

The Semantics and Implementation of CÉU: a Synchronous Reactive Language based on Esterel

Francisco Sant'Anna, Departamento de Informática, PUC-Rio

Adriano Branco, Departamento de Informática, PUC-Rio

Roberto Ierusalimsky, Departamento de Informática, PUC-Rio

Noemi Rodriguez, Departamento de Informática, PUC-Rio

Silvana Rossetto, Departamento de Ciência da Computação, UFRJ

CÉU is a reactive language based on Esterel that targets constrained embedded platforms. Employing the synchronous programming model, it allows for a simple reasoning about shared-memory concurrency. Furthermore, its restricted semantics enables deterministic and memory-safe programs. In this work, we propose a formal semantics for CÉU focusing on its particular control mechanisms, such as parallel compositions, finalization, and stack-based internal events. We also present an implementation with two backends: one aiming for resource efficiency and interoperability with C, and another for code dissemination in sensor networks.

Additional Key Words and Phrases: Concurrency, Determinism, Embedded Systems, Esterel, Synchronous, Reactivity

1. INTRODUCTION

- Relevancia de Esterel
- Ceu
 - Foco inicial em "constrained WSNs"
 - Estendendo para outros dominios? (jogos, artigo Mod'15)
 - Usos: SenSys, Terra, sala de aula, GSoC
- Limitacoes do modelo sincrono
- Semantica e Implementacao (nesse artigo, somente subset estatico)

2. OVERVIEW OF CÉU

CÉU is a synchronous reactive language based on Esterel [Boussinot and de Simone 1991] with support for multiple concurrent lines of execution known as *trails*. By reactive, we mean that programs are stimulated by the environment through input events that are broadcast to all awaiting trails. By synchronous, we mean that all trails at any given time are either reacting to the current event or are awaiting another event; in other words, trails are never reacting to different events.

Figure 1 shows side-by-side the implementations in Esterel and CÉU for the following control specification [Berry 2000]: “*Emit an output O as soon as two inputs A and B have occurred. Reset this behavior each time the input R occurs*”. The first phrase of the specification, awaiting and emitting the events, is translated almost identically in the two languages (lines 4–9, in both implementations), as Esterel’s ‘||’ and CÉU’s *par/and* constructs are equivalent. For the second phrase, the reset behavior, the Esterel version uses a *abort-when* (lines 3–10), which serves the same purpose of CÉU’s *par/or* (lines 3–12): the occurrence of event R aborts the awaiting statements in parallel and restarts the loop.

CÉU, following the Esterel mindset, has a strong imperative flavor, with explicit control flow through sequences, loops, parallels, and also assignments. Being designed for control-intensive applications, it provides support for concurrent lines of execution and broadcast communication through events. In addition, CÉU also employs the synchronous model, in which programs advance in a sequence of discrete reactions to external events. Internal computations within a reaction (e.g. expressions, assignments, and native calls) are considered to take no time in accordance with the synchronous hy-

<pre> 1 // ESTEREL 2 loop 3 abort 4 [5 await A 6 7 await B 8]; 9 emit O 10 when R 11 end </pre>	<pre> 1 // CEU 2 loop do 3 par/or do 4 par/and do 5 await A; 6 with 7 await B; 8 end 9 end 10 emit O; 11 with 12 await R; 13 end 14 end </pre>
---	--

Fig. 1. A control specification implemented in Esterel and C  U: “Emit O after A and B, resetting each R”

pothesis [Potop-Butucaru et al. 2005]. The `await` statements are the only ones that halt a running reaction and allow a program to advance in this notion of time. To ensure that reactions run in bounded time and programs always progress, loops are statically required to contain at least one `await` statement in all possible paths [Sant’Anna et al. 2013; Berry 2000].

In the sections that follow, we review the main differences between C  U and Esterel [Sant’Anna et al. 2013]: deterministic execution (Section 2.1), safe abortion with finalization (Section 2.2), first-class timers (Section 2.3), and stack-based internal events (Section 2.4).

2.1. Deterministic Execution

Esterel is only deterministic with respect to reactive control: “the same sequence of inputs always produces the same sequence of outputs” [Berry 2000]. However, the execution order for operations with side-effects within a reaction is non-deterministic: “if there is no control dependency, as in “`call f1() || call f2()`”, the order is unspecified and it would be an error to rely on it” [Berry 2000]. In C  U, when multiple trails are active at a time, they are scheduled in the order they appear in the program source code. Hence, C  U is deterministic also with respect to the order of execution of side effects within a reaction [Sant’Anna et al. 2013]. Figure 2 compares the two syntactically equivalent code fragments in Esterel and C  U to illustrate the semantic difference regarding (non-)determinism.

On the one hand, enforcing an execution order for concurrent operations may seem arbitrary and also precludes true parallelism. On the other hand, it provides a priority scheme for trails, and makes shared-memory concurrency more tractable. In contrast, Esterel forbids sharing memory: “if a variable is written by some thread, then it can neither be read nor be written by concurrent threads” [Berry 2000]. For constrained embedded development, we believe that deterministic shared-memory concurrency is beneficial, given the extensive use of memory mapped ports for I/O and the lack of hardware support for real parallelism. Other embedded languages made a similar design choice [Karpinski and Cahill 2007; Andalam et al. 2010].

<pre> 1 // ESTEREL 2 [3 call f1(); 4 5 call f2(); 6]; </pre>	<pre> 1 // CEU 2 par/and do 3 _f1(); 4 with 5 _f2(); 6 end </pre>
---	---

Fig. 2. In Esterel, the execution order between `f1` and `f2` is unspecified, whereas in C  U, `_f1` executes before `_f2`.

```

1  input void STOP, RETRANSMIT, SENDACK;
2  par/or do
3    await STOP;
4  with
5    loop do
6      par/or do
7        await RETRANSMIT;
8      with
9        par/and do
10       await 1min;
11       with
12         var _pkt.t buffer;
13         <fill-buffer-info>
14         _send_enqueue(&buffer);
15         await SENDACK;
16       end
17     end
18   end
19 end

11 <...>
12 var _pkt.t buffer;
13 <fill-buffer-info>
14 finalize
15   _send_enqueue(&buffer);
16 with
17   _send_cancel(&buffer);
18 end
19 await SENDACK;
20 <...>

```

Fig. 3. A network protocol in CÉU extended with a finalization clause to cancel a transmission.

