 ...
   SafePointer(): _exists(false){}
   SafePointer(T thing, bool exists = true): _exists(exists), _thing(thing) {}
 ...
};
```

One constructor represents the absence of a value in the container while the other represents its presence. By replacing all occurrences of simple pointers (T*) in the scheduler code with *SafePointer<T*>* values, we have completely avoided passing invalid pointers around.

HLS tools also limit the use of C++ features that rely on dynamic structures in software (e.g. heaps, stacks, virtual method tables), such as recursion and dynamic polymorphism. However, there are means to implement this kind of behavior using static metaprogramming techniques and C++ templates, yielding code supported by HLS tools. Static polymorphism can be implemented as shown below:

```
template <typename derived_class> class Base {
  void interface () {
    static_cast<derived_class*>(this)->implementation();
  }
};

class Derived : public Base<Derived> {
  void implementation();
};
```

Classes can derive from template instantiations of the base class using themselves as template. This is also known as *Curiously Recurring Template Pattern* (CRTP) [30], and allows the static resolution of calls to virtual methods in the base class.

Another issue that needs attention is the use of bit-accurate data types. HLS tools usually provide specific libraries (e.g. the *ac_int/ac_fixed* from CatapultC [1]) that allow the programmer to use only the necessary bit width for each algorithm. However, while bit-accurate data types yield more efficient hardware, their use in software adds overhead since most compilers and processor architectures support only 8/16/32/64-bit integer types, thus requiring additional shifting and masking operations to emulate bit-accurate behavior. In our case studies we use only the standard C++ data types for maximum flexibility of the unified code. However, a possible way to provide bit accuracy while keeping a certain degree of uniformity is to use C++ *typedef* statements to define all allowed data types according to the target domain. For instance, a 5-bit integer can be defined in hardware using `typedef ac_int<5> int5` and in software using `typedef char int5`, since **char** is the smallest native type with the required width. Nonetheless, one must take extra care to keep a semantic consistency when mixing types with different widths.

Other aspects concerning hardware generation using HLS are related to the synthesis process. The same high-level algorithm can span several different hardware implementations. For instance, loops can have each iteration executed in a clock cycle, or can be fully unrolled in order to increase throughput at the cost of additional silicon area. This kind of synthesis decision is usually taken based on directives which are provided separately from the algorithm descriptions. The definition and fine tuning of these directives is part of the design space exploration process and is not in the scope of this work.

4.3 HW/SW aspects encapsulation

We have previously identified two aspects that must be handled differently whether the component is in hardware or in software: *storage allocation* and *method call interface*. Figure 5 shows how we have used scenario adapters to isolate such aspects. The set of aspects which are part of a scenario are incorporated by a class that defines the scenario itself. The scenario adapter then redefines the operations of the base components, wrapping the original behavior with scenario-specific operations. This structure is defined in the form of a metaprogrammed framework implemented in C++ using its OOP e static metaprogramming features. Each part of Figure 5 is explained in more details below.

4.3.1 HW/SW aspects

The class *Static Alloc* generalizes the structure shown in Figure 4 and defines as an aspect a storage allocator used to deal with the absence of dynamic memory allocation

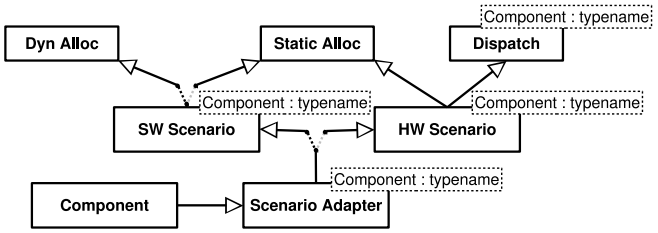


Figure 5. Software and hardware aspect weaving using a scenario adapter

in hardware, therefore it is part of the *HW Scenario*. Component operations may go through this allocator which reserves and releases storage space on demand. In this aspect, the number of storage slots for links and objects is defined at synthesis-time and allocation requests are just mapped to a free slot, therefore yielding a synthesizable allocator. In a *SW Scenario*, dynamic memory allocation is available, thus, storage allocation can be handled by either the *Static Alloc* or the *Dynamic Alloc* aspect, as shown in Figure 5. In this context, the implementation of *Dynamic Alloc* is straightforward and can just convert allocation request to a call using the C++'s *new* operator. A similar approach that uses external allocators for containers such as lists is provided by the *C++ standard template library* (STL) [31]. In principle, we could have relied on STL, however current STL implementations are not synthesizable.

The *Dispatch* aspect is used, in *HW Scenario*, to define an entry point for the component so it will be compliant with HLS tools requirements (a dispatch mechanism is not necessary in the software scenario since operations are requested using direct method calls). Since each tool require a different coding style, this mechanism must be designed to cope with such variation. Figure 6 shows how we have addressed this issue. The class *Dispatch Common* is used to encapsulate dependencies from specific HLS tools and IO protocols. The template parameter *Config* is used to enable the specialization of this class for these different tools in a straightforward way. Figures 6c and 6d provide code snippets that illustrate two possible specializations compliant with Calypto's CatapultC [1] HLS tool. The first one defines the method *top_level* as the entry-point. This method reads the invocation parameters (using a tool-specific library for IO) interprets their values and calls the appropriate method of the component. The second snippet follows an analogous approach which defines the entry-point as a SystemC module.

The actual implementation of the dispatching mechanism is implemented within the *Dispatcher* class. Since each component has a different method call interface, this class must be specialized for each one. Static polymorphism is employed (Section 4.2.3) so that the component-specific dispatching defined at *Dispatch* can be called from within *Dispatch Common*. The binding between *Dispatch* and the desired specialization of *Dispatch Common* is defined through a system-wide *Trait* class at the inheritance.

4.3.2 Scenario definition

The hardware scenario is defined by the *HW Scenario* class and aggregates the static allocation and dispatching aspects through multiple inheritance. The *SW Scenario* class aggregates only the allocation aspects using a *metaprogrammed conditional inheritance*. This is shown using the ⋱ notation in Figure 5, which denotes that only one of the two generalizations occur at the same time. The code below shows how the conditional inheritance is defined in the generic implementation of the *SW scenario*:

```
template <typename Component> public SW_Scenario :
  public
    IF<Traits<Component>::static_alloc,
        Static_Alloc ,
        Dyn_Alloc>::Result
```

The base aspect of the scenario is the result of the *IF* metaprogram which depends on static configurations defined in the component's *Traits*. The code sample below shows the *Trait* class used above, along with the implementation of the *IF* metaprogram:

```
template <> struct Traits<Component>{
    static const bool static_alloc = true;
     ...
};

// IF metaprogram
template<bool condition, typename Then, typename Else>
struct IF { typedef Then Result; };
template<typename Then, typename Else>
struct IF<false, Then, Else> { typedef Else Result; };
```

In the specialization of *Traits* for *Component*, the configuration *static_alloc* is equal to *true*, in other words, static allocation will be used instead of dynamic allocation when the software scenario is applied to *Component*. The *IF* metaprogram that choses between possible options is implemented using C++ partial template specialization. If *condition* is *true*, it defines *Result* as the type *Then*. This is encoded in the base definition of the template. The template is partially specialized for the false condition, defining *Result* as the type *Else*.

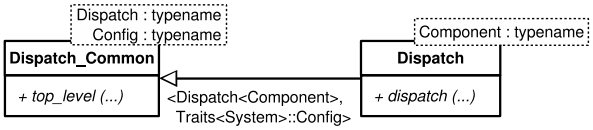
4.3.3 Scenario adapter definition

The scenario adapter incorporates the hardware/software aspects using conditional inheritance as well. The code sample below shows the generic declaration of the scenario adapter:

```
template <typename Component> class Scenario_Adapter :
  private
    IF<Traits<Component>::hardware,
        HW_Scenario<Component>,
        SW_Scenario<Component> >::Result,
  private
    Component
{
   ...
};

// Trait class for a component
template <> struct Traits<Component> {
  static const bool hardware = true;
  static const bool static_alloc = true;
};
```

The adapter inherits the component behavior directly from its unified implementation, while the base scenario is chosen by the *IF* metaprogram according to the component's traits. Components that can be implemented as either hardware or software have a trait named *hardware*



(a) Dispatcher definition

```
template<>
class Dispatch<Component> :
        public Dispatch_Common<Dispatch<Component>,
                               Traits<System>::Config> {
  void dispatch (...) {
    switch(op){
    case OP_0:
      static_cast<Scenario_Adapter<Component>*>(this)->op0();
      break;
    case OP_1: ...
    }
  }
};
```

(b) Component-specific dispatcher

