Synchronous Programming of Reactive Systems*
A Tutorial and Commented Bibliography

Nicolas Halbwachs

Vérimag**, Grenoble – France
e-mail: Nicolas.Halbwachs@imag.fr

1 Reactive Systems

The term "reactive system" was introduced by David Harel and Amir Pnueli [HP85], and is now commonly accepted to designate permanently operating systems, and to distinguish them from "transformational systems" — i.e., usual programs whose role is to terminate with a result, computed from an initial data (e.g., a compiler). In synchronous programming, we understand it in a more restrictive way, distinguishing between "interactive" and "reactive" systems:

Interactive systems permanently communicate with their environment, but at their own speed. They are able to synchronize with their environment, i.e., making it wait. Concurrent processes considered in operating systems or in data-base management, are generally interactive.

Reactive systems, in our meaning, have to react to an environment which cannot wait. Typical examples appear when the environment is a physical process. The specific features of reactive systems have been pointed out many times [Hal93,BCG88,Ber89]:

- In contrast with most interactive systems, they are generally intended to be deterministic.
- Their description involves concurrency, for several different reasons:
  1. They run in parallel with their environment;
  2. They are often implemented on distributed architectures, for reasons of speed, fault-tolerance, or physical distribution requirements;
  3. Most of the time, it is convenient to describe them as sets of concurrent processes.

Cases (2) and (3) must be distinguished. In the later case, concurrency is nothing but a description facility; we call it logical concurrency. Generally, it has nothing to do with physical concurrency involved in case (2), and is not submitted to the same constraints.

- They are submitted to critical reliability requirements. In fact, most critical systems either are reactive, or contain reactive parts.

* This work has been partially supported by the ESPRIT-LTR project "SYRF".
** Verimag is a joint laboratory of Université Joseph Fourier (Grenoble I), CNRS and INPG
2 Synchronous Programming

All control engineers know a simple way to implement a reactive system by a single loop, of the form shown by Fig. 1.a. This program scheme is "event driven" since each reaction is triggered by an input event.

Fig. 1.b shows an even simpler and more common scheme, which consists in periodically sampling the inputs. This "sampling" scheme is mainly used in numeric systems which solve, e.g., systems of differential equations. These two schemes do not deeply differ, but they correspond to different intuitive points of view. In both cases, the program typically implements an automaton: the states are the valuations of the memory, and each reaction corresponds to a transition of the automaton. Such a transition may involve many computations, which, from the automaton point of view, are considered atomic (i.e., input changes are only taken into account between two reactions). This is the essence of the synchronous paradigm, where such a reaction is often said to take no time. An atomic reaction is called an instant (logical time), and all the events occurring during such a reaction are considered simultaneous.

Now, automata are useful tools — from their simplicity, expressive power, and efficiency —, but they are very difficult to design by hand. Synchronous languages aim at providing high level, modular, constructs, to make the design of such an automaton easier. The basic construct that all these languages provide, is a notion of synchronous concurrency, inspired by Milner’s synchronous product [Mil81,Mil83]: in the sampling scheme, when automata are composed in parallel, a transition of the product is made of “simultaneous” transitions of all of them; in the event-driven scheme, some automata can stay idle, when not triggered by events coming either from the environment or from other automata. In any case, when participating in such a compound transition, each automaton considers the outputs of others as being part of its own inputs. This “instantaneous” communication is called the synchronous broadcast [BCG88,Ber89,BB91].

The important point is that, in contrast with the asynchronous concurrency considered in asynchronous languages like ADA [ADA83,Coh96], this synchronous product can preserve determinism, a highly desirable feature in reactive systems design.

There are two fields where this synchronous model has been used for years:

1 Consider, e.g., scanners and parsers, and the usefulness of tools like LEX and YACC!
In synchronous circuit design, it is the usual model of communicating Mealy machines (FSM). Most hardware description formalisms (e.g., [Bli90,CLM91]) are naturally synchronous, or contain a significant synchronous subset [Per93]. As a matter of fact, the compilation and verification of synchronous programs borrow many techniques from circuit CAD. However, while hardware description languages can be directly used to describe the data part of a circuit, they are of little help in designing complex hardware controllers. This explains the success of synchronous imperative languages, like ESTEREL, in this field.