2.2. Safe Abortion with Finalization

The introductory example of Figure 1 illustrates how synchronous languages can abort awaiting lines of execution without tweaking them with synchronization primitives. In contrast, traditional (asynchronous) multi-threaded languages cannot express thread termination safely [Berry 1993; ORACLE 2011]. However, handling abortion when dealing with external resources can be challenging, given that external entities are not subject to the same synchronous execution discipline.

To illustrate threats related to abortion, consider the example in the left of Figure 3 with hierarchical `par/or` and `par/and` compositions describing the state machine of a data collection protocol for sensor networks [Gnawali et al. 2009; Sant’Anna et al. 2013]. The input events `STOP`, `RETRANSMIT`, and `SENDACK` (line 1) represent the external interface of the protocol with a client application. The protocol has two trails: one simply monitors the stopping event (line 3); the other periodically transmits a packet (lines 5–18). The periodic transmission is a loop that starts two other trails (lines 6–17): one handles an immediate retransmission request (line 7); the other transmits the packet¹ and waits for a confirmation (lines 9–16). The actual transmission (lines 12–15) is enclosed with a `par/and` that takes at least one minute before looping, to avoid flooding the network with packets. At any time, the client may request a retransmission (line 7), which terminates the `par/or` (line 6), aborts the ongoing transmission (line 14, if not idle), and restarts the loop (line 5). The client may also request to stop the whole protocol at any time (line 3). Note that the `buffer` packet (line 12) is a local variable whose address is passed to function `_send_enqueue` for transmission. The call enqueues the pointer in the radio driver, which holds it up to the emission of `SENDACK` acknowledging the packet transmission. In the meantime, if the sending trail is aborted by the `STOP` or `RETRANSMIT` requests, the packet buffer goes out of scope, leaving behind a *dangling pointer* in the radio driver.

The CÉU compiler tracks the interaction of `par/or` compositions with local variables and stateful *C* functions (e.g., device drivers) in order to preserve safe abortion of trails [Sant’Anna et al. 2013; 2015]. For instance, CÉU refuses to compile the program in the left of Figure 3, enforcing the programmer to write a *finalization* clause to accompany the stateful *C* call. The code in the right of Figure 3 properly cancels the

¹The underline prefix marks (e.g., `_send_enqueue`) make interactions with external *C* libraries explicit and are required in CÉU.

```

var int v;
await 10ms;
v = 1;
await 1ms;
v = 2;

par/or do
  await 10ms;
  <...> // any non-awaiting sequence
  await 1ms;
  v = 1;
with
  await 12ms;
  v = 2;
end

```

Fig. 4. First-class timers in C  U.

packet transmission when the block of buffer goes out of scope, i.e., the finalization clause (after the `with`) executes automatically on external abortion.²

2.3. First-class Timers

Activities that involve reactions to *wall-clock time*³ appear in typical patterns of embedded development, such as timeout watchdogs and sensor samplings. However, support for wall-clock time is somewhat low-level in existing languages, usually through timer callbacks or “sleep” blocking calls. Furthermore, in any concrete timer implementation, a requested timeout does not expire precisely without delays, a fact that is usually ignored in the development process. We define the difference between the requested timeout and the actual expiring time as the *residual delta time (delta)*. Without explicit manipulation, the recurrent use of timed activities in sequence (or in a loop) may accumulate a considerable amount of deltas that can lead to incorrect behavior in programs.

The `await` statement of C  U supports wall-clock time and handles deltas automatically [Sant’Anna et al. 2013], resulting in more robust applications. For the example in the left of Figure 4, suppose that after the first `await` request, the underlying system gets busy and takes 15ms to check for expiring awaits. The C  U scheduler will notice that the `await 10ms` has not only already expired, but is delayed with `delta=5ms`. Then, the awaiting trail awakes, sets `v=1`, and invokes `await 1ms`. As the current delta is higher than the requested timeout (i.e. $5ms > 1ms$), the trail is rescheduled for execution, now with `delta=4ms`.

C  U also takes into account the fact that time is a physical quantity that can be added and compared. For instance, for the program in the right of Figure 4, although the scheduler cannot guarantee that the first trail terminates exactly in 11ms, it can at least ensure that the program always terminates with `v=1`: Given that any non-awaiting sequence is considered to take no time in the synchronous model, the first trail is guaranteed to terminate before the second trail, because $10 + 1 < 12$. A similar program in a language without first-class support for timers, would depend on the execution timings for the code marked as `<...>`, making the reasoning about the execution behavior more difficult.

2.4. Internal Events

Esterel makes no semantic distinctions between internal and external signals, both having only the notion of either presence or absence during the entire reaction [Berry 1993]. In C  U, however, internal events follow a stack-based execution policy, similar to subroutine calls in typical programming languages [Sant’Anna et al. 2013], contrasting with queue-based event handling adopted for external events. Figure 5 illustrates

²Note that the compiler only enforces the programmer to write the finalization clause, but cannot check if it actually handles the resource properly.

³By wall-clock time we mean the passage of time from the real world, measured in hours, minutes, etc.

