Certified Simulation: What You Get Is What You Want

Vania Joloboff[1], Jean-François Monin[2], and Xiaomu Shi[3]

[1] INRIA - LIAMA
[2] Université de Grenoble - Verimag
[3] Tsinghua University

Abstract. This paper presents an approach to construct a certified virtual prototyping framework of embedded software. The machine code executed on a target architecture can be proven to provide the same results as the initial algorithm, and the proof is verified with an automated theorem prover. This requires the establishment of a tool chain in which each step can be certified. The paper presents the proof of an ARM architecture Instruction Set Simulator, with all of the proofs being verified with the Coq theorem prover.

1 Introduction

In many applications, it is desirable to exhibit a proof of program to certify that the execution of a given algorithm is correct on the specific computer architecture on which the program is executed. For example in the area of distributed algorithms it is interesting to be certain that the implementation on real devices indeed behave the same as the abstract description of the algorithm.

Obtaining such certification is not a one step proof, it consists in a series of steps requiring a tool chain. In this paper, we present an attempt towards building such a certification chain, including a certified simulator of processor architecture. Our approach towards this certification of programs is the following: given the formal specification of an algorithm, one can derive a C program using formal methods such as APTS [1] or event-B [2] or BIP [3]. In this paper we assume that this phase has been achieved, that there is an existing C source code program that is a provably correct implementation of the specification. However, well-known issues concern the specification itself: typically, does it include all desirable features? A number of improvements of the specification can be achieved by reasoning on it: one may state expected conjectures and then try to prove (or disprove) them. This approach is especially relevant for abstract properties, such as the algebraic properties of arithmetics functions. Still, some features are more naturally expressed on the implementation, because they involve interactions with the environment. Alternatively, a piece of software may be well-behaved in some environment, but not so well when the latter changes a little bit. Specifying the precise conditions which are needed or checking that the real environment fits the expectations is not that easy. In all of these situations,

it is important to work with the implementation, not only with the specification. Instead of considering an implementation on real hardware, a *simulated* implementation is more convenient and less expensive. Then, it is important to use a very faithful simulator in order to ensure that no bias is introduced.

Our purpose is to show how it can be certified that the execution of a binary program on the Instruction Set Simulator of a target architecture indeed produces the expected results. This requires sequential steps, to prove first that the translation from C code to machine code is correct, and second that the simulation of the machine code is also correct, that is, they all preserve the semantics of the source code; together with the fact that all of these proofs are verified using a theorem prover or proof checker, not subject to human error in the proof elaboration or verification.

The technique presented here relies on already existing tools, in particular the Coq proof assistant, the Compcert C compiler, a certified compiler for the C language, combined with our own work to prove the correctness of an ARM Instruction Set Simulator, integrated within an existing virtual prototyping framework.

The sequel of the paper presents the background existing tools, related work, our method to combine these tools and our additional work to reach the objectives.

2 Related Work

Program certification has to be based on a formal model of the program under study. Such a formal model is itself derived from a formal semantics of the programming language. Formal semantics are important because they provide an abstract, mathematical, and standard interpretation of a programming language. There are three traditional ways to express how programs perform computations: axiomatic semantics, denotational semantics, and operational semantics. An overview of programming language semantics can be found in [4–6].

Denotational semantics [7] constructs mathematical objects which describe the meaning of expressions of the language using stateless partial *functions*. All observably distinct programs have distinct denotations.

Axiomatic semantics and Hoare logic[8, 9] have been widely used for proving the correctness of programs. For imperative programming languages such as C, a possible approach is to consider tools based on axiomatic semantics, like `Frama-C` [10], a framework for a set of interoperable program analyzers for C. Most of the modules integrated inside rely on ACSL (ANSI/ISO C Specification Language), a specification language based on an axiomatic semantics for C. ACSL is powerful enough to express axiomatizations directly at the level of the C program. State labels can be defined to denote a program control point, and can then be used in logic functions and predicates.

`Frama-C` software leverages off from `Why` technology [11, 12], a platform for deductive program verification, which is an implementation of Dijkstra's calculus of weakest preconditions. `Why` compiles annotated C code into an intermediate

language. The result is given as input to the VC (Verification Conditions) generator, which produces formulas to be sent to both automatic provers or interactive provers like Coq.

Paolo Herms's [13] [14] has provided a certified verification condition generator for several provers, called `WhyCert`. This VC generator was implemented and proved sound in Coq. To make it usable with arbitrary theorem provers as back-ends, it is generic with respect to a logical context, containing arbitrary abstract data types and axiomatisations.

In our case of verifying an instruction set, we have to deal with a very large specification including complex features of the C language. A framework is required that is rich enough to make the specification manageable, using abstraction mechanisms for instance, and in which an accurate definition of C features is available. Automated computations of weakest preconditions and range of variation are not relevant in our case. We need to verify more specific properties referring to a formal version of the ARM rchitecture. It was unclear that `Frama-C` would satisfy those requirements, even with Coq as a back-end.

Operational semantics is a more concrete approach to program semantics as it is based on states, although in contrast with a low-level implementation, operational semantics considers abstract states. The behavior of a piece of program corresponds to a transition between abstract states. This transition relation makes it possible to define the execution of programs by a mathematical computation *relation*. This approach is quite convenient for proving the correctness of compilers, using operational semantics for the source and target languages (and, possibly intermediate languages).

Operational semantics can be presented in two styles. *Small-step* semantics, often known as structural operational semantics, is used to describe how the single steps of computations evaluate. The other is *big-step* semantics, or natural semantics, which returns the final results of an execution in one big step. The corresponding transition relation is defined by rules, according to the syntactic constructs of the language, in a style which is inspired by natural deduction. The book [4] discusses the alternatives between small-step and big-step semantics depending upon the objective. They sometimes can be equivalent. But in general, they provide different views of the same language and one has to choose an appropriate one for a particular usage. Moreover, some language constructs can be hard or even impossible to define with one of these semantics, whereas it may be easy with the other style. In general, when big-step semantics can be used, it is simpler to manage than small-step semantics. A tutorial on programming language semantics made by Benjamin C. Pierce's [15] is mainly dedicated to operational semantics and the material presented in this tutorial is formalized in the Coq proof assistant.

Operational semantics are used in `CompCert` to define the execution of C programs, or more precisely programs in the subset of C considered by the `CompCert` project. The work presented in this paper is based on this approach. Interesting examples are given by Brian Campbell in the CerCo project [16], in

order to show that the evaluation order constraints in C are very lax and not uniform.

