



HAL
open science

Formal Verification of Transformations on Abstract Clocks in Synchronous Compilers

van Chan Ngo, Jean-Pierre Talpin, Paul Le Guernic

► **To cite this version:**

van Chan Ngo, Jean-Pierre Talpin, Paul Le Guernic. Formal Verification of Transformations on Abstract Clocks in Synchronous Compilers. [Research Report] RR-8064, 2012. hal-00730926v1

HAL Id: hal-00730926

<https://inria.hal.science/hal-00730926v1>

Submitted on 11 Sep 2012 (v1), last revised 30 Jan 2013 (v5)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Formal Verification of Transformations on Abstract Clocks in Synchronous Compilers

Van Chan Ngo, Jean-Pierre Talpin, Paul Le Guernic

**RESEARCH
REPORT**

N° 8064

September 2012

Project-Team ESPRESSO



Formal Verification of Transformations on Abstract Clocks in Synchronous Compilers

Van Chan Ngo, Jean-Pierre Talpin, Paul Le Guernic

Project-Team ESPRESSO

Research Report n° 8064 — September 2012 — 21 pages

Abstract: Translation validation was introduced in the 90's by Pnueli et al. as a technique to formally verify correctness of code generated from the data-flow synchronous language SIGNAL. Rather than certifying the code generator (by writing it entirely using a theorem prover) or exhaustively qualifying it (by obeying to the 27 documentations required as per the DO-178C) translation validation provides a scalable approach to assessing the functional correctness of automatically generated code. By revisiting transition validation, which in the 90's suffered from the limitations of theorem proving and model checking technologies available then, we aim at developing a scalable and flexible approach that can apply to an existing 500k-lines implementation of SIGNAL, POLYCHRONY, and handle large-scale, possibly automatically generated, SIGNAL programs, while using off-the-shelf, efficient, model-checkers and SAT/SMT-solving libraries.

We implement translation validation in step-by-step fashion, by proving each transformation of an initial specification separately from the others, until the latest step of actual C-code generation. In this paper, we focus on proving the preservation of timing properties during the compilation process. We use an SMT-solver (Satisfiability Modulo Theory) for checking that the, so-called, implicit clock model of a specification conforms to its implementation as a Boolean function. This Boolean function (or hierarchization) is an important transformation performed by the compiler, as it deterministically characterizes the presence/absence status of all discrete signals manipulated by the specification (inputs and outputs) given that of its trigger.

Key-words: Formal Verification, Translation Validation, Certified Compiler, SMT Solver, Multi-clocked Synchronous Programs, Embedded Systems

RESEARCH CENTRE
RENNES – BRETAGNE ATLANTIQUE

Campus universitaire de Beaulieu
35042 Rennes Cedex

Vérification Formelle des Transformations sur Les Horloges dans Compilateurs Synchrones des Données de Flux

Résumé : Translation validation was introduced in the 90's by Pnueli et al. as a technique to formally verify correctness of code generated from the data-flow synchronous language SIGNAL. Rather than certifying the code generator (by writing it entirely using a theorem prover) or exhaustively qualifying it (by obeying to the 27 documentations required as per the DO-178C) translation validation provides a scalable approach to assessing the functional correctness of automatically generated code. By revisiting transition validation, which in the 90's suffered from the limitations of theorem proving and model checking technologies available then, we aim at developing a scalable and flexible approach that can apply to an existing 500k-lines implementation of SIGNAL, POLYCHRONY, and handle large-scale, possibly automatically generated, SIGNAL programs, while using of-the-shelf, efficient, model-checkers and SAT/SMT-solving libraries.

We implement translation validation in step-by-step fashion, by proving each transformation of an initial specification separately from the others, until the latest step of actual C-code generation. In this paper, we focus on proving the preservation of timing properties during the compilation process. We use an SMT-solver (Satisfiability Modulo Theory) for checking that the, so-called, implicit clock model of a specification conforms to its implementation as a Boolean function. This Boolean function (or hierarchization) is an important transformation performed by the compiler, as it deterministically characterizes the presence/absence status of all discrete signals manipulated by the specification (inputs and outputs) given that of its trigger.

Mots-clés : Formal Verification, Translation Validation, Certified Compiler, SMT Solver, Multi-clocked Synchronous Programs, Embedded Systems

1 Introduction

Adhering to the synchronous paradigm, synchronous data-flow languages such as LUSTRE [12] or SIGNAL [11] have been introduced and successfully used to design and implement embedded software architectures and critical real-time systems. Each synchronous data-flow language, it is generally associated with a compiler which transforms, compiles synchronous programs written in synchronous languages. For safety-critical, high-assurance systems, to obtain the reliability, these systems are verified by the use of formal methods (e.g. model checking, program proof, and static analysis). However, and before code can be generated, the compilation of high-level, synchronous, specification is a complex process that involves many analysis and program transformation stages. Some transformations may introduce additional informations or constraints, to refine the meaning of the original specification and/or remove or specialize the behavior of the source specification, such as optimization, static scheduling. Thus, and even if compliant with a "five-nines" (99.999%) reliability, large-scale use of compilers for large specifications may improbably yet not uncertainly yield bugs. One, nonetheless, expects the formally verified behavior of the source specification to be preserved in the code automatically generated from it and naturally requires it to be formally checked as well. In addition, the formal verification are applied to the source code of a synchronous program. The semantic preservation from the source code to its compiled program guarantees that all the formally verified properties are carried from the source code to the compiled program.

Means to circumvent compiler bugs are to entirely rewrite the code generator (in our case, e.g., the 500k C-code lines SIGNAL compiler) using a theorem proving tool such as COQ, or qualify its compliance to DO-178C recommendations for a particular execution platform, or to formally verify the conformance of its output to its input for each run of the code generator. The first solutions yield a situation where the code generator can either hardly or impossibly be further optimized and updated, whereas the last one provides ideal separation between the tool under verification and its checker.

In this aim, *translation validation* was introduced in the 90's by Pnueli et al.[19, 20] as a technique to formally verify the correctness of code generated from the data-flow synchronous language SIGNAL using model checking. Rather than certifying the code generator (by writing it entirely using a theorem prover) or qualifying it (by obeying to the 27 documentations required as per the DO-178C) translation validation provides a scalable approach to assessing the functional correctness of automatically generated code. By revisiting transition validation, which in the 90's suffered from the limitations of theorem proving and model checking techniques available then, we aim at developing a scalable and flexible approach that applies to an existing 500k-lines implementation of SIGNAL, POLYCHRONY, and is capable of handling large-scale, possibly automatically generated, SIGNAL programs, while using of-the-shelf, efficient, model-checkers and SAT/SMT-solving libraries [6, 16].

Our approach is to apply formal methods to the compiler transformations itself in order to automatically generate formal evidence that the semantics of the source program is preserved during program transformation and compilation, as per applicable qualification standards (DO-178C). Moreover, and on the contrary to previous or related approaches, our aim is to provide means for implementing this approach in a scalable way which, unlike specifications of SMT-solving techniques within theorem provers, uses modern model checking tools or efficient SMT libraries to achieve the expected goals: traceability and formal evidence.

We implement translation validation in step-by-step fashion, by proving each transformation of an initial specification separately from the others, until the latest step of actual C-code generation. In this paper, we focus on proving the preservation of timing properties during the compilation process. We use an SMT-solver (Satisfiability Modulo Theory) for checking that the,

so-called, implicit clock model of a specification conforms to its implementation as a Boolean function. This Boolean function (or hierarchization) is an important transformation performed by the compiler, as it deterministically characterizes the presence/absence status of all discrete signals manipulated by the specification (inputs and outputs) given that of its trigger.

