

Towards a verified compiler for Distributed PlusCal

Presented at WPTE '25 on July 20, 2025

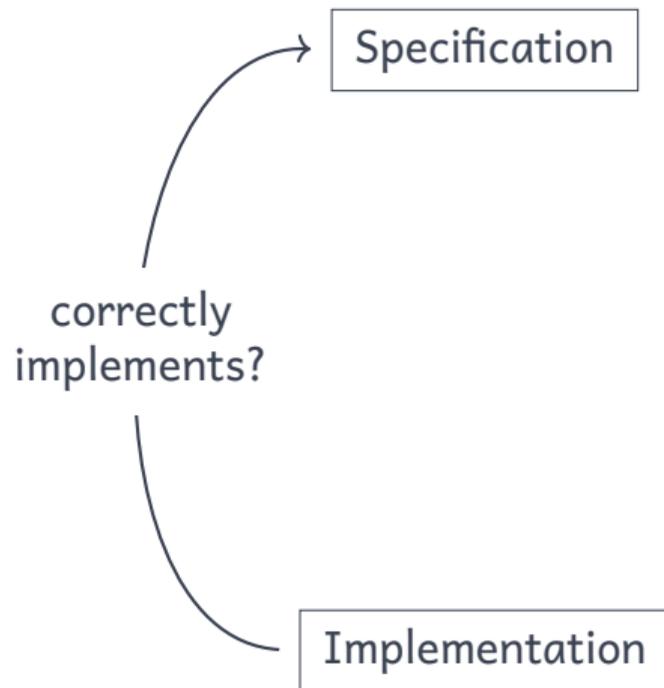
By Ghilain BERGERON, Horatiu CIRSTEA, Stephan MERZ
(Université de Lorraine, CNRS, Inria, LORIA)