In control engineering, high level specification formalisms are often data-flow synchronous formalisms, inherited from earlier analog technology: differential or finite-difference equations, block-diagrams, analog networks. Interpreted in a discrete world, these models can be formalized using the data-flow paradigm [Kah74,AW85,PP83]. However, these formalisms are seldom used as programming languages, and automatic code generation is not available. On the other hand, more imperative languages used for programming automatic controllers (e.g., Sequential Function Charts [LM93,AG96]) generally follow the same cyclic execution scheme.

3 Synchronous Languages

[Hal93,IEE91] are general references on synchronous languages.

Statecharts [Har87] is probably the first, and the most popular, formal language designed in the early eighties for the design of reactive systems. However, they were proposed more as a specification and design formalism, rather than as a programming language. Many features (synchronous product and broadcast) of the synchronous model are already present in Statecharts, but determinism is not ensured, and many semantic problems were raised [vdB94]. Almost at the same time, three programming languages were proposed by French academic groups:

- ESTEREL² [BCG88,BS91,BG92,Ber93,Ber98] is an imperative language developed at the “Ecole des Mines” and Inria, in Sophia Antipolis.
- SIGNAL³ [LGLL91,BL90] and LUSTRE⁴ [HCRP91,CPHP87] are data-flow languages, respectively designed at Inria (Rennes) and CNRS (Grenoble). SIGNAL is more “event-driven”, while LUSTRE mainly corresponds to the “sampled” scheme.

Also, following the formal definition of the synchronous model, a purely synchronous variant of the Statecharts was proposed: ARGOS⁵ [Mar92]. The ideas of ARGOS are currently used as a basis for a graphical version of ESTEREL, named SYMCHARTS⁶ [And96], proposed at the University of Nice.

² see http://www.inria.fr/meije/esterel/esterel-eng.html
³ see http://www.inria.fr/Equipes/EPATR-eng.html
⁴ see http://www.imag.fr/VERIMAG/SYNCHROME/lustre-english
⁵ see http://www-verimag.imag.fr/SYNCHROME/argonaute-english.html
⁶ see http://www.inria.fr/meije/esterel/syncCharts/
In this section, we use simple examples to give a flavor of the programming styles in LUSTRE and ESTEREL.

### 3.1 Overview of the synchronous data-flow language Lustre

LUSTRE is based on the synchronous data-flow model, i.e., on a synchronous interpretation of block-diagrams. A block diagram may be viewed as a network of operators (or as a system of equations, see opposite) running in parallel at the rate of their inputs.

The synchronous interpretation of such a description consists in considering each variable as taking a value at each cycle of the program. According to this interpretation, the above description means: "at any cycle $n$, $s_n = 2 \times (x_n + y_n)$".

A LUSTRE program defines its output variables as functions of its input variables. Each variable or expression $E$ denotes a function of discrete time, giving its value $E_n$ at each "instant" $n$. Variables are defined by means of equations: an equation "$X = E$", specifies that the variable $X$ is always equal to expression $E$.

Expressions are made of variable identifiers, constants (considered as constant functions), usual arithmetic, boolean and conditional operators (considered as applying pointwise to functions) and only two specific operators: the "previous" operator — which refers to the previous value of its argument — and the "followed-by" operator — which is used to define initial values: If $E$ and $F$ are LUSTRE expressions, so are "pre($E$)" and "$E \rightarrow F$", and we have at any instant $n > 0$:

- $(\text{pre}(E))_n = E_{n-1}$, while $(\text{pre}(E))_0$ has the undefined value nil.
- $(E \rightarrow F)_n = F_n$, while $(E \rightarrow F)_0 = E_0$.

