Verifying an Effect-Based Cooperative Concurrency Scheduler in Iris

Adrian Dapprich
Department of Computer Science
Saarland University

Advisors: Prof. Derek Dreyer & Prof. François Pottier

March 16, 2024

Contents

1 Introduction (WIP)

- As a motivation for the work: program verification, **safety** and why we care about it.
- Iris is a new separation logic which allows proving safety for programs using mutable shared state.
- Many programs nowadays user user-level concurrency to handle a big number of tasks. As an example for OCaml 5 there exists the Eio library which provides concurrency primitives using effect handlers.
- Effect handlers are a versatile concept which allow a modular treatment of effects, the implementation in form of a handler is separated from the code using the effect, and it's more lightweight than monads. Give a simple example of state.
  - The biggest upside is that they are more composable than monads which often require rewriting of parts of the program into monadic style
  - In theory effect can be tracked by the type system, although OCaml 5 does not do that yet.
  - Explain the concept of **effect safety** here.
  - Explain how performing and handling effects is implemented using delimited continuations.
  - Mention that continuations can only be invoked once? (not really necessary info)
- We want to verify some parts of the Eio library but the standard Heaplang language for Iris does not support effect handlers.
  - Hazel is an Iris language formalizing effect handlers using protocols.
  - Syntax and semantics of protocols.
  - Since OCaml 5 allows both effect handlers and mutable shared state we had to add a multi-threaded semantics to Hazel.
- Inherent part of a scheduler is liveness, because it is responsible for running all fibers to completion. Unfortunately it is hard to prove liveness properties in Iris, so we just focus on safety and effect safety.

1.1 The Eio Library (WIP)

- Library for cooperative concurrency in OCaml 5.
- Implements switching between tasks using effect handlers.
- A fiber is a normal OCaml function which may perform effects that are handled by a scheduler.
- Each scheduler is only responsible for a single thread, more can be spawned.
- It offers abstractions to operating system resources to fibers, e.g. network, file system, timers etc.
- It also offers synchronization and message passing constructs like mutexes & channels which are specialized to handle fibers, i.e. a mutex does not suspend the system-level thread, but the fiber.

1.2 Focus and Structure of the Thesis

Eio aims to be the standard cooperative concurrency library for OCaml 5, so it includes many functions for structured concurrency of fibers (e.g. `Fiber.{first, any, both, all}`, which run two or more fibers and combine their results), support for cancelling fibers, abstractions for operating system resources, a different scheduler implementation per OS, and synchronization constructs like promises and mutexes. But for this work we restrict ourselves to verifying the safety and effect safety of Eio's core functionalities:

1. Running fibers in a "common denominator" scheduler that does not interact with any OS resources,
2. awaiting the result of other fibers using the *promise* synchronization construct,

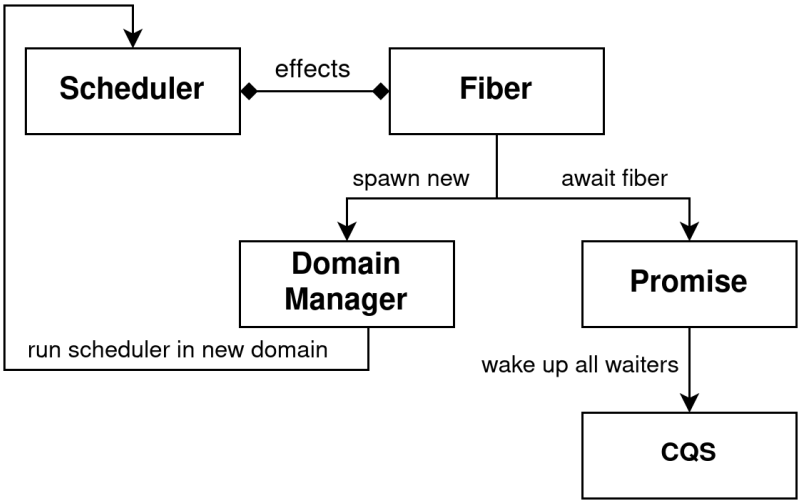

Figure 1: Eio Module Hierarchy

3. and spawning new schedulers to run fibers in another thread.

Figure 1 shows the simplified module hierarchy of the concepts we focus on. A standard arrow stands for a direct source code dependency from one module to another. The diamond arrow between `Scheduler` and `Fiber` stands for the implicit dependency that code in the fiber module performs effects that are handled by code in the scheduler module.

Fibers can fork off new fibers using the *Fork* effect and suspend execution using the *Suspend* effect, which are both handled by the scheduler. The implementation of the fiber and scheduler functions are discussed in section 2.1. *Promises* are built on top of the *CQS* data structure, which is a lock-free condition variable that is used by fibers to suspend execution until a promise is fulfilled. The specification of promises is discussed in section 2.2. The CQS specification is already verified using Iris, but Eio uses a custom implementation for which we had to adapt the proof and we discuss this process in section **??**. Fibers in Eio also have access to *thread-local variables* by performing a *GetContext* effect, which is discussed in section 4. They are thread-local in the sense that they are shared between all fibers of one scheduler. Finally, we discuss our addition of multi-threading to the Hazel operational semantics in order to model running schedulers in different threads. This turned out to be technically trivial, so we only discuss it in appendix B and take a multithreaded semantics and support for Iris *shared invariants* as a given in the reminder of the main text.

1.3 Contributions

To summarize our contributions, in this thesis we verify the **safety** and **effect safety** of a simplified model of Eio which serves as an extended case study on the viability of Hazel for verifying programs with effect handlers. This includes:

- The verification of the basic Eio **fiber abstraction** running on a common denominator scheduler.
- An adaptation of the existing verification of CQS to the customized version used by Eio.
- Adding multi-threading to Hazel's operational semantics, which shows we can reason about programs that use both **multi-threading** and **effect handlers**.