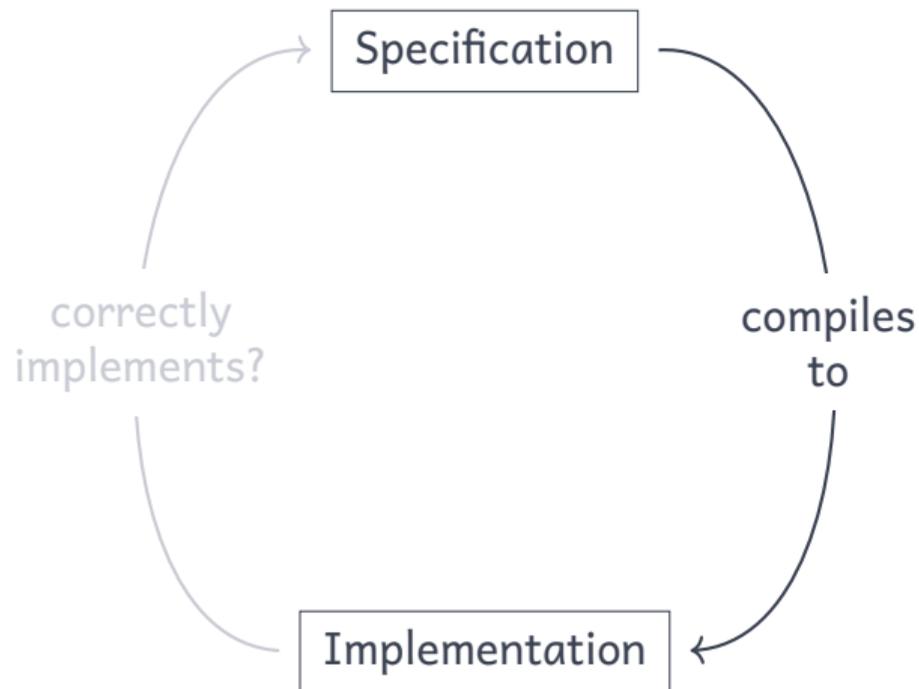


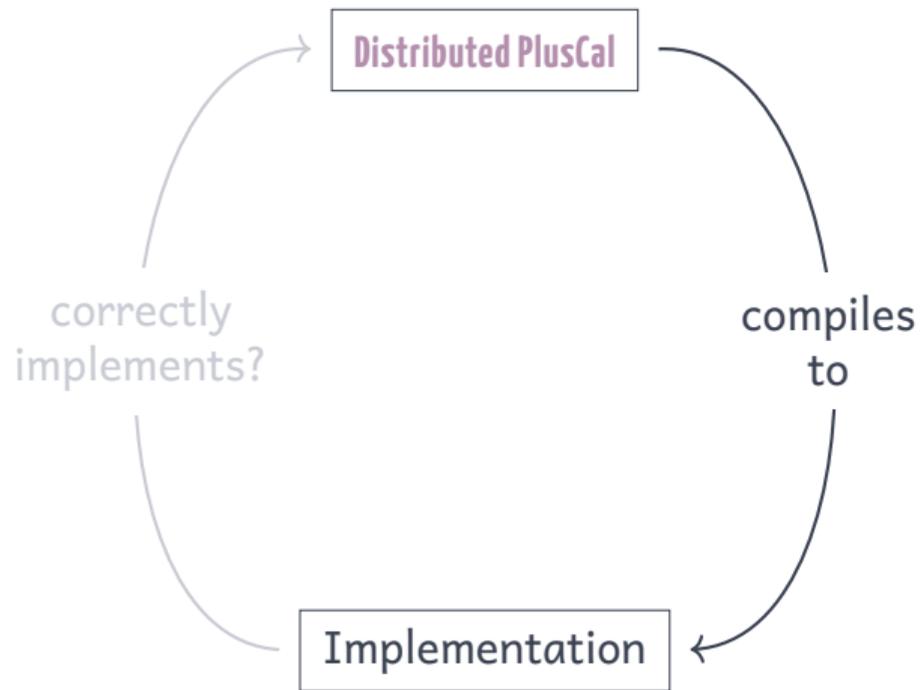
Implementation

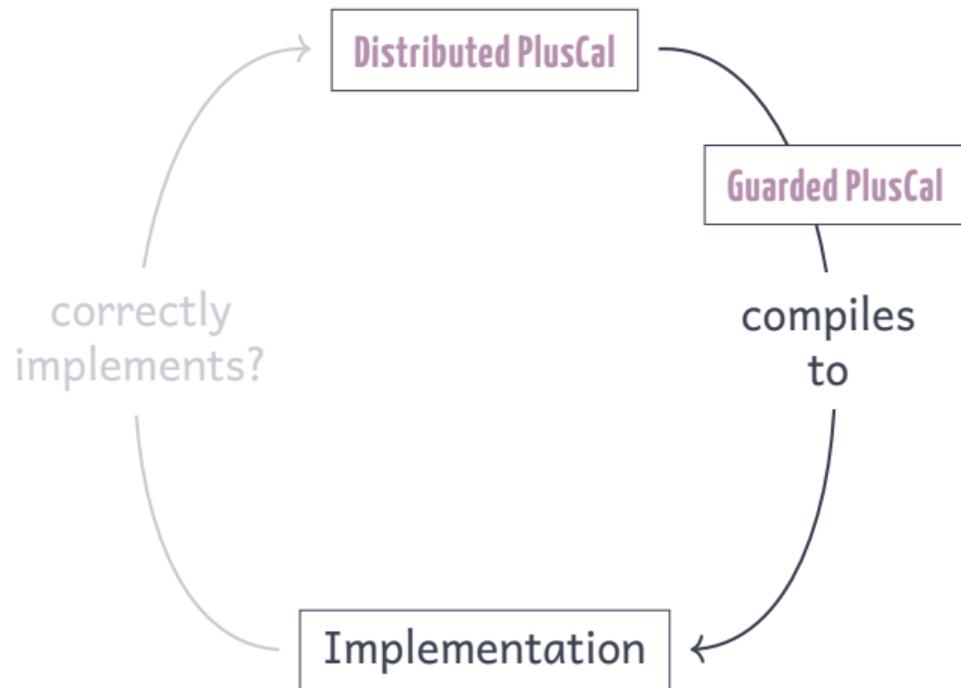
Specification

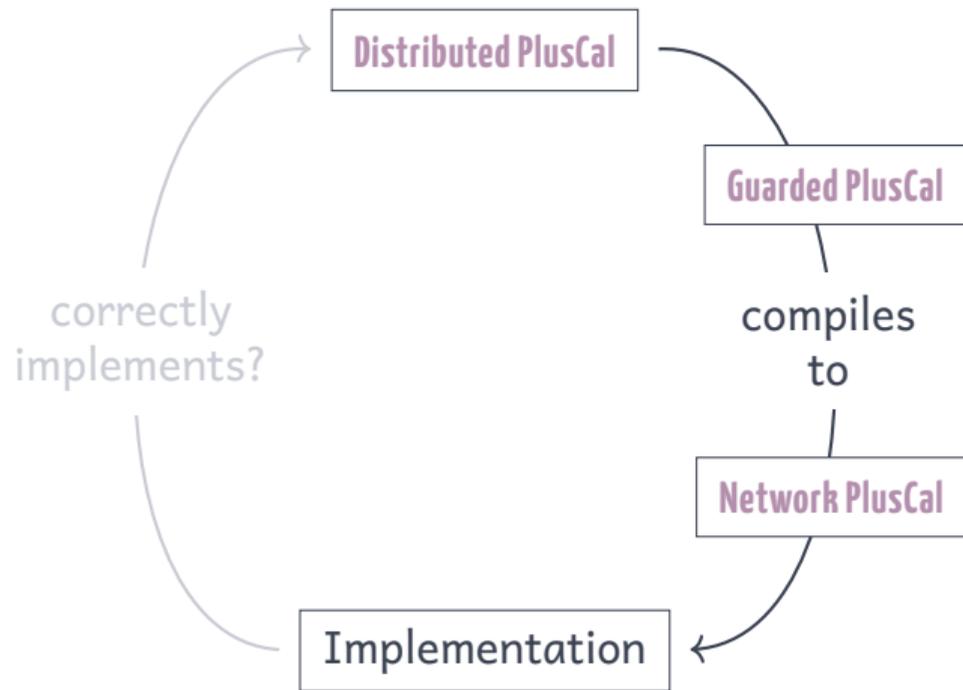
Implementation

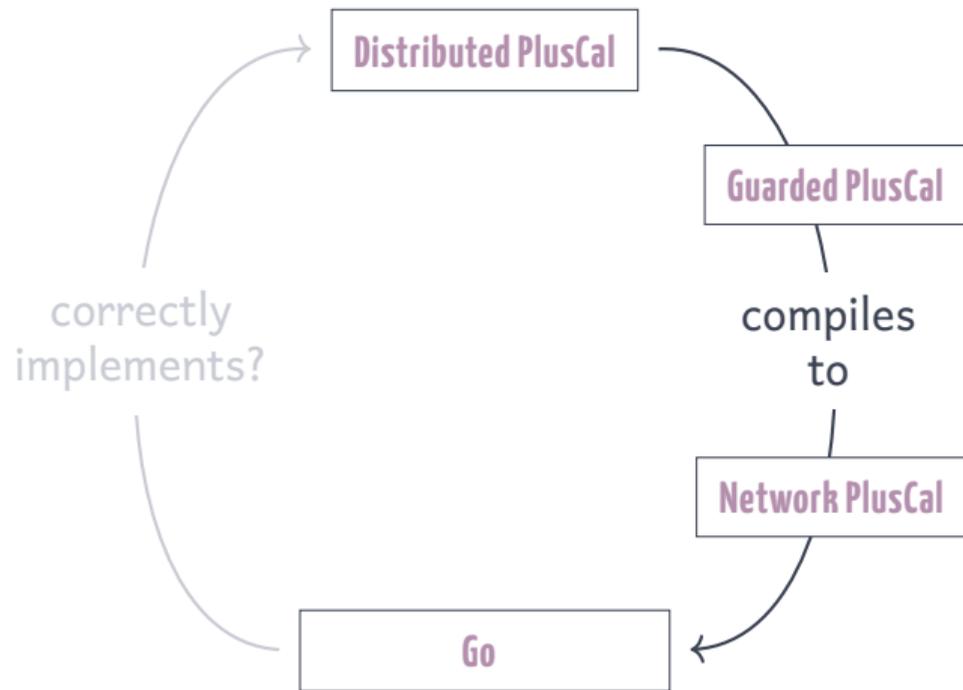












A specification language for distributed/concurrent algorithms...

A specification language for distributed/concurrent algorithms...

- * ...translated to a TLA⁺ specification.

A specification language for distributed/concurrent algorithms...

- * ...translated to a TLA⁺ specification.
- * ...featuring message-passing and threading primitives.

A specification language for distributed/concurrent algorithms...

- * ...translated to a TLA⁺ specification.
- * ...featuring message-passing and threading primitives.
- * ...that looks like the C language.

```
algorithm PingPong {
  fifos ping, pong;

  process (Ping = 0)
    variable tmp1 = "";
  {
    rcv1:  receive(ping, tmp1);
          await tmp1 = "Ping";
          goto pong;
    pong:  send(pong, "Pong");
          goto rcv1;
  }
```

```
process (Pong = 1)
  variable tmp2 = "";
{
  ping:  send(ping, "Ping");
        goto rcv2;
  rcv2:  receive(pong, tmp2);
        await tmp2 = "Pong";
        goto ping;
}
```

```

algorithm PingPong {
  fifos ping, pong;

  process (Ping = 0)
    variable tmp1 = "";
  {
  rcv1:  receive(ping, tmp1);
        await tmp1 = "Ping";
        goto pong;
  pong:  send(pong, "Pong");
        goto rcv1;
  }
}

```

```

process (Pong = 1)
  variable tmp2 = "";
{
  ping:  send(ping, "Ping");
        goto rcv2;
  rcv2:  receive(pong, tmp2);
        await tmp2 = "Pong";
        goto ping;
}
}

```



```

algorithm PingPong {
  fifos ping, pong;