```
template<typename Dispatcher>
class Dispatch_Common<Dispatcher,Config::CatapultC_CPP> {
 ..
   #pragma hls_design top
   void top_level(ac_channel<char> &data_in,
                  ac_channel<char> &data_out){
     // deserialize input parameters
     ...
     static_cast<Dispatch<Component>*>(this)->dispatch(...);
     // serialize operation return value
     ...
   }
};
```

(c) Specialization for a function-based entry-point

```
#pragma hls_design top
template<typename Dispatcher>
class Dispatch_Common<Dispatcher,Config::CatapultC_SC>:
                      public sc_module {
  sc_fifo_in<char> data_in;
  sc_fifo_out<char> data_out;
 ...
  SC_CTOR(Dispatch_Common){
      SC_THREAD(top_level);
  }
  //Analogous to the C++ version
  void top_level() ;
};
```

(d) Specialization for a SystemC-based entry-point

Figure 6. Specialization of the dispatching aspect for different entry-point requirements

that is used to define to which domain the component should be adapted.

4.3.4 Example: adapting the Scheduler

This example illustrates how a component can be adapted using scenario adapters. Considering, for instance, the scheduler case study, whose part of the base implementation (the *unified implementation*) is shown below:

```
template <typename T> class Scheduler {
 ...
  void insert(Element<T>* node);
  T* remove(Element<T>* node);
 ...
};
```

The *insert*/*remove* methods are responsible for insert/removing resources to/from the scheduling queue. As described in Section 4.2.2, in the unified implementation, these methods expect to receive the externally allocated nodes. The code sample below summarizes the definition of the scenario adapter for the scheduler:

```
template <typename T> class Scenario_Adapter<Scheduler<T> > :
  private
    IF<Traits<Scheduler<T> >::hardware,
       HW_Scenario<Scheduler<T> >,
       SW_Scenario<Scheduler<T> > >::Result,
  private
    Scheduler<T>
{
 ...
  Link insert(T *obj, Scheduler<T>::Criterion rank) {
    Link link = Scenario::allocate(obj, rank);
    Scheduler<T>::insert(Scenario::get(link));
    return link;
  }

  T* remove(Link link) {
    T* obj = Scheduler<T>::remove(Scenario::get(link));
    Scenario::free(link);
    return obj;
  }
 ...
};
```

The adapter declaration follows the approach described previously: conditional inheritance to choose between the scenarios and common inheritance to incorporate the unified implementation of the component. The body of the scenario adapter implementation contains the necessary adaptations of the components methods. The *insert* and *remove* methods of the *Scheduler* class are redefined to forward the allocation of nodes to the aspects. *Scenario::allocate(obj,crit)* creates an element for the scheduling list and receives as argument the object being scheduled and its scheduling criterion. After the list element is created, it is obtained through *Scenario::get* and inserted in the scheduler. Depending on the configuration defined in the scheduler's traits, *Scenario::** methods map to either *HW Scenario* or *SW Scenario*, which inherits *allocate*, *get* and *free* from the allocation aspects.

4.4 Summary and discussion

We have shown that a single C++ implementation can be used to generate both hardware and software provided that it follows a careful design process. By careful design process, we mean a process that assumes the later weaving of hardware/software dependencies using a scenario adapter. These dependencies and their influence in the unified implementation are summarized below:

Communication interface: limiting the components interactions to method calls allow the creation of a generic external mechanism (the dispatch aspect) that can be used by HLS tools to infer the final hardware interface.

Resource allocation: components that require resource allocation can be designed in order to deal with links to the expected resources, as shown in the implementation

of the lists in Section 4.2. The actual allocation can then be performed externally using the most suitable approach for the target domain.

In this work we have applied the concept of aspects and scenario adapters to introduce hardware/software dependencies and implemented the mechanisms using C++ template metaprogramming. However, other methods can be used for the same purpose. For instance, C's preprocessor macros (e.g. *#define* and *#ifdef* blocks) allow one to introduce software- or hardware-dependent code and to switch between different implementations. This a common approach to distinguish executable from synthesizable C++ in current industrial designs. Nevertheless, C++ templates offer two significant advantages over old-style C macros: 1) macros are basically a text replacement mechanism and do not offer any kind of syntax and type checking, whereas templates are type-safe and its misuse generates compile-time errors; 2) the implementation of a template can be partially or fully specialized to different types. Template specialization is what in fact allows the metaprogrammed mechanisms shown in the previous sections. Metaprogramming using C++ templates offers some disadvantages, however. Since templates are parsed internally by the C++ compiler, debugging template metaprograms would require a debugger for the C++ compilation process itself. One must rely only on source code analysis and error messages issued by the compiler.

Another approach is to fully rely on mechanisms implemented in EDA tools to generate both hardware and software from a single input model [12], [14], [15]. Although this usually provides a more automated design flow, it creates a strong dependence between the source code and very specific tools. Providing a systematic way to describe the semantics of the design process in the source code level reduces such limitations. ANSI C++ and its OOP features are extensively supported by hardware HLS tools and software compilers, thus increasing the applicability of our approach over different design flows and tools.

It is also worth recalling that, as mentioned in Section 4.2.3, in HLS, the design space can also be explored by using different synthesis directives (e.g. loop unrolling/pipelining), yielding different implementations for the same C++ construct. Although we do not deal with synthesis directives in this work, the techniques described in this paper can be used to encode multiple sets of implementation options. For instance, directives defined using *pragmas*[3] can be specified within an aspect. Indeed, this aspect would have to be specialized for each component, HLS tool, and intended hardware microarchitecture, since each of these would require a different set of pragmas.

3. the *#pragma* directive is the method specified by the C standard for providing specific additional information to the compiler. It is platform and compiler dependent.

5 Deployment of Unified Components

A uniform communication infrastructure is a requirement for design strategies in which the same component can have different implementations in different domains. In this section we describe our approach to support the seamless communication between hardware and software and the current design flow used to deploy our unified components.

5.1 Wrapping communication

As discussed in Section 4.1, the component *C2*, shown in the OO model in Figure 1, is an attribute of *C1*. Considering that *C1* is the top-level in a HLS/compilation step, *C2* would be compiled/synthesized together with *C1*. However, if the designer intends to have *C1* and *C2* as two independent components implemented in different domains, then an additional mechanism is necessary.

A way to overcome this issue is to use concepts from distributed object platforms [18]. Figure 7 illustrates *SW–>HW–>SW* interactions between components *C1*, *C2*, and *C3*. The callee is represented in the domain of the caller by a *proxy*. When an operation is invoked on the components' proxy, the arguments supplied are marshaled in a request message and sent through a *communication channel* to the actual component. In the *HW–>SW* interaction, an *agent* receives requests, unpacks the arguments and performs local method invocations. The whole process is then repeated in the opposite direction, producing reply messages that carry eventual return arguments back to the caller. An *agent* is not explicitly defined in the *SW–>HW* interaction since the *dispatcher aspect* can already play its role for components in hardware.

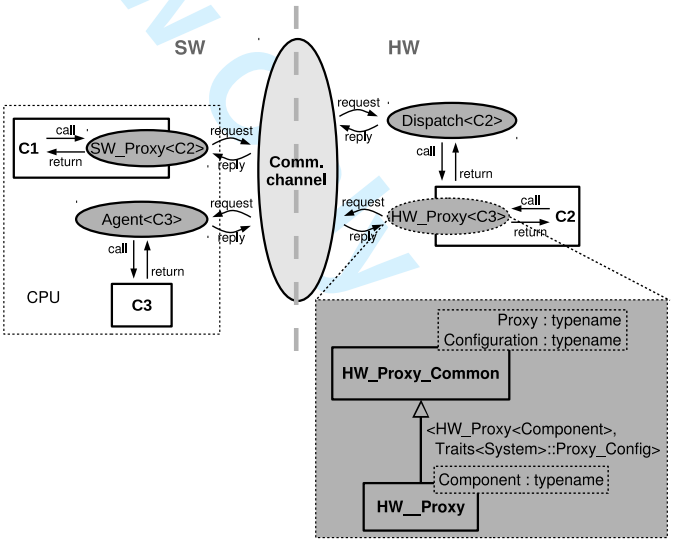


Figure 7. Communication between components in different domains (scenario adapters are omitted for simplicity)

The implementation of *channels*, *proxies* and *agents* can be realized in several different ways. For instance,

when a bus-based communication channel is used, a proxy in hardware may be implemented as slave device with memory mapped registers, and then notify the CPU through interrupt requests when it has a request message ready to be read. Alternatively, on a *Network-on-Chip* (NoC) based implementation, a packet-oriented interface would be used to transmit request messages.

Such variability is related to choices regarding the hardware/software architecture of the chosen implementation platform and should not affect the high-level components. Therefore, it is important to separate the component-specific implementation that is part from the behavioral model from the implementation related to artifacts of the underlying platform. Figure 7 also illustrates how we handle this separation of concerns. The platform-specific implementation of *HW_Proxy* is encapsulated in the class *HW_Proxy_Common*. *HW_Proxy_Common* is specialized for each platform following the same approach employed in the *Dispatch* aspect (see Figure 6). *HW_Proxy* then only defines the binding between the component's interface methods with the marshaling operations defined by the specialized *HW_Proxy_Common*.

A key remaining issue is now how to make these mechanisms transparent within the components' descriptions. An efficient solution is to use template metaprograms to replace the definition of a component by its proxy only when necessary:

```
//C2 definition
class C2 :
  public MAP<Scenario_Adapter<C2>, HW_Proxy<C2>, SW_Proxy<C2>,
         Traits<C2>::hardware>::Result {};

//MAP metaprogram
template<typename Implementation, typename HW_Proxy, typename
      SW_Proxy,
         bool hardware>
struct MAP {
  typedef IF<hardware==Traits<System>::hardware_domain,
             Implementation,
           IF<Traits<System>::hardware_domain,
             HW_Proxy,
             SW_Proxy>::Result> >::Result
    Result;
};
```

In the final implementation domain, *C2* can be defined as an empty class that inherits from its actual implementation depending on the configuration defined by *C2*'s traits. In the example above, the *MAP* metaprogram selects between *Scenario_Adapter<C2>*, *HW_Proxy<C2>*, and *SW_Proxy<C2>*. *MAP* uses the value of *Traits<System>::hardware_domain* to determine if the code is being submitted to a HLS tool or to a software compiler. If *Traits<System>::hardware_domain* is equal to *hardware*, which receives the value of *Traits<C2>::hardware* in the example above, then it must map to the actual implementation; therefore, *Result* is equal to *Scenario_Adapter<C2>* (*Scenario_Adapter<C2>* is also adapting C2's implementation to the correct scenario). Otherwise, it maps to one of the proxies according to the current implementation domain.

5.2 A deployment platform

In order to validate our mechanisms, we have assembled a platform for the deployment of our unified components. Figure 8 shows the general structure of the chosen hardware/software architecture. We rely on a NoC as the main communication link between hardware and software components. The *Real-time Star Network-on-Chip* (RTSNoC) [32] is the core component of our SoC platform. RTSNoC consists of *routers* with a star topology that can be arranged forming a 2-D mesh. Each router has eight bidirectional channels that can be connected to cores or to channels of other routers. Each hardware component is deployed as a node connected to a router. Software components are compiled with an RTOS and run in a *CPU node* that consists of a softcore CPU and memories. The internal structure of CPU nodes is based on the AXI4 family of protocols, which is becoming the industry's standard for bus-based interconnection. In our current implementation we have relied on IPs available at Opencores [33]. Our current CPU node, for instance, is based on the *Plasma* softcore, an implementation of the MIPS32 ISA.

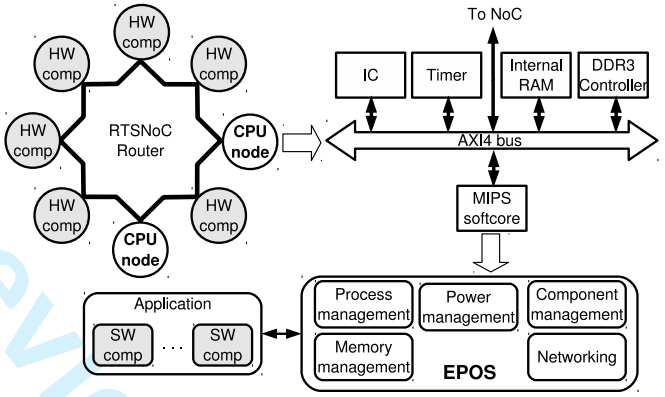


Figure 8. System-on-Chip platform

The software components run on EPOS [27]. EPOS provides the necessary run-time support to implement the platform-specific concerns of proxies and agents. A *component manager* handles communication between software and hardware components and keeps lists of all existing proxies to hardware and agents to software. Each component is associated to a unique ID that is mapped by a static resource table to a physical address in the NoC. Upon a call request, this address is used to build packets containing the target method ID and its arguments. These packets are sent through the NoC using the OS networking stack. If the method has return values, the component manager blocks until it receives the return values from the hardware component.

On the hardware side, the entry-point defined by the *dispatch aspect* (Section 4.3) blocks until a packet containing a method ID is received, followed by packets containing the arguments. It parses the packets and performs the local method invocation, sending back through the NoC eventual return values.

Packets sent to the CPU node trigger interrupts that are handled by an ISR defined by the component manager. The ISR reads all pending packets and performs the

necessary operations. When the manager receives a packet containing return values, it searches a list of blocked proxies and forwards the packet to the correct one. When a packet contains data from a method call request, the information is forwarded to the respective agent.

5.3 Design flow summary

This work focuses only on general implementation guidelines and not on a whole design flow. Issues such as design space exploration and design verification are not in the scope of this work. Nevertheless, our mechanisms based on standard C++ solutions allow for a straightforward integration with current synthesis tools and compilers. Figure 9 shows the steps we are currently using to go from C++ unified descriptions to components deployed in a physical platform.

The first steps show the artifacts of our proposed approach. Aspects, scenarios, and the unified implementation of components made up the inputs for the implementation flows. Proxies and agents can be automatically generated from the components' interface by a tool (EPOS's syntactical analyzer [34] can be used to obtain the operation signatures of all components). The final steps comprises the system generation. Once the hardware/software partitioning is defined (by the components' Traits), the adapted components are fed to either the hardware or the software generation flows.

In the software flow, components are compiled along with the run-time support implemented in EPOS using the GCC C++ compiler. In the hardware flow, CatapultC is used to generate RTL descriptions. These descriptions are bound with the hardware platform library and fed to the RTL synthesis tool. The platform library contains the hardware IPs described in Section 5.2.

6 Case Study

In order to evaluate our approach, we have analyzed an industrial PABX application and designed some of its basic building blocks using the proposed implementation guidelines. The main component of the PABX system is a commutation matrix that switches connections amongst different input/output data channels. These channels are connected to phone lines (through an AD/DA converter), tone generators, and tone detectors. The system also supports the transmission of voice data through an Ethernet network. Figure 10 shows the block diagram of the digital part of the PABX system, with is deployed as a SoC in an FPGA.

The components we have selected to be reimplemented using unified C++ are highlighted and appear in as hardware and software, since they can move between both domains depending on the final hardware/software partitioning. These components are described in more detail below:

EPOS's scheduler: the case described in Section 4.2.

ADPCM codec: an IMA ADPCM encoder/decoder that performs data compression using an *adaptive differential pulse-code modulation* (ADPCM) algorithm to convert 16-bit samples to 4-bit samples. The *encode* and *decode* operations are implemented in the same component so they can share a local lookup table.

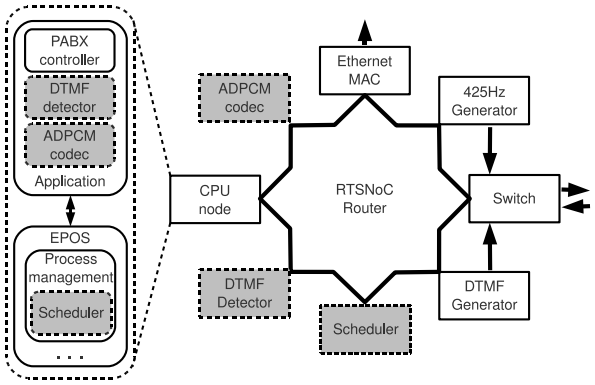


Figure 10. PABX SoC block diagram

DTMF detector: the *Dual-Tone Multi-Frequency* (DTMF) detector receives signals sampled from the phone lines connected to the central and detects DTMF tones. The entry point of the original C implementation was a single function that received a pointer to a frame of samples and returned the detected tone. The redesigned unified implementation defines two public operations: *add_sample* updates a sample of the current frame; and *do_dtmf* performs the detection.

6.1 Results

In order to demonstrate that unified implementations can be compared to dedicated ones in terms of efficiency, we compare software scenario-adapted components against the original C implementations (C++ for the scheduler), and hardware scenario-adapted components against components manually tailored for high-level synthesis.