Regarding formalization and proofs related to instruction set, a Java byte code verifier has been proved by Cornelia Pusch[17], the Power architecture semantics has been formally specified in [18], and closer to our work, the computer science laboratory in Cambridge University has used HOL4 to formalize the instruction set architecture of ARM [19]. The objective of their work was to verify an implementation of the ARM architecture with *logical gates*, whereas we consider a ARM architecture simulator coded in C. Reusing the work done at Cambridge was considered for our work. However, as our approach is based on `CompCert`, which is itself coded in Coq instead of HOL4, it was more convenient to develop our formal model and our proofs in Coq.

There is abundant literature covering Instruction Set Simulation (ISS). Using *interpretive simulation*, such as used in Insulin [20], each instruction of the target program is fetched from memory, decoded, and executed. With *static translation*, the target application program is decoded at compile time and translated into a new program for the simulation host. The simulation speed is vastly improved [21, 22], but it is not suitable for application programs that dynamically modify the code, or dynamically load code at run-time. Most ISS'es today use some kind of *dynamic binary translation*, initiated with systems such as Shade [23] and Embra [24], and later enhanced with various types of optimization techniques [25–32].

Our work is based on the SimSoC simulation framework [33], which uses dynamic binary translation, and is available as open source software [34].

3 Background

3.1 Coq

Coq [35] is an interactive theorem prover, implemented in OCaml. It allows the expression of mathematical assertions, mechanically checks proofs of these assertions, helps to discover formal proofs, and may extract a certified program from the constructive proof of its formal specification. Coq can also be presented as a dependently typed λ-calculus (or functional language). For a detailed presentation, the reader can consult [36] or [35]. Coq proofs are typed functions and checking the correctness of a proof boils down to type-checking. For example, a proof term of type $\forall\ n : nat,\ P\,n\ \rightarrow Q\,n$ is a function $fun\ (n : nat)(p : P\,n)$ which takes a natural number n and a proof p of $P\,n$ as arguments and returns a proof of $Q\,n$.

Coq is not an automated theorem prover: the logic supported by Coq includes arithmetic; therefore it is too rich to be decidable. However, type-checking (in particular, checking the correctness of a proof) is decidable. As full automation is not possible for finding proofs, human interaction is essential. The latter is realized by *scripts*, which are sequences of commands for building a proof step by step. Coq also provides built-in *tactics* implementing various decision procedures

for suitable fragments of the calculus of inductive constructions and a language which can be used for automating the search of proofs and shortening scripts.

An interactive proof assistant, such as Coq, requires man-machine collaboration to develop a formal proof. Human input is needed to create appropriate auxiliary definitions, choose the right inductive property and, more generally, to define the architecture of the proof. Automation is used for non-creative proof steps and checking the correctness of the resulting formal proof. A rich logic can be handled in an interactive proof assistant for a variety of problems.

On the other hand, fully automated theorem provers have been developed. They can perform the proof tasks automatically, that is, without additional human input. Automated theorem prover can be efficient in some cases. However being able to automatically prove a formula means that the problems solved belongs to a decidable, or at least semi-decidable, class. Decidable logic are less powerful expressive than higher-order logic, hence the range of problems one can easily or at least conveniently model with an automated theorem prover is smaller than with an interactive proof assistant. In practice, both approaches are important in the fields of computer science and mathematical logic. In our project, as a rich logical system is needed in order to manage the complexity of the ARM specification and of the proofs of C programs, it was decided to use Coq.

3.2 Compert-C

`CompCert` is a formally verified compiler for the C programming language provided by INRIA [37, 38], which currently targets Power, ARM and 32-bit x86 architectures. The compiler is specified, programmed, and proved in Coq. It aims to be used for programming embedded systems requiring high reliability. The performance of its generated code is often close to that of gcc (version 3) at optimization level O1, and is always better than that of gcc without optimizations.

It has a complex translation chain of eleven steps, from C source code into assembly code. Every internal representation has its own syntax and semantics defined in Coq. It is formally verified in the sense that the generated assembly code is proved to behave exactly the same as the input C program, according to a formally defined operational semantics of the language, and assumptions on the instruction set.

Of course, if the input consists of C language expressions that are ill-defined, according to the ISO-C specification, (i.e. a[i++] = i;) Compert-C generates ill-formed code. It generates verified code when the input is what ISO-C defines as 'correct C program', which is expected if the C code has been generated using formal specifications.

Three parts of `CompCert` C are used in this work. The first is that we use the correct machine code generated by the C compiler. The second is the C language operational semantics in Coq (its syntax and semantics), from which we get a formal model of the program. Third, we use the `CompCert` basic library. It defines Coq data types for words, half-words, bytes etc., and bitwise operations

and lemmas to describe their properties. In our Coq model, we also use these low level representations and operations to describe the instruction set model.

3.3 SimSoC

As mentioned above, our certification target ISS is integrated within `SimSoC` [39], a full system simulator of System-on-Chip that can simulate various processors, available as open source software. `SimSoC` takes as input real binary code and executes simulation models of the complete embedded system: processor, memory units, interconnect, and peripherals. It is developed in SystemC [40] and uses transaction level modeling (TLM) to model communications between the simulation modules. It includes Instruction Set Simulators (ISS) to execute embedded applications on simulated platforms. Our work is introducing a new ISS in the `SimSoC` virtual prototyping framework, for the ARM architecture.

4 Certified Simulation

Our general objective to obtain a certified simulatoris illustrated in Figure 1. Considering the ARM architecture (Version 6), we need to have the following:

- a simulator of the ARM instruction set in (compcert) C that we can obtain from the ARM Reference manual.
- obtain formal operational semantics of the ISS. Given some source code in C, one can obtain through CompCertC the verified machine code, or alternatively the Coq formal semantics of the compiled program constructed by CompCert.
- prove, using a theorem prover, that the resulting ISS semantics indeed implement an ARM processor, to verify that the semantics of the simulator accurately modifies the processor state at each step. For that, we need to prove that the results are compliant with a formal model of the ARM architecture.

These steps are described in the following paragraphs.