<pre>// ESTEREL input A; // external signal B; // internal [[await A; emit B; call f(); await B; call g();]]</pre>	<pre>1 // CEU 2 input void A; // external (in uppercase) 3 event void b; // internal (in lowercase) 4 par/and do 5 await A; 6 emit b; 7 _f(); 8 with 9 await b; 10 _g(); 11 end</pre>
--	--

Fig. 5. Internal signals (events) in Esterel and CÉU.

```
1 event int* inc; // subroutine 'inc'
2 par/or do
3     loop do // definitions are loops
4         var int* p = await inc;
5         *p = *p + 1;
6     end
7 with
8     var int v = 1;
9     await A;
10    emit inc => &v; // call 'inc'
11    _assert(v==2); // after return
12 end
```

Fig. 6. Subroutine `inc` is defined in a loop (lines 3-6), in parallel with the caller (lines 8-11).

the use of internal signals (events) in Esterel and CÉU. In the version in Esterel, on the occurrence of A, B is emitted and they both become active, resulting in the invocation of `f()` and `g()` in no particular order. In the version in CÉU, the occurrence of A makes the program behave as follows (with the stack contents in italics):

- (1) 1st trail awakes (line 5), emits b, and pauses.
stack: [1st]
- (2) 2nd trail awakes (line 9), calls `_f()`, and terminates.
stack: [1st]
- (3) 1st trail (on top of the stack) resumes, calls `_g()`, and terminates.
stack: []
- (4) Both trails have terminated, so the `par/and` rejoins, and the program also terminates;

Internal events bring support for a limited form of subroutines, as depicted in Figure 6. The subroutine `inc` is defined as a loop (lines 3-6) that continuously awaits its identifying event (line 4), incrementing the value passed as reference (line 5). A trail in parallel (lines 8-11) invokes the subroutine in reaction to event A through an `emit` (line 10). Given the stacked execution for internal events, the calling trail pauses, the subroutine awakes (line 4), runs its body (yielding `v=2`), loops, and awaits the next “call” (line 4, again). Only after this sequence that the calling trail resumes and passes the assertion test.

On the one hand, this form of subroutines has a significant limitation that it cannot express recursive calls: an `emit` to itself is always ignored, given that a running body cannot be awaiting itself. On the other hand, this very same limitation brings some important safety properties to subroutines: first, they are guaranteed to react in bounded time; second, memory for locals is also bounded, not requiring data stacks. Also, this form of subroutines can use the other primitives of CÉU, such as parallel compositions and the `await` statement. In particular, they await keeping context information such as locals and the program counter, similarly to coroutines [Moura and Ierusalimsky 2009].

3. FORMAL SEMANTICS

In this section, we introduce a reduced syntax of CÉU and propose an operational semantics to formally describe the language. We describe a small synchronous kernel with broadcast communication highlighting the main differences to Esterel, in particular the stack-based execution for internal events. For the sake of simplicity, we focus on the control aspects of the language, leaving out side effects and *C* calls (which behave like in any conventional imperative language).

3.1. Abstract Syntax

<code>p ::= mem(id)</code>	// primary expressions
<code>await(id)</code>	(any memory access to 'id')
<code>emit(id)</code>	(await event 'id')
<code>break</code>	(emit event 'id')
	(loop escape)
	// compound expressions
<code>if mem(id) then p else p</code>	(conditional)
<code>p ; p</code>	(sequence)
<code>loop p</code>	(repetition)
<code>p and p</code>	(par/and)
<code>p or p</code>	(par/or)
<code>fin p</code>	(finalization)
	// derived by semantic rules
<code>awaiting(id,n)</code>	(awaiting 'id' since sequence number 'n')
<code>emitting(n)</code>	(emitting on stack level 'n')
<code>p @ loop p</code>	(unwinded loop)

Fig. 7. Reduced syntax of CÉU.

Figure 7 shows the BNF-like syntax for a subset of CÉU that is sufficient to describe all semantic peculiarities of the language. Except for *fin* and all semantic-derived expressions (i.e., *awaiting*, *emitting*, and *p @ loop;p*), which are discussed further, all other expressions map to their counterparts in the concrete language.

The *mem(id)* primitive represents all accesses, assignments, and *C* function calls that affect a memory location identified by *id*. As the challenging parts of CÉU reside on its control structures, we are not concerned here with a precise semantics for side effects, but only with their occurrences in programs. The special notation *nop* is used to represent an innocuous *mem* expression (it can be thought as a synonym for *mem(ε)*, where *ε* is an unused identifier). Note that *mem* and *await/emit* expressions do not share identifiers, i.e., an identifier is either a variable or an event.

3.2. Operational Semantics

The core of our semantics is a relation that, given a sequence number *n* identifying the current reaction chain, maps a program *p* and a stack of events *S* in a single step to a modified program and stack:

$$\langle S, p \rangle \xrightarrow[n]{} \langle S', p' \rangle \quad \textbf{(relation-inner)}$$

where

$$\begin{aligned} S, S' \in id^* & \quad (\text{sequence of event identifiers : } [id_{top}, \dots, id_{bottom}]) \\ p, p' \in P & \quad (\text{program as described in Figure 7}) \\ n \in \mathbb{N} & \quad (\text{unique identifier for the reaction chain}) \end{aligned}$$

At the beginning of a reaction chain, the stack is initialized with the occurring external event *ext* (*S* = [*ext*]), but *emit* expressions can push new events on top of it (we

discuss how they are popped further). The ever-increasing sequence number n will prevent that awaiting expressions awake in the same reaction they are reached (ensuring that the program can progress).

We describe this relation with a set of *small-step* structural semantics rules, which are built in such a way that at most one transition is possible at any time, resulting in deterministic reaction chains. The transition rules for the primary expressions are as follows:

$$\begin{aligned}
\langle S, \text{await}(id) \rangle &\xrightarrow[n]{} \langle S, \text{awaiting}(id, n+1) \rangle && \textbf{(awaiting)} \\
\langle id : S, \text{awaiting}(id, m) \rangle &\xrightarrow[n]{} \langle id : S, \text{nop} \rangle, \quad m \leq n && \textbf{(awake)} \\
\langle S, \text{emit}(id) \rangle &\xrightarrow[n]{} \langle id : S, \text{emitting}(|S|) \rangle && \textbf{(emitting)} \\
\langle S, \text{emitting}(|S|) \rangle &\xrightarrow[n]{} \langle S, \text{nop} \rangle && \textbf{(pop)}
\end{aligned}$$

An *await* is simply transformed into an *awaiting* (rule **awaiting**) as an artifact to remember the external sequence number n it can awake. An *awaiting* can only transit to a *nop* (rule **awake**) if its referred event id matches the top of the stack and its sequence number is smaller or equal than the current one ($m \leq n$). An *emit* transits to an *emitting* holding the current stack level ($|S|$ stands for the stack length), and pushing the referred event on the stack (rule **emitting**). With the new stack level $|S| + 1$, the *emitting*($|S|$) itself cannot transit, as rule **pop** expects its parameter to match the current stack level. This trick provides the desired stack-based semantics for internal events.