Figure 11 compares the original software-only C (ADPCM and DTMF) and C++ (Scheduler) with the software scenario-adapted C++. The footprints were obtained from the object files generated after compiling each component in isolation, while the execution times were measured from the running application using a time-stamp counter. Everything was compiled with *gcc 4.0.2* targeting the MIPS32 ISA and using *level 2* optimizations. Figure 11a shows an average increase of about 4.9% in the total memory footprint. In the case of the scheduler, the overhead can be explained by the introduction of the *option types* and the use of a more generic mechanism for storage allocation (the *Static/Dyn Alloc* aspect). For the remaining case studies, most of the overhead comes from additional code required to encapsulate the behavior into more reusable OOP classes with a clear method interface.

Figure 11b shows the execution time of each component operation and compare the average values. The execution time of the Scheduler and the ADPCM codec is about 2.5% higher in the unified implementation. The execution

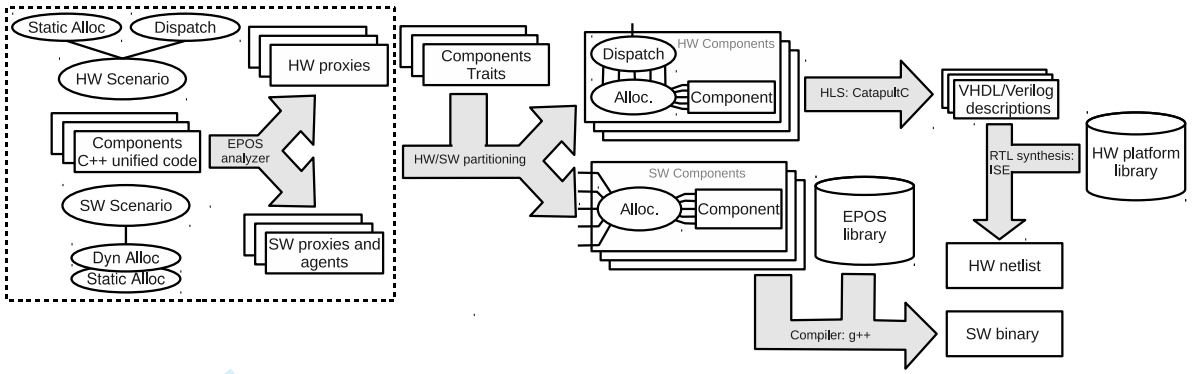
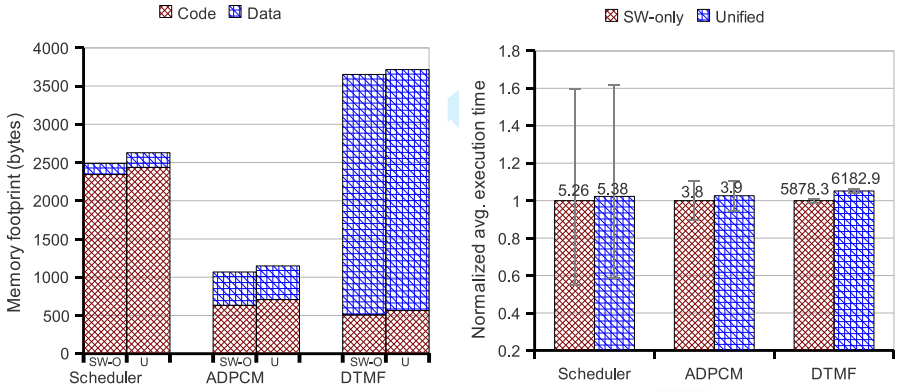


Figure 9. Summary of the implementation flow



(a) Memory footprint

(b) Average and operations execution time

Component		Execution time (μs)	
		SW-only	Unified
Scheduler	insert	6.0	6.0
	remove	2.9	3.3
	suspend	3.0	3.1
	resume	6.0	6.0
	choose	8.4	8.5
ADPCM	encode	4.2	4.2
	decode	3.4	3.6
DTMF	do_dtmf	5878.3	6182.9

Figure 11. Memory footprint (a) and execution time (b) of software-only C++ (SW-O) *vs.* Unified C++ adapted to the software domain (U). The normalized average execution times are plotted with absolute values above their respective bars.

time of the scheduler varies significantly according to the number of threads in the system (see error bars in Figure 11b). In this evaluation, we have experimented with 8 threads and a round-robin scheduling criterion. For the DTMF detector, the difference increases to 5%. The original DTMF detector requires a single call to *do_dtmf* to analyze a frame of samples, while the refactored DTMF detector requires several calls to *add_sample* before performing the same task, which results in a more significant increase in the execution time.

Table 2 and Figures 12–13 compare the hardware generated from unified C++ against hardware-only C++. *Calypto's CatapultC UV 2011a* was used to obtain RTL descriptions of the components. The descriptions were then synthesized using *Xilinx's ISE 13.4* targeting a *Virtex6 XC6VLX240T* FPGA. CatapultC and ISE were configured to minimize circuit area considering a target operating frequency of 100 MHz (the operating frequency of the SoC).

Table 2 shows the amount of FPGA resources required for each component and Figure 12 plots the *average resource utilization*. The *average resource utilization* is the arithmetic mean of the amount of each specific resource weighted by its total amount available. This resulting

Table 2
FPGA synthesis results of hardware-only C++ *vs.* unified C++

Resource	Scheduler		ADPCM		DTMF	
	HW-only	Unified	HW-only	Unified	HW-only	Unified
6-input LUT	2119	2540	524	615	387	443
Flip-flops	1849	2766	208	368	331	431
36Kb BRAM	0	0	1	1	2	1
DSP slice	0	0	0	0	6	6
Max. Freq.	126 MHz	114 MHz	118 MHz	124 MHz	102 MHz	101 MHz

value estimates the total amount of FPGA area (in %) required and is used as an area comparison metric. The results vary significantly in each case study. The apparent smaller area of the unified DTMF detector is due to different mapping of resources between *RAM blocks* and *flip-flops*, which resulted in smaller *average resource utilization*. On the other hand, the unified Scheduler and ADPCM require about 39% and 27% more area, respectively. Since the implementations of the HW-only and unified ADPCMs are very similar, we conclude that, in this case, most of the extra area comes from the *Dispatch* aspect. The aspect implements a generic mechanism

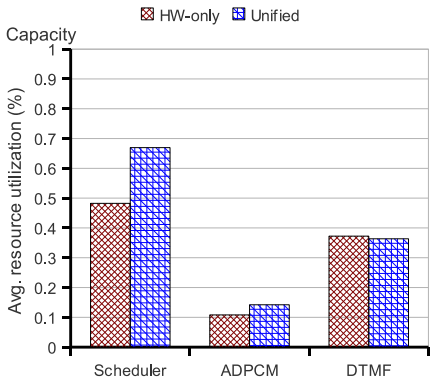


Figure 12. Average FPGA resource utilization

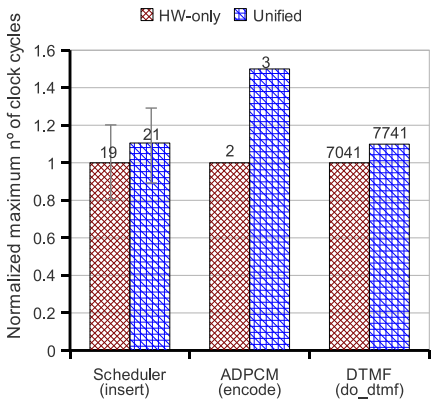


Figure 13. Performance of hardware components

for parsing and issuing operation requests, while the hardware-only descriptions focus on more specific and optimized interfaces. This overhead can be considered significant and increases with the number of supported operations. Nevertheless, this overhead is expected to be significantly reduced in future implementations of *Dispatch*.

Figure 13 shows the maximum number of clock cycles required to perform an operation. For the unified implementations of the Scheduler and the ADPCM, the *agents'* overhead is proportional to the number of arguments in the operation (2 for *Scheduler::insert* and 1 for *ADPCM_Codec::encode*). The high absolute overhead of the unified DTMF detector comes from the several invocations of the *add_sample* method to fill its internal buffer. This operation is implemented in a stream-like fashion in the HW-only detector.

To conclude our analysis, we have evaluated the total overhead of the communication infrastructure required to handle transparent hardware/software communication. EPOS's internal components are implemented using static metaprogramming to achieve high reusability with low overhead and its final footprint depends on the application configuration. Therefore, it does not make sense to evaluate certain components in isolation (e.g. the *Component Manager*—Section 5.2, which implements the runtime support). In order to provide such evaluation, Table 3 thus shows the total footprint considering the following different partitionings of our PABX SoC: all the case studies are in software; all the case studies are in hardware; and a hybrid partitioning in which only the scheduler is in software. The values in the *System* column are the total footprints subtracted by the footprints of all case studies in the respective domain. For instance, the software footprint shown in the *System* column for a specific partitioning is the total software footprint subtracted by the ones of all case studies implemented as software in that partitioning. This value allows us to evaluate the overhead added by the component management mechanisms.

Table 3
SoC area footprint. In the hybrid partitioning, only the Scheduler is in software.

Partitioning	Memory (code+data)		FPGA Avg. rsc. util.	
	Total	System	Total	System
All SW	19553 b	12201 b	3.70%	3.70%
All HW	16479 b	16479 b	5.05%	3.75%
Hybrid	19093 b	16605 b	4.24%	3.74%

As can be seen in Table 3, moving all cases to hardware reduced the total footprint. By comparing the values in the *System* column, we can see that, in the *All HW* partitioning, the introduction of the component management mechanism added an overhead of 4278 bytes. By moving the Scheduler back to software in the *Hybrid* partitioning, the aforementioned overhead is reduced to 4170 bytes. Considering the number of operations implemented in each proxy, we estimate an initial overhead of about 4 Kbytes plus 30 bytes for each additional operation. We expect to significantly reduce this overhead in future implementations, however, the current results are acceptable if compared with similar infrastructures [35]. On the hardware side, the calculated area overhead when all cases are in hardware represents only 1.3% of the total circuit area and negligible when represented in terms of the total amount of resources available in the target device (only 0.05%).

7 Conclusion

In this paper we have explored a methodology based on AOP and OOP concepts in order to produce unified descriptions of hardware and software components. We have shown that components designed following the principles presented in this work are susceptible to both software and hardware generation using standard compilers and HLS tools. This is possible through the isolation of specific hardware and software characteristics (resource allocation and communication interface) into aspect programs which are weaved with the unified descriptions only during the final stages of the design process.