For instance, if $x_n$, $y_n$ denote the respective values of $x$ and $y$ at "instant" $n$, the equation "$z = 0 \rightarrow (\text{pre}(x) + y)$" means that the initial value $z_0$ of $z$ is 0, and that, at any non initial instant $n$, $z_n = x_{n-1} + y_n$.

A LUSTRE program is structured into nodes: a node is a subprogram defining its output parameters as functions of its input parameters. This definition is given by an unordered set of equations, possibly involving local variables. Once declared, a node may be freely instanciated in any expression, just as a basic operator.

As an illustration, Figure 2 shows an extremely simple node describing a counter: it receives two integer inputs, $\text{init}$ and $\text{incr}$, and a boolean input, $\text{reset}$. It returns an integer output, $\text{count}$, which behaves as follows: at the initial instant and whenever the input $\text{reset}$ is true, the output is equal to the current value of the input $\text{init}$. At any other instant, the value of $\text{count}$ is equal to its previous value incremented by the current value of $\text{incr}$. One can make use of this node elsewhere, for instance in the equation

\[
\text{mod5} = \text{Counter}(0, 1, \text{pre(mod5)}=4);
\]
node Counter (init, incr: int; reset: bool) returns (count: int);

let count = init -> if reset then init
    else pre(count)+incr;

tel

(a) Program

(b) Behavior

Fig. 2. Example of LUSTRE program: A counter

which instanciates the node Counter, with 0 and 1 as constant initial and increment values, and resets it whenever the previous value of its output is 4. The variable mod5 is then the cyclic sequence of integers modulo 5.

So, through the notion of node, LUSTRE naturally offers hierarchical description and component reuse. Data traveling along the “wires” of an operator network can be complex, structured informations.

From a temporal point of view, industrial applications show that several processing chains, evolving at different rates, can appear in a single system. LUSTRE offers a notion of boolean clock, allowing the activation of nodes at different rates.

3.2 Overview of the synchronous imperative language Esterel

Being an imperative language, ESTEREL looks more familiar at first glance, since it provides usual constructs, like assignments, sequences, loops, .... However, its synchronous semantics makes this apparent friendliness somewhat deceptive: one must keep in mind that, apart from a few statements that explicitly take time (e.g., “await < signal >”), most ESTEREL statements are conceptually instantaneous, i.e., are executed in the same reaction than other statements that sequentially precede or follow them in the program.

ESTEREL provides a lot of constructs that we cannot present in detail. We only comment a small example, which is a speed supervisor (see Fig. 3): the program is intended to measure the speed of a vehicle, and to detect when this speed exceeds a maximum bound.

Fig. 3.a describes a speedometer: it is an ESTEREL module, receiving two signals, Second and Meter, which occur, respectively, whenever the vehicle has travelled for 1 meter and a second has elapsed. It emits a valued signal Speed, carrying the current value of the speed, an integer, measured in m/s. The body of the module is an infinite loop (lines 4–13) which initializes a local variable Distance — that will measure the number of Meters received within a Second — and enters a “do ... upto Second” construct (lines 6–10). This construct executes its body — a loop incrementing Distance on every occurrence of Meter (lines 7–9) —, until being interrupted by the next occurrence of the signal Second. So, on the first occurrence of Second following the entering in the global loop, the “do ... upto Second” statement is terminated and the counter Distance contains
the number of Meters received during this time. The signal Speed, carrying the value of Distance, is simultaneously emitted (line 11), and the loop is entered again for a new Second. So, the signal Speed is emitted exactly each Second.

Fig. 3.b shows the speed supervisor, which makes use of the Speedometer module. Here the input signals are Second and Meter, and the output is a signal TooFast that is emitted whenever the speed exceeds the bound MaxSpeed. A local signal Speed is used to transmit the result of the speedometer. Within the scope of this signal, the speedometer is instanciated, through the run construct (line 5), in parallel with a process comparing the speed to the bound: this process is triggered whenever the speedometer emits a Speed signal, whose current carried value ?Speed is compared with the bound, with the possible effect of emitting the signal TooFast.