2 Verifying a Simplified Eio Scheduler With Promises

Mention that all code examples are an OCaml rendering of the verified Hazel code, based on but not equal to the Eio code

Cooperative concurrency schedulers for user-level threads (i.e. *fibers*) are commonly treated in the literature on effect handlers [2, 4, 5] because they are a lucid example for their usefulness. Generally, the architecture contains an effect handler as the scheduler and fibers are normal functions which perform effects to yield execution. This is because performing an effect causes execution to jump to the enclosing effect handler (i.e. the scheduler), providing it with the rest of the fiber's computation in the form of a delimited continuation. The scheduler keeps track of a collection of these continuations and by invoking one of them the next fiber is scheduled. This approach is also used in Eio.

We can therefore use the simple cooperative concurrency scheduler case study from the dissertation of de Vilhena [1] as a starting point for our verification work. In the following section we first discuss the implementation of our simplified model of Eio in more detail. Using this implementation we give an intuition about what specifications the functions should satisfy and what kind of logical state is needed to prove these specifications. Based on this intuition we will then build a formalization in section 2.2.

2.1 Implementation

Let us first get an idea of how different components of the core Eio fiber abstraction interact by looking at their types. `Scheduler.run`[1] is the main entry point to Eio. It runs a scheduler and is provided a function which represents the first fiber to be executed. A scheduler runs the main fiber and all forked-off fibers in a single thread. However, a fiber can also spawn new schedulers in separate threads to run other fibers in parallel as detailed in appendix B. The `Fiber.fork_promise` function is used to spawn fibers in the current scheduler. The function returns a promise holding the eventual return value of the new fiber. The promise is thread-safe so that it can be shared with fibers running in different threads. The `Promise.await` function can be used by any fiber to wait until the value of a promise is available. Common problems like deadlocks are not prevented in any way and are the responsibility of the programmer.

```
(* Basic interface of the Eio library. *)
Scheduler.run : (() -> 'a) -> 'a option
Fiber.fork_promise : (() -> 'a) -> 'a Promise.t
Promise.await : 'a Promise.t -> 'a
```

We present code examples in a simplified syntax[2] because the concrete syntax of effect handlers in OCaml 5 is verbose. We use an overloading of the `match` expression, which includes cases for handled effects, that is common in the literature.

```
(* declares an effect E that carries an int and has a bool return value. *)
effect E : int -> bool

(* Evaluates the expression e and if the effect E is performed
 * control is transferred to the second branch.
 * The continuation k captures the rest of the computation of e.
 * The match acts as a deep handler, i.e. even if during the
 * evaluation of e the effect E is performed multiple times, the
 * second branch will be evaluated every time.
 * When e is reduced to a value, the non-effect branches are
 * used for pattern matching as usual. *)
match e with
| v -> ...
| effect (E v) k -> ...
```

2.1.1 `Scheduler.run`

As mentioned above this is the main entry point to the Eio library and its code is shown in figure 2. It sets up the scheduler environment and runs the main fiber that is passed as an argument.

The `run_queue` contains closures that will immediately invoke the continuation of an effect. This represents ready fibers which can continue execution from the point where they performed an effect. The `next` function pops one fiber (i.e. function) from the `run_queue` and executes it. If no more ready fibers

[1] The scheduler's result is optional because the main fiber might deadlock.

[2] This syntax is planned to be implemented in OCaml 5 in the future: https://github.com/ocaml/ocaml/pull/12309

```
effect Fork : (() -> 'a) -> ()
type 'a waker : 'a -> ()
effect Suspend : ('a waker -> ()) -> 'a

let run (main : () -> 'a) : option 'a =
  let run_queue = Queue.create () in
  let next () =
    match Queue.pop run_queue with
    | None -> ()
    | Some cont -> cont ()
  let rec execute fiber =
    match fiber () with
    | () -> next ()
    | effect (Fork fiber) k ->
      Queue.push run_queue (fun () -> invoke k ());
      execute fiber
    | effect (Suspend register) k =>
      let waker = fun v -> Queue.push run_queue (fun () -> invoke k v) in
      register waker;
      next ()
  in
  let result = ref None in
  execute (fun () -> result := main ());
  !result
```

Figure 2: Implementation of `Scheduler.run`

remain – either because all fibers terminated or there is a deadlock – the next function just returns and the scheduler exits.

The inner `execute` function is called once on each fiber to evaluate it and handle any performed effects.

Value Case

The only non-effect case of the `match` just runs the next fiber because there are two types of fibers and their return value is always `()`.

- The main fiber is wrapped in a saves its return value in a reference and returns `()`.
- All other fibers are forked using `Fiber.fork_promise`, which wraps them in a closure that saves their return value in a promise and returns `()`.

This emphasizes the fact that an Eio scheduler is only used for running fibers. The interaction between fibers waiting for values of other fibers is handled separately by promises.

Fork Case

Handling a *Fork* effect is simple because it carries a new `fiber` to be executed, so the handler recursively calls the `execute` function to execute it immediately. The execution of the original fiber is paused due to performing an effect and its continuation `k` is placed in the run queue so that it can be scheduled again. This prioritizes the execution of a new fiber and is a design decision by Eio. It would be equally valid to place the `fiber` argument in the run queue instead.

Suspend Case

Handling a *Suspend* effect may look complicated at first due to the higher-order `register` function. This effect is used by fibers to suspend execution until a condition is met. The fiber defines this condition by constructing a `register` function which in turn receives a wake-up capability by the scheduler in form of the `waker` function. The key point is that as long as the continuation `k` is not invoked, the fiber will not continue execution. The `waker` function places `k` into the run queue so that the fiber can continue execution by a call to the scheduler's `next` function. The `register` function is called by the scheduler right after the fiber suspends execution and is responsible for installing `waker` as a callback at a suitable

place (or even call it directly). For example, to implement promises, the `waker` function is installed in a data structure that will call `waker` after the promise is fulfilled.