The main contribution of this work is to show how the unified implementation of components, aspects, and aspect weaving mechanisms can be implemented only using standard C++ and its metaprogramming features. Therefore, the extraction of hardware/software implementation from the unified implementation happens directly through language-level transformations, thus facilitating compatibility with different C++-based HLS tools and design flows. We have demonstrated our methods by redesigning some functional blocks of a PABX system. The resulting components confirmed that, at an acceptable cost in area and performance, we can use C++ as a unified language to implement both hardware and software in a straightforward way.

Our future works are twofold. As the comparison between hardware implementations suggests, new features must be considered in order to provide a better support for high throughput data-oriented applications. Our main concern with this aspect is on the seamless integration of stream-oriented interfaces with our object-oriented modeling approach. Furthermore, though the application of our design strategy to new cases, we also aim to identify new hardware/software aspects that can be encapsulated using the proposed mechanisms.

REFERENCES

[1] Calypto Design Systems, "CatapultC Synthesis," 2011, http://www.calypto.com/.
[2] Forte Design Systems, "Cynthesizer," 2011, http://www.forteds.com.
[3] Xilinx, "AutoESL High-Level Synthesis," 2012, http://www.xilinx.com/tools/autoesl.htm.
[4] F. V. Polpeta and A. A. Fröhlich, "On the Automatic Generation of SoC-based Embedded Systems," in *Proc. of the 10th IEEE Int. Conf. on Emerging Technologies and Factory Automation*, Catania, Italy, 2005.
[5] K. Czarnecki and U. W. Eisenecker, *Generative programming: methods, tools, and applications*. New York, USA: ACM Press/Addison-Wesley Publishing Co., 2000.
[6] S. Ha, S. Kim, C. Lee, Y. Yi, S. Kwon, and Y.-P. Joo, "PeaCE: A hardware-software codesign environment for multimedia embedded systems," *ACM Trans. Des. Autom. Electron. Syst.*, vol. 12, pp. 24:1–24:25, May 2008.
[7] J. Eker, J. W. Janneck, E. A. Lee, J. Liu, X. Liu, J. Ludvig, S. Neuendorffer, S. Sachs, and Y. Xiong, "Taming Heterogeneity - The Ptolemy Approach," in *Proc. of the IEEE*, 2003, pp. 127–144.
[8] M.-A. Dziri, W. Cesario, F. Wagner, and A. Jerraya, "Unified component integration flow for multi-processor SoC design and validation," in *Proc. of the Design, Automation and Test in Europe Conf. and Exhibition*, vol. 2, 2004, pp. 1132–1137.
[9] A. Davare, D. Densmore, T. Meyerowitz, A. Pinto, A. Sangiovanni-Vincentelli, G. Yang, H. Zeng, and Q. Zhu, "A Next-Generation Design Framework for Platform-based Design," in *Proc. of the Design & Verification Conf. & Exhibition*, February 2007.
[10] A. Sangiovanni-Vincentelli and G. Martin, "Platform-Based Design and Software Design Methodology for Embedded Systems," *IEEE Design & Test of Computers*, vol. 18, pp. 23–33, 2001.
[11] J. Cong, B. Liu, S. Neuendorffer, J. Noguera, K. Vissers, and Z. Zhang, "High-Level Synthesis for FPGAs: From Prototyping to Deployment," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 30, no. 4, pp. 473 –491, 2011.
[12] A. Schallenberg, W. Nebel, A. Herrholz, P. A. Hartmann, and F. Oppenheimer, "OSSS+R: a framework for application level modelling and synthesis of reconfigurable systems," in *Proc. of the Conf. on Design, Automation and Test in Europe*, Nice, France, 2009, pp. 970–975.
[13] F. Mischkalla, D. He, and W. Mueller, "Closing the gap between UML-based modeling, simulation and synthesis of combined HW/SW systems," in *Proc. of the Conf. on Design, Automation and Test in Europe*, Dresden, Germany, 2010, pp. 1201–1206.
[14] J. Keinert, M. Streubühr, T. Schlichter, J. Falk, J. Gladigau, C. Haubelt, J. Teich, and M. Meredith, "SystemCoDesigner—an automatic ESL synthesis approach by design space exploration and behavioral synthesis for streaming applications," *ACM Trans. Des. Autom. Electron. Syst.*, vol. 14, pp. 1:1–1:23, 2009.
[15] D. Rainer, G. Andreas, P. Junyu, S. Dongwan, C. Lukai, Y. Haobo, A. Samar, D. Daniel *et al.*, "System-on-chip environment: a SpecC-based framework for heterogeneous MPSoC design," *EURASIP Journal on Embedded Systems*, 2008.
[16] E. Lubbers and M. Platzner, "A portable abstraction layer for hardware threads," in *Field Programmable Logic and Applications, 2008. FPL 2008. Int. Conf. on*, 2008, pp. 17 –22.
[17] H. K.-H. So and R. Brodersen, "A unified hardware/software runtime environment for FPGA-based reconfigurable computers using BORPH," *ACM Trans. Embed. Comput. Syst.*, vol. 7, pp. 14:1–14:28, 2008.
[18] F. Rincón, J. Barba, F. Moya, F. J. Villanueva, D. Villa, J. Dondo, and J. C. López, "Transparent IP Cores Integration Based on the Distributed Object Paradigm," in *Intelligent Technical Systems*, ser. Lecture Notes in Electrical Engineering, 2009, vol. 38, pp. 131–144.
[19] G. Booch, R. A. Maksimchuk, M. W. Engle, B. J. Young, J. Connallen, and K. A. Houston, *Object-oriented analysis and design with applications*. Addison-Wesley, 2007.
[20] G. Kiczales, J. Lamping, A. Mendhekar, C. Maeda, C. V. Lopes, J.-M. Loingtier, and J. Irwin, "Aspect-Oriented Programming," in *Proc. of the European Conf. on Object-oriented Programming*, vol. 1241, Jyväskylä, Finland, 1997, pp. 220–242.
[21] T. R. Mück, M. Gernoth, W. Schröder-Preikschat, and A. A. Fröhlich, "Implementing OS Components in Hardware using AOP," *SIGOPS Operating Systems Review*, vol. 46, no. 1, pp. 64–72, 2012.
[22] C. Larman, *Applying UML And Patterns: An Introduction To Object-Oriented Analysis And Design And Iterative Development*. Prentice Hall PTR, 2005.
[23] A. Alexandrescu, *Modern C++ Design: Generic Programming and Design Patterns Applied*, ser. C++ in-Depth Series. Addison-Wesley, 2001.
[24] O. Spinczyk, A. Gal, and W. Schröder-Preikschat, "AspectC++: an aspect-oriented extension to the C++ programming language," in *Proc. of the Fortieth Int. Conf. on Tools Pacific: Objects for internet, mobile and embedded applications*, Sydney, Australia, 2002, pp. 53–60.
[25] J. P. P. Flor, T. R. Mück, and A. A. Fröhlich, "High-level Design and Synthesis of a Resource Scheduler," in *Proc. of the 18th IEEE Int. Conf. on Electronics, Circuits, and Systems*, Beirut, Lebanon, 2011, pp. 736–739.
[26] N. Abel, "Design and Implementation of an Object-Oriented Framework for Dynamic Partial Reconfiguration," in *Proc. of the Int. Conf. on Field Programmable Logic and Applications*, 2010, pp. 240 –243.
[27] The EPOS Project, "Embedded Parallel Operating System," 2011, http://epos.lisha.ufsc.br/.
[28] H. Marcondes, R. Cancian, M. Stemmer, and A. A. Fröhlich, "On the Design of Flexible Real-Time Schedulers for Embedded Systems," in *Proc. of the Int. Conf. on Computational Science and Engineering - Volume 02*, Washington, USA, 2009, pp. 382–387.
[29] N. C. Myers, "Traits: a new and useful template technique," *C++ Report*, June 1995. [Online]. Available: http://www.cantrip.org/traits.html
[30] J. O. Coplien, "Curiously recurring template patterns," *C++ Rep.*, vol. 7, pp. 24–27, 1995.
[31] A. Stepanov and M. Lee, "The Standard Template Library," HP Laboratories, Tech. Rep., 1995.
[32] M. D. Berejuck, "Dynamic Reconfiguration Support for FPGA-based Real-time Systems," Federal University of Santa Catarina, Florianópolis, Brazil, Tech. Rep., 2011, PhD qualifying report.
[33] "Opencores," April 2011, http://opencores.org/.
[34] A. Schulter, R. Cancian, M. R. Stemmer, and A. A. M. Fröhlich, "A Tool for Supporting and Automating the Development of Component-based Embedded Systems," *Journal of Object Technology*, vol. 6, no. 9, pp. 399–416, Oct 2007.
[35] G. Gracioli and A. A. Fröhlich, "ELUS: A dynamic software reconfiguration infrastructure for embedded systems," in *Proc. of the IEEE 17th Int. Conf. on Telecommunications*, april 2010, pp. 981 –988.

Tiago Rogério Mück and
Prof. Dr. Antônio Augusto Fröhlich
Federal University of Santa Catarina
Florianópolis, 88040-000, Brazil
Phone: +55 (48) 3721-9516
E-Mail: {tiago,guto}@lisha.ufsc.br

March 16, 2013

IEEE Transactions on Computers