4 Compilation of Synchronous Languages

This section is an overview of the various approaches related to the compilation of synchronous languages into sequential or distributed code. Beforehand, we have to tackle a static semantic problem, which is specific to synchronous languages: causality.

4.1 Causality analysis

Generally speaking, the problem of causality comes from the fact that not all synchronous programs have a unique, deterministic meaning. In the data-flow model,
this problem has a very simple statement, since it boils down to the well known problem of **combinational loops** in synchronous circuits [Mal93, Kau70, Sto92]: consider the following LUSTRE equations:

\[
\begin{align*}
(\text{a}) & \quad x = \text{not } x \\
(\text{b}) & \quad y = y \\
(\text{c}) & \quad z = (z^2 + 1.0)/2.0 \\
(\text{d}) & \quad u = \text{if } c \text{ then } v \text{ else } w; v = \text{if } c \text{ then } w \text{ else } u
\end{align*}
\]

Case (a) is clearly a nonsense: the equation doesn’t have any solution. Case (b) can be viewed as non deterministic, since y can have any value. Case (c) is an equation with one and only one solution (z=1), but solving such implicit algebraic equations is clearly unfeasible, in general. Case (d) can be disputed: apparently, u depends on v, that depends on u, so there is a combinational loop. Now, whatever be the value of the condition c, this loop is semantically cut.

In data-flow languages, all these situations are quite unnatural — because the user keeps the data dependences in mind — and generally easy to avoid. This is why, in LUSTRE, all the preceding examples are rejected by the compiler. However, in imperative languages like ESTEREL, ARGOS, or Statecharts, it is extremely easy to write programs with apparent causality problems — i.e., where, in some states, the presence of a signal seems to depend on itself —, to which users want to give meaning. More precise criteria must be applied to identify really problematic programs. Most of the various semantics that have been proposed for Statechart [vdB94] differ from each other by the way these problems are solved.

Let’s go further into these problems, by means of some ESTEREL examples (see Fig. 4), taken from [Ber95]. A simple way of examining the correctness of these examples is by considering all the cases of presence/absence of signals: we want to have one (reactivity) and only one (determinism) consistent solution, for each configuration of input signals. For the module P1 of Fig. 4, either O is present, in which case the else part of the “present ... else ...” statement is not executed, so O is not emitted, and O is not present; this assumption is not consistent; or O is absent, the else part is executed, so O is emitted, and the assumption is again violated. This module doesn’t have any consistent behavior. In fact, this example shows exactly the same kind of inconsistency as the equation “O = not O” in LUSTRE. Consider now the module P2 of Fig. 4. If we assume that O is present, the then part of the “present ... then ...” statement is executed, so O is emitted, and our assumption is satisfied. Now, assuming that O is absent, O is not emitted, and our assumption is satisfied again. Here, we have two consistent behaviors, it is a case of non determinism similar to the LUSTRE equation “O = O”. The module P3 is a similar case of non determinism, showing that the problem can result from dependence paths of arbitrary length. For P4, if I is present, the first process in the parallel construct emits S, and the presence of S makes the second process to emit O. Conversely, if I is absent, neither S nor O is emitted. So P4 is correct; it corresponds to the LUSTRE fragment “S = I ; O = S”, which doesn’t show any loop. The case of P5 is more questionable: the first process in the parallel seems to be non deterministic (like in P2). Now, if we assume that O1 is present, we find that the second process in
module P1:
output O;
  present 0
  else emit 0
end present
end module

module P2:
output O;
  present 0
  then emit 0
end present
end module

module P3:
output O1, O2;
  present O1 then emit O2 end
  II
  present O2 then emit O1 end
end module

module P4:
input I;
output O;
signal S in
  present I then emit S end
  II
  present S then emit O end
end signal
end module

module P5:
output O1, O2;
  present O1 then emit O1 end
  II
  present O1 then
    present O2 else emit O2 end
  end present
end module

Fig. 4. Causality problems in ESTEREL

the parallel doesn’t have any behavior (like in P1). So, P5 has one and only one consistent behavior, where neither O1 nor O2 is emitted.