4.1 Constructing the ISS

The whole process starts with the ARM reference manual available from ARM web site (version 6) `ARM DDI 0100I` [41]. The relevant chapters for us are mostly:

- `Programmer's Model` introduces the main features in ARMv6 architecture, the data types, registers, exceptions, etc;
- `The ARM Instruction Set` explains the instruction encoding in general and puts the instructions in categories;
- `ARM Instructions` lists all the ARM instructions in ARMv6 architecture in alphabetical order and the `ARM Addressing modes` section explains the five kinds of addressing modes;

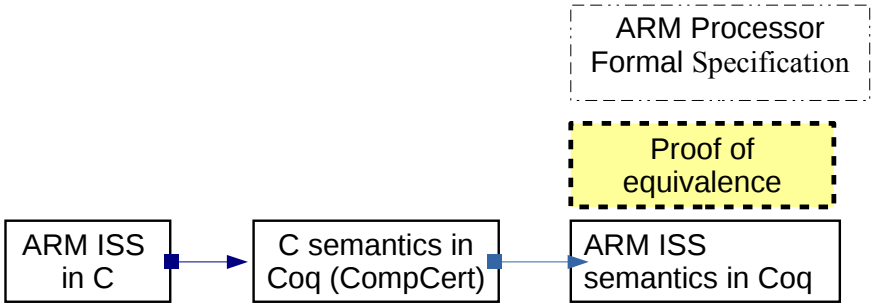

Fig. 1. Overall goal

There are 147 ARM instructions in the ARM V6 architecture. For each instruction, the manual provides its encoding table, its syntax, a piece of pseudo-code explaining its own operation, its exceptions, usage, and notes. Except the semi-formal pseudo-code, everything else is written in natural language. Three files are extracted from the reference manual: a 2100 lines file containing the pseudo-code, a 800 lines file containing the binary encoding tables, and a 500 lines file containing the assembly syntax. Other than these three extracted files, there is still useful information left in the document which cannot be automatically extracted, such as validity constraints information required by the decoder generator. The corresponding information is manually encoded into a 300 lines file. The overall architecture of the generator is given in figure 2.

Three kinds of information are extracted for each ARM operation: its binary encoding format, the corresponding assembly syntax, and the instruction semantics, which is an algorithm operating on the various data structures representing the state of an ARM processor, mostly registers and memory, according to the purpose of the instruction considered. This algorithm may call basic functions defined elsewhere in the manual, for which we provide a `CompCert` C library to be used by the simulator and a Coq library defining their semantics. The latter relies on libraries from the `CompCert` project that allows us, for instance, to manipulate 32-bits representations of words. The result is a set of abstract syntax trees (ASTs) and binary coding tables. These ASTs follow the structure of the (not formally defined) pseudo-code.

In the end, two files are generated: a `CompCert` C file to be linked with other components of `SimSoC` (each instruction can also be executed in stand-alone mode, for test purposes for instance) and files representing each instructions in `CompCert` C abstract syntax to be used for correctness proof. As the C language accepted by Compcert is a subset of the full ISO C language, the generator has been constructed such that it only generates C code in the subset accepted by the compcert compiler. Nonetheless it can be compiled with other C compilers such as GCC to obtain better performance. Even though in this case, the resulting machine code is not guaranteed to be correct (there are well known

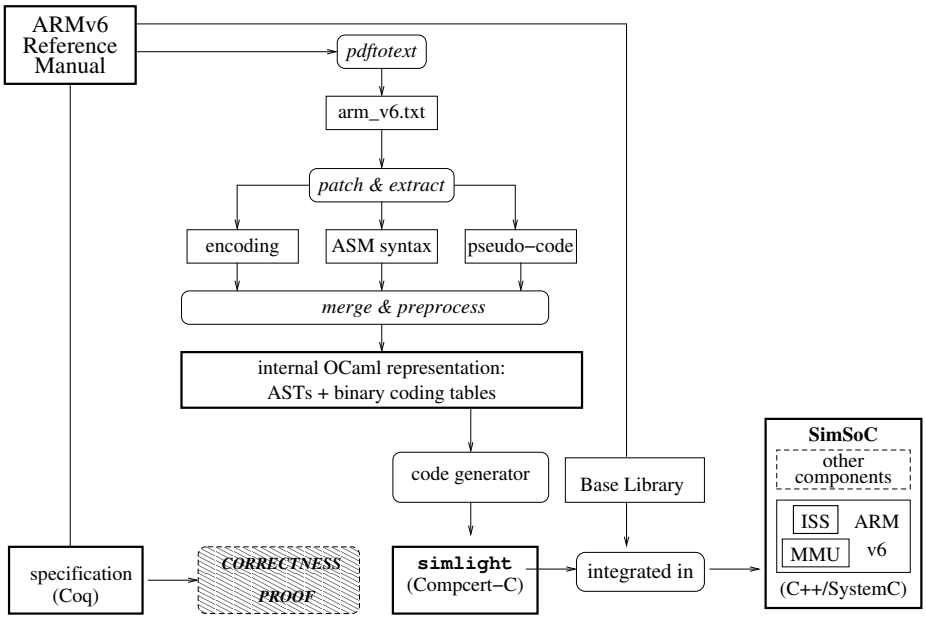

Fig. 2. Overall generation chain

GCC optimization bugs...), at least the original C code has been proven by our technique to be conformant with the ARM semantics.

The ARM V6 code generator not only generates the semantics functions, it also generates the decoder of binary instructions supported in V6 architectures. This decoder is obtained by compiling the opcodes information. The generated decoder is not optimal in performance, but as `SimSoC` uses a cache to store the decoded instructions, the performance penalty is marginal.

4.2 Simulator Semantics

In order to formally reason on the correctness of the simulator, we need to have a formal model of the C implementation of the ARM architecture as described above. It is provided by `CompCert`, which defines operational semantics of C formalized in Coq. As the simulator C program also has the objective to achieve a high speed simulator, the generator makes optimizations. In particular, states in the model of the C implementation are complex, not only due to the inherent complexity of the C language memory model, but also because of optimization and design decisions targeting efficiency. In more detail:

– The C implementation uses large embedded *struct*s to express the ARM processor state. Consequently the model of the state is a complex Coq record type, including not only data fields but also proofs to guaranteed access permission, next block pointer, etc.

– Transitions are defined with a relational style. In general, the relational style is more flexible but functional definitions have some advantages: reasoning steps can be replaced by computations; existence and unicity of the result are automatically ensured. However, the functional style is not always convenient or even possible. It is the case here, where the transitions defined by the C implementation are relations which happen to be functions. This comes first from the operational semantics, which needs to be a relation for the sake of generality. Furthermore in our case, the kind of record type mentioned in the previous item is too complex to execute calculation with it, so it is more convenient to describe the state transformation for memory with a relation.
– The global state is based on a complex memory model with load and store functions that are used for read/write operations.

However, thanks to `CompCert` operational semantics, we are able to formally express the semantics of the C implementation of the ARM processor.

4.3 Proof

As a result of the generator described above, we have a C implementation of the ARM instruction set. To prove that it is correct we need to have a formal specification of that architecture, and prove that, given the initial state of the system, the execution of an instruction as implemented by a C function results in the same state as the formal specification. Ideally the formal specification of the ARM architecture should be provided by the vendor. But it is not the case, such a formal model is not available on their web site, hence we had to build one. As mentioned in the related work section, we considered using the ARM formal model defined in [19], but it would have required us to translate all of the C operational semantics as well, which would not have been error prone, not to mention the effort. So, we chose to define a formal model of ARM architecture in Coq as well, derived from the architecture reference manual. As Coq models are executable, we could validate the model with real programs.

