Mechanized Relational Verification of Concurrent Programs with Continuations

AMIN TIMANY, imec-Distrinet, KU Leuven, Belgium
LARS BIRKEDAL, Aarhus University, Denmark

Concurrent higher-order imperative programming languages with continuations are very flexible and allow for the implementation of sophisticated programming patterns. For instance, it is well known that continuations can be used to implement cooperative concurrency. Continuations can also simplify web server implementations. This, in particular, helps simplify keeping track of the state of server's clients. However, such advanced programming languages are very challenging to reason about. One of the main challenges in reasoning about programs in the presence of continuations is due to the fact that the non-local flow of control breaks the *bind* rule, one of the important modular reasoning principles of Hoare logic.

In this paper we present the first completely formalized tool for interactive mechanized relational verification of programs written in a concurrent higher-order imperative programming language with continuations (call/cc and throw). We develop novel logical relations which can be used to give mechanized proofs of relational properties. In particular, we prove correctness of an implementation of cooperative concurrency with continuations. In addition, we show that that a rudimentary web server implemented using the continuation-based pattern is contextually equivalent to one implemented without the continuation-based pattern. We introduce context-local reasoning principles for our calculus which allows us to regain modular reasoning principles for the fragment of the language without non-local control flow. These novel reasoning principles can be used in tandem with our (non-context-local) Hoare logic for reasoning about programs that do feature non-local control flow. Indeed, we use the combination of context-local and non-context-local reasoning to simplify reasoning about the examples.

CCS Concepts: • **Theory of computation → Logic and verification**; **Hoare logic**; **Separation logic**; **Program specifications**; **Program verification**; *Invariants*; *Pre- and post-conditions*; • **Software and its engineering → Formal software verification**; *Semantics*.

Additional Key Words and Phrases: Logical relations, Continuations, Concurrency

ACM Reference Format:
Amin Timany and Lars Birkedal. 2019. Mechanized Relational Verification of Concurrent Programs with Continuations. *Proc. ACM Program. Lang.* 3, ICFP, Article 105 (August 2019), 28 pages. https://doi.org/10.1145/3341709

1 INTRODUCTION

In a programming language with continuations, a computation can be suspended into a continuation object which can be resumed later. Continuations enable interesting programming patterns. For instance, it is well-known that they can be used to implement cooperative concurrency [Haynes et al. 1984]: switching between threads can be implemented by suspending the running thread, storing the suspension and running another thread. Another notable application of continuations

Authors' addresses: Amin Timany, Department of Computer Science, imec-Distrinet, KU Leuven, Leuven, Belgium, amin.timany@cs.kuleuven.be; Lars Birkedal, Department of Computer Science, Aarhus University, Aarhus, Denmark, birkedal@cs.au.dk.

2475-1421/2019/8-ART105
https://doi.org/10.1145/3341709

is the implementation of continuation-based web servers [Flatt 2017; Krishnamurthi et al. 2007; Queinnec 2004]. Web servers store the state of their communication with each client in order to provide a coherent experience for returning clients. For this purpose continuation-based web servers store the continuation of the server-side program. This helps simplify the web server program because the only thing that the server needs to do in order to serve a returning client appropriately is to resume the continuation corresponding to the last communication with the client.

Both of the aforementioned programming patters involve concurrency, higher-order and imperative aspects of the programming language in combination with continuations in sophisticated and interesting ways. Such expressive and advanced programming languages and programs written in them are known to be challenging to model and reason about. In this paper we present novel techniques for relational reasoning about such programming languages and programs written in them.

Specifically, we develop a new logical relations model for proving contextual refinement of programs written in $\mathsf{F}^{\mu,ref}_{conc,cc}$, a call-by-value programming language featuring concurrency (*conc*), impredicative polymorphism (F), recursive types (μ), dynamically allocated higher-order store (*ref*) and first-class continuations (*cc*) with call/cc and throw primitives. We employ this logical relations model to prove (1) contextual equivalence of two simple web server implementations: a continuation-based one and a state-storing one; and (2) that one-shot continuations (continuations that can only be used once) can be used to simulate ordinary continuations [Friedman and Haynes 1985]. The latter is a well-known result for sequential programs, see, *e.g.*, Dreyer et al. [2012]. Here we show that it also holds in $\mathsf{F}^{\mu,ref}_{conc,cc}$, that is, in the presence of concurrency.

In addition, we develop a relational model for showing the correctness of a continuation-based implementation of cooperative concurrency. We consider two programming languages: a language, $\mathsf{F}^{\mu,ref}_{cc,coop}$, with built-in cooperative concurrency (*coop*) and a sequential language, $\mathsf{F}^{\mu,ref}_{cc}$, featuring continuations. We develop a cross-language logical relation between $\mathsf{F}^{\mu,ref}_{cc}$ and $\mathsf{F}^{\mu,ref}_{cc,coop}$ and use it to show correctness of a translation of cooperative concurrency into one based on continuations.

We define our logical relations models in a variant of the Iris program logic framework [Jung et al. 2016; JUNG et al. 2018; Jung et al. 2015; Krebbers et al. 2017a]. Iris is a framework for state-of-the-art higher-order concurrent separation logics. We use Iris because (1) it allows us to define our logical relations and reason about them at a higher level of abstraction compared to an explicit model construction; (2) we side-step the well-known type-world-circularity problems [Ahmed 2004; Ahmed et al. 2002; Birkedal et al. 2011] involved in defining logical relations for programming languages with higher-order store (since that is already "taken care of" by the model of Iris); and (3) we can leverage the Coq implementation of the Iris base logic [Krebbers et al. 2017a] and the Iris Proof Mode [Krebbers et al. 2017b] when mechanizing our development in Coq. Indeed, accompanying this paper is a tool for mechanized relational verification of concurrent programs with continuations. The mechanization has been done in Coq and all the results in the paper have been formally verified.

Context-local reasoning principles

Some of the most important features of concurrent separation logics for *modular*/*local* reasoning about concurrent imperative programs, *e.g.* da Rocha Pinto et al. [2014]; Dinsdale-Young et al. [2013, 2010]; Jung et al. [2016, 2015]; Krebbers et al. [2017a,b]; Ley-Wild and Nanevski [2013]; Nanevski et al. [2014]; O'Hearn [2007]; Sergey et al. [2015]; Svendsen and Birkedal [2014]; Turon et al. [2013a] are *thread-local* reasoning and *context-local* reasoning. Thread-local reasoning means that we can reason about each thread in isolation: when we reason about a particular thread, we need not

explicitly consider interactions from other concurrently executing threads. Similarly, context-local means that when we reason about a particular expression, we need not consider under which evaluation context it is being evaluated. The latter is sometimes codified by the soundness of a proof rule such as the following:

$$\textsc{Hoare-Bind (inadmissible in presence of continuations)}\quad \frac{\{P\}\, e\, \{\Psi\} \qquad \forall w.\ \{\Psi(w)\}\, K[w]\, \{\Phi\}}{\{P\}\, K[e]\, \{\Phi\}}$$

The Hoare-triple $\{P\}\, e\, \{\Psi\}$ intuitively means that, given precondition P, expression e is *safe* and, whenever it reduces to a value v, we are guaranteed that $\Psi(v)$ holds. Intuitively, the above rule expresses that to prove a Hoare triple for an expression e in an evaluation context K, it suffices to prove a property for e in isolation from K, and then show that the desired postcondition Φ can be obtained when substituting a value w satisfying the postcondition Ψ for e into the evaluation context. In a programming language with non-local control operators, *e.g.* call/cc and throw, the context under which a program is being evaluated is of utmost importance, and thus the above proof rule is *not* sound in general for languages with non-local control operators. In spite of this, we ought be able to reason about those parts of the program that do not feature any non-local control flow operators in a context-local way. We explain this via an example which we use to demonstrate how non-local control flow operators can break context-local reasoning. We then discuss how our context-local program logic allows us to regain context-local reasoning in a sound way, and, how we can employ our context-local program logic to derive a context-local specification for the this example.

To illustrate how non-local control flow operators can break context-local reasoning, consider the following program:

$$\mathit{CallIncr} \triangleq \lambda f.\ \mathsf{let}\, x = \mathsf{ref}(0)\, \mathsf{in}\, \mathsf{let}\, g = f\ ()\, \mathsf{in}\, x \leftarrow\, !\, x + 2; g\ (); !\, x$$

(In Section 2 we present the syntax and semantics of $\mathsf{F}^{\mu,ref}_{conc,cc}$ formally; here we simply explain informally what is intended by our ML-like syntax above.) The program *CallIncr* takes a function f as argument, allocates a local reference with value 0, calls f, and binds its result (again a function) to g. It then increments the internal reference and subsequently calls g before returning the value stored in the internal reference. Intuitively, in a programming language without non-local control flow this program should always return 2 (if it terminates, of course). Hence, using the bind rule we would be able to derive the following (incorrect) specification for *CallIncr*:

$$\{\mathsf{true}\}\, h\, ()\, \{x.\ \{\mathsf{true}\}\, x\, ()\, \{y.\ y = ()\}\} \text{ implies } \{\mathsf{true}\}\, \mathit{CallIncr}\, h\, \{x.\ x = 2\} \qquad (\mathit{CallIncr}\text{-incorrect-spec})$$

This specification says that, if h is a function that upon call returns a function x for which we know the Hoare-triple $\{\mathsf{true}\}\, x\, ()\, \{y.\ y = ()\}$ holds, then *CallIncr* applied to h should return 2. This specification *CallIncr*-incorrect-spec does not hold though. For a concrete counter-example, let h to be

$$h = \lambda\_.\ \mathsf{call/cc}\, (x.\ \ \lambda\_.\ \mathsf{throw}\, (\lambda.\ ())\, \mathsf{to}\, x)$$

This function, when called, captures the current continuation x and then returns a function that returns to the point x, this time, with the result $(\lambda.\ ())$, a function that simply returns the unit value $()$. Hence, when *CallIncr* is called with h, the internal reference of *CallIncr* wil be incremented twice, and thus return value be 4 and not 2. Notice that for h above we can prove the spec $\{\mathsf{true}\}\, h\, ()\, \{x.\ \{\mathsf{true}\}\, x\, ()\, \{y.\ y = ()\}\}$ since this spec considers h in isolation, *i.e.*, the captured continuation will be the empty continuation.

Hence, as expected, the program *CallIncr* only behaves context-locally if the argument it is applied to does. This is indeed reflected in the context-local specification of *CallIncr* expressed in terms of *context-local Hoare triples* (Specification *CallIncr*-context-local-spec below). We introduce context-local Hoare triples in Section 4. A context-local Hoare triple $\{P\}^{\mathsf{cl}}\, e\, \{x.\ Q\}$ not only implies safety of e (as ordinary Hoare triples do), but also implies that e behaves context-locally. Hence, they can be used to reason about programs context-locally, and thus the bind rule holds for them. In particular, we can prove the following context-local specification for the program *CallIncr* above:

$$\{\mathsf{true}\}^{\mathsf{cl}}\, h\, ()\, \{x.\ \{\mathsf{true}\}^{\mathsf{cl}}\, x\, ()\, \{y.\ y = ()\}\}\ \text{implies}\ \{\mathsf{true}\}^{\mathsf{cl}}\, \mathit{CallIncr}\, h\, \{x.\ x = 2\} \qquad (\mathit{CallIncr}\text{-context-local-spec})$$

This specification captures exactly the intuitive idea explained above: if we know that h, and the function it returns, behave context-locally, then so does *CallIncr* h.

Since context-local Hoare-triples cannot be used to reason about all programs, particularly about non-trivial uses of call/cc and throw, we have to express our logical relations model using ordinary (non-context-local) Hoare-triples, following ideas from earlier work [Krebbers et al. 2017b; Turon et al. 2013a]. In Section 4, we discuss how ordinary (non-context-local) Hoare-triples interact with their context-local counter parts. Using a combination of non-context-local and context-local triples, we can simplify reasoning about contextual equivalence of concurrent programs with continuations using our logical relations model. We achieve this by reasoning about parts of programs that do not use non-local control operators using context-local Hoare-triples.

Contributions. In this paper, we make the following contributions:

- We present a program logic (weakest preconditions and Hoare-triples) for reasoning about programs written in $\mathsf{F}^{\mu,ref}_{conc,cc}$, a programming language with impredicative polymorphism, recursive types, higher-order functions, higher-order store, concurrency and first-class continuations.
- We present context-local weakest-preconditions and Hoare-triples which simplify reasoning about programs without non-local control flow.
- We present a novel logical relations model for $\mathsf{F}^{\mu,ref}_{conc,cc}$.
- We use our logical relations model and context-local reasoning to prove equivalence of two simple web server implementations: a continuation-based one and a state-storing one.
- We further use our logical relations model to prove correctness of Friedman and Haynes [1985] encoding of continuations by means of one-shot continuations in a concurrent programming language.
- We develop a cross-language logical relations model between $\mathsf{F}^{\mu,ref}_{cc}$ and $\mathsf{F}^{\mu,ref}_{cc,coop}$ for proving program refinement.
- We use our cross-language logical relations model to prove correctness of a continuation-based implementation of cooperative concurrency.
- We have developed a fully formalized tool for mechanized interactive relational verification of concurrent programs with continuations. Our tool is developed on top of Iris, a state-of-the-art program logic framework, and we have used it to mechanize all of our contributions in the Coq proof assistant.

2 THE LANGUAGE: $\mathsf{F}^{\mu,ref}_{conc,cc}$

The language that we consider in this paper, $\mathsf{F}^{\mu,ref}_{conc,cc}$, is a typed lambda calculus with a standard call-by-value small-step operational semantics. It features impredicative polymorphism (F), recursive types (μ), higher-order mutable references (*ref*), fine-grained concurrency (*conc*) and first-class

continuations (*cc*). The types of $\mathsf{F}^{\mu,ref}_{conc,cc}$ are as follows:

$$\tau ::= \alpha \mid \mathbf{1} \mid \mathbb{B} \mid \mathbb{N} \mid \tau \to \tau \mid \forall\alpha.\ \tau \mid \mu\alpha.\ \tau \mid \tau \times \tau \mid \tau + \tau \mid \mathsf{ref}(\tau) \mid \mathsf{cont}(\tau)$$

Here α ranges over type variables. The types $\mathbf{1}$, $\mathbb{B}$ and $\mathbb{N}$ are the unit type, the type of Boolean values and the type of natural numbers respectively. The type $\mathsf{ref}(\tau)$ is the type of references with contents of type τ and $\mathsf{cont}(\tau)$ is the type of continuations that can be resumed by throwing them a value of type τ. The symbols $\times$, $+$, μ, $\forall$ and $\to$ are the usual product, sum, recursive type, polymorphic type and function type formers, respectively.