The remainder of this paper is organized as follows. Section 3 introduces the formal model describes the abstract clocks, and clock relations, called *clock model*. Section 4 presents the translation of synchronous program into the clock model. In Section 5, we consider the definition of correct transformation on abstract clocks which formally proves conformance between the original specification and that reverse-engineered from its compiled program. Section 6 addresses the application of the verification process to the SIGNAL compiler, and its implementation integrated in the POLYCHRONY toolset [17]. Section 7 presents some related works, concludes our work and outlines future directions.

2 Preliminaries

In this section, we will recall some basic elements of propositional and first-order logics, their semantics and validity, satisfiability checking problems [10, 13].

2.1 Propositional logic

The expressions in propositional logic are called propositional *formulas* which can be any string whose evaluation is either true or false. Propositional formulas are expressed in a propositional language which consists a countable set P whose elements are called *boolean variables* and denoted by p, q, r .

Definition (formula.) Propositional formulas are defined inductively as:

- Every boolean variable is a formula, called *atom*,
- \top and \perp are formulas,
- If A_1, A_2 are formulas, then $A_1 \bowtie A_2$ is a formula.

where $\bowtie \in \{\wedge, \vee, \neg, \rightarrow, \leftrightarrow\}$ used to build formulas are called *connectives*. A formulas of forms $A_1 \vee A_2, A_1 \wedge A_2$ are respectively a *disjunction* and *conjunction*.

The semantics of propositional logic is based on the following assumptions: (i) the meaning of atomic propositions depends on their interpretation, (ii) the meaning of more complex propositions depends on the meaning of their components as it is shown in Table 1.

Definition (boolean value, interpretation, truth.) A boolean value (or *truth* value) is either 1 or 0. An *interpretation* (or truth assignments) for a set of boolean variables P is a mapping from P to the set of boolean values $\{1, 0\}$.

Interpretations can be extended to arbitrary propositional formulas inductively as following:

- $I(\top) = 1$ and $I(\perp) = 0$.
- $I(A_1 \wedge \dots \wedge A_n) = 1$ iff $I(A_i) = 1$ for all i .
- $I(A_1 \vee \dots \vee A_n) = 1$ iff $I(A_i) = 1$ for some i .
- $I(\neg A) = 1$ iff $I(A) = 0$.
- $I(A \rightarrow B) = 1$ iff $I(A) = 0$ or $I(B) = 1$.
- $I(A \leftrightarrow B) = 1$ iff $I(A) = I(B)$.

Given an interpretation I , we say that formula A is true (respectively false) in I if $I(A) = 1$ (respectively $I(A) = 0$), denoted by $I \models A$ ($I \not\models A$).

The following definition defines the satisfiability, validity, and equivalence of formulas.

Definition (model, satisfiability, validity, equivalence) An interpretation I *satisfies* a formula A if A is true in I , and I is called a *model* of A . A formula A is *satisfiable* (*valid*) if it is true in some (every) interpretation. A valid formula is called *tautology*. Two formulas A and B are *equivalent* ($A \equiv B$) if every model of A is a model of B , and vice versa.

The above definition can be generalized to sets of formulas as follows. We say that an interpretation I is a model of a set of formulas S if it satisfies every formula in S , denoted as $I \models S$. A set of formulas is satisfiable if there exists some model. We consider here some main lemmas which are useful for the problems of checking equivalence, validity, and satisfiability.

\wedge	1	0	\vee	1	0	\neg		\rightarrow	1	0	\leftrightarrow	1	0
1	1	0	1	1	1	1	0	1	1	0	1	1	0
0	0	0	0	1	0	0	1	0	1	1	0	0	1

Table 1: Operators semantics

Lemma 2.1 (i) A formula A is valid iff $\neg A$ is unsatisfiable. (ii) A formula A is satisfiable iff $\neg A$ is not valid. (iii) A formula A is valid iff A is equivalent to \top . (iv) Formula A and B are equivalent iff the formula $A \leftrightarrow B$ is valid.

Proof The proofs of properties (i), (ii), (iii), (iv) use the same method. Thus, we will only prove property (iv).

\Rightarrow) Assume that $A \leftrightarrow B$ is valid, given any interpretation I , we have $I \models (A \leftrightarrow B)$, following the truth table of \leftrightarrow we can see that $I \models A$ iff $I \models B$, so A and B are equivalent.

\Leftarrow) Assume that A and B are equivalent, given any interpretation I . If $I \models A$, then by the equivalence $I \models B$, thus $I \models A \leftrightarrow B$. In the similar way, if $I \not\models A$, then by the equivalence $I \not\models B$, and hence $I \models A \leftrightarrow B$. Therefore, $A \leftrightarrow B$ is valid.

The following lemma can help us reduce satisfiability checking for set of formulas to satisfiability checking for a formula.

Lemma 2.2 Let $S = \{A_1, \dots, A_n\}$ be a finite set of formulas. Then S is satisfiable iff the formula $A_1 \wedge \dots \wedge A_n$ is satisfiable.

Proof We prove in the same manner as the proof of Lemma 2.1 by using the interpretation property of $A_1 \wedge \dots \wedge A_n$ and the truth table of \wedge . Hence, we omit the detailed proof here.

Evaluating a formula in an interpretation can be formalized as the following decision problem:

Definition (formula evaluation.) Formula evaluation a decision problem whose instance is a pair (A, I) , where A is formula and I is an interpretation. The answer is "yes" if $I \models A$.

We can evaluate formulas in interpretations by using straightforward the above definition. In other hand, we can first evaluate its sub-formulas, then using the truth tables to evaluate the formula.

Example We consider the formula $A = (p \rightarrow q) \wedge (p \wedge q \rightarrow r) \rightarrow (p \rightarrow r)$ and the interpretation $I = \{p \mapsto 1, q \mapsto 0, r \mapsto 1\}$. The decision problem using the evaluations of sub-formulas is described as Table 2. Verification engines that can reason on propositional formulas to answer whether $I \models A$ are called SAT solvers.

2.2 First-order logic

As with propositional logic, expressions in first-order logic are made up of sequences of symbols which divided into two categories: logical symbols and non-logical symbols or parameters. Logical symbols consists of *parentheses* ($(,)$), *propositional connectives* ($\neg, \vee, \wedge, \rightarrow, \leftrightarrow$), *variable*, and *quantifiers* (\forall, \exists). Parameters consists of *equality symbol* ($=$), *predicate symbols* (e.g. $x > y$), *constant symbols* (e.g. $0, \pi$), *function symbols* (e.g. $x + y * z$). Each predicate and function symbol has an associated *arity* which is a natural number indicating how many arguments it takes. Equality and constant symbols can be considered as a special predicate symbol of arity 2,

	subformula	value
1	$(p \rightarrow q) \wedge (p \wedge q \rightarrow r) \rightarrow (p \rightarrow r)$	1
2	$(p \rightarrow r)$	1
3	$(p \rightarrow q) \wedge (p \wedge q \rightarrow r)$	0
4	$(p \wedge q \rightarrow r)$	1
5	$(p \rightarrow q)$	0
6	$p \wedge q$	0
7	p	1
8	q	0
9	r	1

Table 2: Evaluation of formula using truth tables

and a function of arity 0, respectively. We denote \mathcal{P} is the set of predicate symbols, \mathcal{F} is the set of function symbols, \mathcal{C} is the set of constant symbols. And $\Sigma = \mathcal{P} \cup \mathcal{F} \cup \mathcal{C}$ is called a *signature*.

As propositional logic, the expressions in first-order logic are called *formulas* which can be any sequences of symbols (logical and parameter symbols) whose evaluation is either true or false. First-order formulas are expressed in a first-order language which must first specify its parameters. First, we consider the *terms* of a first-order language which made up of logical and parameter symbols as following:

Definition (terms.) Terms are defined as follows.

- Any variable is a term.
- If $c \in \mathcal{C}$, then c is a term
- If t_1, \dots, t_n are terms and $f \in \mathcal{F}$ with arity $n > 0$, then $f(t_1, \dots, t_n)$ is a term.
- Nothing else is a term.