  process (Ping = 0)
    variable tmp1 = "";
  {
  rcv1:  receive(ping, tmp1);
        await tmp1 = "Ping";
        goto pong;
  pong:  send(pong, "Pong");
        goto rcv1;
  }
}

```

```

process (Pong = 1)
  variable tmp2 = "";
{
  ping:  send(ping, "Ping");
        goto rcv2;
  rcv2:  receive(pong, tmp2);
        await tmp2 = "Pong";
        goto ping;
}
}

```



```

algorithm PingPong {
  fifos ping, pong;

  process (Ping = 0)
    variable tmp1 = "";
  {
  rcv1:  receive(ping, tmp1);
        await tmp1 = "Ping";
        goto pong;
  pong:  send(pong, "Pong");
        goto rcv1;
  }
}

```

```

process (Pong = 1)
  variable tmp2 = "";
{
  ping:  send(ping, "Ping");
        goto rcv2;
  rcv2:  receive(pong, tmp2);
        await tmp2 = "Pong";
        goto ping;
}
}

```



```

algorithm PingPong {
  fifos ping, pong;

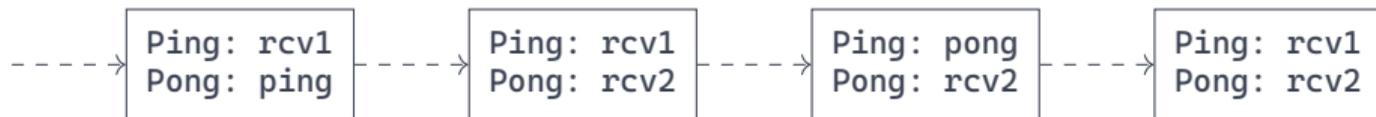
  process (Ping = 0)
    variable tmp1 = "";
  {
  rcv1:  receive(ping, tmp1);
        await tmp1 = "Ping";
        goto pong;
  pong:  send(pong, "Pong");
        goto rcv1;
  }
}

```

```

process (Pong = 1)
  variable tmp2 = "";
{
  ping:  send(ping, "Ping");
        goto rcv2;
  rcv2:  receive(pong, tmp2);
        await tmp2 = "Pong";
        goto ping;
}
}

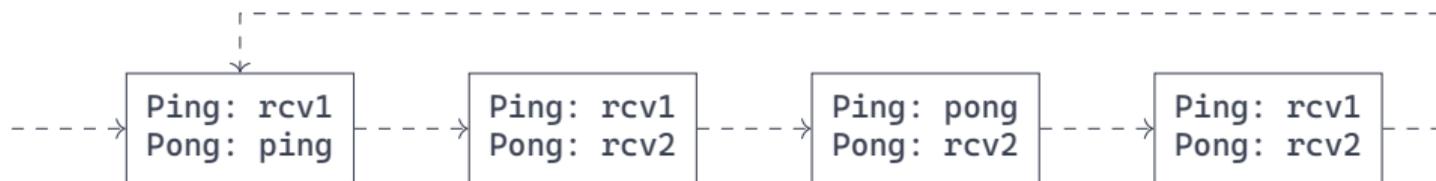
```



```
algorithm PingPong {
  fifos ping, pong;
```

```
  process (Ping = 0)
    variable tmp1 = "";
  {
  rcv1:  receive(ping, tmp1);
        await tmp1 = "Ping";
        goto pong;
  pong:  send(pong, "Pong");
        goto rcv1;
  }
```

```
  process (Pong = 1)
    variable tmp2 = "";
  {
  ping:  send(ping, "Ping");
        goto rcv2;
  rcv2:  receive(pong, tmp2);
        await tmp2 = "Pong";
        goto ping;
  }
```