The state of the ARM V6 processor defined in the formal model is called the *abstract state*. On the other hand, the same state is represented by the data structures corresponding to C semantics that we shall call the *concrete state*. In order to establish correctness theorems we need to relate these two models. Executing the same instruction on the two sides produces a pair of new processor states which should be related by the same correspondence. Informally, executing the same instruction on a pair of equivalent states should produce a new pair of equivalent states.

Our theorems are schematized by Figure 3. The complete proof is too lengthy for this article, and we only provide here an outline of the method. The correctness proof is based on one hand on the semantics of the ARM architecture formal model, and on the other hand on the `CompCert` C representation of the ARM instructions. We need to prove that the operational semantics of that C code correspond to the ARM formal specification. The proof consists in defining provable projections from the C state to the formal model to prove their equivalence,

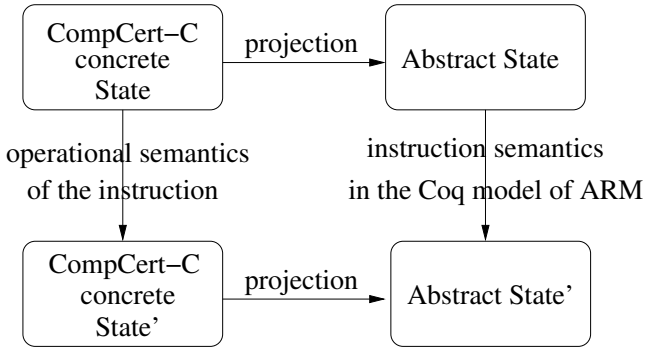

Fig. 3. Theorem statement for a given ARM instruction

as represented in Figure 4. To state the correctness theorem, we compare the `CompCert` C semantics of a function corresponding to an ARM instruction with its formal definition.

In the C instruction set simulator, there is a standalone C function for each ARM V6 instruction. Each function (instruction) has its own separate correctness proof. Every function is composed of its return type, arguments variables, local variables, and the function body. The function body is a sequence of statements including assignments and expressions. Let us consider as an example the ARM instruction `BL` (`Branch and Link`). The generated C code is:

```
void B(struct SLv6_Processor *proc,
       const bool L,
       const SLv6_Condition cond,
       const uint32_t signed_immed_24){
 if (ConditionPassed(&proc->cpsr, cond)){
  if ((L == 1))
   set_reg(proc,14,
           address_of_next_instruction(proc));
   set_pc_raw(proc,
         reg(proc,15)+
         (SignExtend_30(signed_immed_24)<<2));
 }
}
```

The `CompCert` C representation of that C code is an internal representation that can be externalized in a human readable form as:

```
Definition fun_internal_B :=
{|
 fn_return := void;
 fn_params := [
  proc -: `*` typ_SLv6_Processor;
  L -: int8;
  cond -: int32;
  signed_immed_24 -: uint32];
```

```
 fn_vars := [];
 fn_body :=
 'if (call (ConditionPassed':T1)
       E[&(('*(proc':T2)':T3)
        |cpsr':T4)':T5;cond':T6] T7)
 then
  'if ((L':T7)==(#1':T6)':T6)
     then (call (set_reg':T8)
       E[proc':T2; #14':T6;
        (call (address_of_next_instruction':T9)
               E[proc':T2] T10)]
         T11)
  else skip;;
   (call (set_pc_raw':T12)
     E[proc':T2;
      (call (reg':T13) E[proc':T2; #15':T6] T10)+
       ((call (SignExtend_30':T14)
        E[signed_immed_24':T10] T10)
          $<<$(#2':T6)':T10)':T10]
     T11)
 else skip
|}.
```

The main strategy is to define a projection from the concrete state to the abstract state. On both sides, the execution of an instruction is described by a state transition. For the two representations, "State" refers to the full description of the system. This correspondance is expressed by a projection relating the two models of the state.

According to the type of the argument of the projection, the definitions of projections are different. For example, the projection of a register performs a case analysis on a value of type `register`, whereas the projection of SPSR (Save Processor Status Register) depends on the type of exception modes. We define a specific projection for each type. Although Coq is rich enough to allow for the definition of a general projection for all types of elements, using dependent types, we chose to define a projection relation for each instance to improve readability of the proofs. The proofs start from the abstract state described by the formal specification. An issue is that C representations introduce a complex memory model designed for high simulation speed, and we need a mapping to the formal model. In order to verify the projection of the pair of original states, we need the following data: the initial memory state, the local environment, and the formal initial processor state. The projection is meaningful only after the C memory state is prepared for evaluating the current function body representing a ARM instruction. In the abstract Coq model, we directly use the processor state `st`. But on the C side, the memory state must provide the contents of every parameter, especially the memory representation of the processor state. We also need to observe the modification of certain blocks of memory corresponding to local variables.

Fortunately `CompCert` formalizes the C memory model. The semantics of `CompCert` C consider two environments. The global environment *genv* maps global function identifiers, global variables identifiers to their blocks in memory, and function pointers to a function definition body. The local environment *env* maps local variables of a function to their memory blocks reference. It maps each variable identifier to its location and its type, and its value is stored in the associated memory block. The value associated

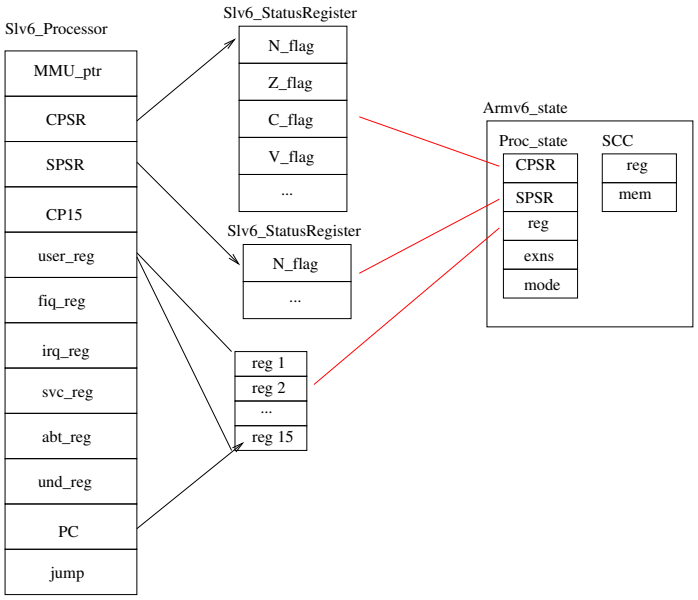

Fig. 4. Projection

to a C variable or a parameter of a C function is obtained by applying (abstract) `load` to the suitable reference block in memory. These two operations are performed when a function is called, building a local environment and an initialized memory state. When the program starts its execution, *genv* is built. On the other hand, *env* is built when the associated function starts to allocate its variables. Therefore, on `CompCert` C side, a memory state and a local environment is prepared initially using two steps. First, allocate function variables: from an empty local environment, all function parameters and local variables ara allocated into the memory state, yielding a new memory state and the local environment. Second, initialize function parameters: using a function `bind_parameters` to initialize parameters with a list of argument values and a new memory state is created.