Or we can write in Backus Naur form: $t ::= x \mid c \mid f(t, \dots, t)$ where x is a variable. Given the definition of terms, we can now define the formulas in first-order language.

Definition Given the set of terms, first-order formulas are defined inductively as follows:

- If $P \in \mathcal{P}$ whose arity $n \geq 1$, and t_1, \dots, t_n are terms, then $P(t_1, \dots, t_n)$ is a formula (*atomic formula*).
- If ϕ is a formula, then $\neg\phi$ is a formula.
- If ϕ and ψ are formulas, then so are $(\phi \vee \psi)$, $(\phi \wedge \psi)$, $(\phi \rightarrow \psi)$ and $(\phi \leftrightarrow \psi)$.
- If ϕ is a formula and x is a variable, then $(\forall x.\phi)$ and $(\exists x.\phi)$ are formulas.
- Nothing else is a formula.

The above definition can be represented in the Backus Naur form as:

$$\phi ::= P(t_1, \dots, t_n) \mid (\neg\phi) \mid (\phi \vee \phi) \mid (\phi \wedge \phi) \mid (\phi \rightarrow \phi) \mid (\phi \leftrightarrow \phi) \mid (\forall x.\phi) \mid (\exists x.\phi)$$

The set of *well-formed formulas* is the set of formulas generated inductively from the atomic formulas by using the operations \mathcal{E}_\neg , \mathcal{E}_\rightarrow , and \mathcal{E}_\forall , where $\mathcal{E}_\neg(\phi) = (\neg\phi)$, $\mathcal{E}_\rightarrow(\phi, \psi) = (\phi \rightarrow \psi)$ and $\mathcal{E}_\forall(\phi) = \forall v_i.\phi, i = 1, 2, \dots$. Given a well-formed formula ϕ , a variable x is said free in ϕ :

- If ϕ is an atomic formula, then x is free iff x occurs in ϕ .
- x is free in $(\neg\phi)$ iff x is free in ϕ .
- x is free in $\phi \rightarrow \psi$ iff x is free in ϕ or ψ .
- x is free in $\forall v_i.\phi$ iff x is free in ϕ and $x \neq v_i$.

If $\forall v_i$ appears in ϕ , then v_i is said to be bound in ϕ . A formula without free variable is called a *sentence*.

In first-order logic, we use a *model* (also called a *structure*) to determine the truth of a formula. Given a signature Σ . A model \mathcal{M} of the pair $(\mathcal{F}, \mathcal{P})$ consists of the following set of data:

- A non-empty set A , the universe of concrete values,
- for each constant $c \in \Sigma$, a concrete element $c^{\mathcal{M}}$ of A ,
- for each function $f \in \mathcal{F}$ with arity $n > 0$, a concrete function $f^{\mathcal{M}} : A^n \rightarrow A$, and
- for each $P \in \mathcal{P}$ with arity $n > 0$, a subset of $P^{\mathcal{M}} \subseteq A^n$.

Here, it is totally different between f and $f^{\mathcal{M}}$ and between P and $P^{\mathcal{M}}$. The symbols f and P are just that symbols, whereas $f^{\mathcal{M}}$ and $P^{\mathcal{M}}$ denote a concrete function and relation in a model \mathcal{M} , respectively.

Example Let $\mathcal{F} = \{i\}$ and $\mathcal{P} = \{T, F\}$ [13] where i is a constant, F and T are predicate symbols with arities one and two, respectively. A model \mathcal{M} a set of concrete values A which may be considered as a states of an automata. The interpretations $i^{\mathcal{M}}$, $T^{\mathcal{M}}$, and $F^{\mathcal{M}}$ would be an initial state, a translation relation, and a set of final states, respectively. For instance, let $A = \{a, b, c\}$, $i^{\mathcal{M}} = a$, $T^{\mathcal{M}} = \{(a, a), (a, b), (a, c), (b, c), (c, c)\}$, and $F^{\mathcal{M}} = \{b, c\}$ where (a, b) means that there exists a transition from state a to state b . This model can be used to check a formula of first-order logic $\exists y.T(i, y)$. This formula says that there is a transition from the initial state to some state, and it is true in our model since there exists transitions from the initial state a to states a, b , and c .

It remains the value assignments of variables in our model. Given a model \mathcal{M} , a *variable assignment* is a mapping which assigns to each variable x a value of \mathcal{M} . Finally, we are able to give a semantics to first-order logic formulas as follows:

Definition Given a model \mathcal{M} for a signature Σ and a variable assignment l , we define the satisfaction relation, denoted by $\mathcal{M} \models_l \phi$ for each formula ϕ over the signature Σ and the variable assignment l by using structural induction on ϕ . If $\mathcal{M} \models_l \phi$ holds, we say that ϕ computes to true in the model \mathcal{M} with respect to the environment l .

The structural induction on formula ϕ is described as the following:

- P : If ϕ of the form $P(t_1, \dots, t_n)$, then we interpret the terms t_1, \dots, t_n in the set A by replacing all variables with their values according to l . Assume that concrete values a_1, \dots, a_n of A for each of these terms, where any function symbol f is interpreted by $f^{\mathcal{M}}$. Then $\mathcal{M} \models_l P(t_1, \dots, t_n)$ holds iff $(a_1, \dots, a_n) \in P^{\mathcal{M}}$.
- $\forall x$: $\mathcal{M} \models_l \forall x\phi$ holds iff $\mathcal{M} \models_{l[x \mapsto a]} \phi$ holds for all $a \in A$.
- $\exists x$: $\mathcal{M} \models_l \exists x\phi$ holds iff $\mathcal{M} \models_{l[x \mapsto a]} \phi$ holds for some $a \in A$.

- \neg : $\mathcal{M} \models_l \neg\phi$ holds iff $\mathcal{M} \models_l \phi$ does not hold.
- \vee : $\mathcal{M} \models_l \phi_1 \vee \phi_2$ holds iff $\mathcal{M} \models_l \phi_1$ or $\mathcal{M} \models_l \phi_2$ holds.
- \wedge : $\mathcal{M} \models_l \phi_1 \wedge \phi_2$ holds iff $\mathcal{M} \models_l \phi_1$ and $\mathcal{M} \models_l \phi_2$ hold.
- \rightarrow : $\mathcal{M} \models_l \phi_1 \rightarrow \phi_2$ holds iff $\mathcal{M} \models_l \phi_2$ holds whenever $\mathcal{M} \models_l \phi_1$ holds.

We use $\mathcal{M} \not\models \phi$ to denote the fact that $\mathcal{M} \models \phi$ does not hold. Given a model \mathcal{M} for a signature Σ and a variable assignment l , verification engines that can reason on formulas of first-order logic to answer whether $\mathcal{M} \models \phi$ are called *Satisfiability Modulo Theories* (SMT) solvers. A primary goal of a SMT solver is to create a verification engine that can reason natively at a higher level of abstraction, while still retaining the speed and automation of boolean engines.

3 A clock model of synchronous program

In this section, we will present an approach to model the abstract clock semantics of a synchronous program. The clock semantics consists of the abstract clock information of data-flows in which define the status of the data-flows (present or absent) and the explicit and/or implicit abstract clock relations in the program. First, we have an overview of a synchronous data-flow language, SIGNAL.

3.1 Overview of the SIGNAL language features

In SIGNAL language [15, 11], data in the system is represented by a data-flow variable, called *signal*. A signal x is a sequence of values with the same type $x(t_i)_{i \in \mathbb{N}}$, which are present at some logic instants. The set of instants (or time tags) where a signal is present is the *abstract clock* of the signal, noted as \hat{x} . A particular type of signals called *event* is characterized only by its presence, and always has the value *true*. The constructs of the language use an equational style to specify the relations and dependencies of data and clock between signals in the form $\mathcal{R}(x_1, \dots, x_k)$. Systems of equations on signals are built using a composition which construct a *process*. A whole program is a process which runs infinitely taking parameters, input signals for computing the output signals to react to the environment.