The syntax for expressions and values is:

$$\begin{aligned} e ::= {} & x \mid () \mid \mathsf{true} \mid \mathsf{false} \mid n \mid e \circledcirc e \mid \mathsf{rec}\, f(x) = e \mid e\ e \mid \Lambda\, e \mid e\ \_ \mid \mathsf{fold}\, e \mid \mathsf{unfold}\, e \mid (e, e) \mid \pi_i\, e \\ & \mid \mathsf{inj}_i\, e \mid \mathsf{match}\, e\, \mathsf{with}\, \mathsf{inj}_i\, x \Rightarrow e_i\, \mathsf{end} \mid \ell \mid \mathsf{ref}(e) \mid\ !\, e \mid e \leftarrow e \mid \mathsf{cas}(e, e, e) \mid \mathsf{fork}\, \{e\} \\ & \mid \mathsf{cont}(K) \mid \mathsf{call/cc}\, (x.\ e) \mid \mathsf{throw}\, e\, \mathsf{to}\, e \\ v ::= {} & () \mid \mathsf{true} \mid \mathsf{false} \mid n \mid \mathsf{rec}\, f(x) = e \mid \Lambda\, e \mid \mathsf{fold}\, v \mid (v, v) \mid \mathsf{inj}_i\, v \mid \ell \mid \mathsf{cont}(K) \end{aligned}$$

We write n for natural numbers and the symbol $\circledcirc$ stands for binary operations on natural numbers (both basic arithmetic operations and basic comparison operations). We consider both recursive functions $\mathsf{rec}\, f(x) = e$ and polymorphic type abstractions $\Lambda\, e$ to be values. We write $e\ \_$ for type level application (e is a polymorphic expression). We use fold and unfold to fold and unfold elements of recursive types. Memory locations ℓ are values of reference types. The expression $!\, e$ reads the memory location e evaluates to, and $e \leftarrow e'$ is an assignment of the value computed by e' to the memory location computed by e. The expression $\mathsf{fork}\, \{e\}$ is for forking off a new thread to compute e and we write $\mathsf{cas}(e, e', e'')$ for the compare-and-set operation. A continuation, $\mathsf{cont}(K)$, is essentially a suspended evaluation context (see the operational semantics below).

Evaluation contexts of $\mathsf{F}^{\mu,ref}_{conc,cc}$ are as follows:

$$\begin{aligned} K ::= {} & [\ ] \mid K \circledcirc e \mid v \circledcirc K \mid K\ e \mid v\ K \mid K\ \_ \mid \mathsf{fold}\, K \mid \mathsf{unfold}\, K \mid \mathsf{if}\, K\, \mathsf{then}\, e\, \mathsf{else}\, e \mid (K, e) \mid (v, K) \\ & \mid \pi_i\, K \mid \mathsf{inj}_i\, K \mid \mathsf{match}\, K\, \mathsf{with}\, \mathsf{inj}_i\, x \Rightarrow e_i\, \mathsf{end} \mid \mathsf{ref}(K) \mid\ !\, K \mid K \leftarrow e \mid v \leftarrow K \\ & \mid \mathsf{cas}(K, e, e) \mid \mathsf{cas}(v, K, e) \mid \mathsf{cas}(v, v, K) \mid \mathsf{throw}\, K\, \mathsf{to}\, e \mid \mathsf{throw}\, v\, \mathsf{to}\, K \end{aligned}$$

The evaluation context $[\ ]$ is the empty evaluation context.

2.1 Typing

An excerpt of the typing rules is depicted in Figure 1. The context $\Xi = \alpha_1, \ldots, \alpha_n$ is a list of distinct type variables and the context $\Gamma = x_1 : \tau_1, \ldots x_n : \tau_n$ assigns types to program variables.

2.2 Operational semantics

We define the call-by-value small-step operational semantics of $\mathsf{F}^{\mu,ref}_{conc,cc}$ in two stages. We first define a head-step relation $\to_K$. Here, K is the context under which the head step is being performed. Based on this, we define the operational semantics of programs by what we call *the thread-pool step relation* $\to$. A thread pool reduces by making a head reduction step in one of the threads, by forking off a new thread, or by resuming a captured continuation in one of the threads. These rules are depicted in Figure 2. In this figure, σ is the physical state of the program, *i.e.*, the program heap, which is a finite partial map from memory locations to values. An excerpt of the head-step relation is given in Figure 3. Notice that the head-step for $\mathsf{call/cc}$ captures the continuation that is the index of the head-step relation.

$\boxed{\Xi \mid \Gamma \vdash e : \tau}$

$$\text{T-Var}\ \frac{x : \tau \in \Gamma}{\Xi \mid \Gamma \vdash x : \tau} \qquad \text{T-Unit}\ \Xi \mid \Gamma \vdash () : \mathbf{1} \qquad \text{T-Nat}\ \Xi \mid \Gamma \vdash n : \mathbb{N} \qquad \text{T-TLam}\ \frac{\Xi, \alpha \mid \Gamma \vdash e : \tau}{\Xi \mid \Gamma \vdash \Lambda\, e : \forall \alpha.\, \tau}$$

$$\text{T-Rec}\ \frac{\Xi \mid \Gamma, x : \tau, f : \tau \to \tau' \vdash e : \tau'}{\Xi \mid \Gamma \vdash \mid \mathsf{rec}\, f(x) = e : \tau \to \tau'} \qquad \text{T-App}\ \frac{\Xi \mid \Gamma \vdash e : \tau \to \tau' \qquad \Xi \mid \Gamma \vdash e' : \tau}{\Xi \mid \Gamma \vdash e\ e' : \tau'} \qquad \text{T-TApp}\ \frac{\Xi \mid \Gamma \vdash e : \forall \alpha.\, \tau}{\Xi \mid \Gamma \vdash e\ \_ : \tau[\tau'/\alpha]}$$

$$\text{T-Fold}\ \frac{\Xi \mid \Gamma \vdash e : \tau[\mu\alpha.\, \tau/\alpha]}{\Xi \mid \Gamma \vdash \mathsf{fold}\, e : \mu\alpha.\, \tau} \qquad \text{T-UnFold}\ \frac{\Xi \mid \Gamma \vdash e : \mu\alpha.\, \tau}{\Xi \mid \Gamma \vdash \mathsf{unfold}\, e : \tau[\mu\alpha.\, \tau/\alpha]} \qquad \text{T-Ref}\ \frac{\Xi \mid \Gamma \vdash e : \tau}{\Xi \mid \Gamma \vdash \mathsf{ref}(e) : \mathsf{ref}(\tau)}$$

$$\text{T-DeRef}\ \frac{\Xi \mid \Gamma \vdash e : \mathsf{ref}(\tau)}{\Xi \mid \Gamma \vdash\, !\, e : \tau} \qquad \text{T-Fork}\ \frac{\Xi \mid \Gamma \vdash e : \tau}{\Xi \mid \Gamma \vdash \mathsf{fork}\, \{e\} : \mathbf{1}} \qquad \text{T-Assign}\ \frac{\Xi \mid \Gamma \vdash e : \mathsf{ref}(\tau) \qquad \Xi \mid \Gamma \vdash e' : \tau}{\Xi \mid \Gamma \vdash e \leftarrow e' : \mathbf{1}}$$

$$\text{T-CAS}\ \frac{\Xi \mid \Gamma \vdash e_1 : \mathsf{ref}(\tau) \qquad \Xi \mid \Gamma \vdash e_2 : \tau \qquad \Xi \mid \Gamma \vdash e_3 : \tau}{\Xi \mid \Gamma \vdash \mathsf{cas}(e_1, e_2, e_3) : \mathbb{B}} \qquad \text{T-Call/cc}\ \frac{\Xi \mid \Gamma, x : \mathsf{cont}(\tau) \vdash e : \tau}{\Xi \mid \Gamma \vdash \mathsf{call/cc}\, (x.\, e) : \tau}$$

$$\text{T-Throw}\ \frac{\Xi \mid \Gamma \vdash e : \tau \qquad \Xi \mid \Gamma \vdash e' : \mathsf{cont}(\tau)}{\Xi \mid \Gamma \vdash \mathsf{throw}\, e\, \mathsf{to}\, e' : \tau'}$$

Fig. 1. An excerpt of the typing rules

$\boxed{(\vec{e}; \sigma) \to (\vec{e}'; \sigma')}$

$$\frac{(e, \sigma) \to_K (e', \sigma')}{(\vec{e}_1, K[e], \vec{e}_2; \sigma) \to (\vec{e}_1, K[e'], \vec{e}_2; \sigma')} \qquad (\vec{e}_1, K[\mathsf{fork}\, \{e\}\,], \vec{e}_2; \sigma) \to (\vec{e}_1, K[()], \vec{e}_2, e; \sigma)$$

$$(\vec{e}_1, K[\mathsf{throw}\, v\, \mathsf{to}\, \mathsf{cont}(K')], \vec{e}_2; \sigma) \to (\vec{e}_1, K'[v], \vec{e}_2; \sigma)$$

Fig. 2. The thread-pool step relation.

Contextual refinement/equivalence. A program e contextually refines a program e' if both programs have type τ and no *well-typed* context (a closed top-level program with a hole) can distinguish a situation where e' is replaced by e. We write $C : (\Xi \mid \Gamma; \tau) \rightsquigarrow (\Xi' \mid \Gamma'; \tau')$ for a context (a term with a hole) such that $\Xi' \mid \Gamma' \vdash C[e] : \tau'$ holds whenever $\Xi \mid \Gamma \vdash e : \tau$ does. We define contextual refinement of e' by e, written $\Xi \mid \Gamma \vdash e \le_{\text{ctx}} e' : \tau$, as follows:

$$\Xi \mid \Gamma \vdash e \le_{\text{ctx}} e' : \tau \triangleq \Xi \mid \Gamma \vdash e : \tau \text{ and } \Xi \mid \Gamma \vdash e' : \tau \text{ and}$$
$$\text{for any } C \text{ such that } C : (\Xi \mid \Gamma; \tau) \rightsquigarrow (\cdot \mid \cdot; \mathbf{1}) \text{ holds } C[e] \Downarrow \text{ implies } C[e'] \Downarrow$$

where $e \Downarrow$ stands for termination of e when run under the empty heap defined as follows:

$$e \Downarrow \triangleq \exists v, \vec{e}, \sigma.\, (e; \emptyset) \to^* (v, \vec{e}; \sigma)$$

$\boxed{(e,\sigma) \to (e',\sigma')}$

$$((\mathsf{rec}\, f(x) = e)\, v, \sigma) \to_K (e[v, (\mathsf{rec}\, f(x) = e)/x, f], \sigma) \qquad (\mathsf{unfold}\,(\mathsf{fold}\, v), \sigma) \to_K (v, \sigma)$$

$$((\Lambda\, e)\, \_, \sigma) \to_K (e, \sigma) \qquad (\mathsf{if\ true\ then}\ e_2\ \mathsf{else}\ e_3, \sigma) \to_K (e_2, \sigma)$$

$$(\mathsf{if\ false\ then}\ e_2\ \mathsf{else}\ e_3, \sigma) \to_K (e_3, \sigma) \qquad (\pi_1\,(v_1, v_2), \sigma) \to_K (v_1, \sigma) \qquad (\pi_2\,(v_1, v_2), \sigma) \to_K (v_2, \sigma)$$

$$(\mathsf{match}\ \mathsf{inj}_1\, v\ \mathsf{with}\ \mathsf{inj}_1\, x \Rightarrow e_1 \mid \mathsf{inj}_2\, x \Rightarrow e_2\ \mathsf{end}, \sigma) \to_K (e_1[v/x], \sigma)$$

$$(\mathsf{match}\ \mathsf{inj}_2\, v\ \mathsf{with}\ \mathsf{inj}_1\, x \Rightarrow e_1 \mid \mathsf{inj}_2\, x \Rightarrow e_2\ \mathsf{end}, \sigma) \to_K (e_2[v/x], \sigma)$$

$$\frac{\ell \notin \mathrm{dom}(\sigma)}{(\mathsf{ref}(v), \sigma) \to_K (\ell, \sigma \uplus \{\ell \mapsto v\})} \qquad \frac{\sigma = \sigma' \uplus \{\ell \mapsto v'\}}{(\ell \leftarrow v, \sigma) \to_K ((), \sigma' \uplus \{\ell \mapsto v\})} \qquad \frac{v = \sigma(\ell)}{(!\,\ell, \sigma) \to_K (v, \sigma)}$$

$$\frac{\sigma = \sigma' \uplus \{\ell \mapsto v\}}{(\mathsf{cas}(\ell, v, v'), \sigma) \to_K (\mathsf{true}, \sigma' \uplus \{\ell \mapsto v'\})} \qquad \frac{\sigma = \sigma' \uplus \{\ell \mapsto v''\} \qquad v \neq v''}{(\mathsf{cas}(\ell, v, v'), \sigma) \to_K (\mathsf{false}, \sigma)}$$

$$(\mathsf{call/cc}\,(x.\, e), \sigma) \to_K (e[\mathsf{cont}(K)/x], \sigma)$$

Fig. 3. An excerpt of the head-reduction rules

The intuitive explanation above for contextual refinement is the reason why in a contextual refinement $e \leq_{\mathrm{ctx}} e'$ or in a logical relatedness relation $e \leq_{\log} e'$, usually, the program on the left hand side, e, is referred to as the implementation side and the program on the right hand side, e', is referred to as the specification side. Two programs are contextually equivalent, if each contextually refines the other:

$$\Xi \mid \Gamma \vdash e \approx_{\mathrm{ctx}} e' : \tau \triangleq \Xi \mid \Gamma \vdash e \leq_{\mathrm{ctx}} e' : \tau \wedge \Xi \mid \Gamma \vdash e' \leq_{\mathrm{ctx}} e : \tau$$

3 LOGICAL RELATIONS

It is challenging to construct logical relations for languages with higher-order store because of the so-called type-world circularity [Ahmed 2004; Ahmed et al. 2002; Birkedal et al. 2011]. The logic of Iris is rich enough to allow for a direct inductive specification of the logical relations for programming languages with advanced features such as higher-order references, recursive types, and concurrency [Krebbers et al. 2017b; Krogh-Jespersen et al. 2017; Timany et al. 2018].