Note that the `waker` function's argument `v` has a *locally abstract type*, which is a typical pattern in effect handlers. From the point of view of the fiber, the polymorphic type `'a` of the *Suspend* effect is instantiated depending on how the effect's return value is used. But the scheduler does not get any information about this so the argument type of the continuation `k` and the `waker` function is abstract.

Waking up must be possible across thread boundaries, which is why the `run_queue` in the scheduler is thread-safe queue. For the verification we assume the specification of a suitable `Queue` module that supports thread-safe push and pop operations.

2.1.2 `Fiber.fork_promise`

```
(* promise.ml *)
let fulfill p result =
  match Atomic.get p with
  | Done _ -> error "impossible"
  | Waiting bcst ->
      Atomic.set p (Done result);
      Broadcast.signal_all bcst

(* fiber.ml *)
let fork_promise (f : () -> 'a) : 'a Promise.t =
  let p = Promise.create () in
  let fiber = fun () ->
    let result = f () in
    Promise.fulfill p result
  in
  perform (Fork fiber)
  p
```

Figure 3: Implementation of `Fiber.fork_promise`

This function is the basic way to create a new fiber in Eio and the only one we model in our development. The code is presented in figure 3. It will create a promise and spawn the provided function as a new fiber using the *Fork* effect. When `f` is reduced to a value `result`, it will fulfill the promise with that value and signal all fibers waiting for that result to wake up. The major difference to the implementation of de Vilhena is that promises in Eio are entirely handled by the fiber, and not in the effect handler code of the scheduler. This achieves a better separation of concerns and simplifies the logical state needed for the proof.

2.1.3 `Promise.await`

This is the most complicated looking function in our development which is partly due to the *Suspend* effect and also due to the use of *broadcast* functions. Its code is presented in figure 4. The purpose of `Promise.await p` is to suspend execution of the calling fiber until `p` is fulfilled with a value and then return this value. The "suspend execution" part is handled by performing a *Suspend* effect. Then, the "until `p` is fulfilled" part is implemented by using a *broadcast* data structure.

In Eio, a *broadcast* is an implementation of the observer pattern and functionally similar to condition variables[3] in languages like C++ (as defined by the POSIX standard), allowing fibers to register callbacks that will be called when a condition is signalled. The difference is that traditional condition variables are always used together with a mutex to enable synchronization between different threads, broadcast is a lock-free data structure implementing a similar API.

In figure 5 we show the public API of the `Broadcast` module. The `Broadcast.register` function registers a given callback with the data structure while `Broadcast.signal_all` calls all registered callbacks. The return value of `Broadcast.register` depends on whether the function detects a parallel execution of `Broadcast.signal_all`. If a parallel execution is detected, `Broadcast.register` will directly call the

[3] https://en.cppreference.com/w/cpp/thread/condition_variable

```
type 'a t = Done of 'a | Waiting of Broadcast.t

let create () : 'a t =
  let bcst = Broadcast.create () in
  Atomic.create (Waiting bcst)

let make_register (p: 'a t) (bcst: Broadcast.t) : (() waker -> ()) =
  fun waker ->
    let register_result = Broadcast.register bcst waker in
    match register_result with
    | None -> ()
    | Some register_handle ->
      match Atomic.get p with
      | Done result ->
        if Broadcast.try_cancel register_handle
        then waker ()
        else ()
      | Waiting _ -> ()

let await (p: 'a t) : 'a =
  match Atomic.get p with
  | Done result -> result
  | Waiting bcst ->
    let register = make_register p bcst
    perform (Suspend register);
    match Atomic.get p with
    | Done result -> result
    | Waiting _ -> error "impossible"
```

Figure 4: Implementation of `Promise.await`

callback and return `Called`[4]. Otherwise, `Broadcast.register` returns a `Registered handle` value, where `handle` can be used to call the `Broadcast.try_cancel` function. `Broadcast.try_cancel` attempts to cancel a registered callback and returns whether the cancellation was successful. If the cancellation was successful, the previously registered callback will not be called when `Broadcast.signal_all` is executed. The implementation and specification of the functions will be expanded upon in section 3, for now we just explain their usage in the context of `Promise.await`.

```
type callback = () -> ()
type register_result = Called | Registered of register_handle
type register_handle

val create : () -> t
val register : t -> callback -> register_result
val try_cancel : register_handle -> bool
val signal_all : t -> ()
```

Figure 5: Interface of the `Broadcast` module.

In the `Promise.await` function if the promise is not fulfilled initially the fiber should wait until that is the case, so it performs a *Suspend* effect. The `register` function passed to the effect will register the `waker` function using `Broadcast.register`. When at some point the `Broadcast.signal_all` function is called – this happens in `Fiber.fork_promise` – all registered `wakers` will be called in turn. Recall that calling a `waker` function will enqueue the fiber that performed the *Suspend* effect in the scheduler's run queue so that it can continue execution.

In the default case the following simplified chain of events happens:

1. The fiber suspends execution at the point of evaluating `perform (Suspend register)`.

2. The `waker` function is registered with a broadcast.

[4]This is purely an optimization. It would be equally valid for functional correctness to register the `waker` function in the data structure and let `Broadcast.signal_all` call it later.

3. The promise is fulfilled.
4. The `waker` function is called.
5. The fiber resumes execution at the point of evaluating `perform (Suspend register)`.

Therefore, after the *Suspend* effect returns we know the state of the promise is `Done` and the final value can be returned.

But because broadcast is a lock-free data structure and promises can be shared between different threads there are a number of possible interleavings that the `register` function must take care of as well. The definition of the register function is interesting enough that we split it out into `make_register` and give a separate specification, which is not part of the public API of the module. First, there could be a race on the state of the promise itself. Right after the state is read in line 21 of figure 4 another thread might change the state to `Done` and go on to call `Broadcast.signal_all`. If that happens there is another race between the `Broadcast.register` in line 9 and the `Broadcast.signal_all` in the other thread. If `Broadcast.register` detects that there is a racing `Broadcast.signal_all` it will directly call the `waker`. Otherwise, the `waker` is registered but in fact the `Broadcast.signal_all` might have already finished before `Broadcast.register` even started. In this case the `waker` would be "lost" in the broadcast, never to be called. To avoid this, `register` must check the state of the promise again in line 13, and if it is fulfilled try to cancel the `waker` registration. The cancel will fail if the `waker` function was already called. If it succeeds the `register` function has the responsibility of calling `waker` itself, which is done in line 16.