The language is based on seven different types of equations to construct primitive processes or equations specifying computations over signals. And a composition operation is used to build more elaborate processes in the form of systems of equations. We will present each equation along with its semantic meaning and the implicit relationships between the clocks of the input and output signals.

- *Equation on Data*: The equation $y := f(x_1, \dots, x_n)$ where f is an n -ary relation over numerical or Boolean data types, defines a process whose output $y(t)$ for tag $t \in \hat{y}$ is $y(t) = f(x_1(t), \dots, x_n(t))$. The clock constraint of the input and output signals is $\hat{y} = \hat{x}_1 = \dots = \hat{x}_n$.
- *Delay*: The equation $y := x\$1$ init a defines a process whose output $y(t_i) = a$ if t_i is the initial time tag, and for every other tag, $y(t_i) = x(t_{i-1})$. The clock constraint of the input and output signals is $\hat{y} = \hat{x}$.
- *Merge*: The merge equation $y := x$ default z defines a process whose output at time tag t is $y(t) = x(t)$ when $t \in \hat{x}$ and $y(t) = z(t)$ if $t \notin \hat{x} \wedge t \in \hat{y}$. The clock constraint of the merge equation is $\hat{y} = \hat{x} \cup \hat{z}$.
- *Sampling*: The sampling equation $y := x$ when b defines a process whose output signal $y(t)$ has value $x(t)$ when the signal x is present and the Boolean signal b is present with the value *true*. The clock constraint of input and output signals is $\hat{y} = \hat{x} \cap [b]$ where $[b] = \{t \in \hat{b} | b(t) = true\}$.
- *Composition*: $P \triangleq P_1 | P_2$ where P_1 and P_2 are processes. P consists of the composition of the systems of equations. The composition operator is commutative and associative.
- *Restriction*: $P \triangleq P_1$ where x , where P_1 and x are a process and a signal, respectively. It enables local declarations in the process P_1 , and leads to the same constraints as P_1 .
- *Equation on clocks*: The language allows clock constraints to be defined *explicitly* by equations. The signal's clock is represented by a special signal of type *event* which carries

only a single value *true*. It specifies the presence of the signal, denoted \hat{x} . Thus, equations on clocks over signals are equations over their corresponding event signals. They are: (i) the synchronization relation $x \hat{=} y \triangleq \hat{x} = \hat{y}$, (ii) clock union relationship $x \hat{+} y \triangleq \hat{x} \text{ default } \hat{y}$, (iii) clock intersection relationship $x \hat{*} y \triangleq \hat{x}$ when \hat{y} , (iv) difference relationship $x \hat{-} y \triangleq \text{when}(\text{not}(\hat{y}) \text{ default } \hat{x})$.

3.2 A clock model of the synchronous program

In the synchronous language such as ESTEREL [2], LUSTRE, and SIGNAL the logical time is completely determined by the system reactions on the occurrences of observed events in which on the system is supposed to react fast enough to produce the corresponding output events on the occurrence of input event before the next input event arrive. Each reaction denotes a single *logical instant* in the synchronous model where the relations between observed events and the data dependencies are expressed [1]. Synchronous data-flow languages represent data as an infinite sequence of values called data-flow, and each data-flow is combined with an associated abstract clock as a means of discrete time to define the presence or absence of the data in its data-flow. Thus, the principle of our encoding scheme is that, at a particular instant, the abstract clock can be represented as a variable whose values are **true** (the corresponding data-flow is present) or **false** (the corresponding data-flow is absent).

Consider a synchronous program P , we denote by $X_P = \{x_1, x_2, \dots, x_n\}$ the set of all data-flow variables of P . With each data-flow x_i with its type numerical, Boolean, or event, we encode its clock with a Boolean variable \hat{x}_i . Almost the structure of synchronous program is described as a set of equational definitions. And the whole system is represented as systems of these equations. This original structure makes that it is natural to represent the relations between abstract clocks described implicitly or explicitly by an equation in terms of first-order logic formulas. And then the combination of equations can be represented by the conjunction of the corresponding first-order logic formulas. We assume that all considered programs are supposed to be written with the primitive operators, meaning that derived operators are replaced by their corresponding primitive ones, and there is no nested operators such as $z := x \text{ default } (y \text{ when } b)$. The nested operators are broken by using fresh variables. These formulas use the usual logic operators and numerical comparison functions [13]. Consider a general equation $y := R(x_1, x_2, \dots, x_n)$, where R is an operator define in synchronous languages (e.g. **suspend** in ESTEREL, **when** in SIGNAL,...) or it can be an usual Boolean and numerical operators, our abstraction uses the following subset of numerical and Boolean expressions in synchronous programs:

$$\begin{aligned} nexpr &\triangleq \text{const} \mid nexpr \diamond nexpr \mid \text{var} \\ bexpr &\triangleq \text{true} \mid \text{false} \mid \text{var} \mid \neg bexpr \mid bexpr \vee bexpr \mid \\ &\quad bexpr \wedge bexpr \mid nexpr \bowtie nexpr \end{aligned}$$

where *const* and *var* denote a constant and a data-flow variable or an abstract clock (e.g. x, \hat{y}, \dots), $\bowtie \in \{<, >, =, >=, =>, / =\}$ and $\diamond \in \{*, /, +, -\}$, then the abstract clock semantics of this equation can be represented as a first-order logic formula over the clocks and/or the Boolean value of the involved signals $\Phi(\hat{y}, \hat{x}_1, \hat{x}_2, \dots, \hat{x}_n, x_1, \dots)$. With the Boolean expression defined by numerical comparison functions $bexpr \triangleq nexpr \bowtie nexpr$ and numerical expressions, to ensure the result formulas are Boolean, we only encode the fact that the clocks of Boolean and numerical expressions are synchronized, and we avoid encoding the numerical comparison function which defines the value of the Boolean expression, the numerical expressions. For each equation i^{th} in program P , we denote by Φ_{eq_i} its abstract clock semantics, then the abstract clock semantics of

P can be represented by a first-order logic formula, called its *clock model*, denoted as:

$$\Phi_P = \bigwedge_i^n \widehat{\Phi}_{eq_i} \quad (1)$$

where n denotes the number of equations composed in P .

We consider a simple illustration example of a synchronous program consists of only one equation written in SIGNAL syntax: $z := x \text{ default } (y \text{ when } b)$. It specifies the relations between the data-flow output three integer data-flows x, y, z and a Boolean data-flow b . The clock of z is implicitly defined by the clocks of data-flows x, y and the value of the Boolean data-flow b . First, the equation is rewritten by introducing a fresh variable ny to break the nested operation `default` as:

```
z := x default ny
| ny := y when b
```

The clock information of the first and second equations can be encoded as $\hat{z} \leftrightarrow \hat{x} \vee \widehat{ny}, \widehat{ny} \leftrightarrow \hat{y} \wedge \hat{b} \wedge b$, respectively. Then the clock model of the above program is:

$$\begin{aligned} \Phi &= && \hat{z} \leftrightarrow \hat{x} \vee \widehat{ny} \\ &\wedge && \widehat{ny} \leftrightarrow \hat{y} \wedge \hat{b} \wedge b \end{aligned}$$