3.1 An Iris primer

Iris [Jung et al. 2016; JUNG et al. 2018; Jung et al. 2015; Krebbers et al. 2017a] is a state-of-the-art higher-order concurrent separation logic designed for verification of programs. In Iris one can quantify over the Iris types κ:

$$\kappa ::= \mathbf{1} \mid \kappa \times \kappa \mid \kappa \to \kappa \mid \mathit{Ectx} \mid \mathit{Var} \mid \mathit{Expr} \mid \mathit{Val} \mid \mathbb{N} \mid \mathbb{B} \mid \kappa \xrightarrow{\mathrm{fin}} \kappa \mid \mathsf{finset}(\kappa) \mid \mathit{Monoid} \mid \mathit{Names} \mid \mathit{iProp} \mid \ldots$$

Here *Ectx*, *Var*, *Expr* and *Val* are Iris types for evaluation contexts, variables, expressions and values of $\mathsf{F}^{\mu,\mathit{ref}}_{\mathit{conc},cc}$. Natural numbers, $\mathbb{N}$, and Booleans $\mathbb{B}$ are also included among the base types of Iris. Iris also features partial maps with finite support, $\kappa \xrightarrow{\mathrm{fin}} \kappa$, and finite sets, $\mathsf{finset}(\kappa)$. Resources in Iris are represented using partial commutative monoids, *Monoid*, and instances of resources are

named using so-called ghost-names, *Names*. Finally, and most importantly, there is a type of Iris propositions *iProp*. The grammar for Iris propositions is as follows:

$$P ::= \top \mid \bot \mid P * P \mid P \mathrel{-\!\!*} P \mid P \wedge P \mid P \Rightarrow P \mid P \vee P \mid \forall x : \kappa.\ \Phi(x) \mid \exists x : \kappa.\ \Phi(x)$$
$$\mid \triangleright P \mid \mu r.P \mid \Box P \mid \mathsf{wp}\ e\ \{x.\ P\} \mid \{P\}\ e\ \{x.\ Q\} \mid \Rightarrow\!\!\!\!\Rightarrow P \mid \boxed{P}^{\mathcal{N}} \mid \ldots$$

Here, $\top$, $\bot$, $\wedge$, $\vee$, $\Rightarrow$, $\forall$, $\exists$ are the standard higher-order logic connectives. The predicates Φ are Iris predicates, *i.e.*, terms of type $\kappa \to \mathit{iProp}$.

The connective $*$ is the separating conjunction. Intuitively, $P * Q$ holds if resources can be split into two *disjoint* pieces such that one satisfies P and the other Q. The magic wand connective $P \mathrel{-\!\!*} Q$ is satisfied by resources such that when these resources are combined with some resource satisfying P the resulting resources would satisfy Q.

The $\triangleright$ modality, pronounced "later" is a modality that intuitively corresponds to some abstract form of step-indexing [Appel and McAllester 2001; Appel et al. 2007; Dreyer et al. 2011]. Intuitively, $\triangleright P$ holds if P holds one step into the future. Iris has support for taking fixed points of *guarded* propositions, $\mu r.P$. This fixed point can only be defined if all occurrences of r in P are guarded, *i.e.*, appear under a $\triangleright$ modality. We use guarded fixed points for defining the interpretation of recursive types in $\mathsf{F}^{\mu,ref}_{conc,cc}$. For any proposition P we have $P \vdash \triangleright P$.

When the modality $\Box$ is applied to a proposition P, the non-duplicable resources in P are forgotten, and thus $\Box P$ is "persistent." In general, we say that a proposition P is *persistent* if $P \dashv\vdash \Box P$ (where $\dashv\vdash$ is the logical equivalence of Iris propositions). A key property of persistent propositions is that they are duplicable: $P \dashv\vdash P * P$. The type system of $\mathsf{F}^{\mu,ref}_{conc,cc}$ is not a sub-structural type system and variables (in the typing environment) may be used multiple times. Therefore when we interpret types as logical relations in Iris, those relations should be duplicable. We use the persistence modality $\Box$ to ensure this.

The Iris program logic for $\mathsf{F}^{\mu,ref}_{conc,cc}$. Iris facilitates specification and verification of programs by means of weakest-preconditions $\mathsf{wp}\ e\ \{x.\ P\}$, which intuitively hold whenever e is *safe* and, moreover, whenever e terminates with a resulting value v, then $P[v/x]$ holds. When x does not appear in P we write $\mathsf{wp}\ e\ \{x.\ P\}$ as $\mathsf{wp}\ e\ \{P\}$. Also, we sometimes write $\mathsf{wp}\ e\ \{\Phi\}$ for $\mathsf{wp}\ e\ \{x.\ \Phi(x)\}$

In Iris, Hoare triples are defined in terms of weakest preconditions:

$$\{P\}\ e\ \{x.\ Q\} \triangleq \Box\left(P \mathrel{-\!\!*} \mathsf{wp}\ e\ \{x.\ Q\}\right)$$

Note that the $\Box$ modality ensures that the Hoare triples are persistent and hence duplicable (in separation logic jargon, Hoare triples should just express "knowledge" and not claim ownership of any resources).

A key feature of Iris (as for other concurrency logics) is that specification and verification is done *thread-locally*: the weakest precondition only describes properties of execution of a single thread. Concurrent interactions are abstracted and reasoned about in terms of resources (rather than by explicit reasoning about interleavings). For programming languages that do not include continuations or other forms of non-local control flow, the weakest precondition is not only thread-local, but also what we may call *context-local*. Context-local means that to reason about an expression in an evaluation context, it suffices to reason about the expression in isolation, and then separately about what the context does to the resulting value. This form of context-locality is formally expressed by the soundness of the following bind rule

$$\frac{\mathsf{wp}\ e\ \{x.\ \mathsf{wp}\ K[x]\ \{\Phi\}\}}{\mathsf{wp}\ K[e]\ \{\Phi\}}\ \textsc{inadmissible-bind}$$

This rule is not sound when expressions include call/cc since call/cc captures the evaluation context and hence its behaviour depends on it. We discussed inadmissibility of this rule in Introduction. Thus, for reasoning about $\mathsf{F}^{\mu,ref}_{conc,cc}$ we cannot use the "standard" Iris rules [Jung et al. 2016; JUNG et al. 2018; Jung et al. 2015; Krebbers et al. 2017a,b] for weakest preconditions. Instead, we use new rules such as the following:

$$\textsc{fst-wp}\ \frac{\triangleright\, \mathsf{wp}\, K[v]\, \{\Phi\}}{\mathsf{wp}\, K[\pi_1\,(v,w)]\, \{\Phi\}} \qquad \textsc{if-true-wp}\ \frac{\triangleright\, \mathsf{wp}\, K[e]\, \{\Phi\}}{\mathsf{wp}\, K[\mathsf{if\ true\ then}\ e\ \mathsf{else}\ e']\, \{\Phi\}} \qquad \textsc{rec-wp}\ \frac{\triangleright\, \mathsf{wp}\, K[e[\mathsf{rec}\, f(x) = e, v/f, x]]\, \{\Phi\}}{\mathsf{wp}\, K[(\mathsf{rec}\, f(x) = e)\, v]\, \{\Phi\}}$$

$$\textsc{callcc-wp}\ \frac{\triangleright\, \mathsf{wp}\, K[e[\mathsf{cont}(K)/x]]\, \{\Phi\}}{\mathsf{wp}\, K[\mathsf{call/cc}\,(x.\, e)]\, \{\Phi\}} \qquad \textsc{throw-wp}\ \frac{\triangleright\, \mathsf{wp}\, K'[v]\, \{\Phi\}}{\mathsf{wp}\, K[\mathsf{throw}\, v\, \mathsf{to}\, \mathsf{cont}(K')]\, \{\Phi\}}$$

The difference from the standard rules is that our new rules include an explicit context K. Earlier, such rules could be derived using the bind rule, but that is not sound in general in our settings. Note that the context is used in the rules CALLCC-WP and THROW-WP for call/cc and throw. These two rules directly reflect the operational semantics of call/cc and throw.

To demonstrate how the rules of our program logic are used to reason about a program we discuss the following program:

$$\mathsf{call/cc}\,(x.\,(\mathsf{throw}\,5\,\mathsf{to}\,x) + 4) + 2$$

This program first captures the current continuation, $[\,] + 2$, and then continues with evaluating the throw operation. At that point, the throw operation causes the program to forget about the rest of computation, *i.e.*, adding 4 and 2, and jumps to the captured continuation, *i.e.*, $[\,] + 2$, with value 5. Hence, the overall result is 7. We prove this fact, *i.e.*, $\mathsf{wp}\ \mathsf{call/cc}\,(x.\,(\mathsf{throw}\,5\,\mathsf{to}\,x) + 4) + 2\,\{x.\ x = 7\}$, in our program logic as follows:

$$\dfrac{\dfrac{\dfrac{}{\mathsf{wp}\ 5 + 2\,\{x.\ x = 7\}}\ \text{trivial}}{\mathsf{wp}\ ((\mathsf{throw}\,5\,\mathsf{to}\,\mathsf{cont}([\,] + 2)) + 4) + 2\,\{x.\ x = 7\}}\ \textsc{throw-wp}\ \text{with}\ K = ([\,] + 4) + 2,\ \text{and},\ K' = [\,] + 2}{\mathsf{wp}\ \mathsf{call/cc}\,(x.\,(\mathsf{throw}\,5\,\mathsf{to}\,x) + 4) + 2\,\{x.\ x = 7\}}\ \textsc{callcc-wp}\ \text{with}\ K = [\,] + 2$$

Notice the importance of the context K in applying the rule CALLCC-WP in this proof.

In summa, for $\mathsf{F}^{\mu,ref}_{conc,cc}$ we use new non-context-local rules for reasoning about weakest preconditions, and the non-context-local rules allow us to reason about call/cc and throw.

Because of the explicit context K, the non-context-local rules for weakest preconditions are somewhat more elaborate to use than the corresponding context-local rules. However, that is the price we have to pay to be able to reason in general about non-local control flow. In Section 4 we will see how we can still recover a form of context-local weakest precondition for reasoning about those parts of the program that do not use non-local control flow.

In the rules above for weakest preconditions, the antecedent is only required to hold a step of computation later ($\triangleright$) — that is because these rules correspond to expressions performing a reduction step.

The update modality and invariants. The update modality $\Rightarrow$ accounts for updating (allocation, deallocation and mutation) of resources.[1] Intuitively, $\Rightarrow P$ is satisfied by resources that can be updated to new resources for which P holds. For any proposition P, we have that $P \vdash \Rightarrow P$. If P holds, then resources can be updated (trivially) so as to have that P holds. The update modality is

[1] This modality is called the fancy update modality in Krebbers et al. [2017a].

idempotent, $\Rrightarrow\Rrightarrow P \dashv\vdash \Rrightarrow P$. We write $P \Rrightarrow\!\!* Q$ as a shorthand for $P \mathrel{-\!\!*} \Rrightarrow Q$. Crucially, resources can be updated throughout a proof of weakest preconditions:

$$\Rrightarrow \mathsf{wp}\, e\, \{\Phi\} \dashv\vdash \mathsf{wp}\, e\, \{\Phi\} \dashv\vdash \mathsf{wp}\, e\, \{x.\ \Rrightarrow \Phi(x)\}$$

Iris features invariants $\boxed{P}^{\mathcal{N}}$ for enforcing concurrent protocols. Each invariant $\boxed{P}^{\mathcal{N}}$ has a name, $\mathcal{N}$, associated to it. Names are used to keep track of which invariants are open.[2] Intuitively, $\boxed{P}^{\mathcal{N}}$ states that P always holds. The following rules govern invariants.

$$\textsc{inv-alloc}\ \frac{\triangleright P}{\Rrightarrow \boxed{P}^{\mathcal{N}}} \qquad \textsc{inv-open-wp}\ \frac{\boxed{R}^{\mathcal{N}} \qquad (\triangleright R) \mathrel{-\!\!*} \mathsf{wp}\, e\, \{y.\ (\triangleright R) * \mathsf{wp}\, K[y]\, \{x.\ Q\}\}}{\mathsf{wp}\, K[e]\, \{x.\ Q\}}\ e \text{ is atomic}$$

These rules say that invariants can always be allocated by giving up the resources being protected by the invariant and they can be kept opened only during the execution of *physically atomic* operations. Iris invariants are impredicative, *i.e.*, they can state P holds invariantly for any proposition P, including invariants. This is why the later operator is used as a guard to avoid self-referential paradoxes [Krebbers et al. 2017a]. Invariants essentially express the knowledge that some proposition holds invariantly. Hence, invariants are always persistent, *i.e.*, $\boxed{P}^{\mathcal{N}} \dashv\vdash \Box \boxed{P}^{\mathcal{N}}$.

3.2 Resources used in defining logical relations

We need some resources in order to define our logical relations in Iris. We need resources for representing memory locations of the implementation side, the memory locations of the specification side and the expression being evaluated on the specification side. These resources are written as follows:

- $\ell \mapsto_i v$: memory location ℓ contains value v on the implementation side.
- $\ell \mapsto_s v$: memory location ℓ contains value v on the specification side.
- $j \Rightarrow e$: the thread j on the specification side is about to execute e.

These resources are defined using more primitive resources in Iris, but we omit such details here. What is important is that we can use these resources to reason about programs. In particular, we can derive the following rules (and similarly for other basic expressions) for weakest preconditions and for execution on the specification side.

$$\frac{\forall \ell.\ \ell \mapsto_i v \mathrel{-\!\!*} \mathsf{wp}\, K[\ell]\, \{\Phi\}}{\mathsf{wp}\, K[\mathsf{ref}(v)]\, \{\Phi\}} \qquad \frac{\ell \mapsto_i v \mathrel{-\!\!*} \mathsf{wp}\, K[v]\, \{\Phi\} \qquad \triangleright\, \ell \mapsto_i v}{\mathsf{wp}\, K[!\,\ell]\, \{\Phi\}}$$

$$\frac{\ell \mapsto_i w \mathrel{-\!\!*} \mathsf{wp}\, K[()]\, \{\Phi\} \qquad \triangleright\, \ell \mapsto_i v}{\mathsf{wp}\, K[\ell \leftarrow w]\, \{\Phi\}} \qquad \frac{j \Rightarrow K[\mathsf{ref}(v)]}{\Rrightarrow \exists \ell.\ \ell \mapsto_s v * j \Rightarrow K[\ell]} \qquad \frac{\ell \mapsto_s v \qquad j \Rightarrow K[!\,\ell]}{\Rrightarrow \ell \mapsto_s v * j \Rightarrow K[v]}$$

$$\frac{\ell \mapsto_s v \qquad j \Rightarrow K[\ell \leftarrow w]}{\Rrightarrow \ell \mapsto_s w * j \Rightarrow K[()]}$$

These resources are all exclusive in the sense that:

$$\ell \mapsto_i v * \ell \mapsto_i v' \vdash \bot \qquad \ell \mapsto_s v * \ell \mapsto_s v' \vdash \bot \qquad j \Rightarrow e * j \Rightarrow e' \vdash \bot$$

[2] Officially in Iris, the update modality is in fact annotated with so-called masks (sets of invariant names), which are used to ensure that invariants are not re-opened. For simplicity, we do not include masks in this paper.