The only **safety** concerns in the above implementation are `Fiber.fork_promise` expecting the promise to be unfulfilled after the fiber has finished execution, and `Promise.await` expecting the promise to be fulfilled in the last match. In both cases, the program would crash (signified by the `error` expression) if the expectation is violated. So to establish the safety of Eio we wish to prove that the expectations always hold, and the two `error` expressions are never reached. In the next section we show how the first situation is addressed by defining a unique resource that is needed to fulfill a promise, and the latter is a consequence of the protocol of the *Suspend* effect.

2.2 Specification

To prove specifications for an effectful program in Hazel we have to define not only ghost state constructs to track program state as usual but also protocols which describe the behavior of the program's effects. For our Eio development we adapt both the ghost state and the effect protocols from the cooperative concurrency scheduler development from chapter 4 of de Vilhena's dissertation [1].

2.2.1 Protocols

First we look at the protocols for the *Fork* and *Suspend* effect that are shown in figure 6 In Hazels' protocol syntax they are formalized in the following way, where the precondition of *Suspend* is given the name *isRegister* to describe the behavior of the fiber-defined `register` function.

$$
\begin{aligned}
\mathit{isWaker}\ \mathit{wkr}\ P &\triangleq \forall v.\ P\ v \mathrel{-\!\!*} \mathit{ewp}\ (\mathit{wkr}\ v)\ \langle\bot\rangle\ \{\top\} \\
\mathit{isRegister}\ \mathit{reg}\ P &\triangleq \forall \mathit{wkr}.\ (\mathit{isWaker}\ \mathit{wkr}\ P) \mathrel{-\!\!*} \triangleright\, \mathit{ewp}\ (\mathit{reg}\ \mathit{wkr})\ \langle\bot\rangle\ \{\top\} \\
\mathit{Coop} \triangleq\ & \mathit{Fork}\ \#\ !\ e\ (e)\ \left\{\triangleright\, \mathit{ewp}\ (e)\ \langle \mathit{Coop}\rangle\ \{\top\}\right\}.?\ ()\ \{\top\} \\
& \mathit{Suspend}\ \#\ !\ \mathit{reg}\ P\ (\mathit{reg})\ \{\mathit{isRegister}\ \mathit{reg}\ P\}.?\ y\ (y)\ \{P\ y\}
\end{aligned}
$$

Figure 6: Definition of *Coop* Protocol with *Fork* & *Suspend* Effects.

The *Fork* effect accepts an arbitrary expression e which represents the computation that a new fiber executes. To perform the effect one must prove that e acts as a function that can be called on unit and obeys the *Coop* protocol itself. This means spawned off fibers can again perform *Fork* and *Suspend* effects. The $\mathit{ewp}\ (e)\ \langle \mathit{Coop}\rangle\ \{\top\}$ is guarded behind a later modality because of the recursive occurrence of the *Coop* protocol. Since promise handling is done entirely in the fibers and the *Fork* effect just hands off the fiber to the scheduler, the protocol is simplified in two ways compared to the original from the case study of de Vilhena. First, the scheduler does not interact with the return value of the fiber, so the *ewp* has a

PS-Create

$$\frac{}{\models\!\!\Rightarrow \exists \gamma.\ \mathit{promiseWaiting}\ \gamma * \mathit{promiseWaiting}\ \gamma}$$

PS-Combine

$$\frac{\mathit{promiseWaiting}\ \gamma * \mathit{promiseWaiting}\ \gamma}{\square\ \mathit{promiseDone}\ \gamma}$$

PS-Contra

$$\frac{\mathit{promiseWaiting}\ \gamma * \mathit{promiseDone}\ \gamma}{\bot}$$

Figure 7: Rules for the *promiseWaiting* γ and *promiseDone* γ.

trivial postcondition. Second, because the scheduler does not create and return the promise, the protocol itself also has a trivial postcondition.

From the type of the *Suspend* effect we already know that some value can be transmitted from the party that calls the `waker` function to the fiber that performed the effect. The *Suspend* protocol now expresses the same idea on the level of resources. To suspend, a fiber must supply a function `register` that satisfies the *isRegister* predicate. This predicate says that `register` must be callable on a `waker` function and in turn gets to assume that the `waker` function is callable on an arbitrary value `v`, which satisfies the predicate `P`. Both `register` and `waker` must not perform effects. The predicate `P` appears twice in the definition of the protocol, once in the precondition of `waker` and then in the postcondition of the whole protocol. It signifies the resources that are transmitted from the party that calls the `waker` function to the fiber that performed the effect.

By appropriately instantiating `P`, we can enforce that some condition holds before the fiber can be signalled to continue execution, and we get to assume the resources `P v` for the rest of the execution. For example, in the `Promise.await` specification below, we ensure that the promise must be fulfilled before the effect returns by instantiating `P` with a resource that says the promise is fulfilled.

2.2.2 Logical State

The most basic ghost state we track is whether a promise is fulfilled or not. If a promise `p` is unfulfilled, two copies of *promiseWaiting* γ exist, one owned by the fiber and one by the invariant that tracks the state of all promises. When fulfilling the promise, both copies can be combined and converted to a persistent *promiseDone* γ resource. The *promiseWaiting* γ and *promiseDone* γ resources cannot exist at the same time. This design allows us to deduce the current state of the promise depending on if we own a *promiseWaiting* γ or a *promiseDone* γ. This is formalized in the rules in figure 7.