Proceeding to compound expressions, the rules for conditionals and sequences are straightforward:

$$\begin{aligned}
&\frac{val(id, n) \neq 0}{\langle S, (if \text{ mem}(id) \text{ then } p \text{ else } q) \rangle \xrightarrow[n]{} \langle S, p \rangle} && \textbf{(if-true)} \\
&\frac{val(id, n) = 0}{\langle S, (if \text{ mem}(id) \text{ then } p \text{ else } q) \rangle \xrightarrow[n]{} \langle S, q \rangle} && \textbf{(if-false)} \\
&\frac{\langle S, p \rangle \xrightarrow[n]{} \langle S', p' \rangle}{\langle S, (p ; q) \rangle \xrightarrow[n]{} \langle S', (p' ; q) \rangle} && \textbf{(seq-adv)} \\
&\langle S, (\text{mem}(id) ; q) \rangle \xrightarrow[n]{} \langle S, q \rangle && \textbf{(seq-nop)} \\
&\langle S, (\text{break} ; q) \rangle \xrightarrow[n]{} \langle S, \text{break} \rangle && \textbf{(seq-brk)}
\end{aligned}$$

Given that our semantics focuses on control, rules **if-true** and **if-false** are the only to query *mem* expressions. The “magical” function *val* receives the memory identifier and current reaction sequence number, returning the current memory value. Although the value here is arbitrary, it is unique in a reaction chain, because a given expression can execute only once within it (remember that *loops* must contain *awaits* which, from rule **awaiting**, cannot awake in the same reaction they are reached).

The rules for loops are analogous to sequences, but use ‘@’ as separators to properly bind breaks to their enclosing loops:

$$\begin{aligned}
 \langle S, (loop\ p) \rangle &\xrightarrow[n]{} \langle S, (p\ @\ loop\ p) \rangle && \textbf{(loop-expd)} \\
 \frac{\langle S, p \rangle \xrightarrow[n]{} \langle S', p' \rangle}{\langle S, (p\ @\ loop\ q) \rangle \xrightarrow[n]{} \langle S', (p'\ @\ loop\ q) \rangle} && \textbf{(loop-adv)} \\
 \langle S, (mem(id)\ @\ loop\ p) \rangle &\xrightarrow[n]{} \langle S, loop\ p \rangle && \textbf{(loop-nop)} \\
 \langle S, (break\ @\ loop\ p) \rangle &\xrightarrow[n]{} \langle S, nop \rangle && \textbf{(loop-brk)}
 \end{aligned}$$

When a program encounters a *loop*, it first expands its body in sequence with itself (rule **loop-expd**). Rules **loop-adv** and **loop-nop** are similar to rules **seq-adv** and **seq-nop**, advancing the loop until they reach a *mem(id)*. However, what follows the loop is the loop itself (rule **loop-nop**). Note that if we used ‘;’ as a separator in loops, rules **loop-brk** and **seq-brk** would conflict. Rule **loop-brk** escapes the enclosing loop, transforming everything into a *nop*.

Proceeding to parallel compositions, the semantic rules for *and* and *or* always force transitions on their left branches *p* to occur before their right branches *q*:

$$\begin{aligned}
isBlocked(n, a : S, awaiting(b, m)) &= (a \neq b \vee m = n) \\
isBlocked(n, S, emitting(s)) &= (|S| \neq s) \\
isBlocked(n, S, (p ; q)) &= isBlocked(n, S, p) \\
isBlocked(n, S, (p @ loop q)) &= isBlocked(n, S, p) \\
isBlocked(n, S, (p and q)) &= isBlocked(n, S, p) \wedge isBlocked(n, S, q) \\
isBlocked(n, S, (p or q)) &= isBlocked(n, S, p) \wedge isBlocked(n, S, q) \\
isBlocked(n, S, _) &= false \quad (mem, await, \\
&\quad emit, break, if, loop)
\end{aligned}$$

Fig. 8. The recursive predicate *isBlocked* is true only if all branches in parallel are hanged in *awaiting* or *emitting* expressions that cannot transit.

$$\begin{aligned}
&\frac{\langle S, p \rangle \xrightarrow{n} \langle S', p' \rangle}{\langle S, (p and q) \rangle \xrightarrow{n} \langle S', (p' and q) \rangle} \quad \textbf{(and-adv1)} \\
&\frac{isBlocked(n, S, p) , \quad \langle S, q \rangle \xrightarrow{n} \langle S', q' \rangle}{\langle S, (p and q) \rangle \xrightarrow{n} \langle S', (p and q') \rangle} \quad \textbf{(and-adv2)} \\
&\frac{\langle S, p \rangle \xrightarrow{n} \langle S', p' \rangle}{\langle S, (p or q) \rangle \xrightarrow{n} \langle S', (p' or q) \rangle} \quad \textbf{(or-adv1)} \\
&\frac{isBlocked(n, S, p) , \quad \langle S, q \rangle \xrightarrow{n} \langle S', q' \rangle}{\langle S, (p or q) \rangle \xrightarrow{n} \langle S', (p or q') \rangle} \quad \textbf{(or-adv2)}
\end{aligned}$$

The deterministic behavior of the semantics relies on the *isBlocked* predicate, defined in Figure 8 and used in rules **and-adv2** and **or-adv2**, requiring the left branch *p* to be blocked in order to allow the right transition from *q* to *q'*. Basically, the *isBlocked* predicate determines that an expression becomes blocked when all of its trails in parallel hang in *awaiting* and *emitting* expressions.