3.3 Logical relations in Iris

Figure 4 presents our binary logical relation for $\mathsf{F}^{\mu,ref}_{conc,cc}$. We define the logical relation in several stages. The first thing we define is the relation of observational refinement. Intuitively, an expression e observationally refines an expression e' if, whenever e reduces to a value so does e'. We define this in Iris using magic wand and weakest precondition. The whole formula reads as follows: assuming that there is some thread j on the specification side that is about to execute e' (represented in Iris by $j \Mapsto e'$) then, after execution of e, we know that thread j on the specification side has also been executed to some value w.

We then use the notion of observational refinement defined above to define the value relation, the expression relation and the evaluation context relation for each type. In contrast to earlier definitions of logical relations in Iris [Krebbers et al. 2017b; Krogh-Jespersen et al. 2017; Timany et al. 2018], our logical relation is an example of so-called biorthogonal logical relations [Pitts 2005], also known as top-top closed logical relations. That is, we define two expressions to be related if plugging them into related evaluation contexts results in observationally related expressions. Two evaluation contexts are defined to be related if plugging related values into them results in observationally related expressions.

The value relation interpretation $[\![\Xi \vdash \tau]\!]_\Delta$ of a type τ in context Ξ is defined by induction on τ. Here Δ is an environment mapping type variables in Ξ to Iris relations. For all the non-continuation types, the definition is exactly as in for the language without call/cc [Krebbers et al. 2017b].

Pairs of values of base types, $\mathbf{1}$, $\mathbb{B}$, and $\mathbb{N}$, are related if they are equal values of their corresponding type. A pair of values are related at the sum type if they are both formed by applying the same injection to values that are in turn related at the appropriate type. A pair of values are related at the product type, $\tau \times \tau'$, if they are each a pair of values such that their first components are related at τ and their second components are related at τ'. The value interpretation of recursive types are defined using Iris's guarded fixpoint operator $\mu r.P$. It intuitively states that two values of are related at a recursive types if they are both folded values with their underlying values again related at recursive types. The value relation for functions types $\tau \to \tau'$ expresses that two values of the function type are related if whenever applied to related values of the domain type, τ, the resulting *expressions* are related at the codomain type, τ'. The use of the *persistently modality*, $\Box$, is to make sure that value interpretations of polymorphic programs are persistent. The value relation for polymorphic values requires two related values to produce two related expressions when instantiated, regardless of what predicate f we take as the interpretation of the type variable; provided that f is a persistent predicate. The use of the persistently modality here, as well as the side-condition on f being persistent, are required to ensure that the value interpretations are persistent. Two values of a reference type are related if they are both locations that always (expressed using Iris invariants) store related values. Finally, the relational interpretation of $\mathrm{cont}(\tau)$ expresses that two continuations are related whenever their corresponding evaluation contexts are related at the evaluation context relation for the type in question.

The evaluation context relation $\mathcal{K}[\![\Xi \vdash \tau]\!]_\Delta$ relates evaluation contexts K and K' if plugging related values of type τ in them results in observationally related expressions.

The expression relation is the standard biorthogonal expression relation. It states that $\mathcal{E}[\![\Xi \vdash \tau]\!]_\Delta(e, e')$ holds whenever, for any two related evaluation contexts K and K', the expressions $K[e]$ and $K'[e']$ are observationally related.

The notion of logical relatedness states, as usual for call-by-value languages, that two expressions e and e' are logically related if substituting related values for their free variables results in related expressions.

Observational refinement: $O : Expr \times Expr \to iProp$

$$O(e, e') \triangleq \forall j.\ j \Mapsto e' \mathrel{-\!\!*} \mathrm{wp}\ e\ \{\exists w.\ j \Mapsto w\}$$

Value interpretation of types: $[\![\Xi \vdash \tau]\!]_\Delta : Val \times Val \to iProp$

$$[\![\Xi \vdash \alpha]\!]_\Delta \triangleq \Delta(\alpha)$$

$$[\![\Xi \vdash \mathbf{1}]\!]_\Delta(v, v') \triangleq v = v' = ()$$

$$[\![\Xi \vdash \mathbb{B}]\!]_\Delta(v, v') \triangleq v = v' = \mathsf{true} \lor v = v' = \mathsf{false}$$

$$[\![\Xi \vdash \mathbb{N}]\!]_\Delta(v, v') \triangleq \exists n.\ v = v' = n$$

$$[\![\Xi \vdash \tau_1 + \tau_2]\!]_\Delta(v, v') \triangleq \bigvee_{i \in \{1,2\}} (\exists w, w'.\ v = \mathsf{inj}_i\ w \land v' = \mathsf{inj}_i\ w' \land [\![\Xi \vdash \tau_i]\!]_\Delta(w, w'))$$

$$[\![\Xi \vdash \tau \times \tau']\!]_\Delta(v, v') \triangleq \exists w_1, w_2, w_1', w2'.\ v = (w_1, w_2) \land v' = (w_1', w_2') \land [\![\Xi \vdash \tau]\!]_\Delta(w_1, w_1') * [\![\Xi \vdash \tau']\!]_\Delta(w_2, w_2')$$

$$[\![\Xi \vdash \mu\alpha.\ \tau]\!]_\Delta \triangleq \mu f : Val \times Val \to iProp.\ \lambda(v, v').\ \exists w, w'.\ v = \mathsf{fold}\ w \land v' = \mathsf{fold}\ w' \land \triangleright [\![\alpha, \Xi \vdash \tau]\!]_{\Delta, \alpha \mapsto f}(w, w')$$

$$[\![\Xi \vdash \tau \to \tau']\!]_\Delta(v, v') \triangleq \Box\left(\forall w, w'.\ [\![\Xi \vdash \tau]\!]_\Delta(w, w') \Rightarrow \mathcal{E}[\![\Xi \vdash \tau']\!]_\Delta(v\ w, v'\ w')\right)$$

$$[\![\Xi \vdash \forall\alpha.\ \tau]\!]_\Delta(v, v') \triangleq \forall f : Val \times Val \to iProp.\ \mathsf{persistent}(f) \Rightarrow \Box\left(\mathcal{E}[\![\alpha, \Xi \vdash \tau]\!]_{\Delta, \alpha \mapsto f}(v\ \_, v'\ \_)\right)$$

$$[\![\Xi \vdash \mathsf{ref}(\tau)]\!]_\Delta(v, v') \triangleq \exists \ell, \ell'.\ v = \ell \land v' = \ell' \land \boxed{\exists w, w'.\ \ell \mapsto_i w * \ell' \mapsto_s w' * [\![\Xi \vdash \tau]\!]_\Delta(w, w')}^{\mathcal{N}.\ell.\ell'}$$

$$[\![\Xi \vdash \mathsf{cont}(\tau)]\!]_\Delta(v, v') \triangleq \exists K, K'.\ v = \mathsf{cont}(K) \land v' = \mathsf{cont}(K') \land \mathcal{K}[\![\Xi \vdash \tau]\!]_\Delta(K, K')$$

Evaluation context interpretation of types: $\mathcal{K}[\![\Xi \vdash \tau]\!]_\Delta : Ectx \times Ectx \to iProp$

$$\mathcal{K}[\![\Xi \vdash \tau]\!]_\Delta(K, K') \triangleq \forall v, v'.\ [\![\Xi \vdash \tau]\!]_\Delta(v, v') \Rightarrow O(K[v], K'[v'])$$

Expression interpretation of types: $\mathcal{E}[\![\Xi \vdash \tau]\!]_\Delta : Expr \times Expr \to iProp$

$$\mathcal{E}[\![\Xi \vdash \tau]\!]_\Delta(e, e') \triangleq \forall K, K'.\ \mathcal{K}[\![\Xi \vdash \tau]\!]_\Delta(K, K') \Rightarrow O(K[e], K'[e'])$$

Logical relatedness: $\Xi \mid \Gamma \vDash e \le_{\log} e' : \tau : iProp$ for $\Gamma = x_1 : \tau_1, \ldots, x_n : \tau_n$

$$\Xi \mid \Gamma \vDash e \le_{\log} e' : \tau \triangleq \forall \Delta, \vec{v}, \vec{v'}.\ \left(\mathop{\Large *}_{x_i:\tau_i} [\![\Xi \vdash \tau_i]\!]_\Delta(v_i, v_i')\right) \Rightarrow \mathcal{E}[\![\Xi \vdash \tau]\!]_\Delta\left(e[\vec{v}/\vec{x}], e'[\vec{v'}/\vec{x}]\right)$$

The map $\Delta : Var \to (Val \times Val) \to iProp$ maps free type variables to their value interpretations. The λ in the definition of $[\![\Xi \vdash \mu\alpha.\ \tau]\!]_\Delta$ is the meta-level λ used for forming an Iris predicate.

Fig. 4. An excerpt of the logical relations for $\mathsf{F}^{\mu,ref}_{conc,cc}$

We can now state and prove the fundamental theorem of logical relations for $\mathsf{F}^{\mu,ref}_{conc,cc}$. The theorem expresses that any well-typed expression is logically related to itself.

Theorem 3.1 (Fundamental theorem of logical relations).

$$\Xi \mid \Gamma \vdash e : \tau \Rightarrow \Xi \mid \Gamma \vDash e \leq_{log} e : \tau$$

This theorem is proven by induction on the typing derivation using the basic rules for weakest-preconditions and executions on the specification side.

The above theorem, together with some basic properties of observational refinement, implies the soundness of our logical relations, *i.e.*, that logical relatedness implies contextual refinement:

Theorem 3.2 (Soundness of logical relations).

$$\Xi \mid \Gamma \vDash e \leq_{log} e' : \tau \Rightarrow \Xi \mid \Gamma \vDash e \leq_{ctx} e' : \tau$$

Our logical relation is expressed in terms of weakest preconditions and the proofs of the above theorems use the earlier presented proof rules for weakest preconditions. Before turning to applications, we pause to present *context-local weakest preconditions*, which we can use to simplify reasoning about program fragments, which do not use non-local control flow.

4 CONTEXT-LOCAL WEAKEST PRECONDITIONS

To make it simpler to reason about expressions that do not use non-local control flow, we define a new notion of *context-local weakest precondition* (CLWP). The definition is given in terms of the earlier weakest precondition, which, as we will explain below, means that we will be able to mix and match reasoning steps using (non-context local) weakest preconditions and context-local weakest preconditions.

Definition 4.1. The *context-local weakest precondition* of e with respect to Φ is defined as follows:

$$\mathsf{clwp}\ e\ \{\Phi\} \triangleq \forall K, \Psi.\ (\forall v.\ \Phi(v) \twoheadrightarrow \mathsf{wp}\ K[v]\ \{\Psi\}) \twoheadrightarrow \mathsf{wp}\ K[e]\ \{\Psi\}$$

Based on this, we also define context-local Hoare triples:

$$\{P\}^{\mathsf{cl}}\ e\ \{x.\ Q\} \triangleq \square\ (P \twoheadrightarrow \mathsf{clwp}\ e\ \{x.\ Q\})$$

Note how the above definition essentially says that $\mathsf{clwp}\ e\ \{\Phi\}$ holds if the bind rule holds for e, which intuitively means that e behaves context-locally. Another way to look at the definition of context-local weakest preconditions above is the following. In order to prove $\mathsf{clwp}\ e\ \{\Phi\}$, we have to show that given any evaluation context K and predicate Ψ we have to show that running e under K can satisfy postcondition Ψ. Since K and Ψ are universally quantified and the *only thing* we know about them is that $\forall v.\ \Phi(v) \twoheadrightarrow \mathsf{wp}\ K[v]\ \{\Psi\}$, the *only way* for the program $K[e]$ to guarantee Ψ as its postcondition is that it would, in its execution, eventually reach a point $K[v]$ for a value v such that $\Phi(v)$. This means that intuitively e does not tamper with the evaluation context it is running under. In other words, for all intents and purposes e behaves context-locally. Therefore, the bind rule below is sound for context-local weakest preconditions.

$$\frac{\mathsf{clwp}\ e\ \{x.\ \mathsf{clwp}\ K[x]\ \{\Phi\}\}}{\mathsf{clwp}\ K[e]\ \{\Phi\}}\ \textsc{bind}$$

Moreover, the "standard" rules for the basic language constructs (excluding call/cc and throw, of course) can also be derived for context-local weakest preconditions, as shown in Figure 5. We

FST-CLWP
$$\frac{\triangleright\ \mathsf{clwp}\ v\ \{\Phi\}}{\mathsf{clwp}\ \pi_1\ (v, w)\ \{\Phi\}}$$

IF-TRUE-CLWP
$$\frac{\triangleright\ \mathsf{clwp}\ e\ \{\Phi\}}{\mathsf{clwp}\ \mathsf{if\ true\ then}\ e\ \mathsf{else}\ e'\ \{\Phi\}}$$

REC-CLWP
$$\frac{\triangleright\ \mathsf{clwp}\ e[v/x][\mathsf{rec}\ f(x) = e/f]\ \{\Phi\}}{\mathsf{clwp}\ (\mathsf{rec}\ f(x) = e)\ v\ \{\Phi\}}$$

ALLOC-CLWP
$$\mathsf{clwp}\ \mathsf{ref}(v)\ \{x.\ \exists \ell. x = \ell * \ell \mapsto_i v\}$$

LOAD-CLWP
$$\frac{\triangleright\ \ell \mapsto_i v}{\mathsf{clwp}\ !\ell\ \{x.\ x = v * \ell \mapsto_i v\}}$$

Fig. 5. An excerpt of inference rules for CLWP's.

can also use invariants during atomic steps of computation while proving context-local weakest preconditions.

INV-OPEN-CLWP
$$\frac{\boxed{R}^{\mathcal{N}} \qquad (\triangleright R) \twoheadrightarrow \mathsf{clwp}\ e\ \{x.\ (\triangleright R) * Q\} \qquad e\ \text{ is atomic}}{\mathsf{clwp}\ e\ \{x.\ Q\}}$$

Now we have both (non-context-local) weakest preconditions and context-local weakest preconditions. What is the upshot of this? The key point is that when we prove correctness / relatedness of programs, we can use the simpler context-local weakest preconditions for reasoning about those parts of the program which are context local (do not use call/cc or throw) and only use the (non-context-local) weakest preconditions for reasoning about those parts that may involve non-local control flow. This fact is expressed formally by the rule CLWP-WP below, which establishes a connection between weakest-preconditions and context-local weakest preconditions.