Maybe use meta_tokens to hide gamma

Other pieces of ghost state are *PInvInner*, *isPromise*, and *Ready* described in figure 8. *PInvInner* tracks additional resources for all existing promises by using an authoritative map which contains for each promise: a location p holding its current program value, a ghost name γ that is used for the *promiseWaiting* γ and *promiseDone* γ resources, and a predicate Φ that describes the value the promise will eventually hold.

Additionally, for each promise in the map we own some resources as part of *PInvInner* that depend on the current state of the promise. As long as the promise is not fulfilled we own a broadcast, one copy of *promiseWaiting* γ, and a *signalAllPermit*. The *signalAllPermit* is used to call the `Broadcast.signal_all` function which must only be called once. When the promise has been fulfilled, we instead own a *promiseDone* γ and the knowledge that the final value satisfies the given postcondition Φ.

PromiseInv is an invariant that wraps *PInvInner* so that we can share it.

isPromise is a persistent resource that denotes that p exists as a promise in *PInvInner*. The *pn* ghost name is globally unique and included in the resource algebra we use for the proofs.

The *Ready* predicate describes fibers in a scheduler's `run_queue`. It expresses that f is safe to be executed and is used as the invariant for a scheduler's run queue, i.e. it should hold for all fibers in the run queue that they can be executed.

In the next sections we discuss the specifications we proved for the three functions. We show a detailed proof of the specification only for `Promise.await` because it is the most involved.

$$
\begin{aligned}
\mathit{PromiseState}\ p\ \gamma\ \Phi &\triangleq (\exists\, v.\ p \mapsto \mathit{Done}\ v * \mathit{promiseDone}\ \gamma * \square\, \Phi\ v) \\
&\quad \vee (\exists\, \mathit{bcst}.\ p \mapsto \mathit{Waiting}\ \mathit{bcst} * \mathit{isBroadcast}\ \mathit{bcst} * \mathit{promiseWaiting}\ \gamma * \mathit{signalAllPermit}) \\
\mathit{PInvInner} &\triangleq \exists M.\ \boxed{\bullet\ M}^{pn} * \forall\, (p,\gamma) \mapsto \Phi \in M.\ \mathit{PromiseState}\ p\ \gamma\ \Phi \\
\mathit{PromiseInv} &\triangleq \boxed{\mathit{PInvInner}}^{\mathcal{N}} \\
\mathit{isPromise}\ p\ \Phi &\triangleq \exists \gamma.\ \boxed{\circ\ \{[(p,\gamma) \mapsto \Phi]\}}^{pn} \\
\mathit{Ready}\ f &\triangleq \mathit{ewp}\ (f\ ())\ \langle \bot \rangle\ \{\top\}
\end{aligned}
$$

Figure 8: Logical State Definitions for the Verification of Scheduler & Promise Modules

2.2.3 `Scheduler.run`

The interesting part about the scheduler specification Spec-Run is that it proves **effect safety** of the fiber runtime, i.e. no matter what a fiber does it will not crash the scheduler due to an unhandled effect. This is expressed by allowing the fiber *main* to perform effects according to the *Coop* protocol, but running the scheduler on the main fiber (*run main*) obeys the empty protocol, so no effects escape. Of course, the *ewp* itself also implies **safety** of running both the main fiber and the scheduler.

Don't ignore the return value.

$$
\textsc{Spec-Run}\quad \frac{\mathit{ewp}\ (\mathit{main}\ ())\ \langle \mathit{Coop} \rangle\ \{\top\}}{\mathit{ewp}\ (\mathit{run}\ \mathit{main})\ \langle \bot \rangle\ \{\top\}}
$$

However, the specification only talks about effect safety and not about handling fibers correctly in any other way, e.g. regarding fairness of scheduling or just not dropping fibers. For example, a trivial *run* function which ignores the *main* argument and immediately returns satisfies the same specification. For a scheduler it would be desirable to prove these properties, too, but since they are liveness properties it is hard to do in Iris and not a focus of this thesis.

Explain why it's hard

2.2.4 `Fiber.fork_promise`

The specification Spec-ForkPromise expresses that we receive from $\mathit{fork\_promise}$ a promise value p that will eventually hold a value satisfying Φ. It has two preconditions, for one we must give it an arbitrary expression f representing the new fiber. When called, f obeys the *Coop* protocol and returns some value v satisfying Φ. Also, $\mathit{fork\_promise}$ needs the *PromiseInv* invariant to interact with the global collection of promises, because it creates a new promise and fulfills it after f has finished executing.

$$
\textsc{Spec-ForkPromise}\quad \frac{\mathit{PromiseInv} * \mathit{ewp}\ (f\ ())\ \langle \mathit{Coop} \rangle\ \{v,\ \square\, \Phi\ v\}}{\mathit{ewp}\ (\mathit{fork\_promise}\ f)\ \langle \mathit{Coop} \rangle\ \{p,\ \mathit{isPromise}\ p\ \Phi\}}
$$

2.2.5 `Promise.await`

The specification Spec-Await is the direct counterpart to Spec-ForkPromise. It shows that *await* consumes a promise p and eventually returns its value v satisfying the predicate Φ. The precondition *PromiseInv* is again necessary to interact with the global collection of promises and *isPromise* is used to identify the promise p in that collection.

If p is still unfulfilled the first time *await* checks the promise state, it will call *make_register* to create a `register` function which it passes to the *Suspend* effect. As the Spec-MakeRegister specification shows, `make_register` returns a suitable function that satisfies the *isRegister* predicate, instantiating P with $(\lambda\, v,\ \ulcorner v = () \urcorner * \mathit{promiseDone}\ \gamma)$ so that we receive a $\mathit{promiseDone}\ \gamma$ resource when the effect returns.