Boolean signals		Non-Boolean signals	
$y := \text{not } x$	$\hat{y} \leftrightarrow \hat{x}$ $\wedge y \leftrightarrow \neg x$	$y := f(x_1, \dots, x_n)$	$\hat{y} \leftrightarrow \hat{x}_1 \leftrightarrow \dots \leftrightarrow \hat{x}_n$
$y := x \text{ and } z$	$\hat{y} \leftrightarrow \hat{x} \leftrightarrow \hat{z}$ $\wedge y \leftrightarrow x \wedge z$		
$y := x \text{ or } z$	$\hat{y} \leftrightarrow \hat{x} \leftrightarrow \hat{z}$ $\wedge y \leftrightarrow x \vee z$		
$y := x \text{ default } z$	$\hat{y} \leftrightarrow \hat{x} \vee \hat{z}$ $\wedge y \leftrightarrow (\hat{x} \wedge x \vee \neg \hat{x} \wedge \hat{y} \wedge y)$	$y := x \text{ default } z$	$\hat{y} \leftrightarrow \hat{x} \vee \hat{z}$
$y := x \text{ when } b$	$\hat{y} \leftrightarrow (\hat{x} \wedge \hat{b} \wedge b)$ $\wedge y \leftrightarrow (\hat{x} \wedge x \wedge \hat{b} \wedge b)$	$y := x \text{ when } b$	$\hat{y} \leftrightarrow (\hat{x} \wedge \hat{b} \wedge b)$
$y := x\$1 \text{ init } a$	$\hat{y} \leftrightarrow \hat{x}$ $\wedge y \leftrightarrow (\hat{x} \wedge m.x)$ $\wedge m.x_0 \leftrightarrow a$ $\wedge m.x' \leftrightarrow (\hat{x} \wedge x \vee \neg \hat{x} \wedge m.x)$	$y := x\$1 \text{ init } a$	$\hat{y} \leftrightarrow \hat{x}$
$P_1 \mid P_2$	$\Phi_{P_1} \wedge \Phi_{P_2}$		
$P \text{ where } x$	$\exists x. \Phi_P$		

Table 3: Translation of the primitive equations

4 SIGNAL to clock model

We will provide in-depth representation of the abstract clock semantics of the synchronous data-flow language, SIGNAL. We use the method which has been discussed before that means for each signal x , we use a Boolean variable \hat{x} to encode its abstract clock. The language uses seven primitive equations as in Section 3 to construct programs. Therefore, we only need to define the translation of these primitive equations to first-order logic formulas encoding the abstract clocks, and the implicit or explicit clock relations of the signal involved in the equation. The composition of equations is simply translated as the conjunction of the corresponding first-order logic formulas. For the delay operator $\$$ (e.g. $x\$1$), it requires memorizing the past value of the signal, that is done by introducing a new variable $m.x$, where $m.x$ stores the previous value of signal x and $m.x'$ stores the current value of signal x . Table 3 shows the translation of the primitive equations of the language. For instance, the primitive equation $y := x_1 \text{ and } x_2$ is represented by a first-order logic formulas; $\hat{y} \leftrightarrow \hat{x}_1 \leftrightarrow \hat{x}_2$ and $y \leftrightarrow x_1 \wedge x_2$. SIGNAL allows clock constraints to be defined explicitly by equations as in Subsection 3.1, in this context, the signal clock is represented by a special signal of type event, our abstraction encodes the clock by using a Boolean variable. By applying the above translation scheme, the following translations are obtained for equation on clocks:

- $x^{\wedge} = y = \hat{x} \leftrightarrow \hat{y}$ (synchronization)
- $z := x^{\vee} + y = \hat{z} \leftrightarrow (\hat{x} \vee \hat{y})$ (union)
- $z := x^{\wedge} * y = \hat{z} \leftrightarrow (\hat{x} \wedge \hat{y})$ (intersection)
- $z := x^{\wedge} - y = \hat{z} \leftrightarrow (\hat{x} \wedge \neg \hat{y})$ (difference)

For example the SIGNAL program `Bathtub` [4] shown in Figure 1 specifies the level of a bathtub. It takes no input, three output signals at lines 2 and 3, respectively. The water level (signal `level`) in the bathtub is defined by the previous level (signal `zlevel`), the increase (signal `faucet`), and


```

1 process Bathtub =
2 (?
3 ! integer level; boolean alarm, ghost_alarm;)
4 (|(| level := zlevel + faucet - pump
5   | zlevel := level$1 init 1
6   | faucet := zfaucet + (1 when zlevel <= 4)
7   | zfaucet := faucet$1 init 0
8   | pump := zpump + (1 when zlevel >= 7)
9   | zpump := pump$1 init 0 |)
10 |(| overflow := level >= 9
11   | scarce := 0 >= level
12   | alarm := scarce or overflow
13   | ghost_alarm := (true when scarce when overflow)
14                       default false |)|)
15 where
16   integer zlevel, zfaucet, zpump, faucet, pump;
17   boolean overflow, scarce;
18 end;

```

Figure 1: Bathub model in SIGNAL

the decrease (signal `pump`) of the water level. The increase and decrease are modified by one unit under some conditions of the previous water level value at line 6 and 8 to keep the water level in a appropriate values. The system emits a warning at line 12 (signal `alarm`) whenever the water level overflows at line 10 or be scarce at line 11. Signal `ghost_alarm` is defined at line 13-14 is expected to never happen.

Following our translation scheme above to program `Bathtub` which consists of two processes P_1 (from line 4 to line 9) and P_2 (from line 10 to line 14). Applying the translation of the process composition, we have the clock model $\Phi_{\text{Bathtub}} = \Phi_{P_1} \wedge \Phi_{P_2}$, where:

$$\begin{aligned}
\Phi_{P_1} = & \widehat{level} \leftrightarrow \widehat{zlevel} \leftrightarrow \widehat{faucetpump} \\
& \wedge \widehat{faucet} \leftrightarrow \widehat{zfaucet} \leftrightarrow \widehat{x1} \wedge x1 \\
& \wedge \widehat{x1} \leftrightarrow \widehat{zlevel} \\
& \wedge \widehat{pump} \leftrightarrow \widehat{zpump} \leftrightarrow \widehat{x2} \wedge x2 \\
& \wedge \widehat{x2} \leftrightarrow \widehat{zlevel}
\end{aligned}$$

We rewrite lines 6 and 8 with $x1 := (zlevel \leq 4)$ and $x2 := (zlevel \geq 7)$. For Φ_{P_2} , we rewrite equations at line 13 and 14 as: $y1 := \text{true when scarce} \mid y2 := y1 \text{ when overflow}$

$$\begin{aligned}
\Phi_{P_2} = & \widehat{overflow} \leftrightarrow \widehat{level} \leftrightarrow \widehat{scarce} \\
& \wedge \widehat{alarm} \leftrightarrow \widehat{scarce} \leftrightarrow \widehat{overflow} \\
& \wedge \widehat{y1} \leftrightarrow \widehat{scarce} \wedge \widehat{scarce} \\
& \wedge \widehat{y2} \leftrightarrow \widehat{y1} \wedge \widehat{overflow} \wedge \widehat{overflow} \\
& \wedge \widehat{ghost_alarm} \leftrightarrow \widehat{y2}
\end{aligned}$$

5 The correct clock transformation

5.1 Clock semantics of synchronous program

Let $X = \{x_1, \dots, x_n\}$ be a finite set of typed data-flow variables. For each $x_i \in X$, we write \mathcal{D}_{x_i} to denote its value domain, then we have $\mathcal{D}_X = \bigcup_{i \in [1, n]} \mathcal{D}_{x_i}$ is the value domain of X . We base on the basic elements defined of *trace semantics* [14] in data-flow computing to define the *clock semantics* of a synchronous program.

Definition (clock events). Given a non-empty set X , the set of clock events on X , denoted by \mathcal{E}_{c_X} , is the set of all interpretations I for the set of clocks X .

An interpretation I for the set of variables X is a mapping from X to the set of boolean values $\{0(\text{false}), 1(\text{true})\}$ means that the variables are *interpreted* as boolean values. $I(x) = 1$ means data-flow x holds a value $v \in \mathcal{D}_x$ while $I(x) = 0$ means that it holds no value or it is absent.

Definition (clock traces). Given a non-empty set of clocks X , the set of clock traces on X , denoted by \mathcal{T}_{c_X} , is defined by the set of functions T_c defined from the set \mathbb{N} of natural numbers to \mathcal{E}_{c_X} .