Distributed PlusCal is syntactically close to a programming language.

Distributed PlusCal is syntactically close to a programming language. Yet its semantics is very cumbersome to preserve:

Distributed PlusCal is syntactically close to a programming language. Yet its semantics is very cumbersome to preserve:

```
l: x := f();  
   receive(c, y);  
   print y;  
   await y > 8;  
   goto k;
```

Distributed PlusCal is syntactically close to a programming language. Yet its semantics is very cumbersome to preserve:

```
l: x := f();  
   receive(c, y);  
   print y;  
   await y > 8;  
   goto k;
```

 Receiving from a channel means consuming a message!

Distributed PlusCal is syntactically close to a programming language. Yet its semantics is very cumbersome to preserve:

```
l: x := f();  
   receive(c, y);  
   print y;  
   await y > 8;  
   goto k;
```

- ⚠ Receiving from a channel means consuming a message!
- ⚠ No printing when $y \leq 8$!

Distributed PlusCal is syntactically close to a programming language. Yet its semantics is very cumbersome to preserve:

```
l: x := f();  
   receive(c, y);  
   print y;  
   await y > 8;  
   goto k;
```

- ⚠ Receiving from a channel means consuming a message!
- ⚠ No printing when $y \leq 8$!
- ⚠ No rollback mechanisms please!

- * Each process may only receive messages from a single (globally unique) channel (called “mailbox”).

- * Each process may only receive messages from a single (globally unique) channel (called “mailbox”).
- * Each atomic block must start with all its `receive` and `await` statements.

- * Each process may only receive messages from a single (globally unique) channel (called “mailbox”).
- * Each atomic block must start with all its `receive` and `await` statements.
 - ✓ No more printing if the conditions are not satisfied.

- * Each process may only receive messages from a single (globally unique) channel (called “mailbox”).
- * Each atomic block must start with all its `receive` and `await` statements.
 - ✓ No more printing if the conditions are not satisfied.
 - ✓ No need to rollback (most) assignments.

Guarded PlusCal does not avoid all rollbacks.

Guarded PlusCal does not avoid all rollbacks. Consider this example again:

```
l: receive(c, y);  
   await y > 8;  
   x := f();  
   print y;  
   goto k;
```

Guarded PlusCal does not avoid all rollbacks. Consider this example again:

```
l: receive(c, y);  
   await y > 8;  
   x := f();  
   print y;  
   goto k;
```

What if the first message in the channel *c* is 3?

Guarded PlusCal does not avoid all rollbacks. Consider this example again:

```
l: receive(c, y);  
   await y > 8;  
   x := f();  
   print y;  
   goto k;
```

What if the first message in the channel c is $\neq 10$?

Guarded PlusCal but **receive** statements are replaced with local message buffering.

Guarded PlusCal but **receive** statements are replaced with local message buffering.

✓ No more mandatory consuming!

Guarded PlusCal but **receive** statements are replaced with local message buffering.

- ✓ No more mandatory consuming!
- ✓ No need for rollback mechanisms anymore!

1. Compile `receive` into buffer indexing.

1. Compile `receive` into buffer indexing.

```
process (Ping = 0)
  variable tmp1 = "";
{
rcv1:  receive(ping, tmp1);
      await tmp1 = "Ping";
      goto pong;
pong:  send(pong, "Pong");
      goto rcv1;
}
```

1. Compile `receive` into buffer indexing.

```
process (Ping = 0)
  variable tmp1 = "";
{
rcv1:  receive(ping, tmp1);
      await tmp1 = "Ping";
      goto pong;
pong:  send(pong, "Pong");
      goto rcv1;
}
```

1. Compile `receive` into buffer indexing.

```

process (Ping = 0)
  variable tmp1 = "";
{
rcv1: receive(ping, tmp1);
      await tmp1 = "Ping";
      goto pong;
pong: send(pong, "Pong");
      goto rcv1;
}

```

 $\Rightarrow \mathcal{E}$

```

process (Ping = 0)
  variables tmp1 = "", inbox = <<>>;
{
rcv1: await Len(inbox) > 0;
      tmp1 := Head(inbox);
      inbox := Tail(inbox);
      await tmp1 = "Ping";
      goto pong;
pong: send(pong, "Pong");
      goto rcv1;
}
{
rx†: receive(ping, rcv†);
      inbox := Append(inbox, rcv†);
      goto rx†;
}

```

1. Compile `receive` into buffer indexing.

```

process (Ping = 0)
  variable tmp1 = "";
{
rcv1: receive(ping, tmp1);
      await tmp1 = "Ping";
      goto pong;
pong: send(pong, "Pong");
      goto rcv1;
}