For a parallel *and*, if one of the sides terminates, the composition is simply substituted by the other side (rules **and-nop1** and **and-nop2**). The last two rules **and-brk1** and **and-brk2** deal with a *break* in each of the sides in parallel. A *break* should terminate the whole composition in order to escape the innermost loop (*aborting* the other side):

$$\begin{aligned}
\text{clear}(\text{fin } p) &= p \\
\text{clear}(p ; q) &= \text{clear}(p) \\
\text{clear}(p @ \text{loop } q) &= \text{clear}(p) \\
\text{clear}(p \text{ and } q) &= \text{clear}(p) ; \text{clear}(q) \\
\text{clear}(p \text{ or } q) &= \text{clear}(p) ; \text{clear}(q) \\
\text{clear}(_) &= \text{mem}(\text{id})
\end{aligned}$$

Fig. 9. The function *clear* extracts *fin* expressions in parallel and put their bodies in sequence.

$$\langle S, (\text{mem}(\text{id}) \text{ and } q) \rangle \xrightarrow{n} \langle S, q \rangle \quad \textbf{(and-nop1)}$$

$$\langle S, (p \text{ and } \text{mem}(\text{id})) \rangle \xrightarrow{n} \langle S, p \rangle \quad \textbf{(and-nop2)}$$

$$\langle S, (\text{break} \text{ and } q) \rangle \xrightarrow{n} \langle S, (\text{clear}(q) ; \text{break}) \rangle \quad \textbf{(and-brk1)}$$

$$\frac{\text{isBlocked}(n, S, p)}{\langle S, (p \text{ and } \text{break}) \rangle \xrightarrow{n} \langle S, (\text{clear}(p) ; \text{break}) \rangle} \quad \textbf{(and-brk2)}$$

The *clear* function, defined in Figure 9, concatenates all active *fin* bodies of the side being aborted (to execute before the *and* rejoins). Note that there are no transition rules for *fin* expressions. This is because once reached, a *fin* expression halts and will only execute when it is aborted by a trail in parallel and is expanded by the *clear* function. In Section 3.3.3, we show how a *fin* is mapped to a finalization block in the concrete language. Note that there is a syntactic restriction that a *fin* body can only contain *mem* expressions, i.e., they are guaranteed to execute entirely within a reaction chain.

For a parallel *or*, if one of the sides terminates, the whole composition terminates and applies the function *clear* to the aborted side in order to properly finalize it (rules **or-nop1** and **or-nop2**). Finally, a *break* (rules **or-brk1** and **or-brk2**) behaves in the same way as in a parallel *and*:

$$\langle S, (mem(id) \text{ or } q) \rangle \xrightarrow[n]{} \langle S, clear(q) \rangle \quad \textbf{(or-nop1)}$$

$$\frac{isBlocked(n, S, p)}{\langle S, (p \text{ or } mem(id)) \rangle \xrightarrow[n]{} \langle S, clear(p) \rangle} \quad \textbf{(or-nop2)}$$

$$\langle S, (break \text{ or } q) \rangle \xrightarrow[n]{} \langle S, (clear(q) ; break) \rangle \quad \textbf{(or-brk1)}$$

$$\frac{isBlocked(n, S, p)}{\langle S, (p \text{ or } break) \rangle \xrightarrow[n]{} \langle S, (clear(p) ; break) \rangle} \quad \textbf{(or-brk2)}$$

A reaction chain eventually blocks in *awaiting* and *emitting* expressions in parallel trails. If all trails hang only in *awaiting* expressions, it means that the program cannot advance in the current reaction chain. However, *emitting* expressions should resume their continuations of previous *emit* in the ongoing reaction, they are just hanged in lower stack indexes (see rule **pop**). Therefore, we define another relation that behaves as **relation-inner** (presented above) if the program is not blocked, and, otherwise, pops the stack:

$$\frac{\langle S, p \rangle \xrightarrow[n]{} \langle S', p' \rangle}{\langle S, p \rangle \xRightarrow[n]{} \langle S', p' \rangle} \quad \frac{isBlocked(n, s : S, p)}{\langle s : S, p \rangle \xRightarrow[n]{} \langle S, p \rangle} \quad \textbf{(relation-outer)}$$

To describe a *reaction chain* in CÉU, i.e., how a program behaves in reaction to a single external event, we use the reflexive transitive closure of **relation-outer**:

$$\langle S, p \rangle \xRightarrow[n]{*} \langle S', p' \rangle$$

Finally, to describe the complete execution of a program, we need multiple “invocations” of reaction chains, incrementing the sequence number:

$$\begin{aligned} \langle [e1], p \rangle &\xRightarrow[1]{*} \langle [], p' \rangle \\ \langle [e2], p' \rangle &\xRightarrow[2]{*} \langle [], p'' \rangle \\ &\dots \end{aligned}$$

Each invocation starts with an external event at the top of the stack and finishes with a modified program and an empty stack. After each invocation, the sequence number is incremented.

3.3. Concrete Language Mapping

Most statements from CÉU map directly to those presented in the reduced syntax of Figure 7, namely, the *if*, *‘;’*, *loop*, *par/and*, and *par/or*. For instance, the *if* in the concrete language behaves exactly like the *if* in the described semantics. However, there are some significant mismatches between CÉU and the formal semantics, which we provide a mapping in this section.

Again, we are not considering side-effects, which are all mapped to the *mem* semantic construct.

<pre> par/or do <...> emit e => 1; with v = await e; _printf("%d\n", v); end </pre>	<pre> par/or do <...> e_ = 1; emit e; with await e; v = e_; _printf("%d\n", v); end </pre>	<pre> <...> ; mem ; emit(e) or await(e) ; mem ; mem </pre>
--	--	--

Fig. 10. Two-step translation from concrete to abstract emit and await expressions.

<pre> dt = await 10ms; </pre>	<pre> var int tot = 10000; // 10ms loop do await TICK; tot = tot - TICK_; if tot <= 0 then dt = tot; break; end end </pre>	<pre> mem; loop(await(TICK); mem; if mem then mem; break else nop) </pre>
-------------------------------	---	---

Fig. 11. Two-step translation from concrete to abstract timer.

3.3.1. await and emit. The concrete `await` and `emit` primitives of CÉU support communication of values between them. In the two-step translation of Figure 10, we start with the concrete program in CÉU, which communicates the value 1 between the `emit` and `await` in parallel (left-most code). In the intermediate translation, we include the shared variable `e_` to hold the value being communicated between the two trails in order to simplify the `emit`. Finally, we convert the program into the equivalent in the formal semantics, translating side-effect statements into *mem* expressions. External events require a similar translation, i.e., each external event has a corresponding variable that is explicitly set by the environment before each reaction chain.