The natural numbers represent the time instant $t = 0, 1, 2, \dots$, a trace T_c is a chain of clock events along the time instants. For example, consider a set of data-flow variables $X_1 = \{x_1, x_2\}$, then the possible clock events are $\mathcal{E}_{c_{X_1}} = \{(x_1 \mapsto 0, x_2 \mapsto 0), (x_1 \mapsto 0, x_2 \mapsto 1), (x_1 \mapsto 1, x_2 \mapsto 0), (x_1 \mapsto 1, x_2 \mapsto 1)\}$ and $T_c : (0, (x_1 \mapsto 0, x_2 \mapsto 0)), (1, (x_1 \mapsto 1, x_2 \mapsto 0)), \dots$ is one of the possible clock traces on X_1 . The clock semantics of a synchronous program P is a set of constrained traces, denoted by $[[P]]_c$. Table 4 shows the clock semantics of each SIGNAL primitive operator. $T(t)(b)$ denotes the value of the boolean signal b at the time instant t .

5.2 Soundness of the clock model

Let $X = \{x_1, \dots, x_n\}$ be a finite set of typed data-flow variables of a synchronous program P and its clock model Φ_P over the corresponding set of clocks \hat{X} , it is easy to see that an interpretation satisfying Φ_P describes the abstract clocks of P at a given time instant. Given an interpretation $\hat{I} \models \Phi_P$ for \hat{X} , we have the set of clock events computed as: $S_{sat}(\hat{I}) = \{I \in \mathcal{E}_{c_X} \mid \forall i, I(x_i) = \hat{I}(\hat{x}_i)\}$, where I is an interpretation for X . The set of all clock events are union of all clock events for each interpretation satisfying the clock model $S_{sat}(\Phi_P) = \bigcup_{\hat{I} \models \Phi_P} S_{sat}(\hat{I})$. With a set of clock events $S_{sat}(\Phi_P)$, the concretization of the clock model Φ_P is the set of clock traces which satisfy:

$$\Gamma(\Phi_P) = \{T_c \in \mathcal{T}_{c_X} \mid \forall t, T_c(t) \in S_{sat}(\Phi_P)\} \quad (2)$$

The translation is sound in term of preserving the behaviors of the abstracted program: if a property defined over the abstract clocks is true on the clock model, then it is true as well on the program as the following proposition.

Proposition 5.1 *Let P, Φ_P, X, \hat{X} to be a synchronous program, its clock model, data-flow variables, and abstract clocks, respectively, φ is a formula whose atom is $\hat{x}_i \in \hat{X}$. If $\Phi_P \rightarrow \varphi$ then $[[P]]_c \subseteq \Gamma(\varphi)$, and P is said to satisfy φ denoted as $P \vdash \varphi$.*

Proof The proof of Proposition 5.1 is done by using Lemma 5.2. Given a clock trace $T_c \in [[P]]_c$, applying Lemma 5.2, $T_c \in \Gamma(\Phi_P)$ means that $\forall t, T_c(t) \in S_{sat}(\Phi_P)$. Since $\Phi_P \rightarrow \varphi$, then every interpretation \hat{I} satisfying Φ_P also satisfies φ , and any clock event $I \in S_{sat}(\Phi_P)$ also be in $S_{sat}(\varphi)$, thus $\forall t, T_c(t) \in S_{sat}(\varphi)$. Therefore, we have $T_c \in \Gamma(\varphi)$.

Process P	Semantics $[[P]]_c$
$y := R(x_1, \dots, x_n)$	$\{T_c \in \mathcal{T}c_{\{y, x_1, \dots, x_n\}} \mid \forall t \in \mathbb{N}, (\forall i, T_c(t)(x_i) = T_c(t)(y))\}$
$y := x$ default z	$\{T_c \in \mathcal{T}c_{\{x, y, z\}} \mid \forall t \in \mathbb{N}, (T_c(t)(y) = T_c(t)(x) = 1) \vee (T_c(x) = 0 \wedge T_c(t)(y) = T_c(t)(z) = 1) \vee (T_c(t)(y) = T_c(t)(x) = T_c(t)(z) = 0)\}$
$y := x$ when b	$\{T_c \in \mathcal{T}c_{\{x, y, b\}} \mid \forall t \in \mathbb{N}, (T_c(t)(x) = 1 \wedge T_c(t)(b) = 1 \wedge T_c(t)(y) = 1) \vee (T_c(x) = 0 \wedge T_c(t)(y) = 0) \vee (T_c(t)(x) = 1 \wedge T_c(t)(b) = 0 \wedge T_c(t)(y) = 0)\}$
$y := x$ \$!\$ init a	$\{T_c \in \mathcal{T}c_{\{x, y\}} \mid \forall t \in \mathbb{N}, (T_c(t)(x) = T_c(t)(y))\}$
$\hat{x} = y$	$\{T_c \in \mathcal{T}c_{\{x, y\}} \mid \forall t \in \mathbb{N}, (T_c(t)(x) = T_c(t)(y))\}$
$z := \hat{x} + y$	$\{T_c \in \mathcal{T}c_{\{x, y, z\}} \mid \forall t \in \mathbb{N}, (T_c(t)(x) = 1 \wedge T_c(t)(z) = 1) \vee (T_c(t)(x) = 0 \wedge T_c(t)(y) = 1 \wedge T_c(t)(z) = 1) \vee (T_c(t)(x) = 0 \wedge T_c(t)(y) = 0 \wedge T_c(t)(z) = 0)\}$
$z := \hat{x} * y$	$\{T_c \in \mathcal{T}c_{\{x, y, z\}} \mid \forall t \in \mathbb{N}, (T_c(t)(x) = 1 \wedge T_c(t)(y) = 1 \wedge T_c(t)(z) = 1) \vee (T_c(t)(x) = 0 \wedge T_c(t)(z) = 0) \vee (T_c(t)(y) = 0 \wedge T_c(t)(z) = 0)\}$
$z := \hat{x} - y$	$\{T_c \in \mathcal{T}c_{\{x, y, z\}} \mid \forall t \in \mathbb{N}, (T_c(t)(x) = 1 \wedge T_c(t)(y) = 0 \wedge T_c(t)(z) = 1) \vee (T_c(t)(x) = 0 \wedge T_c(t)(z) = 0) \vee (T_c(t)(x) = 1 \wedge T_c(t)(y) = 1 \wedge T_c(t)(z) = 0)\}$

Table 4: Clock semantics of the primitive equations

Lemma 5.2 For all program P , $[[P]]_c \subseteq \Gamma(\Phi_P)$

Proof We prove by induction on the structure of program P means that for all primitive operators of synchronous language we show that the clock semantic of any primitive operator is subset of the corresponding concretization. For instance, here we prove the case of SIGNAL language.

- Equation on data: $P = y := f(x_1, \dots, x_n)$, y . First, consider y is numerical signal, following the translation scheme, we have the clock $\Phi_P = \hat{y} \leftrightarrow \hat{x}_1 \leftrightarrow \dots \leftrightarrow \hat{x}_n$. If an interpretation \bar{I} is model of Φ_P then:

- either $\forall i, \hat{y} = 0$ and $\hat{x}_i = 0$;
- or $\forall i, \hat{y} = 1$ and $\hat{x}_i = 1$.

$S_{sat}(\Phi_P)$ is the set of all interpretations of the form above. Let $T_c \in [[P]]_c$ be a clock trace and any instant $t \in \mathbb{N}$, then either $\forall i, T_c(t)(y) = T_c(t)(x_i) = 0$ or $T_c(t)(y) = T_c(t)(x_i) = 1$, that means $T_c \in \Gamma(\Phi_P)$. When y is boolean signal, the proof is similar.