```

 $\Rightarrow \mathcal{E}$ $\Rightarrow \mathcal{M}$

```

process (Ping = 0)
  variables tmp1 = "", inbox = <<>>;
{
rcv1: await Len(inbox) > 0;
      tmp1 := Head(inbox);
      inbox := Tail(inbox);
      await tmp1 = "Ping";
      goto pong;
pong: send(pong, "Pong");
      goto rcv1;
}
{
rx†: receive(ping, rcv†);
     inbox := Append(inbox, rcv†);
     goto rx†;
}

```

2. Reorder statements to conform to Network PlusCal syntactical restrictions.

2. Reorder statements to conform to Network PlusCal syntactical restrictions.

```
process (Ping = 0)
  variables tmp1 = "", inbox = <<>>;
{
rcv1:  await Len(inbox) > 0;
      tmp1 := Head(inbox);
      inbox := Tail(inbox);
      await tmp1 = "Ping";
      goto pong;
pong:  send(pong, "Pong");
      goto rcv1;
}
{
rx†: receive(ping, rcv†);
      inbox := Append(inbox, rcv†);
      goto rx†;
}
```

2. Reorder statements to conform to Network PlusCal syntactical restrictions.

```
process (Ping = 0)
  variables tmp1 = "", inbox = <<>>;
{
rcv1:  await Len(inbox) > 0;
      tmp1 := Head(inbox);
      inbox := Tail(inbox);
      await tmp1 = "Ping";
      goto pong;
pong:  send(pong, "Pong");
      goto rcv1;
}
{
rx†: receive(ping, rcv†);
      inbox := Append(inbox, rcv†);
      goto rx†;
}
```

2. Reorder statements to conform to Network PlusCal syntactical restrictions.

```

process (Ping = 0)
  variables tmp1 = "", inbox = <<>>;
{
rcv1:  await Len(inbox) > 0;
      tmp1 := Head(inbox);
      inbox := Tail(inbox);
      await tmp1 = "Ping";
      goto pong;
pong:  send(pong, "Pong");
      goto rcv1;
}
{
rx†: receive(ping, rcv†);
      inbox := Append(inbox, rcv†);
      goto rx†;
}

```

 $\Rightarrow \mathcal{S}$

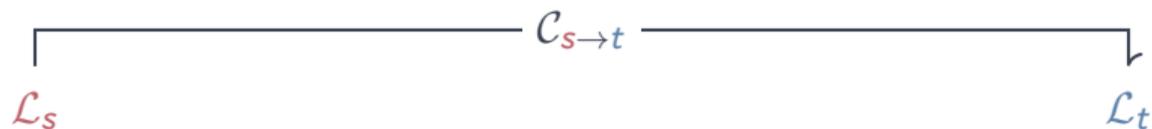
```

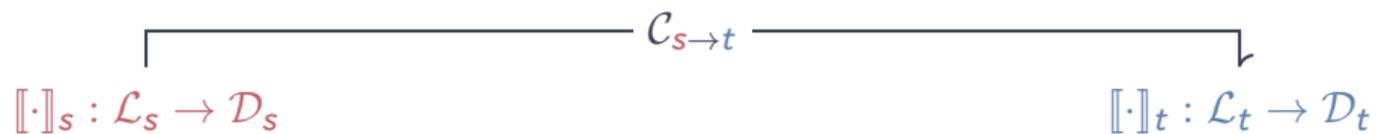
process (Ping = 0)
  variables tmp1 = "", inbox = <<>>;
{
rcv1:  await Len(inbox) > 0;
      await Head(inbox) = "Ping";
      tmp1 := Head(inbox);
      inbox := Tail(inbox);
      goto pong;
pong:  send(pong, "Pong");
      goto rcv1;
}
{
rx†: receive(ping, rcv†);
      inbox := Append(inbox, rcv†);
      goto rx†;
}

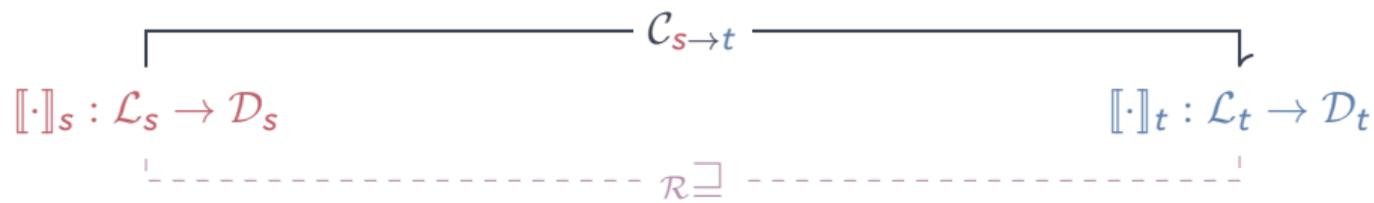
```

\mathcal{L}_s

\mathcal{L}_t







i Definition (Behavior justification)

For $\mathcal{R} \subseteq \mathcal{D}_s \times \mathcal{D}_t$, a behavior $\sigma_t \in [[P_t]]_t$ is \mathcal{R} -justified by a behavior $\sigma_s \in [[P_s]]_s$ if $(\sigma_s, \sigma_t) \in \mathcal{R}$.

i Definition (Partial correctness)

$\mathcal{C}_{s \rightarrow t}$ preserves the semantics of \mathcal{L}_s if for all programs $P \in \mathcal{L}_s$, every (terminating) behavior of $[[\mathcal{C}_{s \rightarrow t}(P)]]_t$ is \mathcal{R} -justified by some behavior of $[[P]]_s$.