Spec-MakeRegister

$$\frac{\mathit{PromiseInv} * \mathit{isPromise}\ p\ \Phi * \mathit{isBroadcast}\ \mathit{bcst}}{\mathit{ewp}\ (\mathit{make\_register}\ p\ \mathit{bcst})\ \langle \bot \rangle\ \{\mathit{reg},\ \mathit{isRegister}\ \mathit{reg}\ (\lambda\ v,\ \ulcorner v = () \urcorner * \mathit{promiseDone}\ \gamma)\}}$$

Spec-Await

$$\frac{\mathit{PromiseInv} * \mathit{isPromise}\ p\ \Phi}{\mathit{ewp}\ (\mathit{await}\ p)\ \langle \mathit{Coop} \rangle\ \{v,\ \square\ \Phi\ v\}}$$

In figures 9 and 10 we give Hoare-style proof annotations for the two functions `make_register` and `Promise.await` from figure 4. The proof of Spec-MakeRegister uses the specifications of some broadcast functions. We briefly explain these specifications and their logical state definitions now and expand upon them in section 3.3.

$$\begin{aligned}
\mathit{isCallback}\ \mathit{cb}\ R &\triangleq R \mathrel{-\!\!*} \mathit{ewp}\ (\mathit{cb}\ ())\ \langle \bot \rangle\ \{\top\} \\
\mathit{isBroadcastRegisterResult}\ r\ \mathit{cb}\ R &\triangleq (\ulcorner r = \mathit{Called} \urcorner) \\
&\quad \vee (\ulcorner r = \mathit{Registered}\ h \urcorner * \mathit{isBroadcastRegisterHandle}\ h\ \mathit{cb}\ R) \\
\mathit{isBroadcastRegisterHandle} &: \mathit{Val} \to \mathit{Val} \to \mathit{iProp} \to \mathit{iProp}
\end{aligned}$$

Spec-BroadcastRegister

$$\frac{\mathit{isBroadcast}\ \mathit{bcst} * \mathit{isCallback}\ \mathit{callback}\ R}{\mathit{ewp}\ (\mathit{register}\ \mathit{bcst}\ \mathit{callback})\ \langle \bot \rangle\ \{r,\ \mathit{isBroadcastRegisterResult}\ r\ \mathit{callback}\ R\}}$$

Spec-BroadcastTryCancel

$$\frac{\mathit{isBroadcastRegisterHandle}\ h\ \mathit{cb}\ R}{\mathit{ewp}\ (\mathit{try\_cancel}\ h)\ \langle \bot \rangle\ \{b,\ \mathit{if}\ b\ \mathit{then}\ \mathit{isCallback}\ \mathit{cb}\ R\ \mathit{else}\ \top\}}$$

The function `Broadcast.register` takes a callback *cb* that satisfies the *isCallback* predicate to register it in the broadcast data structure. This predicate is structurally similar to *isWaker* and, in fact, in the proof of Spec-MakeRegister we instantiate the precondition R with $\mathit{promiseDone}\ \gamma$ and pass as the callback a `waker` function, which has the precondition $(\lambda\ v,\ \ulcorner v = () \urcorner * \mathit{promiseDone}\ \gamma)$ as described above. The result of `Broadcast.register` is either a value *Called*, which expresses that it called the callback directly, or a register handle, which can be used to call `Broadcast.try_cancel`.

`Broadcast.try_cancel` will attempt to cancel a previous registration identified by the given *handle*. If the cancellation is successful, we get back the *isCallback* resource so that we can call the callback in `make_register`.

Hoare-Style Proofs for Spec-MakeRegister and Spec-Await In the proof below an opened invariant *Inv* is represented as ~~*Inv*~~ and resources that are not needed for the rest of the proof are dropped implicitly.

The proof of Spec-MakeRegister is straightforward and follows from the specifications of `Broadcast.register` and `Broadcast.try_cancel`. For Spec-Await, the crux is that we define Spec-MakeRegister so that it returns a *register* function which satisfies $\mathit{isRegister}\ \mathit{register}\ (\lambda\ v,\ \ulcorner v = () \urcorner * \mathit{promiseDone}\ \gamma)$. Then, we get access to the $\mathit{promiseDone}\ \gamma$ resource when the *Suspend* effect returns, and we can refute the case of the promise still being unfulfilled when checking the state of promise again for the last time.

```
let make_register (p: 'a t) (bcst: Broadcast.t) : (() waker -> ()) =
```
$\{PromiseInv * isPromise\ p * isBroadcast\ bcst\}$

```
  fun (waker: () waker) ->
```
[intro waker that satisfies *isWaker*]

$\left\{\begin{array}{l} PromiseInv * isPromise\ p * isBroadcast\ bcst * \\ (promiseDone\ \gamma \mathrel{-\!\!*} ewp\ (waker\ ())\ \langle\bot\rangle\ \{\top\}) \end{array}\right\}$

```
    let regres = Broadcast.register bcst waker in
```
[apply Spec-BroadcastRegister]

$\left\{\begin{array}{l} PromiseInv * isPromise\ p * isBroadcast\ bcst * \\ isBroadcastRegisterResult\ regres \end{array}\right\}$

```
    match regres with
```
[CA on regres]

---

1. $\{regres = None\}$

```
    | None -> ()
```
[by done]

---

2. $\left\{\begin{array}{l} PromiseInv * isPromise\ p * isBroadcast\ bcst * \\ regres = Some\ handle * isBroadcastRegisterHandle\ handle \end{array}\right\}$

```
    | Some handle ->
```
[open *PromiseInv*, lookup p using *isPromis*

$\left\{\begin{array}{l} \cancel{PromiseInv} * isBroadcast\ bcst * \\ isBroadcastRegisterHandle\ handle * PromiseState\ p\ \gamma\ \Phi \end{array}\right\}$

```
      match Atomic.get p with
```
[CA on *PromiseState*]

2.1. $\left\{\begin{array}{l} \cancel{PromiseInv} * isBroadcast\ bcst * \\ isBroadcastRegisterHandle\ handle * \\ p \mapsto Done\ result * promiseDone\ \gamma \end{array}\right\}$

```
        | Done result ->
```
[close *PromiseInv*]