- Delay, sampling, and merging operators, we prove in the same manner.
- Composition: $P = P_1 | P_2$. We have $[[P]]_c \subseteq [[P_1]]_c \subseteq \Gamma(\Phi_{P_1})$ by applying the induction hypothesis. In the same way, we also have $[[P]]_c \subseteq \Gamma(\Phi_{P_2})$. Then, $[[P]]_c \subseteq \Gamma(\Phi_{P_1}) \cap \Gamma(\Phi_{P_2})$. Since $\Gamma(\Phi_{P_1}) \cap \Gamma(\Phi_{P_2}) \subseteq \Gamma(\Phi_{P_1} \wedge \Phi_{P_2})$, we have $[[P]]_c \subseteq \Gamma(\Phi_{P_1} \wedge \Phi_{P_2}) = \Gamma(\Phi_P)$.
- Restriction: $P = P_1$ where x . By definition of clock semantics we have $[[P]]_c \subseteq [[P_1]]_c$ and $\Gamma(\Phi_{P_1}) \subseteq \Gamma(\exists x. \Phi_{P_1})$. Since $[[P_1]]_c \subseteq \Gamma(\Phi_{P_1})$ by induction then we have the proof.

5.3 Definition of correct transformation: Refinement

In this section, we adopt the translation validation approach [19, 20] to verify formally that the abstract clock semantics are preserved for every transformation of synchronous compiler. In order to do that, we propose a formal definition of correct transformation on clock models. Consider the two clock models Φ_{P_1}, Φ_{P_2} , to which we refer respectively as a source program and a compiled program produced by a synchronous data-flow compiler. We assume that they have the same set of input/output data-flow variables (in our case both of them are derived from the original source code program). We say that P_1 and P_2 have the same clock semantics if for every property φ if Φ_{P_1} satisfies this property then Φ_{P_2} also satisfies and vice-versa or:

$$\forall \varphi. ((\Phi_{P_1} \rightarrow \varphi) \leftrightarrow (\Phi_{P_2} \rightarrow \varphi)) \quad (3)$$

Requirement (3) is too strong in general to be practice for synchronous data-flow languages. The source language is usually non-deterministic, compilers are allowed to select one of the possible behaviors of the source program. Additionally, compilers do transformations, optimizations for removing or eliminating some wrong behaviors of the source program (e.g. eliminating sub-expressions, trivial clock constraints). To address these issues, we relax the requirement above as follows:

$$\forall \varphi. ((\Phi_{P_2} \rightarrow \varphi) \rightarrow (\Phi_{P_1} \rightarrow \varphi)) \quad (4)$$

Requirement (4) says that all properties which are satisfied by the clock semantics of Φ_{P_2} which are satisfied by the clock semantics of Φ_{P_1} as well. And we say that the transformation on abstract clocks is *correct transformation* or P_2 *refines* P_1 w.r.t the clock semantics. We write $P_2 \sqsubseteq P_1$ to denote the fact that P_2 refines P_1 . With an unverified synchronous data-flow compiler, each compilation task is followed by our refinement verification process to provide formal guarantees as strong as those provided by a formally verified compiler. Indeed, consider the following process:

$$\begin{aligned} Cp'(P_1) &= \text{if } Cp(P_1) \text{ is} \\ &\quad \text{Error} \rightarrow \text{Error} \\ &| \quad \text{OK}(P_2) \rightarrow \text{if } P_2 \sqsubseteq P_1 \text{ then OK}(P_2) \text{ else Error} \end{aligned}$$

where $Cp(P_1)$ is the compilation task from source program P_1 to either compiled code (written as $Cp(P_1) = \text{OK}(P_2)$) or compilation errors (written as $Cp(P_1) = \text{Error}$).

We now discuss an approach to check the existing of refinement by using a SMT-solver [9]. Based on the above computation of $\Gamma(\Phi_{P_1}), \Gamma(\Phi_{P_2})$ from the corresponding clock models and the soundness property of Φ_{P_1} and Φ_{P_2} , we can now have the following theorem which gives us a method to check the refinement relation.

Theorem 5.3 *Given a source program P_1 and its transformed program P_2 , P_2 is correct transformation of P_1 on abstract clocks if it satisfies that the formula Φ_{P_1} is a logical consequence of the formula Φ_{P_2} , or*

$$\text{If } \models (\Phi_{P_2} \rightarrow \Phi_{P_1}) \text{ then } P_2 \sqsubseteq P_1 \quad (5)$$

Proof To prove Theorem 5.3, we show that if $\models (\Phi_{P_2} \rightarrow \Phi_{P_1})$ then $\forall \varphi. ((\Phi_{P_2} \rightarrow \varphi) \rightarrow (\Phi_{P_1} \rightarrow \varphi))$. For any formula φ such that $\Phi_{P_2} \rightarrow \varphi$. It is easy to see that since $\models (\Phi_{P_2} \rightarrow \Phi_{P_1})$ means for every interpretations I if $I \models \Phi_{P_2}$ then $I \models \Phi_{P_1}$. In addition, $I \models \varphi$, hence we have $\Phi_{P_1} \rightarrow \varphi$.

6 Implementation

In this section, we describe the main components of the implementation which is integrated in the existing POLYCHRONY toolset [17] to prove the correctness of the SIGNAL compiler on abstract clocks. The compiler [5] consists of a sequence of code transformations. Some transformations are optimizations that rewrite the code to eliminate subexpressions, inefficiencies. The compilation process may be seen as a sequence of morphisms rewriting programs to SIGNAL programs. And the final steps (C or Java code generation) are simple morphisms over the ultimately transformed SIGNAL program. For convenience, the transformations of the compiler are classed into three phases as depicted in Figure 2. The optimized final program $*_SEQ_TRA.SIG$ is translated

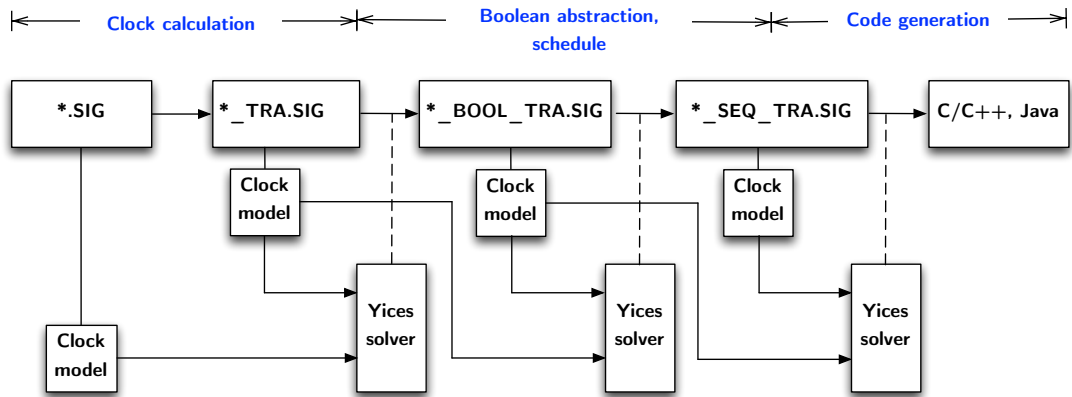


Figure 2: An overview of our integration within POLYCHRONY toolset

directly to executable code. We are interested in the first two stages of the compiler: the clock calculation and boolean abstraction, scheduling. The intermediate forms in the transformations of the compiler may be expressed in the SIGNAL language itself. To prove the correctness of the compiler transformations on abstract clocks our implementation approach takes the input program $P.SIG$ and its transformed program $P_TRA.SIG$. It first computes the clock models based on the above translation scheme by using *flex lexer* and *bison parser* [3]. The clock models of input and transformed programs are combined as a formula $(\Phi_{P_TRA.SIG} \rightarrow \Phi_{P.SIG})$. Then it checks that $\models (\Phi_{P_TRA.SIG} \rightarrow \Phi_{P.SIG})$ or equivalently checks $M \not\models \neg(\Phi_{P_TRA.SIG} \rightarrow \Phi_{P.SIG})$. The result of this checking can be exploited for the correctness of the compiler's transformations. If the result says that the checked formula is not valid (or the negation formula is satisfiable) then the SIGNAL compiler emits an error compilation. Otherwise, the compiler continues its work. The same procedure is applied for other stages of the compiler. Finally, our verification process asserts that $P_SEQ_TRA.SIG \sqsubseteq P_BOOL_TRA.SIG \sqsubseteq P_TRA.SIG \sqsubseteq P.SIG$ along the transformations of the compiler.