CLWP-WP
$$\frac{\mathsf{clwp}\ e\ \{\Psi\} \qquad \forall v.\ \Psi(v) \twoheadrightarrow \mathsf{wp}\ K[v]\ \{\Phi\}}{\mathsf{wp}\ K[e]\ \{\Phi\}}$$

This rule basically says that if we know that e *context-locally* guarantees postcondition Ψ then we can prove $\mathsf{wp}\ K[e]\ \{\Phi\}$ by assuming that, *locally* under the context K, it will only evaluate to values that satisfy Ψ. Moreover, it guarantees that the evaluation of e does not tamper with the evaluation context that we are considering it under.

It is easy to show that $\mathsf{clwp}\ e\ \{\Phi\}$ implies $\mathsf{wp}\ e\ \{\Phi\}$; simply take K to be the empty evaluation context, [ ], and Ψ to be Φ. Hence, similarly to a (non-context-local) weakest precondition, the context-local weakest precondition $\mathsf{clwp}\ e\ \{\Phi\}$ also implies the safety of e, and, that whenever e terminates with a value v, $\Phi(v)$ holds.

The proposition $\mathsf{clwp}\ e\ \{\Phi\}$ only says that e behaves context-locally. Therefore, in principle, we should be able to give context-local specifications to programs that do use call/cc and throw, but still behave context-locally. For instance, the following example from Section 3.1:

$$\mathsf{call/cc}\ (x.\ (\mathsf{throw}\ 5\ \mathsf{to}\ x) + 4) + 2$$

This program does indeed use call/cc and throw, but it does not tamper with the evaluation context that it runs under. That is to say that the effects of call/cc and throw are confined within the program itself. Hence, we can prove that the following *context-local* specification holds for this program.

$$\mathsf{clwp}\ \mathsf{call/cc}\ (x.\ (\mathsf{throw}\ 5\ \mathsf{to}\ x) + 4) + 2\ \{x.\ x = 7\}$$

However, as there are no context-local reasoning rules for call/cc and throw, we have to unfold the definition of context-local weakest precondition proposition and prove this fact manually:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{(\forall v.\ v = 7 \twoheadrightarrow \mathsf{wp}\ K[v]\ \{\Psi\}) \vdash \mathsf{wp}\ K[5+2]\ \{\Psi\}}\ \text{trivial}}{(\forall v.\ v = 7 \twoheadrightarrow \mathsf{wp}\ K[v]\ \{\Psi\}) \vdash \mathsf{wp}\ K[((\mathsf{throw}\ 5\ \mathsf{to}\ \mathsf{cont}(K[[\ ]+2])) + 4) + 2]\ \{\Psi\}}\ \textsc{throw-wp}}{(\forall v.\ v = 7 \twoheadrightarrow \mathsf{wp}\ K[v]\ \{\Psi\}) \vdash \mathsf{wp}\ K[\mathsf{call/cc}\,(x.\ (\mathsf{throw}\ 5\ \mathsf{to}\ x) + 4) + 2]\ \{\Psi\}}\ \textsc{callcc-wp}}{\forall K, \Psi.\ (\forall v.\ v = 7 \twoheadrightarrow \mathsf{wp}\ K[v]\ \{\Psi\}) \twoheadrightarrow \mathsf{wp}\ K[\mathsf{call/cc}\,(x.\ (\mathsf{throw}\ 5\ \mathsf{to}\ x) + 4) + 2]\ \{\Psi\}}\ \text{introducing assumptions}}{\mathsf{clwp}\ \mathsf{call/cc}\,(x.\ (\mathsf{throw}\ 5\ \mathsf{to}\ x) + 4) + 2\ \{x.\ x = 7\}}\ \text{unfolding clwp}$$

5 WEB SERVER REFINEMENT

The combination of continuations and concurrency allows for a simplified implementation of web servers. Such servers store explicitly captured (using call/cc) server-side continuations in order to track the state of communication with the client [Flatt 2017; Hendershott 2017; Krishnamurthi et al. 2007; Might 2017; Queinnec 2004]. To serve a returning user such web servers simply resume the stored continuation for that user. We refer to these servers as *continuation-based* as opposed to the more traditional *state-storing*, where server stores data that it later uses to reconstruct the state of the server for returning clients. In this section we show that two simple servers, one implemented in continuation-based style and one implemented in state-storing style, are equivalent.

Two Servers. The two servers that we consider take a number as a request. They reply to each user with the sum of the numbers that *that user* has submitted. Figure 6 shows implementations of the handler functions of the servers.[3]

The handlers take a connection (of type `serverConnT`) as input which consists of a pair of functions for reading the request and writing the response. The idea is that these functions are an abstraction of a TCP connection and thus the contextual equivalence can be understood as showing that clients cannot distinguish between the two implementations. These servers both internally use a table to associate a *resumption id* (a number to remember the client by) to each client. Here the functions `associate` and `get` are used for storing into the table and looking up resumption ids in the table, respectively. The state-storing implementation associates to a resumption id the sum so far. The continuation-based implementation associates a captured continuation to a resumption id. Each returning client sends its resumption id along with its request. The function `sumloop` in the continuation based implementation is essentially a loop that sends the sum so far to the client and asks for the next number by calling `read_client`. The function `read_client` captures the current continuation and associates it with a resumption id. This resumption id is then sent to the client. After this, the server stops as the request in question is served.[4] Notice that the `read_client` never returns. In practice the control is returned to the point after the call to this function when the client returns with the resumption id associated to the current continuation. In this case, as the connection is new, the server will supply the new connection (`reader`, `writer`) along with the new request.

Since the two servers, apart from their handlers, are identical (they just pass requests to their handlers) we here only discuss the contextual equivalence of the two handlers by showing that each handler refines the other. Here we show one of these refinements:

$$\Xi \mid \Gamma \vDash \mathit{handler2} \leq_{\log} \mathit{handler1} : \mathbf{ServerConnT} \to \mathbf{1}$$

The other is similar.

The two handlers internally use a concurrent (protected by a spin lock) table to store and look-up resumptions. The table and lock implementations are straightforward and thus omitted. Since these

[3]We use an ML-like syntax for the sake of brevity and legibility; our Coq formalization includes the $\mathsf{F}^{\mu,ref}_{conc,cc}$ code for these handlers.

[4]The command **`abort`** is the command that ends the program (thread) and can be written in our programming language as **`throw`** `()` **`to`** `[]` where `[]` is the empty evaluation context.

```
(* In the code below, the function `resumptionid` converts the index of a continuation in
   the table into a session id, i.e., a session cookie. Similarly, the function `result`
   formats the response appropriately, making it ready to be  sent to the client. *)
let handler1 :  ServerConnT -> 1 =
  let tb = newTable () in
  fun (cn : ServerConnT) ->
  let (reader, writer) = cn in
  match reader () with
   (Some cid, n) ->
     begin
       match get tb cid with
        None -> () (* unknown resumption id! *)
       | Some sum -> writer (result (sum + n));
           writer (resumptionid (associate tb (sum + n)));
           abort
     end
  | (None, n) ->
      writer (result n);
      let cid = associate tb n in
      writer (resumptionid cid)
```

```
let read_client tb writer =
 callcc (k. writer (resumptionid (associate tb k)); abort)

let rec sumloop m reader writer =
 writer (result m);
 let (v, reader, writer) = read_client tb writer in
 sumloop (m + v) reader writer

let handler2 : ServerConnT -> 1 =
  let tb = newTable () in
  fun (cn : ServerConnT) ->
    let (reader, writer) = cn in
    match reader () with
     (Some cid, n)->
     begin
       match get tb cid with
         None -> () (*unknown resumption id!*)
       | Some k -> throw (n,reader,writer) to k
     end
   | (None, n)-> sumloop n reader writer
```

Fig. 6. Two server handlers: one state-storing (top) and one continuation-based (bottom).

implementations do not use call/cc and throw, we give their relational specs using context-local weakest preconditions. Given the rule CLWP-WP we can use these specs during the proof of contextual refinement of handlers. We discuss the relational specs of the table before presenting the refinement of handlers.

5.1 Relational spec for the table and the lock

The essence of relating the tables on both sides (specification side and implementation side) is simple. Two tables are related if their contents are. For this purpose we introduce the predicate $relTables(v, v', \gamma, \Phi)$ which states that the table v is related to the table v' where their contents are related by the binary predicate Φ. We ignore γ for now. It is only used for the internal lock protecting the table. With this definition, new tables are related (as they are both empty). The related specs for the **get** and **associate** operations require that only related values can be stored

into tables and guarantee that when looking the table up we are guaranteed to receive related values, if any.

What does it mean for the contents, *i.e.*, a number and a continuation to be related here, *i.e.*, what should we take for Φ? The answer is simple. The number n is related to the continuation K if n is the same as the number that the suspended program in K considers to be the sum. Notice that this relation specifies some details about the program captured into K. This program uses the table itself! Therefore, it is not possible to write down the relation for the contents of the table before the table is created. As a result, the standard relational spec for the table, *e.g.*, the relational spec that we get *for free* from the fundamental theorem of logical relations, does not suffice for our application. Such standard specs require the relation on the contents to be given before the table is created:

$$\begin{array}{l}\forall \Phi.\ \{j \Rrightarrow K[\textbf{newTable}\ ()]\}^{\mathrm{cl}} \qquad \text{(weaker standard spec)}\\ \quad \textbf{newTable}\ ()\\ \quad \{x.\ \exists v'.\ j \Rrightarrow K[v'] * \exists \gamma.\ relTables(x, v', \gamma, \Phi)\}\end{array}$$

Note the quantification over Φ outside the whole triple. Hence, we give the following stronger relational spec to our tables:

$$\begin{array}{l}\{j \Rrightarrow K[\textbf{newTable}\ ()]\}^{\mathrm{cl}}\\ \quad \textbf{newTable}\ ()\\ \{x.\ \exists v'.\ j \Rrightarrow K[v'] * \forall \Phi.\ \Rrightarrow \exists \gamma.\ relTables(x, v', \gamma, \Phi)\}\end{array}$$

$$\begin{array}{l}\{relTables(tb, tb', \gamma, \Phi) * j \Rrightarrow K[\textbf{get}\ tb'\ n]\}^{\mathrm{cl}}\\ \quad \textbf{get}\ tb\ n\\ \left\{\begin{array}{l} x.\ \exists v'.\ j \Rrightarrow K[v'] * \big(x = v' = None \vee \\ \quad (\exists w, w'\ x = Some(w) \wedge v' = Some(w') * \Phi(v, v'))\big)\end{array}\right\}\end{array}$$

$$\begin{array}{l}\{relTables(tb, tb', \gamma, \Phi) * \Phi(v, v') * j \Rrightarrow K[\textbf{associate}\ tb'\ v]\}^{\mathrm{cl}}\\ \quad \textbf{associate}\ tb\ v\\ \{x.\ \exists n.\ x = n * j \Rrightarrow K[n]\}\end{array}$$

Notice that with our stronger specification we can refer to the tables themselves in the predicate Φ that we pick for relating the contents, whereas in the (weaker standard spec) specification one has to pick this relation beforehand, and hence one cannot refer to the tables v and v' because they have not been created yet!

The predicate $relTables(tb, tb', \gamma, \Phi)$ is defined in terms of the *relLocks* predicate, which pertains to the relational specification of spin locks given below.

$$\begin{array}{l}relTables(tb, tb', \gamma, \Phi) \triangleq relLocks(tb.lock, tb'.lock, \gamma, P_\Phi)\\ \qquad P_\Phi \triangleq \exists ls.\ contents(tb, map\ \pi_1\ ls)\ * contents(tb', map\ \pi_2\ ls) * \mathop{\Large *}_{(x,x') \in ls} \Phi(x, x')\end{array}$$

Here $tb.lock$ is the lock protecting the table tb. The proposition P_Φ above simply states that the there is a list of pairs of values, which are pairwise related by Φ and, moreover, that the first projections of these pairs are stored in the implementation side table and the second projections of these pairs are stored in the specification side table. The *contents* predicate simply specifies that the index of an element in the table is its index in the list.

Relational spec for the spin lock. Similar to the table specification the lock specification also needs to be strengthened. The relational specs for locks is given below. In this specs the persistent proposition $\mathit{relLocks}(v, v', \gamma, P)$ states that v is a lock protecting two things: resources P *and* the fact that v' is not acquired; this is crucial as we discuss below. The proposition $\mathit{locked}(\gamma)$ states that both of the locks associated to γ are currently acquired.

$$\{j \Rrightarrow K[\mathbf{newlock}\ ()]\}^{\mathrm{cl}}$$
$$\quad \mathbf{newlock}\ ()$$
$$\{x.\ \exists v'.\ j \Rrightarrow K[v'] * \forall P.\ P \Rrightarrow \exists \gamma.\ \mathit{relLocks}(x, v', \gamma, P)\}$$

$$\{\mathit{relLocks}(v, v', \gamma, P) * j \Rrightarrow K[\mathbf{acquire}\ v']\}^{\mathrm{cl}}$$
$$\quad \mathbf{acquire}\ v$$
$$\{\_.\ j \Rrightarrow K[()] * \mathit{locked}(\gamma) * P\}$$

$$\{\mathit{relLocks}(v, v', \gamma, P) * P * \mathit{locked}(\gamma) * j \Rrightarrow K[\mathbf{release}\ v']\}^{\mathrm{cl}}$$
$$\quad \mathbf{release}\ v$$
$$\{\_.\ j \Rrightarrow K[()]\}$$

The specification captures that whenever we acquire the lock on the implementation side, the lock on the specification side is free and can be acquired. This is necessary for showing contextual refinements because if the implementation side converges, then we need to show that so does the specification side and the acquire operation is potentially non-terminating. This also means that whenever we release the lock on the implementation side, the lock on the specification side is also released.