Define $\llbracket \cdot \rrbracket : \text{PlusCal} \rightarrow \text{State} \times \text{Trace} \times \text{State}$

Define $\llbracket \cdot \rrbracket : \text{PlusCal} \rightarrow \text{State} \times \text{Trace} \times \text{State}$ where

* *State* is a tuple *Memory* \times *Channels*

Define $\llbracket \cdot \rrbracket : \text{PlusCal} \rightarrow \text{State} \times \text{Trace} \times \text{State}$ where

- * *State* is a tuple *Memory* \times *Channels*
- * *Trace* is a sequence of events $\text{out}(v)$ or $c!v$

Define $\llbracket \cdot \rrbracket : \text{PlusCal} \rightarrow \text{State} \times \text{Trace} \times \text{State}$ where

- * *State* is a tuple *Memory* \times *Channels*
- * *Trace* is a sequence of events $\text{out}(v)$ or $c!v$

$$\frac{M \vdash e \Downarrow v \quad \tau = \text{out}(v)}{\langle \langle M, C \rangle, \tau, \langle M, C \rangle \rangle \in \llbracket \text{print } e \rrbracket}$$

$$\frac{M \vdash e \Downarrow \text{TRUE} \quad \tau = \varepsilon}{\langle \langle M, C \rangle, \tau, \langle M, C \rangle \rangle \in \llbracket \text{await } e \rrbracket}$$

$$\frac{\langle \langle M, C \rangle, \tau', \langle M'', C'' \rangle \rangle \in \llbracket S \rrbracket \quad \langle \langle M'', C'' \rangle, \tau'', \langle M', C' \rangle \rangle \in \llbracket B \rrbracket \quad \tau = \tau' * \tau''}{\langle \langle M, C \rangle, \tau, \langle M', C' \rangle \rangle \in \llbracket S; B \rrbracket}$$

$$\frac{C(c) = v \bullet vs \quad C' = C(c := vs) \quad M' = M(x := v) \quad \tau = \varepsilon}{\langle \langle M, C \rangle, \tau, \langle M', C' \rangle \rangle \in \llbracket \text{receive}(c, x) \rrbracket}$$

i Definition ($\sim \subseteq \text{State} \times \text{State}$)

$$\langle M_G, C_G \rangle \sim \langle M_N, C_N \rangle \stackrel{\text{def}}{\iff} \bigwedge \begin{cases} \forall v \neq \text{inbox}. M_G(v) = M_N(v) \\ \forall c \neq \text{mailbox}. C_G(c) = C_N(c) \\ C_G(\text{mailbox}) = M_N(\text{inbox}) ++ C_N(\text{mailbox}) \end{cases}$$

i Definition ($\sim \subseteq \text{State} \times \text{State}$)

$$\langle M_G, C_G \rangle \sim \langle M_N, C_N \rangle \stackrel{\text{def}}{\iff} \bigwedge \begin{cases} \forall v \neq \text{inbox}. M_G(v) = M_N(v) \\ \forall c \neq \text{mailbox}. C_G(c) = C_N(c) \\ C_G(\text{mailbox}) = M_N(\text{inbox}) ++ C_N(\text{mailbox}) \end{cases}$$

i Theorem (Partial correctness of $\mathcal{C}_{G \rightarrow N}$)

$$\forall P \in \text{Guarded PlusCal}. \llbracket \mathcal{C}_{G \rightarrow N}(P) \rrbracket \sqsubseteq \sim \llbracket P \rrbracket$$

i Definition ($\sim \subseteq \text{State} \times \text{State}$)

$$\langle M_G, C_G \rangle \sim \langle M_N, C_N \rangle \stackrel{\text{def}}{\iff} \bigwedge \begin{cases} \forall v \neq \text{inbox}. M_G(v) = M_N(v) \\ \forall c \neq \text{mailbox}. C_G(c) = C_N(c) \\ C_G(\text{mailbox}) = M_N(\text{inbox}) ++ C_N(\text{mailbox}) \end{cases}$$

i Theorem (Partial correctness of $\mathcal{C}_{G \rightarrow N}$)

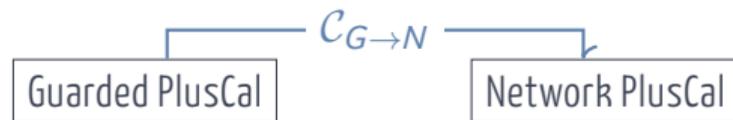
$$\forall P \in \text{Guarded PlusCal}. \llbracket \mathcal{C}_{G \rightarrow N}(P) \rrbracket \sqsubseteq \sim \llbracket P \rrbracket$$

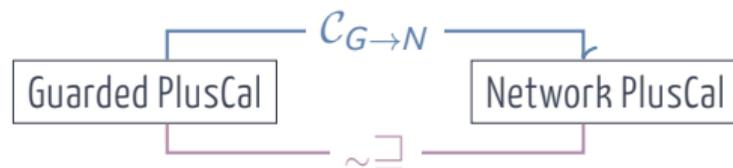
➔ Proof idea:

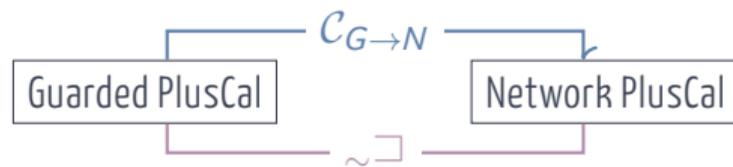
Prove partial correctness of \mathcal{E} , \mathcal{S} and \mathcal{M} separately, then compose.