Editor-in-Chief
Prof. Dr. Albert Y. Zomaya

Dear Prof. Zomaya,

A revised version of the manuscript "Unified Design of Hardware and Software Components", initially submitted to the IEEE Transactions on Computers on July 2012 (log number TC-2012-06-0397.R1), has just been uploaded through the manuscript submission site under the title "Towards Unified Design of Hardware and Software Components Using C++".

First of all, we would like to thank the reviewers for the detailed discussion. In summary, the main changes carried out in this resubmission were:

- the manuscript title was changed to "Towards Unified Design of Hardware and Software Components Using C++"
- Figure 2 and the second paragraph in section 4.1 were changed to explain more clearly the issue of mapping object-oriented design to physical implementations;
- To clarify the points highlighted by the reviewers regarding the general applicability of our approach, major changes were performed in the remaining of section 4 and in section 5:
  - We extended the 6th paragraph of section 4.2 and moved it a new section *4.2.1 Scheduler implementation*. This section gives more details about our example;
  - The explanation of how hardware/software concerns influences component implementations was improved and moved to section a new section *4.2.2 Scheduler implementation for HW/SW*;
  - Former section 4.2.1 became section 4.2.3;
  - For improved readability, the scenario adapter example in section 4.3.1 is now part of a new section *4.3.4 Example: adapting the Scheduler*;
  - The definition of the dispatching aspect was improved to show how different synthesis tools and interfaces can be supported and abstracted;
  - We extended section 5 with more details about the implementation of proxies and agents, emphasizing how different target platforms can be supported;
- Case studies description in section 6 is summarized to reduce the section size and make room for the improvements above;
- Improved section 7 with some direction for future work.

Please find below a more detailed description of how we have addressed the issues pointed out by each reviewer.

Reviewer #1

1. Reviewer's comment: *The examples are not discussed very well. Instead of giving some details, the authors add some basic literature on programming. For me, this is not a satisfactory response to review and I cannot understand why the authors did not simply explain their examples.*

   We have reviewed the previous manuscript considering the reviewer opinion. In the revised version, we have substantially extended the description of our example in section 4.2. The new sections 4.2.1 and 4.2.2 now gives more details about the how hardware/software specific concerns are detached from the base implementation of the scheduler. Sections 4.3 and 5.1 were also extended in order to explain how our design artifacts can be used to support different synthesis approaches and target platforms. We

cover this concern in the next answer and in the answer for Review #3, comment #3.

2. Reviewer's comment: *An important conecern in my review was that the proposed approach is only useful when using a high-level synthesis tool, which supports synthesis of functions. The authors responded to this concern by simply adding a sentence that their proposed approach could be used with other synthesis tools as well. So far, so good, but I am not interested in reading that the proposed approach could be used with other tools I am interested in understanding how it works.*

   We have extended the description of the dispatching mechanism in section 4.3.1 and given more details related to its specialization considering different interface synthesis requirements. In the revised manuscript, Figure 6 illustrates this process and shows specializations for two different entry-point definitions: one based on a single function and one based on a SystemC module. We believe this new example gives the reader a clearer understanding on how our static metaprogrammed approach can be further applied to different C++-based synthesis tools.

Reviewer #2

1. Reviewer's comment: *The only complain I have is a minor style issue. We now have section 4.2.1, but there is no section 4.2.2. The authors may want to merge section 4.2.1 into section 4.2.*

   We thank the reviewer for pointing out this issue. This has been addressed in the revised manuscript.

Reviewer #3

1. Reviewer's comment: *References to Wikipedia should be replaced by well recognised literature.*

   We have changed the previous references about OOP, UML, and C++ to the following ones:

   [6] K. Czarnecki and U. W. Eisenecker, Generative programming: methods, tools, and applications. New York, USA: ACM Press/Addison-Wesley Publishing Co., 2000.

   [34] C. Larman, Applying UML And Patterns: An Introduction To Object-Oriented Analysis And Design And Iterative Development. Prentice Hall PTR, 2005.

   [35] A. Alexandrescu, Modern C++ Design: Generic Programming and Design Patterns Applied, ser. C++ in-Depth Series. Addison-Wesley, 2001.

2. Reviewer's comment: *The approach is technically based on C++. For this reason I would propose to modify the title of the article in the following way: "Towards a Unified C++-based Design of Hardware and Software Components"*

   We would like to thank the reviewer for the suggestion. Although the object/aspect oriented domain decomposition is independent from the implementation language, we agree that our current approach has been driven by technical characteristics of C++ and the support of HLS tools for C++-based languages.

3. Reviewer's comment: *The overall approach is targeted for a specific implemenation platform, as described in Section 5.2. Still it becomes not clear how the proposed approach supports a separation of the behaviour/functionality and the mapping to the implementation platform (including certain artefacts). For instance, a real-time operating system could be considered as platform artefact. For reuseablilty of the behaviour model (i.e. retargeting to another platform with a different or even without an operating system) the separation of functionality and structure should be as explicit as possible. The approach presented in the article seems to be very tailored to the EPOS environment. If this should be a remaining limitation of the work I would appreciate if you would metion this in the future work.*

   We presented the platform described in 5.2 as a possible "implementation instance" of the mechanisms mentioned earlier on section 5, however our approach is not limited to that particular system architecture. Nevertheless, we agree with the reviewer comment since we have skipped some details on how this portability can be provided. In the revised manuscript, we have improved the first paragraphs of section 5 and figure 7 in order to illustrate how different implementation approaches can be encapsulated using templates and static metaprogramming. This encapsulation is performed by first separating in different classes the component-specific and the platform-specific behavior of proxies/agents. Different platform-specific behavior is supported by specializing the related classes in a way in which the correct platform can then be automatically chosen through configurations provided in system trait classes. A comprehensive example of this approach has also been added to section 4.3.1 (please, refer to Review

#1, comment #2)

4. Reviewer's comment: *In general the conclusion could be significantly improved by a summary of the advantages and drawbacks of the proposed approach. In an outlook on future work possible solutions for the mitigation of existing drawbacks could be described. This would definitely further enhance the quality of the article.*

   Although we have chosen to keep a more conceptual discussion at section 4.4, we extended section 7 to stress the contribution and also provide our prospect for future work:

```
  Our future works are twofold. As the comparison
between hardware implementations suggests, new fea-
tures must be considered in order to provide a better
support for high throughput data-oriented applications.
Our main concern with this aspect is on the seamless
integration of stream-oriented interfaces with our object-
oriented modeling approach. Furthermore, though the
application of our design strategy to new cases, we also
aim to identify new hardware/software aspects that can
be encapsulated using the proposed mechanisms.
```

Sincerely,

Tiago Rogério Mück and Antônio Augusto Fröhlich

Annex: Changes in the first resubmission

Main changes carried out in the first resubmission:

- In section 3, we added an explanation about basic OOP concepts and extended the description of ADESD and AOP;
- Minor changes in section 4.1 to make the purpose of Figure 2 clearer;
- To clarify the points highlighted by the reviewers and improve the paper organization, major changes were performed in the remaining of section 4:
  - We extended the original section *4.2.1 Component implementation* and merged its contents in to the top-level section *4.2 Defining C++ unified descriptions*;
  - Concerns related to the limitations of high-level synthesis were moved to a new section *4.2.1 Synthesis considerations*;
  - We added a paragraph about bit-accurate data types in the new section 4.2.1;
  - The original subsection 4.2.2 became subsection 4.3 and was extended to provide a more detailed explanation about our approach to apply hardware/software aspects and to consider alternative solutions;
  - Former sections *4.2.3 Wrapping communication* and *4.3 Implementation flow summary* moved to a new section *5 Deployment of unified components*;
  - We added a new section *4.4 Summary and discussion* that summarizes our approach and discuss other methods for describing hardware and software in a common language;
- We extended section 5 with more details about the implementation of proxies and agents;
- Minor changes in section 6 (former section 5):
  - Details of the base SoC platform moved to section 5;
  - Error bars added to measured execution times;
  - Improved the readability of Table 5;
  - Minor changes to reduce the section size

Please find below a more detailed description of how we have addressed the issues pointed out by each reviewer.

Reviewer #1

1. Reviewer's comment: *In my opinion, the main weakness of the paper lies in the presentation of proposed solution in section 4. Instead of giving general guidelines, the authors only present sample code snippets, which provide solutions for their running example.*

   We have carefully analyzed the way we present our approach, and we agree that the most general guidelines were indeed not very clear. In summary, these guidelines consists in the following: 1) limiting component interaction to method calls allow the creation of a generic external mechanism; 2) components that require resource allocation must be designed in order to deal with links to the expected resources, thus allowing such allocation to be performed externally using the most suitable approach for hardware or software; and 3) the C++ in the unified implementation must be synthesizable. We have reorganized and extended section 4 in order to make the considerations above clearer.

   The latter guideline does not affect the actual design of the components, but consists mostly of coding guidelines that must be followed due to current limitations of high-level synthesis. Such guidelines were all moved to a new section *4.2.1 Synthesis considerations*.

   The former ones must be considered during the domain decomposition process; otherwise the incorporation of hardware/software characteristics using the proposed aspect weaving approach would not be possible. Since explaining the OOP decomposition process itself is not in the scope of this paper, we have showed the *Scheduler* as a case that satisfies the considerations mentioned above. In order to make clearer how this is achieved, we have extended the description of the Scheduler in section *4.2 Defining C++ unified descriptions*. Also, to better demonstrate how hardware/software characteristics can then be applied in a systematic way using the proposed scenario adapters, we have reorganized and extended section *4.3 HW/SW aspects encapsulation*: section 4.3.1 first presents the hardware/software aspects in details; section 4.3.2 then shows their aggregation to build a scenario; and section 4.3.3 describes the scenario adapter definition. We have also replaced the previous scenario diagram of Figure 4 by a more detailed one which shows the inheritance/template specialization structure more clearly.

   We have also created a new section *4.4 Summary and discussion*. This section summarizes our strategy and highlights the points mentioned above. It also includes a discussion suggested in the comments 2 and 4 from reviewer #3.

2. Reviewer's comment: *Often, the code snippets are even not discussed at all, which makes it difficult to even comprehend this particular solution. Especially, I do have problems understanding the Traits example on page six in the right column.*

   We thank the reviewer for pointing out these issues. We have improved the explanation of all code snippets. The provided explanations, however, assumes prior knowledge about object-oriented constructs in C++ and templates. Due to the page limit, we cannot review more general C++ concepts in the paper. Therefore, we have only added the following footnote (page 3, third paragraph) that provides some resources to which the reader may refer:

   ```
   1. this paper assumes prior knowledge about OOP, UML and C++.
   An overview of the relevant concepts is available at [31], [32], and
   [33]. The reader may also refer to [34], [7], and [35] for an indepth
   explanation.
   ```

   The following additional references were added:

   ```
   [31] "Wikipedia - Class diagram," 2012,
        http://en.wikipedia.org/wiki/Class_diagram.
   [32] "Wikipedia - Object-oriented programming," 2012,
        http://en.wikipedia.org/wiki/Object-oriented_programming.
   [33] The C++ Resources Network, "Templates," 2012,
        http://www.cplusplus.com/doc/tutorial/templates/.
   [34] C. Larman, Applying UML And Patterns: An Introduction To Object-
        Oriented Analysis And Design And Iterative Development. Prentice
        Hall PTR, 2005.
   [35] A. Alexandrescu, Modern C++ Design: Generic Programming and
        Design Patterns Applied, ser. C++ in-Depth Series.     Addison-
        Wesley, 2001.
   ```

   Regarding the specific Traits example, we have extended the explanation of the trait concept in the second paragraph of section 4.3.2. Additional code examples are provided in sections 4.3.2 and 4.3.3. We have also added a reference to the original proposal of the trait concept, which provides additional examples:

[43] N. C. Myers, "Traits: a new and useful template technique," C++ Report, June 1995. [Online]. Available: http://www.cantrip.org/traits.html

3. Reviewer's comment: *No alternative solutions are discussed. This would have been interesting especially for the static allocation and dispatching aspects. In particular, the proposed dispatching strategy is very specific for the CatapultC high-level synthesis. Nearly all other HLS tools do not use a single function signature for describing hardware components. They typically require to present hardware components as SystemC modules having signal ports or transaction sockets. In this case, the proposed solution could not be applied directly.*

   We agree with the reviewer in the sense that in our explanation we mentioned only one of the possible approaches; however, this is not a limitation for our proposal. Our current implementation of the dispatcher is compliant only with CatapultC since it is the tool used in our experimental evaluation. Nevertheless, the technical effort to define a top-level SystemC module for another HLS tool is about the same. To highlight the alternative solutions, we have added the following sentences in the last paragraph of section 4.3.1: 1) *In Calypto's CatapultC [1] and Xilinx's AutoESL [5], for instance, the top-level interface of the resulting hardware block (port directions and sizes) is inferred from a single function signature*; and 2) *Some tools require the definition of the entry point as a SystemC module with signal ports used to define the IO protocol. Our current implementation is compliant only with CatapultC, since it is the tool used in our experimental evaluation. Nevertheless, the Dispatch aspect can be specialized to support different entry-point requirements and different IO protocols (e.g. two-way handshaking, bus-based, etc). Upon system generation, the desired dispatcher can be selected using a* Trait.

   We have also added more details to Figure 4, which makes clearer the specialization of the *Dispatch* aspect.

4. Reviewer's comment: *Moreover, why do the authors define their own allocation class while the standard template library is already providing a ready to use implementation? In summary, by only providing some examples, it is not clear what exact solution is proposed. Hence, the applicability of the solution and its limitations remain unclear.*

   In principle, we could have relied on STL; however, current STL implementations are not synthesizable. Nevertheless we agree that STL should be mentioned in the paper as a possible alternative. We have added the following sentence in the third paragraph of section 4.3.1: *A similar approach that uses external allocators for containers such as lists is provided by the C++ standard template library (STL) [42]. In principle, we could have relied on STL; however, current STL implementations are not synthesizable.*

   Also in section 4.3.1, we have extended the explanation of the List code example and added a code snippet showing part of the *Static Alloc* aspect implementation. We believe this will give the reader a clearer idea of the applicability of our allocators.

5. Reviewer's comment: *Although the paper addresses an important topic and the results clearly show the benefits of the proposed approach, the presentation of the key contributions is in my opinion too weak to accept the paper for publication. As only some illustrative examples are given without providing more general guidelines, the paper looks more like a case study presentation than a research paper.*

   In the answer for comment #1 we have described how we have addressed these issues. We would like to thank the reviewer for the constructive comments and we hope that the improvements to the manuscript will be sufficient to answer the main concerns raised by the reviewer.

Reviewer #2

1. Reviewer's comment: *Without reading the previous work of the authors it is very hard to get a deeper understanding of the technical concept and soundness of the approach. Section 3 should better aim on giving an introduction to the general concepts of aspects and aspect weaving. Currently section 3 points to previous work and gives only little high-level/abstract information.*