3.3.2. First-class Timers. To encompass first-class timers, we introduce a special external `TICK` event that is intercalated with each other event occurrence in an application (e.g. *e1*, *e2*):

$$\begin{aligned}
\langle [TICK], p \rangle &\xRightarrow[1]{*} \langle [], p' \rangle \\
\langle [e1], p' \rangle &\xRightarrow[2]{*} \langle [], p'' \rangle \\
\langle [TICK], p'' \rangle &\xRightarrow[3]{*} \langle [], p''' \rangle \\
\langle [e2], p''' \rangle &\xRightarrow[4]{*} \langle [], p'''' \rangle \\
&\dots
\end{aligned}$$

The `TICK` event has an associated variable `TICK_` carrying the wall-clock time elapsed between the two occurrences, as depicted by the two-step translation of Figure 11.

TODO: continuar explicacao

3.3.3. Finalization Blocks. The biggest mismatch between CÉU and the formal semantics is regarding finalization blocks, which require more complex modifications in the program for a proper mapping using the *fin* semantic construct. In the three-step translation of Figure 12, we start with a concrete program (CODE-1) that uses a `finalize` to safely `_release` the reference to `ptr` kept after the call to `_hold`. In the translation to

```

do
  var int* ptr = <...>;
  await A;
  finalize
    _hold(ptr);
  with
    _release(ptr);
  end
  await B;
end
// CODE-1

par/or do
  var int* ptr = <...>;
  await A;
  _hold(ptr);
  await B;
  with
    { fin
      _release(ptr); }
  end
end
// CODE-2

f_ = 0;
par/or do
  var int* ptr = <...>;
  await A;
  _hold(ptr);
  f_ = 1;
  await B;
  with
    { fin
      if f_ then
        _release(ptr);
      end }
  end
end
// CODE-3

mem;
(
  mem;
  await (A);
  mem;
  mem;
  await (B);
  or
  fin
  if mem then
    mem
  else
    nop
  end
)
// CODE-4

```

Fig. 12. Three-step translation from concrete to abstract finalization.

the semantics, we first need to catch the do-end termination to run the finalization code. For this, we translate the block into a *par/or* (CODE-2) with the original body in parallel with a *fin* to run the finalization code. This way, the *fin* body executes whenever the *par/or* terminates, either normally (after the *await B*) or aborted from an outer composition. However, the *fin* still (incorrectly) executes even if the call to *_hold* is not reached in the body due to an abort before awaking from the *await A*. To deal with this issue, for each *fin* we need a corresponding flag to keep track of code that needs to be finalized (CODE-3). The flag is initially set to false, avoiding the finalization code to execute. Only after the call to *_hold* that we set the flag to true and enable the *fin* body to execute. The complete translation substitutes the side-effect operations with *mem* expressions (CODE-4).

4. IMPLEMENTATION

The compilation process of a program in CÉU is composed of three main phases, as illustrated in Figure 13:

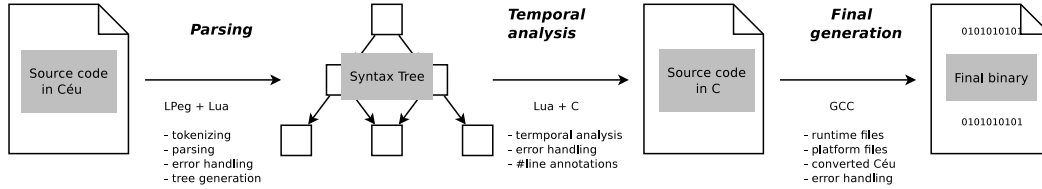


Fig. 13. Compilation process: from the source code in CÉU to the final binary.

Parsing. The parser of CÉU is written in *LPeg* [Ierusalimschy 2009], a pattern matching library that also recognize grammars, making it possible to write the tokenizer and grammar with the same tool. The source code is then converted to an *abstract syntax tree (AST)* to be used in further phases. This phase may be aborted due to syntax errors in the CÉU source file.

Analysis. This phase detects inconsistencies in CÉU programs, such as unbounded loops, and also makes “classical” semantic analysis, such as building a symbol table for checking variable declarations. This phase outputs code in *C*, given the way CÉU is tied to *C* by design. Some type checking is delayed to the last phase to take advantage of *gcc*’s error handling. Therefore, we annotate the *C* output file with *#line* pragmas matching the original file in CÉU.

```

input int A, B, C;
do
  var int a = await A;
end
do
  var int b = await B;
end
par/and do
  await B;
with
  await C;
end

union {
  // sequence
  int a_1; // do_1
  int b_2; // do_2
  struct {
    // par/and
    u8 _and_3: 1;
    u8 _and_4: 1;
  };
} MEM ;

```

Fig. 14. A program with blocks in sequence and in parallel, with corresponding memory layout.

Code Generation. The final phase packs the generated *C* file with the C  U runtime and platform-dependent functionality, compiling them with *gcc* and generating the final binary. The C  U runtime comprehends the scheduler, timer management, and the external *C* API. The platform files include libraries for I/O and bindings to invoke the C  U scheduler on external events.

4.1. Memory Layout

C  U favors a fine-grained use of trails, being common the use of trails that await a single event. For this reason, C  U does not allocate per-trail stacks; instead, all data resides in fixed memory slots—this is true for the program variables as well as for temporary values and flags needed during runtime. Memory for trails in parallel must coexist, while statements in sequence can reuse it. C  U reserves a single static block of memory to hold all memory slots, whose size is the maximum the program uses at a given time. A given position in the memory may hold different data (with variable sizes) during runtime.

Translating this idea to *C* is straightforward [Kasten and R  mer 2005; Bernauer and R  mer 2013]: memory for blocks in sequence are packed in a *struct*, while blocks in parallel, in a *union*. As an example, Figure 14 shows a program with corresponding memory layout. Each variable is assigned a unique *id* (e.g. *a_1*) so that variables with the same name can be distinguished. The *do-end* blocks in sequence are packed in a *union*, given that their variables cannot be in scope at the same time, e.g., *MEM.a_1* and *MEM.b_2* can safely share the same memory slot. The example also illustrates the presence of runtime flags related to the parallel composition, which also reside in reusable slots in the static memory.