5.2 Proving equivalence of handlers

We prove only one direction here: $\Xi \mid \Gamma \vDash \mathit{handler2} \leq_{\log} \mathit{handler1} : \mathbf{ServerConnT} \to \mathbf{1}$. We use the rules for weakest preconditions and executions on the specification side explained above and make use of the relational specification given above for tables, which is justified by the CLWP-WP rule. A key element of the proof is the choice of predicate for relating the contents of the two tables. We use the following predicate:

$$\Phi_{\mathit{handlers}}(w, w') = \exists \mathit{sum} \in \mathbb{N}.\ w' = \mathit{sum} \land$$
$$\exists K.\ w = \mathsf{cont}\left(K\left[\begin{array}{l}\mathsf{let}\,(v, \mathit{reader}, \mathit{writer}) = [\ ]\,\mathsf{in}\\ \mathbf{sumloop}\ (\mathit{sum} + v)\ \mathit{reader}\ \mathit{writer}\end{array}\right]\right)$$

The relation Φ_{handlers} essentially captures what was explained in prose earlier: the two sides consider the *sum* stored (as part of a continuation object in the continuation-based server) to be the same value. The relation Φ_{handlers} above is indeed capturing the essence of the intuitive reason why the two implementations of handlers have contextually equivalent behavior. According to the definition of our logical relations, to show logical relatedness we need to show that given any two related contexts the two programs behave in an observationally related way. Since at the time of picking the predicate above we do not know what contexts we will have to operate under, we have to consider that our code of interest is inside some arbitrary (hence existentially quantified in the continuation-based) evaluation context. The actual value of this evaluation context is not important as the control never reaches this evaluation context; the thread is ended with an **abort** before that. Note the use of the table itself (referenced inside **sumloop**) in the Φ_{handlers}.

6 ONE-SHOT CALL/CC

In this section we consider a more technical verification challenge involving continuations, due to Friedman and Haynes [1985]. The challenge is to show that call/cc can be implemented using references and one-shot continuations, *i.e.*, continuations that can only be called once. This problem has been studied for *sequential* higher-order languages with references in Dreyer et al. [2012]; Støvring and Lassen [2007], with pen-and-paper proofs. Here we show that the equivalence also holds in our concurrent language (subtly so; because we are using *may* contextual equivalence) and we give a mechanized formal proof thereof.

For this verification we introduce two closed programs $\mathfrak{CC}$ (essentially a call/cc) and $\mathfrak{CC}_1$ (encoding a one shot continuation) both of type $\forall\alpha.\,(\mathsf{cont}(\alpha) \to \alpha) \to \alpha$.

$$\mathfrak{CC} \triangleq \Lambda\,\lambda f.\ \mathsf{call/cc}\,(x.\ f\ x)$$

$$\mathfrak{CC}_1 \triangleq \Lambda\,\lambda f.\ \mathsf{let}\ b = \mathsf{ref}(\mathsf{false})\ \mathsf{in}\ \mathsf{call/cc}\left(x.\ f\ \left(\mathsf{cont}\begin{pmatrix}\mathsf{let}\ y = [\,]\ \mathsf{in}\\ \mathsf{if}\ !b\ \mathsf{then}\ \Omega\ \mathsf{else}\ b \leftarrow \mathsf{true};\\ \mathsf{throw}\ y\ \mathsf{to}\ x\end{pmatrix}\right)\right)$$

Here Ω is the trivially diverging expression. When applied, the one-shot continuation, $\mathfrak{CC}_1$, first allocates a *one-shot bit* b and then calls the given function with a continuation that uses b to ensure that the continuation is only called once. Using one-shot continuations, we simulate normal continuations by defining $\mathfrak{CC}'$:

$$\mathfrak{CC}' \triangleq \Lambda\,\lambda f.\ \mathsf{let}\ \ell = \mathsf{ref}(\mathsf{cont}([\,]))\ \mathsf{in}\ G\ f$$

$$G \triangleq \mathsf{rec}\ G(f) = \mathsf{let}\ x = \mathfrak{CC}_1\ \_\ \begin{pmatrix}\lambda y.\ \ell \leftarrow y;\\ f\ (\mathsf{cont}(\mathsf{throw}\ \mathsf{cont}([\,])\ \mathsf{to}\ !\ell))\end{pmatrix}\ \mathsf{in}\ \mathfrak{CC}_1\ \_\ (\lambda y.\ G\ (\lambda z.\ \mathsf{throw}\ x\ \mathsf{to}\ y))$$

The expression $\mathfrak{CC}'$ above has the same type as $\mathfrak{CC}$. $\mathfrak{CC}'$ perhaps looks fairly complex but the intuition is straightforward. It first allocates ℓ with the trivial continuation, then it takes a one-shot continuation and updates ℓ. When the one-shot continuation is used, it will first grab another *fresh* one-shot continuation and update ℓ with it before continuing. Hence, intuitively, every time the one-shot continuation stored in ℓ is used, it is immediately refreshed, thus mimicking the behavior of $\mathfrak{CC}$. We now prove that $\mathfrak{CC}$ is contextually equivalent to $\mathfrak{CC}'$:

Theorem 6.1. $\cdot \mid \cdot \vDash \mathfrak{CC} \approx_{ctx} \mathfrak{CC}' : \forall\alpha.\,(\mathsf{cont}(\alpha) \to \alpha) \to \alpha$

We only discuss one side of the refinement, namely, $\mathfrak{CC}' \leq_{\mathrm{ctx}} \mathfrak{CC}$. The invariant that we need in order to prove this refinement is intuitively that the one-shot bit of the continuation stored in ℓ is always storing the value false indicating that the one-shot continuation is unused. This fact is expressed in terms of Iris invariants as follows:

$$\boxed{\exists b.\ b \mapsto_i \mathsf{false} * \ell \mapsto_i \mathsf{cont}\begin{pmatrix}\mathsf{let}\ y = [\,]\ \mathsf{in}\\ \mathsf{if}\ !\,b\ \mathsf{then}\ \Omega\ \mathsf{else}\ b \leftarrow \mathsf{true};\\ \mathsf{throw}\ y\ \mathsf{to}\ \mathsf{cont}(K[\mathit{restore}(\ell)])\end{pmatrix}}^{\mathcal{N}.\mathfrak{CC}}$$

where

$$\mathit{restore}(\ell) \triangleq \mathsf{let}\ x = [\,]\ \mathsf{in}\ \mathfrak{CC}_1\ \_\ (\lambda y.\ G\ (\lambda\_.\ \mathsf{throw}\ x\ \mathsf{to}\ y))$$

Here K is the continuation that is captured by $\mathfrak{CC}$. The invariant above is exactly the invariant that Dreyer et al. [2012] use when translated to our system.[5]

This invariant suffices for a *sequential* programming language. However, in our concurrent settting, the "continuation" captured by $\mathfrak{CC}'$ may be shared among multiple threads and, if they use it concurrently, it may happen that a thread is using the continuation captured by $\mathfrak{CC}'$ and before this thread manages to capture another one-shot continuation and restore ℓ, another thread attempts to use the *then invalid* one-shot continuation, and hence it diverges.

We prove that the contextual refinement still holds (despite the possibility of divergence). However, because of the possible racing, we need to use a weaker invariant:

$$\boxed{\begin{array}{l}\exists b, M.\ \mathit{OneShotBits}(M) * \mathit{isOneShotBit}(b) * \\ \qquad \left(\ast_{r \in M} \exists v \in \{\mathsf{true}, \mathsf{false}\}\ .\ r \mapsto_i v\right) * \ell \mapsto_i \mathsf{cont}\begin{pmatrix}\mathsf{let}\ y = [\,]\ \mathsf{in}\\ \mathsf{if}\ !b\ \mathsf{then}\ \Omega\ \mathsf{else}\ b \leftarrow \mathsf{true};\\ \mathsf{throw}\ y\ \mathsf{to}\ \mathsf{cont}(K[\mathit{restore}(\ell)])\end{pmatrix}\end{array}}^{\mathcal{N}.\mathfrak{CC}}$$

This invariant says that ℓ stores a one-shot continuation with a one-shot bit b and that we have a set of bits that, intuitively, have been associated to one-shot continuations. We also know that b is one such one-shot bit, $\mathit{isOneShotBit}(b)$. The predicates $\mathit{OneShotBits}()$ and $\mathit{isOneShotBit}()$ are defined using Iris resources.[6] Here, we only need to know two things about them: $\mathit{isOneShotBit}(b)$ is persistent and:

$$\mathit{OneShotBits}(M) * \mathit{isOneShotBit}(b) \vdash b \in M \qquad \text{(in-bits)}$$

Persistence allows us to retain the information $\mathit{isOneShotBit}(b)$ once we have opened the invariant and have read ℓ. Due to the race condition explained above, when we open the invariant we know, by (in-bits), that there is a value $v \in \{\mathsf{true}, \mathsf{false}\}$ stored in b, and this suffices for being able to complete the refinement proof.

7 CORRECTNESS OF CONTINUATION BASED COOPERATIVE CONCURRENCY

Cooperative concurrency, a.k.a. light-weight concurrency, is a form of concurrency where threads cooperate and use a *yield* command to relinquish control to other threads. This is in contrast to *preemptive* concurrency, where the operating system preempts and schedules threads. Cooperative concurrency is often implemented using continuations [Haynes et al. 1984]. Forking a new thread suspends the execution of the current thread, enqueues the suspension in a queue, and runs the forked thread. The yield command dequeues a previously enqueued suspension (thread) and resumes it, after enqueuing the current continuation.

In this section, we prove correctness of a continuation-based implementation of *cooperative concurrency*. It is not entirely obvious how to state the desired correctness property. Here we use a relational approach inspired by compiler correctness, and show that a language with built-in cooperative concurrency can be compiled into a continuation-based implementation of cooperative concurrency. We prove the correctness of this compilation, by showing that a compiled program refines its source program.

The programming language with built-in cooperative concurrency is called $\mathsf{F}^{\mu,\mathit{ref}}_{\mathit{cc},\mathit{coop}}$. This language serves as our specification of cooperative concurrency. Concretely, $\mathsf{F}^{\mu,\mathit{ref}}_{\mathit{cc},\mathit{coop}}$ provides two primitive commands Cfork and yield. The semantics of $\mathsf{F}^{\mu,\mathit{ref}}_{\mathit{cc},\mathit{coop}}$ keeps track of the currently running thread

[5] In the work of Dreyer et al. [2012], invariants were called *islands*.

[6] The one-shot bit predicates are defined using the authoritative resource algebra over the resource algebra of finite sets of locations (where the algebra operation is set union). See JUNG et al. [2018]

and keeps executing that thread until it reaches a fork or a yield command. In the former case, it starts a new thread and starts executing that. In the latter case, the semantics picks another thread and proceeds with executing that. The programming language that we consider as the target of translation of $\mathsf{F}^{\mu,ref}_{cc,coop}$ is $\mathsf{F}^{\mu,ref}_{cc}$, the sequential fragment of $\mathsf{F}^{\mu,ref}_{conc,cc}$, *i.e.*, $\mathsf{F}^{\mu,ref}_{conc,cc}$ without fork and cas.

In the rest of this section, we define the precise syntax and semantics of the source and target languages, the translation of cooperative concurrency, and a cross-language logical relation between the source and target language, which we use to show the correctness of the translation.

7.1 Syntax and semantics of $\mathsf{F}^{\mu,ref}_{cc}$ and $\mathsf{F}^{\mu,ref}_{cc,coop}$

The programming languages $\mathsf{F}^{\mu,ref}_{conc,cc}$, $\mathsf{F}^{\mu,ref}_{cc}$ and $\mathsf{F}^{\mu,ref}_{cc,coop}$ have the same types and similar typing rules for shared parts; we use $\vdash_{\mathsf{coop}}$ and $\vdash_{\mathsf{seq}}$ to distinguish them. The typing rules for Cfork and yield in $\mathsf{F}^{\mu,ref}_{cc,coop}$ are as follows:

$$\textsc{T-CFork}\ \frac{\Xi \mid \Gamma \vdash_{\mathsf{coop}} e : \tau}{\Xi \mid \Gamma \vdash_{\mathsf{coop}} \mathsf{Cfork}\ \{e\} : \mathbf{1}} \qquad \textsc{T-Yield}\ \frac{}{\Xi \mid \Gamma \vdash_{\mathsf{coop}} \mathsf{yield} : \mathbf{1}}$$

The language $\mathsf{F}^{\mu,ref}_{cc}$ is a fragment of $\mathsf{F}^{\mu,ref}_{conc,cc}$ with the same operational semantics. Formally, the head steps for $\mathsf{F}^{\mu,ref}_{cc}$ are identical to the corresponding fragment in $\mathsf{F}^{\mu,ref}_{conc,cc}$. For the general execution, however, we define the *sequential step* relation, $\rightarrow_{\mathsf{seq}}$, in place of the thread-pool step $\rightarrow$ for $\mathsf{F}^{\mu,ref}_{conc,cc}$.

$$\frac{(e,\sigma) \rightarrow_K (e',\sigma')}{(K[e];\sigma) \rightarrow_{\mathsf{seq}} (K[e'];\sigma')} \qquad (K[\mathsf{throw}\ v\ \mathsf{to}\ \mathsf{cont}(K')];\sigma) \rightarrow_{\mathsf{seq}} (K'[v];\sigma)$$

The head-step relation for the fragment of $\mathsf{F}^{\mu,ref}_{cc,coop}$, apart from Cfork and yield, is defined identically to $\mathsf{F}^{\mu,ref}_{conc,cc}$ and $\mathsf{F}^{\mu,ref}_{cc}$. For general reduction, in place of a thread-pool step we define a *cooperative step*, $\rightarrow_{\mathsf{coop}}$. The step $(\vec{e};n;\sigma) \rightarrow_{\mathsf{coop}} (\vec{e'};n';\sigma')$ is to be understood as a step transforming the thread pool $\vec{e}$ into the thread pool $\vec{e'}$ and state σ into σ' while changing the current thread being executed from thread number n to thread number n'. All head steps (and throw) do not change the current thread. For Cfork and yield we have:

$$\frac{\mathit{length}(e_1) = n \qquad m = \mathit{length}(\vec{e}_1, K[\mathsf{Cfork}\ \{e\}\ ], \vec{e}_2)}{(\vec{e}_1, K[\mathsf{Cfork}\ \{e\}\ ], \vec{e}_2; n; \sigma) \rightarrow_{\mathsf{coop}} (\vec{e}_1, K[()], \vec{e}_2, e; m; \sigma)}$$

$$\frac{\mathit{length}(e_1) = n \qquad 0 \leq m < \mathit{length}(\vec{e}_1, K[\mathsf{yield}], \vec{e}_2)}{(\vec{e}_1, K[\mathsf{yield}], \vec{e}_2; n; \sigma) \rightarrow_{\mathsf{coop}} (\vec{e}_1, K[()], \vec{e}_2; m; \sigma)}$$

A quick inspection of the cooperative step relation should make it clear that $\mathsf{F}^{\mu,ref}_{cc,coop}$ does really capture cooperative concurrency. Hence, by showing that $\mathsf{F}^{\mu,ref}_{cc,coop}$ can be *correctly* compiled to $\mathsf{F}^{\mu,ref}_{cc}$ by a compiler that compiles Cfork and yield using continuations, we establish correctness of an implementation of cooperative concurrency using continuations.