Guarded PlusCal

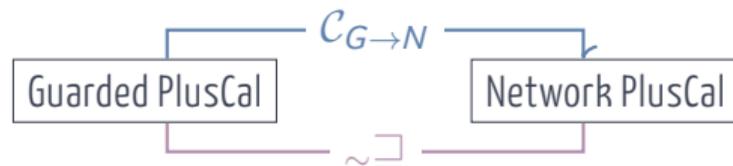
Network PlusCal





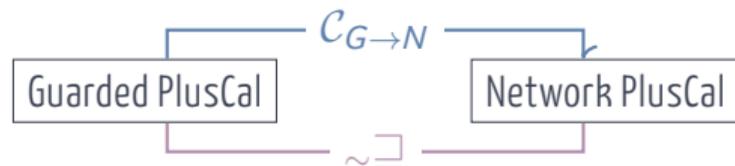


Fully mechanized/developed in the LEAN proof assistant/programming language!



Fully mechanized/developed in the LEAN proof assistant/programming language!

* Totalling **~11k** LoC (**~8k** LoC of proofs)

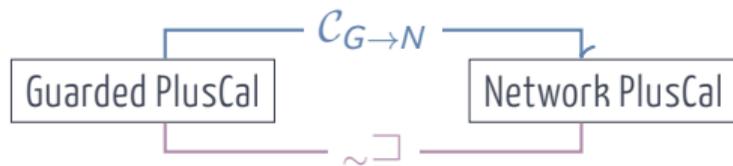


Fully mechanized/developed in the LEAN proof assistant/programming language!

- * Totalling **~11k** LoC (**~8k** LoC of proofs)
- * **~400** lemmas & theorems

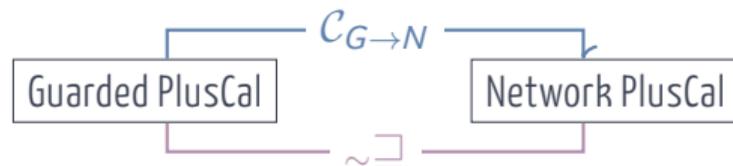
What next?

This is only the beginning!



What next?

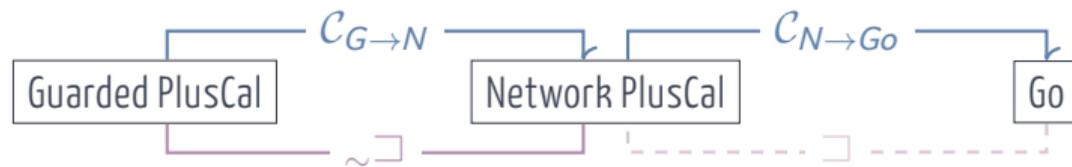
This is only the beginning!



Some challenges remain to be (fully) solved:

What next?

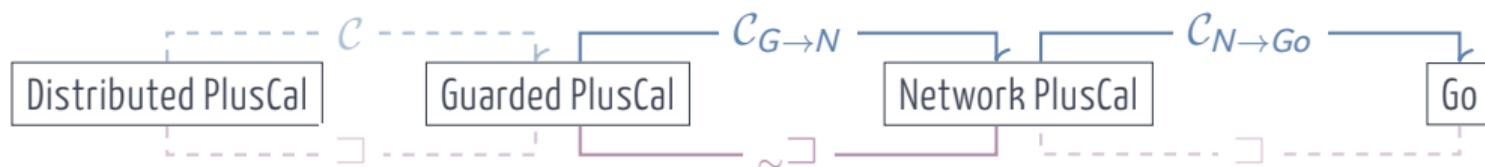
This is only the beginning!



Some challenges remain to be (fully) solved:

- * Correctness of $C_{N \rightarrow Go}$?

This is only the beginning!



Some challenges remain to be (fully) solved:

- * Correctness of $\mathcal{C}_{N \rightarrow Go}$?
- * How to relax the restrictions of Guarded PlusCal to improve on ergonomics?

Appendix

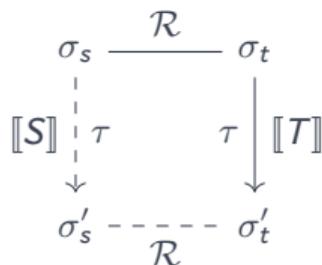


* $[[\cdot]] \subseteq \text{State} \times \text{Trace} \times \text{State}$: terminating (error-free) behaviors

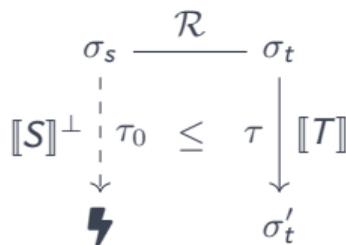
- * $\llbracket \cdot \rrbracket \subseteq \text{State} \times \text{Trace} \times \text{State}$: terminating (error-free) behaviors
- * $\llbracket \cdot \rrbracket^{\perp} \subseteq \text{State} \times \text{Trace}$: erroneous behaviors

- * $\llbracket \cdot \rrbracket \subseteq \mathit{State} \times \mathit{Trace} \times \mathit{State}$: terminating (error-free) behaviors
- * $\llbracket \cdot \rrbracket^{\perp} \subseteq \mathit{State} \times \mathit{Trace}$: erroneous behaviors
- * $|\cdot| \in \mathit{State} \rightarrow \mathbb{N}$: some measure on states