4.2. Trail Allocation

Each line of execution in C  U carries associated data, such as which event it is awaiting and which code to execute when it awakes. The compiler statically infers the maximum number of trails a program can have at the same time and creates a static vector to hold the runtime information about them. Trails that cannot be active at the same time can share memory slots in the static vector.

At any given moment, a trail can be awaiting in one of the following states: *INACTIVE*, *STACKED*, *FIN*, or in any of the events defined in the program:

```

enum {
  INACTIVE = 0,
  STACKED,
  FIN,
  EVT_A,    // input void A;
  EVT_e,    // event int e;
  <...>     // other events
}

```

All terminated or not-yet-started trails stay in the `INACTIVE` state and are ignored by the scheduler. A `STACKED` trail holds its associated stack level and is delayed until the scheduler runtime level reaches that value again. A `FIN` trail represents a hanged finalization block which is only scheduled when its corresponding block goes out of scope. A trail waiting for an event stays in the state of the corresponding event, also holding the minimum sequence number (*seqno*) in which it can awake. A trail is represented by the following struct:

```
struct trail_t {
    state_t evt;
    label_t lbl;
    union {
        unsigned char seqno;
        stack_t      stk;
    };
};
```

The field `evt` holds the state of the trail (or the event it is awaiting); the field `lbl` holds the entry point in the code to execute when the trail is scheduled; the third field depends on the `evt` field and may hold the *seqno* for an event, or the stack level `stk` for a `STACKED` state.

The size of `state_t` depends on the number of events in the application; for an application with less than 253 events (plus the 3 states), one byte is enough. The size of `label_t` depends primarily on the number of `await` statements in the application—each `await` splits the code in two and requires a unique entry point in the code for its continuation. Additionally, `split & join` points for parallel compositions, `emit` continuations, and finalization blocks also require labels. The *seqno* will eventually overflow during execution (every 256 reactions). However, given that the scheduler traverses all trails in each reaction, it can adjust them to properly handle overflows (actually 2 bits to hold the *seqno* is already enough). The stack size depends on the maximum depth of nested emissions and is bounded to the maximum number of trails, e.g., a trail emits an event that awakes another trail, which emits an event that awakes another trail, and so on—the last trail cannot awake any trail, because they will be all hanged in a `STACKED` state. In general, in constrained sensor network applications, the size of `trail_t` is typically only 3 bytes (1 byte for each field), imposing a negligible overhead even for trails that only await a single event and terminate.

4.2.1. Code Generation. The example in Figure 15 illustrates how trails and labels are statically allocated in a program. The program has a maximum of 2 trails, because the `par/and` (line 4) can reuse *TRAIL 0*, and the join point (line 16) can reuse both *TRAIL 0* and *TRAIL 1*. Each label is associated with a unique identifier in the `enum`. The static vector to hold the two trails in the example is defined as

```
trail_t TRLS[2];
```

In the final generated *C* code, each label becomes a *switch case* working as the entry point to execute its associated code. Figure ?? shows the corresponding code for the program of Figure 15. The program is initialized with all trails set to `INACTIVE`. Then, the scheduler executes the *Main* label in *TRAIL-0*. When the *Main* label reaches the `par/and`, it “stacks” *TRAIL-1* (line 5–8) and proceeds to the code in *TRAIL-0* (lines 10–15), respecting the deterministic execution order. The code sets the running *TRAIL-0* to await `EVT_e` on label *Awake_e*, and then halts with a `break`. The next iteration of the scheduler takes *TRAIL-1* and executes its registered label *ParAnd_sub_2* (lines 17–22), which sets *TRAIL-1* to await `EVT_A` and also halts.

Regarding cancellation, trails in parallel are always allocated in subsequent slots in the static vector `TRLS`. Therefore, when a `par/or` terminates, the scheduler sequentially searches and executes `FIN` trails within the range of the `par/or`, and then clears all

<pre> input void A; event void e; // TRAIL 0 - lbl Main par/and do // TRAIL 0 - lbl Main await e; // TRAIL 0 - lbl Awake_e // TRAIL 0 - lbl And_chk with // TRAIL 1 - lbl And-sub.2 await A; // TRAIL 1 - lbl Awake_A.1 emit e; // TRAIL 1 - lbl Emit_cont // TRAIL 1 - lbl And_chk end // TRAIL 0 - lbl And_out await A; // TRAIL 0 - lbl Awake_A.2 </pre>	<pre> 1 enum { 2 Main = 1, // ln 3 3 Awake_e, // ln 7 4 And_chk, // ln 8,15 5 And-sub.2, // ln 10 6 Awake_A.1, // ln 12 7 Emit_cont, // ln 14 8 And_out, // ln 17 9 Awake_A.2 // ln 19 10 }; </pre>	<pre> 1 while (<...>) { 2 trail_t* trail = <...> 3 switch (trail->lbl) { 4 case Main: 5 // activate TRAIL 1 6 TRLS[1].evt = STACKED; 7 TRLS[1].lbl = And-sub.2; 8 TRLS[1].stk = cur.stack; 9 10 // code in the 1st trail 11 // await e; 12 TRLS[0].evt = EVT.e; 13 TRLS[0].lbl = Awake_e; 14 TRLS[0].seq = cur.segno; 15 break; 16 17 case And-sub.2: 18 // await A; 19 TRLS[1].evt = EVT.A; 20 TRLS[1].lbl = Awake_A.1; 21 TRLS[1].seq = cur.segno; 22 break; 23 24 <...> // other labels 25 } 26 } </pre>
---	---	--

Fig. 15. Static allocation of trails, entry-point labels, and generated code.

of them to INACTIVE at once. Given that finalization blocks cannot contain await statements, the whole process is guaranteed to terminate in bounded time. Escaping a loop that contains parallel compositions also trigger the same finalization process.

TODO internal emit?