Now we have all elements ready to make the projection from the local environment to the abstract state to verify their equivalence. Next, we need to consider the execution of the instruction. `CompCert` has designed semantics for `CompCert` C in both small-step and big-step. The big-step inductive type for evaluating expression is enough for our proof. The semantics is defined as a relation between an initial expression and an output expression after evaluation. Then the body of the function is executed. On the `CompCert` C side, the execution is yielding a new memory state **mfin**. On the abstract side, the new processor state is obtained by running the formal model. Finally, we have to verify that the projection from the concrete state **mfin** should provide the latter abstract state. Note that all projections are performed using the same local environment `e`.

The formal model of ARM V6 is defined as a much simpler functional model and computing the value of a component can be performed directly.

The proof is performed in a top-down manner. It follows the definition of the instruction, analyzing the expression step by step. The function body is split into

statements and then into expressions. When evaluating an expression, we search for two kinds of information. the first one is how the memory state changes on `CompCert` C side; the other is whether the results on the abstract and the concrete model are related by the projection. To this effect, we use mostly the following lemmas.

1. *Lemma: Evaluating a `CompCert` expression with no modification on the memory state.*
   This lemma is concerned with the expression evaluation on `CompCert` C side and in particular the C memory state change issue. Asserting that a memory state is not modified has two aspects: one is that the memory contents are not modified; the other is that the memory access permission is not changed. For example, evaluating the boolean expression $Sbit == 1$ returns an unchanged memory state.

   $$\begin{aligned} &\text{if } G, E \vdash \texttt{eval\_binop}_c\ (Sbit == 1), \\ &M \overset{\varepsilon}{\Longrightarrow} vres,\ M' \text{then } M = M'. \end{aligned}$$

   In Coq syntax, the relation in premise is expressed with `eval_binop`, a companion predicate of `exec_stmt` above, devoted to binary operations. In this lemma and the following, E is the local environment, G is the global environment and M is the memory state; ε is the empty event (`Events.E0` in Coq syntax); usually t is used to represent a series of system events; $vres$ is the result. Here, $vres$ is not important. The evaluation is performed under environments G and E. Before evaluation, we are in memory state M. With no event occurring, we get the next memory state M'. According to the definition of `eval_binop`, an internal memory state will be introduced.

   $$\frac{G, E \vdash a_1, M \Rightarrow M' \quad G, E \vdash a_2, M' \Rightarrow M''}{G, E \vdash (a_1\ binop\ a_2), M \Rightarrow M''}$$

   Now, in our example, expression a_1 is the value of $Sbit$ and a_2 is the constant value 1. By inverting the hypothesis of type `eval_binop`, we obtain several new hypotheses, including on the evaluation of the two subexpressions and the introduction of an intermediate memory state M''. Evaluating them has no change on the C memory state. Then we have $M = M'' = M'$. In more detail, from the `CompCert` C semantics definition, we know that, evaluation of an expression will change the memory state if the evaluation contains uses of `store_value_of_type` (in `CompCert` versions before 1.11), which stores the value in memory at a given block reference and memory chunk. In `CompCert`-1.11, the basic store function on memory is represented by an inductive type `assign_loc` instead of `store_value_of_type`. Since `CompCert` version 1.11 introduced volatile memory access, we have to determine whether the object type is volatile before storage, and also type size in addition of the access mode.
2. *Lemma: Result of the evaluation of an expression with no modification on the memory.* Continuing the example above, we now discuss the result of evaluating the binary operation $Sbit == 1$ both in the abstract and the concrete model. At the end of evaluation, a boolean value $true$ or $false$ should be returned. in both the `CompCert` C model and the formal model, using the projection definition.

   $$\begin{aligned} &\text{if } \texttt{Sbit\_related}\ M\ \texttt{Sbit}, \\ &\text{and } G, E \vdash \texttt{eval\_rvalue\_binop}_c\ (Sbit == 1), \\ &M \Rightarrow v, \text{then } v = (Sbit == 1)_{coq} \end{aligned}$$

Intuitively, if the projection corresponding to the parameter `sbit` in the C program yields the right information from the abstract state, then the evaluation will return the same value both in the abstract and in the concrete model. Here, the expression is a so-called "simple expression" that always terminates in a deterministic way, and preserves the memory state. To evaluate the value of simple expressions, `CompCert` provides two other big-step relations `eval_simple_rvalue` and `eval_simple_lvalue` for evaluating respectively their left and right values. The rules have the following shape:

$$\frac{G, E \vdash a_1, M \Rightarrow v_1 \quad G, E \vdash a_2, M \Rightarrow v_2 \quad \texttt{sem\_binary\_operation}(op, v_1, v_2, M) = v}{G, E \vdash (a_1\ op\ a_2), M \Rightarrow v}$$

In order to evaluate the binary expression $a_1\ op\ a_2$, the sub-expressions a_1 and a_2 are first evaluated, and their respective results v_1 and v_2 are used to compute the final result v.

3. *Lemma : Memory state changed by storage operation or side effects of evaluating expression*
   As mentioned before, evaluating some expressions such as `eval_assign` can modify the memory state. Then we need lemmas stating that corresponding variables in the abstract and in the concrete model will evolve consistently. For example, considering an assignment on register Rn, the projection relation `register_related` is used. Expressions with side effects of modifying memory are very similar.

$$\begin{array}{l}\text{if } \texttt{rn\_related}\ M\ rn \\ \text{and } G, E \vdash \texttt{eval\_assign}_c\ (rn := rx), M \Rightarrow M', v \\ \text{then } \texttt{rn\_related}\ M'\ rn\end{array}$$

4. *Lemma: Internal function call.*
   Internal functions are described in an informal manner in the ARM V6 reference manual. No pseudo-code is available for them, which means that the corresponding library functions, both in the abstract model and in C, are coded manually. In order to get a suitable `CompCert` C AST to reason about, we use the parser provided in `CompCert`. When combining the simulation code of an instruction with the code of library functions, we need to take care of the memory allocation problem. In `CompCert` C representation, identifiers are unique positive numbers which indicate the memory block where corresponding variables are allocated. Currently, the extra identifiers introduced by library functions are added manually and assigned with fresh block numbers.