7.2 Translation and its correctness

The translation from $\mathsf{F}^{\mu,ref}_{cc,coop}$ into $\mathsf{F}^{\mu,ref}_{cc}$ is very straightforward. It simply translates Cfork and yield to programs that use the the light-weight thread library LiThr given in Figure 7.[7] This simple and

[7]Note that the specification side, $\mathsf{F}^{\mu,ref}_{cc,coop}$, does not restrict scheduling, *i.e.*, yield non-deterministically chooses another thread. Here, on the implementation side we have chosen to use a queue for scheduling of threads.

$$
\begin{array}{l}
\mathsf{LiThr} \triangleq \mathsf{let}\ Q = \mathit{newQueue}\ \mathsf{in} \\
\quad \mathsf{let}\ \mathit{Frk} = \lambda x.\ \mathsf{call/cc}\left(\begin{array}{l} y.\mathit{enqueue}\ Q\ y; \\ \ \mathsf{throw}\ x\ \mathsf{to}\ \mathsf{cont}([\ ]\ ())\end{array}\right)\ \mathsf{in} \\
\quad \mathsf{let}\ \mathit{Yld} = \lambda\_.\ \mathsf{call/cc}\left(\begin{array}{l} x.\ \mathit{enqueue}\ Q\ x; \\ \ \mathsf{let}\ y = \mathit{dequeue}\ Q\ \mathsf{in} \\ \ \mathsf{match}\ y\ \mathsf{with} \\ \ \ \mathit{Some}(z) \Rightarrow \mathsf{throw}\ ()\ \mathsf{to}\ z \\ \ |\ \mathit{None} \quad \Rightarrow () \\ \ \mathsf{end}\end{array}\right)\ \mathsf{in} \\
\quad (\mathit{Frk}, \mathit{Yld})
\end{array}
$$

Fig. 7. The light-weight thread library LiThr.

minimalistic light-weight thread library provides two functions: one for forking a new thread, *Frk*, and one for relinquishing control to other threads, *Yld*. Notice that the *dequeue* operation can return *None* if the queue empty. However, since in *Yld* the *dequeue* operation is immediately preceded by an *enqueue* operation, this will never happen.

We translate programs in $\mathsf{F}^{\mu,ref}_{cc,coop}$ into programs of $\mathsf{F}^{\mu,ref}_{cc}$ as follows:

$$\mathfrak{Comp}(e) \triangleq \mathsf{let}\ (\mathbb{F}, \mathbb{Y}) = \mathsf{LiThr}\ \mathsf{in}\ \langle\!\langle e \rangle\!\rangle$$

Where $\langle\!\langle e \rangle\!\rangle$ translates $\mathsf{F}^{\mu,ref}_{cc,coop}$ programs into programs that use special free variables $\mathbb{F}$ and $\mathbb{Y}$ as functions for forking and yielding respectively. The translation $\langle\!\langle e \rangle\!\rangle$ is very simple. It only changes Cfork and yield and leaves the rest of the program untouched (similar to throw below):

$$\langle\!\langle \mathsf{throw}\ e\ \mathsf{to}\ e' \rangle\!\rangle \triangleq \mathsf{throw}\ \langle\!\langle e \rangle\!\rangle\ \mathsf{to}\ \langle\!\langle e' \rangle\!\rangle$$

$$\langle\!\langle \mathsf{yield} \rangle\!\rangle \triangleq \mathbb{Y}\ () \qquad \langle\!\langle \mathsf{Cfork}\ \{e\} \rangle\!\rangle \triangleq \mathbb{F}\ (\lambda\_.\ \langle\!\langle e \rangle\!\rangle)$$

Notice that for Cfork we use a λ to turn e into a thunk.

Lemma 7.1 (Typing of translation). *Let e be a program in $\mathsf{F}^{\mu,ref}_{cc,coop}$ such that $\Xi \mid \Gamma \vdash_{\mathsf{coop}} e : \tau$. The following typing judgement holds for the translation of e.*

$$\Xi \mid \Gamma, \mathbb{Y} : \mathbf{1} \to \mathbf{1}, \mathbb{F} : (\mathbf{1} \to \mathbf{1}) \to \mathbf{1} \vdash_{\mathsf{seq}} \langle\!\langle e \rangle\!\rangle : \tau$$

Correctness of translation. We will show correctness of our translation by showing observational refinement. To this end, we define propositions $e \Downarrow_{\mathsf{seq}}$ and $e \Downarrow_{\mathsf{coop}}$, which state when programs of $\mathsf{F}^{\mu,ref}_{cc}$ and $\mathsf{F}^{\mu,ref}_{cc,coop}$ terminate:

$$e \Downarrow_{\mathsf{seq}} \triangleq \exists v, \sigma.\ (e; \emptyset) \to^*_{\mathsf{seq}} (v; \sigma)$$

$$e \Downarrow_{\mathsf{coop}} \triangleq \exists \vec{e}, \sigma.\ (e; 0; \emptyset) \to^*_{\mathsf{coop}} (\vec{e}; n; \sigma) \land e_n \text{ is a value}$$

Notice that programs of $\mathsf{F}^{\mu,ref}_{cc}$ terminate whenever the current thread at the time has terminated.

Theorem 7.2 (Correctness of translation). *Let e be a closed program of type τ, $\cdot \mid \cdot \vdash_{\mathsf{coop}} e : \tau$. The following holds*

$$\textit{if}\ \mathfrak{Comp}(e) \Downarrow_{\mathsf{seq}}\ \textit{then}\ e \Downarrow_{\mathsf{coop}}$$

Intuitively, this theorem expresses that if the compiled program $\mathfrak{Comp}(e)$ produces a result so would the original program (standard observational refinement). We prove this theorem by setting up a cross-language logical relation between $\mathsf{F}^{\mu,ref}_{cc}$ and $\mathsf{F}^{\mu,ref}_{cc,coop}$. We show that the translation $\langle\!\langle e \rangle\!\rangle$ is suitably related to e. Theorem 7.2 then follows essentially because $\mathfrak{Comp}(e)$ reduces to $\langle\!\langle e \rangle\!\rangle[Frk/\mathbb{F}][Yld/\mathbb{Y}]$.

7.3 Cross-language logical relation

The types and basic terms of both $\mathsf{F}^{\mu,ref}_{cc,coop}$ and $\mathsf{F}^{\mu,ref}_{cc}$ are the same as $\mathsf{F}^{\mu,ref}_{conc,cc}$. Hence, our cross-language logical relation is very similar to the logical relation presented in Section 3, except that it is defined for pairs of expressions, values and evaluation contexts where the first component pertains to $\mathsf{F}^{\mu,ref}_{cc}$ and the second to $\mathsf{F}^{\mu,ref}_{cc,coop}$. In fact, the only part of the logical relation in Section 3 that we need to change is the definition of observational refinement. To define the observational refinement relation between $\mathsf{F}^{\mu,ref}_{cc}$ and $\mathsf{F}^{\mu,ref}_{cc,coop}$ we need to introduce resources that we may use for keeping track of the execution on the specification side, *i.e.*, the $\mathsf{F}^{\mu,ref}_{cc,coop}$ side. For this purpose, in addition to propositions $\ell \mapsto_s v$ and $j \Rightarrow e$ for keeping track of the heap and the (light-weight) threads on the specification side, we need a proposition to keep track of the current thread on the specification side. Hence, we introduce the proposition $\mathit{CurTh}(j)$, for asserting that the current thread being run is thread j. For all the basic terms of $\mathsf{F}^{\mu,ref}_{cc,coop}$, the rules for execution on the specification side remain similar to the ones in $\mathsf{F}^{\mu,ref}_{conc,cc}$, except that they require the thread being evaluated to be the current thread. As an example, the rule for storing a value in a reference on the specification side is given below. The only rules that change the current thread are those pertaining to Cfork and yield.

$$\frac{\ell \mapsto_s v \qquad \mathit{CurTh}(j) \qquad j \Rightarrow K[\ell \leftarrow w]}{\Rightarrow \mathit{CurTh}(j) * \ell \mapsto_s w * j \Rightarrow K[()]} \qquad \frac{\mathit{CurTh}(j) \qquad j \Rightarrow K[\mathsf{Cfork}\ \{e\}]}{\Rightarrow \exists j'.\ \mathit{CurTh}(j') * j \Rightarrow K[()] * j' \Rightarrow e}$$

$$\frac{\mathit{CurTh}(j) \qquad j \Rightarrow K[\mathsf{yield}]}{\Rightarrow \mathit{CurTh}(j) * j \Rightarrow K[()]} \qquad \frac{\mathit{CurTh}(j) \qquad j \Rightarrow K[\mathsf{yield}] \qquad j' \Rightarrow e'}{\Rightarrow \mathit{CurTh}(j') * j \Rightarrow K[()] * j' \Rightarrow e'}$$

Note that there are two rules pertaining to the execution of yield as it may result in continuing the execution of the same thread.

For our cross-language logical relations we define the observational refinement relation, O^{cross} as follows:

$$O^{\text{cross}}(e, e') \triangleq \forall j.\ \mathit{CurTh}(j) * j \Rightarrow e' \mathrel{-\!\!*} \text{wp}\ e\ \{\exists j', w.\ j' \Rightarrow w * \mathit{CurTh}(j')\}$$

Note how the specification side is expected to be in a thread that has reached a value in its execution, similarly to how we defined $\Downarrow_{\text{coop}}$ above.

7.4 Proof of correctness of translation

In order to prove correctness of the translation we need to reason about a relation between the internal state of the LiThr library and threads on the specification side. In particular, for each continuation K stored in the internal queue of LiThr there must be a thread $j' \Rightarrow e'$ on the specification side such that $K[()]$ observationally refines e'. We use the proposition $\mathit{LiThrInv}$ for

this purpose.[8]

$$LiThrInv \triangleq \boxed{\exists l.\ isQueue(Q, l) * \mathop{\Large *}_{K \in l} \exists j', e',\ j' \Rrightarrow e' * \square(O^{\text{cross}}(K[()], e'))}^{\mathcal{N}.\text{LiThr}}$$

The proposition $isQueue(Q, l)$ asserts that Q is a queue whose contents are the list l. The two operations of the queue, *enqueue* and *dequeue*, have the following specs:

$\{isQueue(Q, l)\}^{\text{cl}}$
$\quad enqueue\ Q\ v$
$\{x.\ x = () * isQueue(Q, v :: l)\}$

$\{isQueue(Q, l)\}^{\text{cl}}$
$\quad dequeue\ Q$
$\{x.\ (\exists l'.\ l = l' +\!\!+ [v] * x = Some(v) * isQueue(Q, l')) \vee (x = None * l = [] * isQueue(Q, l))\}$

where $+\!\!+$, $::$ and $[\cdot]$ are the usual list operations and [] is the empty list.

To prove the desired correctness theorem (Theorem 7.2), we now prove the fundamental theorem for our cross-language logical relation. It says that, under the assumption that the internal state of the library LiThr is appropriate, any well-typed $\mathsf{F}^{\mu,ref}_{cc,coop}$ program is refined by its translation when linked with the *LightTh* library.

Theorem 7.3 (Fundamental theorem of cross-language logical relations). *Let e be a program in $\mathsf{F}^{\mu,ref}_{cc,coop}$ such that $\Xi \mid \Gamma \vdash_{\text{coop}} e : \tau$. Then we have*

$$LiThrInv \Rightarrow \Xi \mid \Gamma \vDash \langle\!\langle e \rangle\!\rangle[Frk, Yld/\mathbb{F}, \mathbb{Y}] \leq_{log} e : \tau$$

Proof. By induction on the derivation of $\Xi \mid \Gamma \vdash_{\text{coop}} e : \tau$. All cases, except for Cfork and yield follow similarly to their counterpart in the proof of Theorem 3.1. Cases Cfork and yield follow by the fact that their translations are applications to $\mathbb{F}$ and $\mathbb{Y}$ respectively under the assumption that the invariant *LiThrInv* holds for the internal queue of LiThr. □

Theorem 7.2 now follows from Theorem 7.3 (in the same way that Theorem 3.2 followed from Theorem 3.1), using and the fact that the program $\mathfrak{Comp}(e)$ reduces to $\langle\!\langle e \rangle\!\rangle[Frk, Yld/\mathbb{F}, \mathbb{Y}]$.

8 MECHANIZATION IN COQ

Taking advantage of the Coq formalization of Iris and Iris Proof Mode (IPM) [Krebbers et al. 2017b], we have mechanized all the technical development and results presented in this paper in Coq. This includes mechanizing the small-step operational semantics of $\mathsf{F}^{\mu,ref}_{conc,cc}$, $\mathsf{F}^{\mu,ref}_{cc}$ and $\mathsf{F}^{\mu,ref}_{cc,coop}$, and instantiating Iris with them. Our Coq development is about 15800 lines and includes proofs of contextual refinements for pairs of fine-grained/coarse-grained stacks and counters which we omitted discussion of for reasons of space.

For binders, we use the Autosubst library [Schäfer et al. 2015] which facilitates the use of de Bruijn indices by providing support for simplification of substitutions. In $\mathsf{F}^{\mu,ref}_{conc,cc}$, $\mathsf{F}^{\mu,ref}_{cc}$ and

[8] This is a slight simplification. The proposition *LiThrInv* is defined using Iris's non-atomic invariants. These are invariants that can be kept open for multiple steps of computation. They are admissible in our system because we have no real concurrency causing racy behavior. It is crucial to be able to keep the *LiThrInv* open for multiple steps so as to prove that the *dequeue* operation in *Yld* never returns a *None* value. (If the languages had included concurrent racy behaviour, then we could have used a lock in the library and then we would have been able to use standard Iris atomic invariants.) The definition of O^{cross} is also slightly simplified here. It should be slightly adjusted to allow for the use of non-atomic invariants. Iris's weakest preconditions by default only allow atomic invariants.

$\mathsf{F}^{\mu,ref}_{cc,coop}$, evaluation contexts are also values and hence also expressions. This forces us to define these mutually inductively. This means that we need to derive the induction principle for these inductive types in Coq by hand. Furthermore, we have to help Autosubst in deriving substitution and simplification lemmas for $\mathsf{F}^{\mu,ref}_{conc,cc}$ that it should otherwise automatically infer. This is mainly why the definition of $\mathsf{F}^{\mu,ref}_{conc,cc}$, $\mathsf{F}^{\mu,ref}_{cc}$ and $\mathsf{F}^{\mu,ref}_{cc,coop}$ combined takes up about 20% of the whole Coq development.