Here, we delegate the checking of the above formula against the clock models to an SMT solver that efficiently deals with first-order logic formulas over Boolean and numeric expressions. The checking formulas belong to decidable theories, this solver gives two types of answers: *sat* when the formula has a model (there exists an interpretation that satisfies it); or *unsat* otherwise. Our implementation uses the SMTLIB common format [8] to encode the formulas obtained from the previous step as input of SMT solvers. For our implementation, we consider the YICES [9] solver, which is one of the best two solvers at the last SMTCOMP competition [21].

7 Related work and conclusion

The notion of translation validation was introduced in [19, 20] by A. Pnueli et al. to verify the code generator of SIGNAL. In that work, the authors define a language of symbolic models to represent both the source and target programs called *Synchronous Transition Systems (STS)*. A STS is a set of logic formulas which describe the functional and temporal constraints of the whole program and its generated C code. Then they use BDD [7] representations to implement the symbolic models STSS, and their proof method uses a SAT-solver to reason on the signal constraints. It amounts to the mapping for selected states, consisting of the values of input-output-memory variables, for the source and the target code. The drawback of this approach is that it does not capture explicitly the clock semantics and in some cases, the code generator eliminates the use of a local register variable in the generated code and then, the mapping cannot be established. Additionally, for a large SIGNAL programs, the logic formula is asked to SAT-solver to solve is very large that makes some inefficiency. In addition, the whole calculation of a synchronous program or the generated code is considered as one atomic transition in STS, thus it does not capture the scheduling semantics, data dependencies of the programs. Another related work is the approach of J. C. Peralta et al. [18] in which is based on translation validation approach. In particular, they translate both the SIGNAL (multi-clocked) specifications and its generated code C/C++ or Java simulator into LTSS. Then, an appropriate pre-order test on both LTSS can be interpreted as a refinement between a generated code implementation and its specification. The refinement they propose is a bisimulation relation and they use the existing tools to generate the greatest bisimulation relation for the specification and the target generated code in C/C++. In case there is no bisimulation relation, counterexamples are generated automatically. However, this approach has not been fully automated.

The present paper provides a proof of correctness of the multi-clocked synchronous programming language compiler for clock semantics preservation and applies this approach to the synchronous data-flow language SIGNAL compiler. We have presented a technique based on SMT-solving to prove the preservation of timing properties during compilation. Namely, we have shown that implicit clock relations, describing the discrete timing model of a data-flow specification could be check conform with their implementation by a Boolean function (the hierarchy of the program) which deterministically characterizes the presence/absence status of all its input/output signals.

The desired behavior of a given source program and the transformed one are represented as PDSS over the finite field of integers modulo $p = 3$. A refinement relation between source and transformed programs is used to express the preservation. A proof by simulation is presented to establish the refinement relation. All compilation stages are followed by a similar refinement verification process.

We have implemented and integrated our translation validation process within the POLYCHRONY toolset by extending the functionality of the existing model checker SIGALI to prove the correctness of the full compilation phases of the compiler. As future work, to deal with synchronous programs with large number of variables we intend to represent clock information as formulas over Boolean variables, then use a SMT-solver to check that the clock model of compiled program is a model of the source program. Another perspective is to use *Synchronous Data-flow Dependency Graph (SDDG)* to represent the dependency semantics (or schedule semantics) of synchronous programs and verify that the compiler compilation preserves the data dependency semantics.

References

- [1] A. Benveniste, P. Caspi, S.A. Edwards, N. Halbwachs, P. Le Guernic, and R. De Simone: The synchronous languages 12 years later. *Proceedings of the IEEE 91(1)*. pp.64-83, 2003
- [2] G. Berry: The foundations of ESTEREL. In *Proof, Language and Interaction: Essay in Honor of Robin Milner*, MIT Press, 2000.
- [3] Bison - GNU parser generator. <http://www.gnu.org/software/bison>
- [4] F. Besson, T. Jensen, and J-P. Talpin: Polyhedral analysis for synchronous languages. In *Proceedings of the 6th International Symposium on Static Analysis, volume 1694 LNCS*. pp.51-68, Sep 1999.
- [5] L. Besnard, T. Gautier, P. Le Guernic, and J-P. Talpin: Compilation of polychronous data flow equations. In *Synthesis of Embedded Software*, Springer, 2010.
- [6] A. Biere, M. Heule, H. van Maaren, and T. Walsh: Handbook of satisfiability: Volume 185 Frontiers in Artificial Intelligence and Applications. IOS Press, Amsterdam, The Netherlands. ISBN 978-1-5860-3929-5, 2009.
- [7] R. Bryant: Graph-based algorithms for boolean function manipulation. *IEEE Transactions on Computers*, C-35(8):677-691, Aug 1986.
- [8] C. Barrett, S. Ranise, A. Stump, and C. Tinelli: The Satisfiability Modulo Theories Library (SMT-LIB). <http://www.SMT-LIB.org>, 2008.
- [9] B. Dutertre, and L. de Moura: Yices sat-solver. <http://yices.csl.ri.com>, 2009.
- [10] J.H. Gallier: Logic for computer science. John Wiley. 1987.
- [11] A. Gamatié: Designing embedded systems with the SIGNAL programming: Synchronous, Reactive Specification. Springer, New York. ISBN 978-1-4419-0940-4, 2009.
- [12] N. Halbwachs: A synchronous language at work: the story of LUSTRE. In *3th ACM-IEEE International Conference on Formal Methods and Models for Codesign (MEMOCODE'05)*, Jul 2005.
- [13] M. Huth, and M. Ryan: Logic in computer science: Modelling and Reasoning about systems. Cambridge University Press. ISBN 978-0-5215-4310-1, 2004.
- [14] P. Le Guernic, and T. Gautier: Advanced topics in data-flow computing, chapter data-flow to von Neumann: the SIGNAL approach. Prentice-Hall. pp.413-438, 1991.
- [15] P. Le Guernic, J-P. Talpin, and J-C. Le Lann: Polychrony for system design. *Journal for Circuits, Systems and Computers*. 12(3):261-304, Apr 2003.
- [16] L. de Moura, and N. Bjorner: Satisfiability Modulo Theories: An appetizer. In *Brazilian Symposium on Formal Methods (SBMF'2009)*, Gramado, Brazil, Aug 2009.
- [17] Inria/Espresso: Polychrony Toolset. <http://www.irisa.fr/espresso/Polychrony>.
- [18] J. C. Peralta, T. Gautier, L. Besnard, and P. Le Guernic: LTSs for translation validation of (multi-clocked) SIGNAL specifications. In *8th IEEE/ACM International Conference on Formal Method and Models for Codesign*. MEMOCODE, 2010.

- [19] A. Pnueli, M. Siegel, and E. Singerman: Translation validation. *In B. Steffen, editor, 4th Intl. Conf. TACAS'98. LNCS 1384.* pp.151-166, 1998.
- [20] A. Pnueli, O. Shtrichman, and M. Siegel: Translation validation: From SIGNAL to C. *In Correct Sytem Design Recent Insights and Advances. LNCS 1710.* pp.231-255, 2000.
- [21] A. Stump, and M. Deters: <http://www.smtcomp.org/2009>, 2009.



**RESEARCH CENTRE
RENNES – BRETAGNE ATLANTIQUE**

Campus universitaire de Beaulieu
35042 Rennes Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399