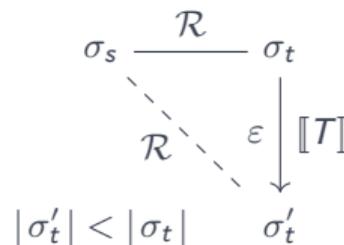
- * $[[\cdot]] \subseteq \text{State} \times \text{Trace} \times \text{State}$: terminating (error-free) behaviors
- * $[[\cdot]]^{\perp} \subseteq \text{State} \times \text{Trace}$: erroneous behaviors
- * $|\cdot| \in \text{State} \rightarrow \mathbb{N}$: some measure on states



or

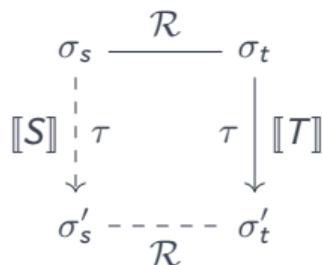


or

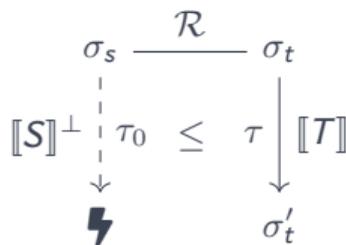


Behavior refinement $\sqsubseteq_{\mathcal{R}}$

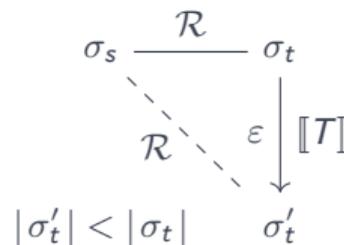
- * $\llbracket \cdot \rrbracket \subseteq \text{State} \times \text{Trace} \times \text{State}$: terminating (error-free) behaviors
- * $\llbracket \cdot \rrbracket^{\perp} \subseteq \text{State} \times \text{Trace}$: erroneous behaviors
- * $|\cdot| \in \text{State} \rightarrow \mathbb{N}$: some measure on states



or



or



- * Writing specifications in Guarded PlusCal is cumbersome.

- * Writing specifications in Guarded PlusCal is cumbersome.
- * We would like the full expressive power of loops and non-deterministic branching as in Distributed PlusCal.

- * Writing specifications in Guarded PlusCal is cumbersome.
- * We would like the full expressive power of loops and non-deterministic branching as in Distributed PlusCal.
- * Sometimes it is more convenient to write assignments before `awaits`.

Relaxing syntactical restrictions of Guarded PlusCal

```
* if (e) { B } else { B' }  $\Rightarrow$   
  either { await e; B } or { await \lnot e; B' }
```

Relaxing syntactical restrictions of Guarded PlusCal

- * `if (e) { B } else { B' } \Rightarrow
 either { await e; B } or { await \lnot e; B' }`
- * `l: while (e) { B }; B' \Rightarrow l: if (e) { B; goto l } else { B' }`

Relaxing syntactical restrictions of Guarded PlusCal

- * `if (e) { B } else { B' } \Rightarrow either { await e; B } or { await \lnot e; B' }`
- * `l: while (e) { B }; B' \Rightarrow l: if (e) { B; goto l } else { B' }`
- * `x := e; await e' \Rightarrow await e'[x\e]; x := e`

Relaxing syntactical restrictions of Guarded PlusCal

- * `if (e) { B } else { B' } \Rightarrow either { await e; B } or { await \lnot e; B' }`
- * `l: while (e) { B }; B' \Rightarrow l: if (e) { B; goto l } else { B' }`
- * `x := e; await e' \Rightarrow await e'[x\e]; x := e`
- * `x := e; receive(c, x) \Rightarrow receive(c, x)`

Relaxing syntactical restrictions of Guarded PlusCal

- * `if (e) { B } else { B' } \Rightarrow either { await e; B } or { await \lnot e; B' }`
- * `l: while (e) { B }; B' \Rightarrow l: if (e) { B; goto l } else { B' }`
- * `x := e; await e' \Rightarrow await e'[x\e]; x := e`
- * `x := e; receive(c, x) \Rightarrow receive(c, x)`
- * `x := e; receive(c, y) \Rightarrow receive(c, y); x := e`

Relaxing syntactical restrictions of Guarded PlusCal

- * `if (e) { B } else { B' } \Rightarrow either { await e; B } or { await \lnot e; B' }`
- * `l: while (e) { B }; B' \Rightarrow l: if (e) { B; goto l } else { B' }`
- * `x := e; await e' \Rightarrow await e'[x\e]; x := e`
- * `x := e; receive(c, x) \Rightarrow receive(c, x)`
- * `x := e; receive(c, y) \Rightarrow receive(c, y); x := e`
- * `print e; await e' \Rightarrow await e'; print e`

Relaxing syntactical restrictions of Guarded PlusCal

- * `if (e) { B } else { B' } \Rightarrow either { await e; B } or { await \neg e; B' }`
- * `l: while (e) { B }; B' \Rightarrow l: if (e) { B; goto l } else { B' }`
- * `x := e; await e' \Rightarrow await e'[x\e]; x := e`
- * `x := e; receive(c, x) \Rightarrow receive(c, x)`
- * `x := e; receive(c, y) \Rightarrow receive(c, y); x := e`
- * `print e; await e' \Rightarrow await e'; print e`
- * `either ... or { B; either B1 or ... or Bn; B' } or ... \Rightarrow either ... or { B; B1; B' } or ... or { B; Bn; B' } or ...`

Relaxing syntactical restrictions of Guarded PlusCal

- * `if (e) { B } else { B' } \Rightarrow either { await e; B } or { await \neg e; B' }`
- * `l: while (e) { B }; B' \Rightarrow l: if (e) { B; goto l } else { B' }`
- * `x := e; await e' \Rightarrow await e'[x\e]; x := e`
- * `x := e; receive(c, x) \Rightarrow receive(c, x)`
- * `x := e; receive(c, y) \Rightarrow receive(c, y); x := e`
- * `print e; await e' \Rightarrow await e'; print e`
- * `either ... or { B; either B1 or ... or Bn; B' } or ... \Rightarrow either ... or { B; B1; B' } or ... or { B; Bn; B' } or ...`