$$\begin{array}{l}\text{if } \texttt{proc\_state\_related}\ M\ st \\ \text{and } G, E \vdash \texttt{eval\_funcall}_c \\ (copy\_StatusRegister)_c, M \Rightarrow v,\ M' \\ \text{and } st' = (copy\_StatusRegister)_{coq}\ st \\ \text{then } \texttt{proc\_state\_related}\ M'\ st'.\end{array}$$

After an internal function is called, a new stack of blocks is allocated in memory. After the evaluation of the function is performed, these blocks will be freed. Unfortunately, this may not bring the memory back to the previous state: the memory contents may stay the same, but the pointers and memory organization

may have changed. For lemmas on evaluation of internal functions, we can observe the returned result on variables, compare it with the corresponding evaluation in the formal specification, and verify some conditions. For example, the lemma above is about the processor state after evaluating an internal function call `copy_StatusRegister` which reads the value of CPSR and then assigns it to SPSR. The evaluation of `copy_StatusRegister` should be protected by a check on the current processor mode. If it is neither system mode nor user mode, the function `copy_StatusRegister` can be called. Otherwise, the result is "unpredictable", which is defined by ARM architecture

It is then necessary to reason on the newly returned states, which should still be related by the projection. This step is usually easy to prove, by calculation on the two representations of the processor state to verify they match.

5. *Lemma: External function call.*
   The `CompCert` C AST of an external function call contains the types of input arguments and of the returned value, and an empty body. `CompCert` provides the expected properties of a few built-in external functions such as `printf`, `malloc` and `free`. We have proceeded similarly for the external functions of the ARM simulator. The general expected properties of an external call are that (i) the call returns a result, which has to be related to the abstract state, (ii) the arguments must comply with the signature. (iii) after the call, no memory blocks are invalidated, (iv) the call does not increase the access permission of any valid block, and finally that the memory state can be modified only when the access permission of the call is granted. For each external call, we verify such required properties.

In addition to the above lemmas we had to prove a fair number of more trivial lemmas that are omitted here. Most of them are related to the semantics of `CompCert` C. With these lemmas we can build the proof scripts for ARM instructions. For that, we are decomposing the ARM instruction execution step by step to perform the execution of the C programs . `CompCert` C operational semantics define large and complex inductive relations. Each constructor describes the memory state transformation of an expression, statement, or function. As soon as we want to discover the relation between memory states before and after evaluating the C code, we have to invert the hypotheses of operational semantics to follow the clue given by its definition, to verify the hypotheses relating concrete memory states according to the operational semantics.

During the development of a proof, if a hypothesis is an instance of an inductive predicate and we want to derive the consequences of this hypothesis, the general logical principle to be used is called *inversion*. An *inversion* is a kind of forward reasoning step that allows for users to extract all useful information contained in a hypothesis. It is an analysis over the given hypothesis according to its specific arguments, that removes absurd cases, introduces relevant premises in the environment and performs suitable substitutions in the whole goal. The practical need for automating inversion has been identified many years ago and most proof assistants (Isabelle, Coq, Matita,...) provide an inversion mechanism. To this effect, the Coq proof assistant provides a useful tactic called `inversion` [36].

Every instruction contains complex expressions, but each use of `inversion` will go one step only. If we want to find the relation between the memory states affected by these expressions, we have to invert many times. For illustration, let us consider the simple example from the ARM reference manual `CPSR = SPSR`. As the status register is not implemented by a single value, but a set of individual fields, the corresponding C code is a call to the function `copy_StatusRegister`, which sets the CPSR field by

field with the values from SPSR. Lemma `same_cp_SR` below states that the C memory state of the simulator and the corresponding formal representation of ARM processor state evolve consistently during this assignment.

```
Lemma same_copy_SR :
  forall e m l b s t m' v em,
  proc_state_related m e
   (Ok tt (mk_semstate l b s)) ->
      eval_expression
       (Genv.globalenv prog_adc)
          e m expr_cp_SR t m' v ->
          forall l b,
          proc_state_related m' e
          (Ok tt
           (mk_semstate l b
             (Arm6_State.set_cpsr s
             (Arm6_State.spsr s em))))
```

From this, we have to invert generated hypotheses until all constructors used in its type are exhausted. On this example, 18 consecutive inversions are needed... The first proofs scripts we wrote were becoming unmanageable, and not robust to version changes of Coq or `CompCert`. In order to reduce the script size and get better maintainability, we studied a general solution to this problem, and developed a new inversion tactics in Coq . We have first expanded the small inversion mechanism from Coq and added a new inversion tactic for inductive types in `CompCert`. The semantics of `CompCert` C tells us how the memory state is transformed by evaluating expressions. Using the built-in constructs of the tactics language, we can define a high-level tactic for each inductive type, gathering all the functions defined for its constructors.

This tactic has two arguments corresponding to the C memory states. The first step of the tactic introduces generated components with new names. The second step is related to previously reverted hypotheses, ensuring that the new names introduced are correctly managed by Coq. The tactic then proceeds as follows. First, it automatically finds the hypothesis to invert by matching the targeted memory states. Then the related hypotheses are reverted and an appropriate auxiliary function is called (all auxiliary functions are gathered into the tactic) and meaningful names are given to derived variables and hypotheses. Next, all other related hypotheses are updated according to the new names, and finally new values and useless variables or hypotheses are cleaned up. The above steps are then repeated until all transitions between the two targeted memory states are discovered. As a result, considering the former example of `same_copy_SR` where 18 standard `inv` were used in the first proof script we developed, our tactics reduced them into one step: `inv_eval_expr m m'`. Thanks to this new tactics, the size of the proofs has become smaller and the proof scripts are more manageable. The size vary with the instructions complexity from less than 200 lines (e.g 170 for LDRB) to over 1000 (1204 for ADC). As a result, for each ARM instruction, we have established a theorem proving that the C code simulating an ARM instruction is equivalent to the formal specification of the ARM processor. All of these lemmas and theorems are verified by the Coq theorem prover.