4.3. The External C API

As a reactive language, the execution of a program in CÉU is guided entirely by the occurrence of external events. From the implementation perspective, there are three external sources of input into programs, which are all exposed as functions in a C API:

- ceu_go_init()*:. initializes the program (e.g. trails) and executes the “boot” reaction (i.e., the Main label).
- ceu_go_event(id,param)*:. executes the reaction for the received event id and associated parameter.
- ceu_go_wclock(us)*:. increments the current time in microseconds and runs a reaction if any timer expires.

Given the semantics of CÉU, the functions are guaranteed to take a bounded time to execute. They also return a status code that says if the CÉU program has terminated after the reactions. Further calls to the API have no effect on terminated programs.

The bindings for the specific platforms are responsible for calling the functions in the API in the order that better suit their requirements. As an example, it is possible to set different priorities for events that occur concurrently (i.e. while a reaction chain is running). However, a binding must never interleave or run multiple functions in parallel. This would break the CÉU sequential/discrete semantics of time.

As an example, Figure 16 shows our binding for TinyOS which maps *nesC* callbacks to input events in CÉU. The file *ceu.h* (included in line 3) contains all definitions for the compiled CÉU program, which are further queried through *#ifdef*’s. The file *ceu.c* (included in line 4) contains the main loop of CÉU pointing to the labels defined in the program. The callback *Boot.booted* (lines 6-11) is called by TinyOS on mote startup,


```

1 implementation
2 {
3     #include "ceu.h"
4     #include "ceu.c"
5
6     event void Boot.booted () {
7         ceu_go_init();
8 #ifdef CEU_WCLOCKS
9         call Timer.startPeriodic(10);
10    #endif
11    }
12
13 #ifdef CEU_WCLOCKS
14     event void Timer.fired () {
15         ceu_go_wclock(10000);
16     }
17 #endif
18
19 #ifdef _EVT_PHOTO_READDONE
20     event void Photo.readDone (uint16_t val) {
21         ceu_go_event(EVT_PHOTO_READDONE, (void*)val);
22     }
23 #endif
24
25 #ifdef _EVT_RADIO_SENDDONE
26     event void RadioSend.sendDone (message_t* msg) {
27         ceu_go_event(EVT_RADIO_SENDDONE, msg);
28     }
29 #endif
30
31 #ifdef _EVT_RADIO_RECEIVE
32     event message_t* RadioReceive.receive (message_t* msg) {
33         ceu_go_event(EVT_RADIO_RECEIVE, msg);
34         return msg;
35     }
36 #endif
37
38     <...>    // other events
39 }

```

Fig. 16. The *TinyOS* binding for CÉU.

so we initialize CÉU inside it (line 7). If the CÉU program uses timers, we also start a periodic timer (lines 8-10) that triggers callback `Timer.fired` (lines 13-17) every 10 milliseconds and advances the wall-clock time of CÉU (line 15)⁴. The remaining lines map pre-defined *TinyOS* events that can be used in CÉU programs, such as the light sensor (lines 19-23) and the radio transceiver (lines 25-36).

5. RELATED WORK

- Kiel Esterel compiler

- We are not concerned with WCET: Predictable system design is concerned with the challenge of building systems in such a way that requirements can be guaranteed from the design. This means that an off-line analysis should demonstrate satisfaction of timing requirements, subject to assumptions made on operating conditions foreseen for the system [Stankovic and Ramamritham 1990]

PRET-C <http://embedded.cs.uni-saarland.de/publications/BuildingTimingPredictableEmbeddedSystems.pdf>

⁴We also offer a mechanism to start the underlying timer on demand to avoid the “battery unfriendly” 10ms polling.

6. FINAL REMARKS

REFERENCES

- Sidharta Andalam and others. 2010. Predictable multithreading of embedded applications using PRET-C. In *Proceeding of MEMOCODE'10*. IEEE, 159–168.
- Alexander Bernauer and Kay Römer. 2013. A Comprehensive Compiler-Assisted Thread Abstraction for Resource-Constrained Systems. In *Proceedings of IPSN'13*. Philadelphia, USA.
- G rard Berry. 1993. Preemption in Concurrent Systems.. In *FSTTCS (LNCS)*, Vol. 761. Springer, 72–93.
- G. Berry. 2000. *The Esterel-V5 Language Primer*. CMA and Inria, Sophia-Antipolis, France. Version 5.10, Release 2.0.
- F. Boussinot and R. de Simone. 1991. The Esterel language. *Proc. IEEE* 79, 9 (Sep 1991), 1293–1304.
- Omprakash Gnawali and others. 2009. Collection tree protocol. In *Proceedings of SenSys'09*. ACM, 1–14.
- Roberto Ierusalimsky. 2009. A text pattern-matching tool based on Parsing Expression Grammars. *Softw. Pract. Exper.* 39 (March 2009), 221–258. Issue 3.
- Marcin Karpinski and Vinny Cahill. 2007. High-Level Application Development is Realistic for Wireless Sensor Networks. In *Proceedings of SECON'07*. 610–619.
- Oliver Kasten and Kay R mer. 2005. Beyond Event Handlers: Programming Wireless Sensors with Attributed State Machines. In *Proceedings of IPSN '05*. 45–52.
- Ana L cia De Moura and Roberto Ierusalimsky. 2009. Revisiting coroutines. *ACM TOPLAS* 31, 2 (Feb. 2009), 6:1–6:31.
- ORACLE. 2011. Java Thread Primitive Deprecation. <http://docs.oracle.com/javase/6/docs/technotes/guides/concurrency/threadPrimitiveDeprecation.html> (accessed in Aug-2014). (2011).
- Dumitru Potop-Butucaru and others. 2005. The Synchronous Hypothesis and Synchronous Languages. In *Embedded Systems Handbook*, R. Zurawski (Ed.).
- Francisco Sant'Anna and others. 2013. Safe System-level Concurrency on Resource-Constrained Nodes. In *Proceedings of SenSys'13*. ACM.
- Francisco Sant'Anna and others. 2015. Structured Synchronous Reactive Programming with C u. In *Proceedings of Modularity'15*. to appear.