$\left\{\begin{array}{l} isBroadcast\ bcst * isBroadcastRegisterHandle\ handle * \\ promiseDone\ \gamma \end{array}\right\}$

```
          if Broadcast.try_cancel handle
```
[apply Spec-BroadcastTryCancel, CA o

2.1.1. $\{promiseDone\ \gamma * (promiseDone\ \gamma \mathrel{-\!\!*} ewp\ (waker\ ())\ \langle\bot\rangle\ \{\top\})\}$
[specialize assumption]

$\{ewp\ (waker\ ())\ \langle\bot\rangle\ \{\top\}\}$

```
          then waker ()
```
[by apply]

2.1.2. $\{\top\}$

```
          else ()
```
[by done]

2.2. $\{\cancel{PromiseInv} * p \mapsto Waiting\ \_\}$

```
        | Waiting _ -> ()
```
[close *PromiseInv*, by done]

Figure 9: Annotated proof of Spec-MakeRegister.

`let await (p: 'a t) : 'a =`	
$\{PromiseInv * isPromise\ p\}$	[open *PromiseInv*, lookup p using *isPromise* p]
$\{\cancel{PromiseInv} * isPromise\ p * PromiseState\ p\ \gamma\ \Phi\}$	
`  match Atomic.get p with`	[CA on *PromiseState*]
1. $\left\{\begin{array}{c}\cancel{PromiseInv}* \\ p \mapsto Done\ result * \square(\Phi\ result)\end{array}\right\}$	
`  | Done result ->`	[close *PromiseInv*]
$\{PromiseInv * \square(\Phi\ result)\}$	
`  result`	[by assumption]
2. $\left\{\begin{array}{c}\cancel{PromiseInv} * isPromise\ p* \\ p \mapsto Waiting\ bcst * isBroadcast\ bcst\end{array}\right\}$	
`  | Waiting bcst ->`	[close *PromiseInv*]
$\left\{\begin{array}{c}PromiseInv * isPromise\ p* \\ isBroadcast\ bcst\end{array}\right\}$	
`  let register = make_register p bcst`	[apply Spec-MakeRegister]
$\left\{\begin{array}{c}PromiseInv * isPromise\ p* \\ isRegister\ register\end{array}\right\}$	
`  perform (Suspend register);`	[protocol of *Suspend* with $(P := \lambda v,\ \ulcorner v = () \urcorner * promiseDone\ \gamma)$]
$\left\{\begin{array}{c}PromiseInv * isPromise\ p* \\ promiseDone\ \gamma\end{array}\right\}$	[open *PromiseInv*, lookup p using *isPromise* p]
$\left\{\begin{array}{c}\cancel{PromiseInv} * promiseDone\ \gamma \\ *PromiseState\ p\ \gamma\ \Phi\end{array}\right\}$	
`  match Atomic.get p with`	[CA on *PromiseState*]
2.1. $\left\{\begin{array}{c}\cancel{PromiseInv}* \\ p \mapsto Done\ result * \square(\Phi\ result)\end{array}\right\}$	
`  | Done result ->`	[close *PromiseInv*]
$\{PromiseInv * \square(\Phi\ result)\}$	
`  result`	[by assumption]
2.2. $\left\{\begin{array}{c}\cancel{PromiseInv} * promiseDone\ \gamma* \\ p \mapsto Waiting\ bcst * promiseWaiting\ \gamma\end{array}\right\}$	
`  | Waiting _ ->`	[specialize PS-Contra]
$\{\cancel{PromiseInv} * \bot\}$	
`  error "impossible"`	[by contradiction]

Figure 10: Hoare-style proof of Spec-Await.

2.2.6 Comparison of Logical State

$$Ready\ q\ \phi\ k \triangleq \forall v.\ \square\phi(v) \mathrel{-\!\!*} \triangleright PromiseInv\ q \mathrel{-\!\!*} \triangleright isQueue\ q\ (Ready\ q\ (\lambda w.\ w = ())) \mathrel{-\!\!*} ewp\ (k\ v)\ \langle\bot\rangle\{\_.\mathrm{True}\}$$

replace png with latex

Since our logical state definitions are based on a case study of de Vilhena, we want to give a short comparison of what had to be changed when adapting it to our model of Eio. First, in the original development the *Ready* predicate fulfills two roles.

1. It expresses that all continuations in the scheduler's run-queue are safe to execute.
2. It expresses that all continuations in a promise's waiting-queue are safe to execute.

PromiseInv and *isQueue* were both necessary as preconditions because they are not persistent and need to be passed around explicitly.

In our development *PromiseInv* could be dropped from the definition of Ready because it is now put into an Iris shareable invariant, and can be passed implicitly. Similarly, the *isQueue* precondition was dropped from the definition of *Ready* because in Eio the run queue must be thread-safe, so the *isQueue* resource is persistent and can be passed to a fiber once when it is spawned. Therefore, our *Ready* is neither recursive nor mutually recursive with *PromiseInv* anymore, which simplifies its usage in Iris. We note that the (mutual) recursion was only necessary because *PromiseInv* was used to track global state but was not put into an Iris shareable invariant, so it had to be passed around explicitly in many places.

We also split up the two uses of *Ready* and only use it under this name for the first role. In the case of a scheduler's run-queue, $\Phi\ v$ degenerates just to $\ulcorner v = () \urcorner$, so we can drop both from the definition and use () directly. This is why our definition of *Ready* only contains an *ewp* without preconditions.

For the second use case of describing the continuations in a promise's waiting-queue we now have another specialized version of *Ready*. As explained in the next section, a broadcast has an invariant $P\ v \mathrel{-\!\!*}$ *ewp* (*callback* ()) $\langle \bot \rangle\ \{\top\}$ for all stored *callback*s. This is just *Ready* where $P\ v$ replaces $\Phi\ v$, and it is the same P as in the definition of the *Suspend* effect.