   We would like to thank the reviewer for pointing out this issue. We agree that a reader with little background on OOP, AOP and C++ may find it difficult to understand the ideas shown in the paper. In the revised manuscript, the first paragraph of section 3 presents some basic concepts of OOP. We also extended the description of AOP concepts and the ADESD methodology in the subsequent paragraphs. As mentioned by reviewer #3, an introduction to UML and C++ templates was also missing. However, due to the page limit, we cannot provide an in-depth explanation of these concepts, but we have added a footnote that provides additional references. Please refer to comment 2 from reviewer #1 for these changes.

2. Reviewer's comment: *In Section 4.1 the comparison of TLM-based communication and communication in object-oriented modes seems to be a little strange. Without giving further details on the methodology behind the presented approach (incl. different models for the application, execution platform and the mapping of application elements to component of the execution platform) this comparison is dangerous. Communication in object-oriented application models describe application specific communication, while TLM models describe how communication is realized in the execution platform through physical channels.*

   Our goal with Figure 2 is not to provide a direct comparison between OOP and TLM as methodologies for the same purpose. We believe the following sentence may induce the reader to this misunderstanding: *This strategy provides a clear separation between communication and behavior, but it is still too hardware-oriented since it basically provides higher-level versions of RTL signals.*

   This sentence was removed and we have done minor changes in the second paragraph of section 4.2 to emphasize our main goal with Figure 2, which is to illustrate the problem that arises when an object-oriented approach is used to describe hardware/software components. As described in the paper: *In OOP the original structure will be "disassembled" if different objects in the same class hierarchy represent components that are to be implemented in different domains. For example, C2 is "inside" C1 in the OO model in Figure 2, but, in the final implementation, C1 could be implemented as a hardware component while C2 could run as software in a processor*. This also motivates the use of the proxy/agent mechanism described in section 5. Such problem does not exists in the TLM model, since, as highlighted by the reviewer, TLM is closer to the execution platform.

3. Reviewer's comment: *Section 4.2 aims at defining C++ unified description. This section appears to focus only on some specific issues like the use of pointers, static polymorphism, allocation, and dispatching. These issues are indeed important but the overall description of the unified description is missing. The presented techniques from template meta programming are interesting, but the description is not sufficient for understanding how these techniques are applied in a systematic way in an overall unified description.*

   We have described how we have addressed this issue in the answer to comments 1 and 2 from Reviewer #1.

4. Reviewer's comment: *Section 4.2.3 presents the idea of a Remote Method invocation with marshaling and unmarshaling services. It remains unclear how these techniques are supported by the presented methodology. Nothing about interrupt handling for software calls is mentioned.*

   We agree with the reviewer. The link between proxies/agents and their actual implementation was indeed unclear since the information was scattered among sections *4 Unified hardware/software design* and *5 Case study*. We have addressed this with the following changes in the revised manuscript: former sections *4.2.3 Wrapping communication* and *4.3 Implementation flow summary* were moved to a new section *5 Deployment of unified components*, becoming sections 5.1 and 5.3 respectively; added a new section *5.2 Implementation platform*.

   The new section 5.2 describes the execution platform that is later used to deploy the case studies. This description provides details of the run-time support, which includes the information requested by the reviewer about interrupt handling for software calls.

5. Reviewer's comment: *A Trait is a very generic template meta programming technique. It remains unclear how this technique solves the problem of HW/SW communication.*

   In this context, the traits only encapsulate the information that the HLS tool or the compiler need to define if the component itself or its proxy is going to be instantiated. In the answer to comment 2 from Reviewer #1, we provide a better description of the trait concept. We have also extended the explanation of the code snippet in the last paragraph of section 5.1.

6. Reviewer's comment: *In the case-study error bars should be added to measured execution times.*

   We have added error bars to Figures 13 and 15b. The figures now show that the execution time of the scheduler varies significantly according to the number of threads in the system, which is an expected result since we have experimented with 8 threads. This explanation was added to the third paragraph of section 6.1.

7. Reviewer's comment: *The main focus of the case-study is on the comparison of component implementations using C++ with the presented unified C++ approach. The evaluation is quite extensive and could be reduced.*

   We have reduced the case study section by moving the paragraphs that described Figure 11 in the original submission to the new section 5.2. Additionally, in order to comply with the page limit, we chose to reduce the description of the PABX system in the first paragraph of section 6 and removed the PABX system diagram from Figure 8.

8. Reviewer's comment: *The results presented in Table 5 are hard to understand.*

   Table 5 and its description were improved. The values in ()'s were moved to the *System* column, since these values are used to define the "system" overhead (e.g. the memory footprint of the run-time support, FPGA area of the RTSNoC interconnect, etc.), while the *Total* column shows the total area footprint.

Reviewer #3

1. Reviewer's comment: *On the general approachability of the text. The authors seem to assume that readers understand many concepts like "aspect-oriented programming", and do not really explain these concepts in the paper. More explanations on the concepts of OOP/UML and C++ template will be helpful. I understand that many basic concepts take much space to explain. Yet the author should at least say something like "we assume that the readers are comfortable with C++ templates and UML diagrams in the reset of this paper. For information on these, please see [xx] and [yy] for more explanation."*

   Please refer to the answer to comment 1 from Reviewer #2.

2. Reviewer's comment: *While the proposed method could work well for the purpose of describing hardware and software in a unified way, comparison with other methods for the same purpose is missing. For example, it is possible to use C (with extensions/pragmas) to describe both hardware and software (I know people in the industry doing so), and I believe what static metaprogramming offers can also be achieved using C language constructs, like macros and #ifdef blocks. Discussion on the advantage of the proposed approach should be elaborated, with comparison to possible alternatives (not necessarily experiemtal comparison). For example, one of the advantages I can see is that the proposed approach is more systematic and thus can potentially offer better checking at the syntax level; i.e., a domain-specific compiler may easily inform the designer about missing constructs. Yet, a disadvantage could be that a coding style using static-metaprogramming can be difficult to debug. Comparisons like this could help the reader better understand the proposed method and appreciate its unique advantages.*

   Indeed, old CPP macros are still widely used in practice and it is important to provide this kind of comparison in the paper. We would like to thank the reviewer for the suggestion. The second and third paragraphs of section 4.4 address these points.

3. Reviewer's comment: *There are more differences in hardware design and software design that may need attention. For example, for high-level synthesis, bit-accurate data types are very useful. By using only enough bit widths for operators, storage elements and interconnects, significant saving in resource/power can be achieved. On the other hand, software compilers usually only use 8/16/32/64 bit for integers. The tool used in your experiments, Catapult-C, supports the ac_int/ac_fixed datatypes in C++. I wonder if bit-accurate datatypes are used in your experimental evaluation? How do you unify the hardware/software description if bit-accurate datatypes are needed? My personal experience is that there can be subtle issues when overflow occurs.*

   We have implemented our current components using the standard C++ data types (for instance, some of the data types can be seen in the UML diagrams in figures 9, 10, and 11) to ensure maximum flexibility of the unified code. However, we agree that bit-accurate data types are crucial to optimize the area of the final hardware design. We have addressed this point in the fifth paragraph of section 4.2.1. A possible way to provide bit accuracy while keeping a certain degree of uniformity is to use C++ typedef statements to define all allowed data types according to the target domain. For instance, a 5-bit integer can be defined in hardware using `typedef ac_int<5> int5` and in software using `typedef char int5`, since `char` is the smallest native type with the required width. We believe this is a reasonable solution since most compilers and processor architectures support only 8/16/32/64-bit integer types, therefore true bit-accuracy in software would require additional shifting and masking operations that may add a significant overhead.

4. Reviewer's comment: *The approach, as a coding style, can be applied to a more general class of "module selection" problem. The scenarios considered do not have to be limited to either "software" or "hardware". In high-level synthesis, the design space can be explored by using different pragmas (like loop unrolling/pipeling), and each implementation could be a scenario. Some of the techniques described in this paper can actually encode multiple sets of implementation options in a unified description.*

   Although we consider that tuning the hardware microarchitecture using synthesis directives is part of the design space exploration and out of the scope of this paper; this is indeed an interesting direction for future works. The hardware scenario could be extended to also provide such kind of semantics in order to aid the design space exploration process. A possible approach would be, for instance, to specify synthesis directives using C's pragmas within an aspect. This aspect would have to be specialized for each component, HLS tool, and intended hardware microarchitecture, since each of these would require a different set of pragmas.

   The last paragraph of section 4.4 address the points highlighted above.