5 Certified Simulation

Based on the work afore mentioned, we can now consider the certified execution of a C program. We take here as an example the DES cryptographic encryption code. The C code for encrypting a block of data is straightforward:

```
#define GET_ULONG_BE(n,b,i)
 (n)=((unsigned long)(b)[(i)]<<24)
    |((unsigned long)(b)[(i)+1]<<16)
    |((unsigned long)(b)[(i)+2]<<8)
    |((unsigned long)(b)[(i)+3] );

#define DES_IP(X,Y)
    T =((X>>4)^Y)&0x0F0F0F0F;
    Y^=T;X^=(T<<4);
    T=((X>>16)^Y)&0x0000FFFF;
    Y^=T;X^=(T<<16);
    T=((Y>>2)^X)&0x33333333;
    X^=T;Y^=(T<<2);
    T=((Y>>8)^X)&0x00FF00FF;
    X^=T;Y^=(T<<8);
    Y=((Y<<1)|(Y>>31))&0xFFFFFFFF;
    T=(X^Y)&0xAAAAAAAA;
    Y ^=T;X^= T;
   X=((X<<1)|(X>>31))&0xFFFFFFFF;

#define DES_ROUND(X,Y)
    T=*key++^X;
    Y^=SB8[(T)&0x3F]
       ^SB6[(T>>8)&0x3F]
       ^SB4[(T>>16)&0x3F]
       ^SB2[(T>>24)&0x3F];
    T=*key++^((X<<28)|(X>>4));
    Y^=SB7[(T)&0x3F]
       ^SB5[(T>>8)&0x3F]
       ^SB3[(T>>16)&0x3F]
       ^SB1[(T>>24)&0x3F];

void des_crypt_ecb(unsigned long *key,
                unsigned char input[8],
              unsigned char output[8] ){
    int i;
    unsigned long X,Y,T;
    GET_ULONG_BE(X,input,0);
    GET_ULONG_BE(Y,input,4);
    DES_IP(X,Y);
    for(i=0;i<8;i++) {
        DES_ROUND(Y,X);
        DES_ROUND(X,Y);
    }
```

```
    DES_FP(Y,X);
    PUT_ULONG_BE(Y,output,0);
    PUT_ULONG_BE(X,output,4);
}
```

Looking at the binary code of that function generated by the compiler, one may observe that this code actually uses only 22 different types of ARM instructions, namely **add, and, asr, b, ble, bne, bx, cmp, eor, ldm, ldr, ldrb, lsl, lsr, mov, orr, pop, push, str, str, strb, sub**.

Given that we have a proof that the machine code generated from C is correct, thanks to `CompCert`, and now a proof of the ARM instruction set for these instructions, we have a proof that the simulation of the DES algorithm on our ARM simulator is conformant with the algorithm.

6 Conclusion

We have constructed a tool chain that makes it possible to certify that the simulation of a program is conformant with the formal definition of the algorithm, by leveraging off from three existing tools namely, Compcert-C, that has defined formal C semantics and a formally proven the code generator, Coq and SimSoc, to which we have added a proven generated simulator of the ARM instruction set. Although we have considered a small example in the paper, there is no limit on the size of the C code that can be certified as long as the instruction set has been certified.

In fact, if there existed a publicly availale formal model of the ARM processor approved by ARM Ltd company, our work could be construed to define a *certified execution of a C program*.

We acknowledge a weak point in our chain, due to the fact that the hardware vendors do not provide formal semantics of their instruction set. Because such formal models are unavailable, we had to define a formal model of the ARM processor ourselves, which may be wrong. As Coq specifications are executable, we have been able to validate our ARM formal model by checking that we obtain identical results with real test programs, but this is not a formal proof... If the vendors would make public formal specifications of their architectures, then our toolchain could become fully verified.

Finally, assuming that the real ARM chips commercialized by the various circuit vendors are indeed implementations of the formal model, an outcome of our work is that, since we prove the execution of programs over a simulator that is itself a proved implementation of the architecture, it means then that the execution of programs that are themselves generated from formal specifications (C programs generated from formal models with a certified generator) on that architecture are proved to be correct with regards to their specification. This assumption remains to be proved but we believe however that it represents a significant step forward.

Acknowledgments

Removed for review.

References

1. E. I. Leonard and C. L. Heitmeyer, "Automatic program generation from formal specifications using apts," in *Automatic Program Development*, O. Danvy, H. Mairson, F. Henglein, and A. Pettorossi, Eds. Springer Netherlands, 2008, pp. 93–113. [Online]. Available: http://dx.doi.org/10.1007/978-1-4020-6585-9_10
2. J.-R. Abrial, *Modeling in Event-B: System and Software Engineering*, 1st ed. New York, NY, USA: Cambridge University Press, 2010.
3. A. Basu, S. Bensalem, M. Bozga, J. Combaz, M. Jaber, T.-H. Nguyen, and J. Sifakis, "Rigorous component-based system design using the bip framework," *IEEE Software*, vol. 28, no. 3, pp. 41–48, 2011.
4. H. R. Nielson and F. Nielson, *Semantics with applications: a formal introduction*. New York, NY, USA: John Wiley & Sons, Inc., 1992.
5. C. A. Gunter, *Semantics of programming languages, Structures and Techniques*. The MIT Press, 1992.
6. G. Winskel, *The formal semantics of programming languages: an introduction*. The MIT Press, 1993.
7. J. E. Stoy, *Denotational Semantics: The Scott-Strachey Approach to Programming Language Theory*. Cambridge, MA, USA: MIT Press, 1977. [Online]. Available: http://gallium.inria.fr/~xleroy/publi/compcert-CACM.pdf
8. R. W. Floyd, "Assigning Meanings to Programs," in *Proceedings of a Symposium on Applied Mathematics*, ser. Mathematical Aspects of Computer Science, J. T. Schwartz, Ed., vol. 19. Providence: American Mathematical Society, 1967, pp. 19–31. [Online]. Available: http://www.eecs.berkeley.edu/\~{}necula/Papers/FloydMeaning.pdf
9. C. A. R. Hoare, "An axiomatic basis for computer programming," *Commun. ACM*, vol. 12, no. 10, pp. 576–580, Oct. 1969. [Online]. Available: http://doi.acm.org/10.1145/363235.363259
10. G. Canet, P. Cuoq, and B. Monate, "A value analysis for c programs," in *Source Code Analysis and Manipulation, 2009. SCAM'09. Ninth IEEE International Working Conference on*. IEEE, 2009, pp. 123–124.
11. F. Bobot, J. Filliâtre, C. Marché, and A. Paskevich, "Why3: Shepherd your herd of provers," *Boogie*, vol. 2011, pp. 53–64, 2011.
12. J.-C. Filliâtre and C. Marché, "The Why/Krakatoa/Caduceus platform for deductive program verification," in *Proceedings of the 19th International Conference on Computer Aided Verification, Lecture Notes in Computer Science 4590*, 2007.
13. P. Herms, "Certification of a Tool Chain for Deductive Program Verification," Ph.D. dissertation, Université de Paris-Sud, January 2013.
14. P. Herms, C. Marché, and B. Monate, "A certified multi-prover verification condition generator," *Verified Software: Theories, Tools, Experiments*, pp. 2–17, 2012.
15. B. C. Pierce, C. Casinghino, M. Greenberg, H. Ctlin, V. Sjberg, and B. Yorgey, "Software foundation tutorial." [Online]. Available: http://www.cis.upenn.edu/~bcpierce/sf
16. B. Campbell, "An executable semantics for compcert c," in *Certified Programs and Proofs*. Springer, 2012, pp. 60–75.
17. C. Pusch, "Proving the soundness of a java bytecode verifier specification in isabelle/hol," in *Tools and Algorithms for the Construction and Analysis of Systems (TACAS99*. Springer-Verlag, 1999, pp. 89–103.
18. J. Alglave, A. Fox, S. Ishtiaq, M. O. Myreen, S. Sarkar, P. Sewell, and F. Z. Nardelli, "The semantics of power and arm multiprocessor machine code," in