All the semantics proposed for synchronous languages reject modules P1, P2, and P3, and accept P4. The Boolean causality considered in [HM95], analyzes the problem in classical logic, and accepts also P5, since it has one and only one solution. The constructive causality [SBT96,Ber95] rejects P5, because the only solution doesn’t have a constructive explanation, by means of causes and effects. Moreover, the constructive causality has been shown [SBT96] to coincide with electric stability: a constructively causal circuit will stabilize whatever be the traversal delays of its gates.

Causality problems are an obstacle to separate compilation and distributed code generation. Several authors [Bou91,BdS96,Bon95] propose a weakenning of the synchronous communication, to get round these problems.

4.2 Sequential code generation

The straightforward way for compiling a data-flow synchronous language like LUSTRE into sequential code, is by generating a single loop, after conveniently sorting the equations according to their dependences. Sequential code generation from an imperative language like ESTEREL is less obvious: in the compilers ESTEREL-V2 and -V3 [BCG88,BG92], the control part of the program was compiled into an explicit automaton, representing the control structure of the code. This approach has also been applied to LUSTRE [CPHP87], with on-the-fly minimization (by bisimulation) of the automaton [BFH+92,HRR91]. The explicit automaton is a very efficient implementation — since the whole internal synchronization of the program is computed at compile time —, with the drawback

8 See also http://www.inria.fr/meije/rc/rc-project.html.
of possibly involving an exponential expansion of the code size. This is why it is now generally abandonned for single loop compilation. However, it played a central role in the development of verification tools.

The single loop code generation for ESTEREL was a side-effect of the development of a silicon compiler [Ber92] (see §4.5): compiling ESTEREL into circuits is mainly a translation into a data-flow network, which can be easily implemented by a single loop program. The ESTEREL-V4 and -V5 compilers [Ber95] are based on this principle.

About the compilation of synchronous imperative languages into data-flow networks, let us mention also the translation of ARGOS [MH96] into the DC common format (see §4.3) and the REGLO tool [Ray96] which produces recognizers (in LUSTRE) for regular expressions.

### 4.3 Common formats

The LUSTRE, SIGNAL, and ESTEREL compilers were developed in tight cooperation. In order to share common tools, and to make the languages integration easier, common formats were defined and used as intermediate codes in the compilers: the Oc (“object code”) format [PS87] was used to encode explicit automata in the earlier versions of the compilers. Another format [CS95], named Dc/DC+ (for “declarative code”) is used now to encode implicit automata, at the equational level.

### 4.4 Distributed code generation

Compiling synchronous languages into code for distributed architectures is obviously a challenge. In [ML94], techniques are proposed to separate a clocked data-flow networks into sub-networks independent enough to be separately compiled into processes. The SYNDEX tool [LSSS91] can be used to schedule the resulting tasks on various architectures, and to study adjust the performances of the resulting system. Another approach is presented in [CGP94,CG95], which starts from the sequential code produced by the standard compilers, and from distribution directives given by the user; it consists of (1) replicating the code on each site of the architecture, (2) simplifying the code on each site according to its assigned role, and (3) adding communications along simple bounded FIFOs, to ensure both communication and synchronization.

### 4.5 Silicon compiling

Compiling synchronous programs into circuits is also an important goal, particularly in a codesign approach. Synchronous data-flow languages are very close to hardware description languages (HDLs), and their compilation to circuits is quite easy (see, e.g., [RH91]). The translation into circuits of ESTEREL was a much more difficult task, but the result is more interesting, since ESTEREL is of much higher level than usual HDLs for describing controllers. The translation proposed
in [Ber92] is structural, and must be completed by deep optimizations, using both standard CAD tools [SSL+92], and specific techniques [STB96,STB97] that take advantage of knowledge coming from the structure of the source code.