9 RELATED WORK

There has been a considerable body of work on (delimited) continuations, but, we are not aware of any logics or relational models for reasoning about concurrent programs with continuations, let alone a mechanized framework for relational verification of concurrent programs with continuations.

Program logics for reasoning about continuations. Delbianco and Nanevski [2013] present a type theory for Hoare-style reasoning about an imperative higher-order programming language with (algebraic) continuations, but without concurrency. The system of Delbianco and Nanevski [2013] does not allow higher-order code (including continuations) to be stored in the heap. Note that storing higher-order code in the heap is essential for all our case studies: implementing cooperative concurrency with continuations, implementing the continuation-based web servers and implementing continuations in terms of one-shot continuations. Crolard and Polonowski [2012] develop a program logic for reasoning about jumps but their sequential programming language features no heap or recursive types. Berger [2010] presents a program logic for reasoning about programs in a programming language which is essentially an extension of PCF [Plotkin 1977] with continuations.

Relational reasoning about continuations. The work most closely related to ours is that of Dreyer et al. [2012] who consider a variety of different stateful programming languages and investigate the impact of the higher-order state and control effects (including call/cc and throw). In contrast to our work, they do not consider concurrency. Moreover, they reason directly in a model, whereas we define our logical relation using a program logic (Iris), which means that we can reason more compositionally and at a higher level of abstraction. Another advantage of using Iris, is that we have been able to leverage its Coq formalization and thus to mechanize all of our development. As mentioned in Section 6, our proof that continuations can be expressed in terms of one-shot continuations is inspired by loc. cit.

There are several other works on relational reasoning for sequential programming languages with continuations, *e.g.*, Felleisen and Hieb [1992]; Laird [1997]; Støvring and Lassen [2007]. These differ from our work at least in that they do not consider concurrency.

Relational reasoning about concurrency. There has been much work on relational reasoning about concurrent higher-order imperative programs, without continuations. The work most closely related to ours also is that of Krebbers et al. [2017b], who develop mechanized logical relations (in Iris) for reasoning about contextual equivalence of programs in $\mathsf{F}^{\mu,ref}_{conc}$, a language similar to the one we consider but without call/cc and throw. The approach in *loc. cit.* is based on earlier, non-mechanized logical relations for fine-grained concurrent programs [Birkedal et al. 2012; Turon et al. 2013a,b]. These relational models give an alternative method to linearizability [Herlihy and Wing 1990] for reasoning about contextual refinement for fine-grained concurrent programs. The logical relations method also works in the presence of higher-order programs, which linearizability traditionally struggles with, although there has been some recent promising developments [Cerone et al. 2014; Murawski and Tzevelekos 2017]. In this paper, we have extended the method of logical relations

for reasoning about contextual refinement for higher-order fine-grained concurrent programs to work for programs that also use continuations.

Correctness of compilation of cooperative concurrency. The only other work that we are aware of which proves correctness of compilation of cooperative concurrency is by Nakata and Saar [2013]. Nakata and Saar [2013] compile a simple while language extended with procedures, first-order store (only integers can be stored) and cooperative concurrency. Their target language is again a simple while language extended with procedures, delimited continuations and a store that can store in each cell either an integer value or a captured continuation. They use a syntactic proof of correctness as opposed to our semantic (using a logical relations model) proof. Such syntactic proof methods do not scale to higher-order programming languages with advanced features like higher-order store and impredicative polymorphism, *e.g.*, $\mathsf{F}^{\mu,ref}_{conc,cc}$ in our case.

10 CONCLUSION AND FUTURE WORK

We have developed a logical relation for $\mathsf{F}^{\mu,ref}_{conc,cc}$, a programming language with advanced features such as impredicative polymorphism à la system F, higher-order mutable references, recursive types, concurrency and most notably continuations. We have devised new non-context-local proof rules for reasoning about weakest preconditions in Iris in the presence of continuations and also introduced context-local weakest preconditions for regaining context-local reasoning about expressions that do not involve non-local control flow. We have defined our relational model and proved properties thereof in the Iris program logic framework. This has greatly simplified the definition of our relational model, the existence of which is non-trivial because of the type-world circularity [Ahmed 2004; Ahmed et al. 2002; Birkedal et al. 2011]. Furthermore, working inside Iris has enabled us to mechanize the entire development presented in this paper on top of the Coq proof assistant.

We have demonstrated how our logical relation can be used to establish contextual equivalence for a pair of simplified web-server implementations: one storing the state explicitly and one storing the current continuation. The application of context local reasoning in the middle of our logical relatedness proofs demonstrates the usefulness and versatility of context-local weakest preconditions. Finally, we have also given the first (mechanized) proof of the correctness of Friedman and Haynes [1985] encoding of continuations by means of one-shot continuations in a concurrent programming language.

We developed a cross-language logical relation between $\mathsf{F}^{\mu,ref}_{cc}$ and $\mathsf{F}^{\mu,ref}_{cc,coop}$. We used this logical relation to give a compiler-correctness-inspired proof of correctness of the continuation-based implementation of cooperative concurrency. This is to the best of our knowledge the first formal proof of correctness of continuation-based cooperative concurrency for a programming language with a rich advanced features and types.

In the future, we wish to extend our mechanization to reason about delimited continuations [Danvy and Filinski 1990; Felleisen 1988]. Currently our mechanized reasoning is done interactively, in the same style as one reasons in Coq. In the future, we would also like to complement that with more automated reasoning methods.

ACKNOWLEDGMENTS

The first author is a postdoctoral fellow of the Flemish research fund (FWO). This project was supported in part by the FWO grant (grant no. G.0962.17N), the FWO travel grant (V435817N), the EU Types (CA15123) short scientific mission (STSM) grant (reference: 40667) and by the ModuRes Sapere Aude Advanced Grant from The Danish Council for Independent Research for the Natural Sciences (FNU).

REFERENCES

Amal Ahmed. 2004. *Semantics of Types for Mutable State.* Ph.D. Dissertation. Princeton University.

Amal J. Ahmed, Andrew W. Appel, and Roberto Virga. 2002. A Stratified Semantics of General References Embeddable in Higher-Order Logic. In *Proceedings of 17th Annual IEEE Symposium Logic in Computer Science*. IEEE Computer Society Press, 75–86.

Andrew Appel and David McAllester. 2001. An Indexed Model of Recursive Types for Foundational Proof-Carrying Code. *TOPLAS* 23, 5 (2001), 657–683.

Andrew Appel, Paul-André Melliès, Christopher Richards, and Jérôme Vouillon. 2007. A Very Modal Model of a Modern, Major, General Type System. In *POPL*.

Martin Berger. 2010. *Program Logics for Sequential Higher-Order Control.* Springer Berlin Heidelberg, Berlin, Heidelberg, 194–211.

Lars Birkedal, Bernhard Reus, Jan Schwinghammer, Kristian Støvring, Jacob Thamsborg, and Hongseok Yang. 2011. Step-Indexed Kripke Models over Recursive Worlds. In *POPL*.

Lars Birkedal, Filip Sieczkowski, and Jacob Thamsborg. 2012. A Concurrent Logical Relation. In *CSL*.

Andrea Cerone, Alexey Gotsman, and Hongseok Yang. 2014. Parameterised Linearisability. In *ICALP*.

T. Crolard and E. Polonowski. 2012. Deriving a Floyd-Hoare logic for non-local jumps from a formulæ-as-types notion of control. *The Journal of Logic and Algebraic Programming* 81, 3 (2012), 181 – 208. The 22nd Nordic Workshop on Programming Theory (NWPT 2010).

Pedro da Rocha Pinto, Thomas Dinsdale-Young, and Philippa Gardner. 2014. TaDA: A Logic for Time and Data Abstraction. In *ECOOP*. 207–231.

Olivier Danvy and Andrzej Filinski. 1990. Abstracting Control. In *Proceedings of the 1990 ACM Conference on LISP and Functional Programming*.

Germán Andrés Delbianco and Aleksandar Nanevski. 2013. Hoare-style reasoning with (algebraic) continuations. *ACM SIGPLAN Notices* 48, 9 (2013), 363–376.

Thomas Dinsdale-Young, Lars Birkedal, Philippa Gardner, Matthew Parkinson, and Hongseok Yang. 2013. Views: Compositional Reasoning for Concurrent Programs. In *POPL*.

T. Dinsdale-Young, M. Dodds, P. Gardner, M. Parkinson, and V. Vafeiadis. 2010. Concurrent abstract predicates. In *ECOOP*. 504–528.

D. Dreyer, A. Ahmed, and L. Birkedal. 2011. Logical Step-Indexed Logical Relations. *LMCS* 7, 2:16 (2011).

Derek Dreyer, Georg Neis, and Lars Birkedal. 2012. The impact of higher-order state and control effects on local relational reasoning. *Journal of Functional Programming* 22, 4-5 (2012), 477–528.

Mattias Felleisen. 1988. The Theory and Practice of First-class Prompts. In *Proceedings of the 15th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL)*.

Matthias Felleisen and Robert Hieb. 1992. The revised report on the syntactic theories of sequential control and state. *Theoretical Computer Science* 103, 2 (1992), 235 – 271.

Matthew Flatt. 2017. More: Systems Programming with Racket. https://docs.racket-lang.org/more/index.html.

Daniel P. Friedman and Christopher T. Haynes. 1985. Constraining Control. In *Proceedings of the 12th ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (POPL '85)*. ACM, New York, NY, USA, 245–254.

Christopher T. Haynes, Daniel P. Friedman, and Mitchell Wand. 1984. Continuations and Coroutines *(LFP '84)*.

Greg Hendershott. 2017. http://www.greghendershott.com/2014/09/written-in-racket.html.

Maurice P. Herlihy and Jeannette M. Wing. 1990. Linearizability: a correctness condition for concurrent objects. *TOPLAS* 12, 3 (1990), 463–492.

Ralf Jung, Robbert Krebbers, Lars Birkedal, and Derek Dreyer. 2016. Higher-order ghost state. In *ICFP*. 256–269.

RALF JUNG, ROBBERT KREBBERS, JACQUES-HENRI JOURDAN, ALEŠ BIZJAK, LARS BIRKEDAL, and DEREK DREYER. 2018. Iris from the ground up: A modular foundation for higher-order concurrent separation logic. *Journal of Functional Programming* 28 (2018), e20. https://doi.org/10.1017/S0956796818000151

Ralf Jung, David Swasey, Filip Sieczkowski, Kasper Svendsen, Aaron Turon, Lars Birkedal, and Derek Dreyer. 2015. Iris: Monoids and Invariants as an Orthogonal Basis for Concurrent Reasoning. In *POPL*. 637–650.

Robbert Krebbers, Ralf Jung, Aleš Bizjak, Jacques-Henri Jourdan, Derek Dreyer, and Lars Birkedal. 2017a. The essence of higher-order concurrent separation logic. In *European Symposium on Programming (ESOP)*.

Robbert Krebbers, Amin Timany, and Lars Birkedal. 2017b. Interactive Proofs in Higher-Order Concurrent Separation Logic. In *POPL*.

Shriram Krishnamurthi, Peter Walton Hopkins, Jay McCarthy, Paul T Graunke, Greg Pettyjohn, and Matthias Felleisen. 2007. Implementation and use of the PLT Scheme web server. *Higher-Order and Symbolic Computation* 20, 4 (2007), 431–460.

Morten Krogh-Jespersen, Kasper Svendsen, and Lars Birkedal. 2017. A Logical Account of a Type-and-Effect System. In *POPL*.

James Laird. 1997. Full Abstraction for Functional Languages with Control. In *Proceedings of the 12th Annual IEEE Symposium on Logic in Computer Science (LICS '97)*. IEEE Computer Society, Washington, DC, USA, 58–. http://dl.acm.org/citation.cfm?id=788019.788859

Ruy Ley-Wild and Aleksandar Nanevski. 2013. Subjective Auxiliary State for Coarse-Grained Concurrency. In *POPL*.

Matt Might. 2017. http://matt.might.net/articles/low-level-web-in-racket/.

Andrzej S. Murawski and Nikos Tzevelekos. 2017. Higher-Order Linearisability. In *CONCUR 2017*.

Keiko Nakata and Andri Saar. 2013. Compiling Cooperative Task Management to Continuations. In *Fundamentals of Software Engineering*, Farhad Arbab and Marjan Sirjani (Eds.).

Aleksandar Nanevski, Ruy Ley-Wild, Ilya Sergey, and Germán Andrés Delbianco. 2014. Communicating State Transition Systems for Fine-Grained Concurrent Resources. In *ESOP*.

Peter W. O'Hearn. 2007. Resources, Concurrency and Local Reasoning. *Theor. Comput. Sci.* 375, 1-3 (2007), 271–307.

Andrew M. Pitts. 2005. Typed Operational Reasoning. In *Advanced Topics in Types and Programming Languages*, B. C. Pierce (Ed.). The MIT Press, Chapter 7, 245–289.

Gordon D. Plotkin. 1977. LCF considered as a programming language. *Theoretical computer science* 5, 3 (1977), 223–255.

Christian Queinnec. 2004. Continuations and web servers. *Higher-Order and Symbolic Computation* 17, 4 (2004), 277–295.

Steven Schäfer, Tobias Tebbi, and Gert Smolka. 2015. Autosubst: Reasoning with de Bruijn Terms and Parallel Substitutions. In *ITP (LNCS)*, Vol. 9236. 359–374.

Ilya Sergey, Aleksandar Nanevski, and Anindya Banerjee. 2015. Mechanized verification of fine-grained concurrent programs. In *PLDI*. 77–87.

Kristian Støvring and Soren Lassen. 2007. A Complete, Co-Inductive Syntactic Theory of Sequential Control and State. In *POPL*.

Kasper Svendsen and Lars Birkedal. 2014. Impredicative Concurrent Abstract Predicates. In *ESOP*. 149–168.

Amin Timany, Léo Stefanesco, Morten Krogh-Jespersen, and Lars Birkedal. 2018. A Logical Relation for Monadic Encapsulation of State: Proving contextual equivalences in the presence of runST. *Proc. ACM Program. Lang.* 2, POPL (Jan. 2018), to appear.

Aaron Turon, Derek Dreyer, and Lars Birkedal. 2013a. Unifying refinement and Hoare-style reasoning in a logic for higher-order concurrency. In *ICFP*.

Aaron Turon, Jacob Thamsborg, Amal Ahmed, Lars Birkedal, and Derek Dreyer. 2013b. Logical relations for fine-grained concurrency. In *POPL*.