Proceedings of the 4th workshop on Declarative aspects of multicore programming, ser. DAMP '09. New York, NY, USA: ACM, 2008, pp. 13–24. [Online]. Available: http://doi.acm.org/10.1145/1481839.1481842

19. A. C. J. Fox and M. O. Myreen, "A Trustworthy Monadic Formalization of the ARMv7 Instruction Set Architecture," in *ITP*, 2010, pp. 243–258.
20. S. Sutarwala, P. G. Paulin, and Y. Kumar, "Insulin: An instruction set simulation environment," in *CHDL '93: Proceedings of the 11th IFIP WG10.2 International Conference sponsored by IFIP WG10.2 and in cooperation with IEEE COMPSOC on Computer Hardware Description Languages and their Applications*. Amsterdam, The Netherlands, The Netherlands: North-Holland Publishing Co., 1993, pp. 369–376.
21. J. Zhu and D. D. Gajski, "An ultra-fast instruction set simulator," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 10, no. 3, pp. 363–373, 2002.
22. M.-K. Chung and C.-M. Kyung, "Improvement of compiled instruction set simulator by increasing flexibility and reducing compile time," in *RSP '04: Proceedings of the 15th IEEE International Workshop on Rapid System Prototyping*. Washington, DC, USA: IEEE Computer Society, 2004, pp. 38–44.
23. B. Cmelik and D. Keppel, "Shade: A fast instruction-set simulator for execution profiling," in *SIGMETRICS '94: Proceedings of the 1994 ACM SIGMETRICS conference on Measurement and modeling of computer systems*. New York, NY, USA: ACM, May 1994, pp. 128–137.
24. E. Witchel and M. Rosenblum, "Embra: fast and flexible machine simulation," in *SIGMETRICS '96: Proceedings of the 1996 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*. New York, NY, USA: ACM, 1996, pp. 68–79.
25. F. Bellard, "Qemu, a fast and portable dynamic translator," in *ATEC '05: Proceedings of the annual conference on USENIX Annual Technical Conference*. Berkeley, CA, USA: USENIX Association, 2005, pp. 41–41.
26. K. Scott, N. Kumar, S. Velusamy, B. Childers, J. W. Davidson, and M. L. Soffa, "Retargetable and reconfigurable software dynamic translation," in *Proceedings of the International Symposium on Code Generation and Optimization (CGO03)*, 2003.
27. V. Bala, E. Duesterwald, and S. Banerjia, "Dynamo: a transparent dynamic optimization system," *SIGPLAN Not.*, vol. 35, no. 5, pp. 1–12, 2000.
28. A. Nohl, G. Braun, O. Schliebusch, R. Leupers, H. Meyr, and A. Hoffmann, "A universal technique for fast and flexible instruction-set architecture simulation," in *DAC '02: Proceedings of the 39th conference on Design automation*. New York, NY, USA: ACM, 2002, pp. 22–27.
29. W. Qin, J. D'Errico, and X. Zhu, "A multiprocessing approach to accelerate retargetable and portable dynamic-compiled instruction-set simulation," in *CODES+ISSS '06: Proceedings of the 4th international conference on Hardware/software codesign and system synthesis*. New York, NY, USA: ACM, 2006, pp. 193–198.
30. H. Shi, Y. Wang, H. Guan, and A. Liang, "An intermediate language level optimization framework for dynamic binary translation," *SIGPLAN Not.*, vol. 42, no. 5, pp. 3–9, 2007.
31. D. Jones and N. Topham, "High speed cpu simulation using ltu dynamic binary translation," in *HiPEAC '09: Proceedings of the 4th International Conference on High Performance Embedded Architectures and Compilers*. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 50–64.

32. F. Brandner, A. Fellnhofer, A. Krall, and D. Riegler, "Fast and accurate simulation using the llvm compiler framework," in *Rapid Simulation and Performance Evaluation: Methods and Tools (RAPIDO'09)*, O. T. Smail Niar, Rainer Leupers, Ed. Paphos, Cyprus: HiPEAC, January 2009, pp. 1–6.
33. C. Helmstetter and V. Joloboff, "Simsoc: A systemc tlm integrated iss for full system simulation," in *Circuits and Systems, 2008. APCCAS 2008. IEEE Asia Pacific Conference on.* IEEE, 2008, pp. 1759–1762.
34. I. forge, "Simsoc open source software." [Online]. Available: http://gforge.inria.fr/projects/simsoc
35. Y. Bertot and P. Castéran, *Interactive Theorem Proving and Program Development. Coq'Art: The Calculus of Inductive Constructions*, ser. Texts in Theoretical Computer Science. Springer-Verlag, 2004.
36. Coq Development Team, *The Coq Reference Manual, Version 8.2*, INRIA Rocquencourt, France, 2008, http://coq.inria.fr/.
37. X. Leroy, *The CompCert C verified compiler. Documentation and user's manual*, INRIA Paris-Rocquencourt, March 2012, http://creativecommons.org/licenses/by-nc-sa/3.0/.
38. ——, "Formal verification of a realistic compiler," *Communications of the ACM*, vol. 52, no. 7, pp. 107–115, 2009. [Online]. Available: http://gallium.inria.fr/~xleroy/publi/compcert-CACM.pdf
39. C. Helmstetter, V. Joloboff, and H. Xiao, "SimSoC: A full system simulation software for embedded systems," in *Open-source Software for Scientific Computation (OSSC), 2009 IEEE International Workshop on*, IEEE, Ed., Sept 2009, pp. 49–55.
40. *SystemC v2.2.0 Language Reference Manual (IEEE Std 1666-2005)*, Open SystemC Initiative, 2006, http://www.systemc.org/.
41. ARM, *ARM Architecture Reference Manual DDI 0100I.* ARM, 2005.