5 Verification of Synchronous Programs

Since synchronous programming is mainly devoted to the field of critical embedded systems, the formal verification of synchronous programs is a particularly important goal. By chance, it can take advantage of some specific features of the application field and of the synchronous model:

- Experience shows that critical properties that must be verified are generally simple, safety properties. By “simple properties”, we mean logical dependency relations between events, in contrast with deep arithmetic properties. As a consequence, these properties can often be model-checked on an abstraction of the program [Hal94], the most natural of which is the control automaton generated by the earlier versions of the compilers (see §4.2).
- The control automaton, being obtained as a synchronous product, is generally much smaller that models obtained by asynchronous composition (no interleaving, no need for partial orders, ...).
- The transition relation is a vectorial function, which allows particularly efficient BDD-based techniques [CBM89a] to be applied for the symbolic construction of the reachable state space.
- Thanks to the synchronous model, a specific approach can be applied to express safety properties by means of synchronous observers, i.e., special programs possibly written in the same language as the program under verification.

The verification methods for synchronous programs are all based on the control automaton.

Reduction methods have been applied [dSR94], mainly to automata compiled from ESTEREL: using the tools AUTO/AUTOGRAPH [RdS90] reduced views of the automaton can be obtained and compared. Other methods are based on model-checking [QS82,CES86], and mainly symbolic model-checking [CBM89b,BCM+90]. TEMPEST [JPV95] is a model-checker of temporal logic formulas dedicated to ESTEREL. SIGALI [LDBL93] model-checks SIGNAL programs, using the symbolic resolution of SIGNAL clocks constraints. The tool LESAR [HLR92] is a symbolic, BDD-based, model-checker of LUSTRE programs; it is based on the use of synchronous observers, to describe both the property to be checked and the assumptions on the program environment under which these properties are intended to hold: an observer of a safety property is a program, taking as inputs the inputs/outputs of the program under verification, and deciding (e.g., by emitting an alarm.
signal) at each instant whether the property is violated. Running in parallel the program, $P$, an observer $Φ$ of the desired property, and an observer $A$ of the assumption made about the environment one has just to check that either the alarm signal of $Φ$ is never emitted (property satisfied) or the alarm signal of $A$ is emitted (assumption violated), which can be done by a simple traversal of the reachable states of the compound program. Besides this only need of considering reachable states (instead of paths) this specification technique has several advantages:

- observers are written in the same language as the program under verification;
- observers are executable; they can be tested, or even kept in the actual implementation (redundancy, autotest).

Notice that, with an asynchronous language, observers would have to be explicitly synchronized with the program under verification, with the risk of changing its behavior. For an application of this technique, see also [WNT96].

Taking into account some numerical aspects — in particular delay counting — is considered in [HPR97], using abstract interpretation techniques [CC77]. In [LHR97], observers and abstract interpretation are used for verifying parameterized networks of synchronous processes.

6 Other Related Topics

[Le94] is an interesting comparative study of formal description techniques, including synchronous languages. This section lists some other works related to synchronous programming.

Program testing will remain a validation technique complementary to formal verification. The automatic testing of synchronous programs is studied in [MHMM95,TMC94,OP94,MRW98].

Integration of synchronous/asynchronous aspects: In general, only some parts of a complex system can be suitably described in the synchronous model. The language ELECTRE [CR95] is devoted to the synchronous control of asynchronous tasks. CRP [BSR93] allows ESTEREL modules to be composed asynchronously.

Combination of formalisms: Data-flow and imperative synchronous languages are complementary, and several attempts have been made, either to combine them [JLRM94], or to introduce imperative concepts in data-flow languages [RM95,MR98]. The most advanced work on combining synchronous languages is probably the “SYNCHRONIE WORKBENCH”⁹, developed at GMD.

Higher order data-flow languages: Data-flow synchronous languages can be viewed as lazy functional languages working on infinite sequence. “Lucid synchrone” [Cas93,CP95,CP96] is a higher order extension of LUSTRE, where the synchronous execution (bounded memory) is preserved.

Aknowledgements: I hope all prominent contributors to synchronous languages have been properly cited. However, the theses on the subject are not referenced, because most of them are written in French. Nevertheless, I would like to acknowledge the significant contribution of the following students:


References


10 Of course, most of them are no longer students!