3 Verifying Eio's Broadcast

In this section we go into detail on the *broadcast* data structure that Eio uses in the implementation of *promises*. The Eio implementation of broadcasts is an adaptation of something called *CQS* [3]. CQS (for CancellableQueueSynchronizer) is a synchronization primitive that allows execution contexts to wait until signalled. Its specification is already formally verified in Iris, so we were able to adapt the proofs to use them in our development. CQS keeps the nature of an execution context abstract, but it is assumed that they support stopping execution and resuming with some value. This is because CQS is designed to be used in the implementation of other synchronization constructs (e.g. mutex, barrier, promise, etc.) which take care of actually suspending and resuming execution contexts as required by their semantics.

In the case of Eio an "execution context" is an Eio fiber. Still, CQS is multithreaded, so fibers can use CQS functions to synchronize with fibers running in another thread. In the following we describe the behavior of Eio's *broadcast*, highlight differences to the *original CQS*, and explain how we adapted the verification of the original CQS for our development. If something applies to both the customized and original version we just use the term *CQS*.

3.1 Operations of Broadcast

The original CQS supports three operations that are interesting to us: *suspend*, *tryCancel*, and *resume*. While we established Eio's broadcast as similar to condition variables where fibers can register callbacks to be notified about events, the original formulation of CQS uses a more abstract future-based interface for the same purpose.

For example, a *suspend* operation is done by an execution context that wants to wait for an event. This operation creates and returns a new future which is used to stop execution because it is assumed that the program runtime supports suspending an execution context until a future is completed. But Eio cannot use this interface because it uses the customized CQS to *build* the runtime that allows its execution contexts (i.e. fibers) to suspend until an event happens (i.e. a promise is fulfilled). So for broadcasts the *suspend* operation is replaced by the *register* operation, that takes a callback as an additional argument and registers it to be called later.

Analogously, a *resume* operation in the original CQS completed a single future which signalled the runtime to resume execution of the associated execution context. This is replaced by the *signalAll* operation, which invokes all callbacks that are registered with the data structure. Eio uses *signalAll* instead of a *signal* operation to make the implementation of promises more straightforward. When a promise is fulfilled, *all* fibers waiting on its value must continue execution, so the fine-grained control of a *signal* operation is not needed.

In the following we focus on the operations of Eio's broadcast, and to understand them it is helpful to view the context in which they are used. An interaction with the original CQS as described in [3] is always guarded by first accessing an atomic variable which encodes some information about CQS like the number of registered futures. This atomic variable is a part of the synchronization construction that is implemented using CQS. Thereby, it is ensured that operations only happen when they make sense (e.g. a *resume* operation is not performed when no futures are available). For Eio's broadcast, figure 11 shows that the atomic variable is the state variable of a promise, which is accessed in the functions `Promise.fulfill` and `Promise.await` as was described in section 2.1. `Promise.await` will then perform *register* and *cancel* operations if necessary, and `Promise.fulfill` will do a *signalAll* operation.

Note that because broadcast is a lock-free data structure and fibers can run on different threads, there can be a race between concurrent *register*, *tryCancel*, and *signalAll* operations. Possible interleavings and the necessity of the *cancel operation* were explained in section 2.1.3.

3.2 Implementation and Logical Interface of Broadcast

CQS is implemented as a queue of *cells* with two pointers pointing to the beginning and end of the active cell range, the *suspend pointer* and the *resume pointer*. Cells not reachable from either pointer are garbage collected, but their logical state is still tracked. There is a stack of operations for manipulating these pointers to implement the higher-level functionality, but they are not part of the public API, so we do not focus on them. Each cell is a container for one callback and the logical state of the queue tracks the logical

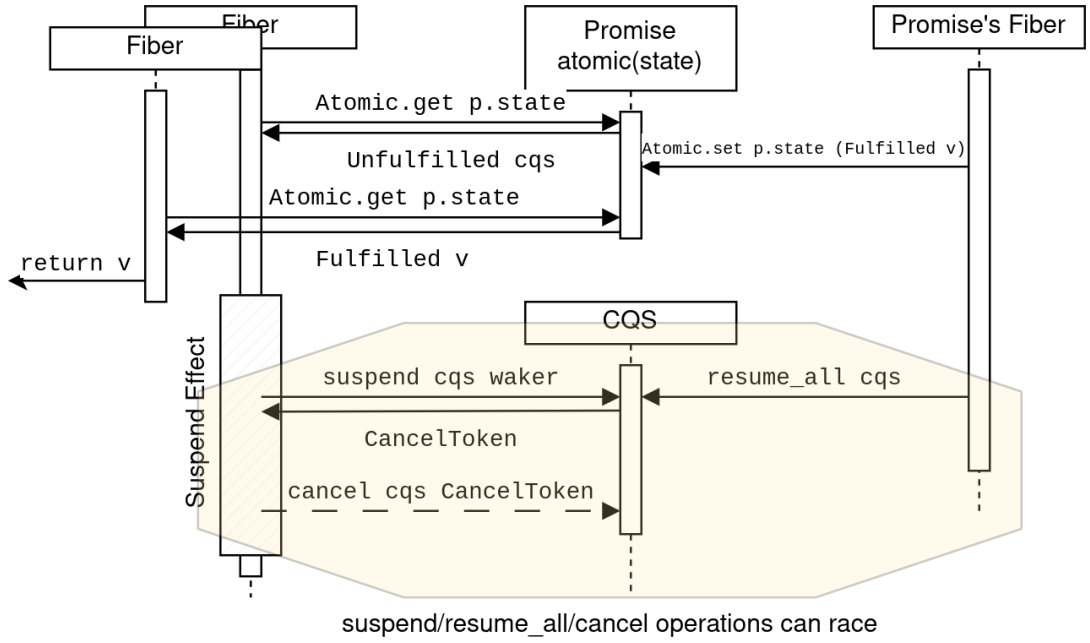

Figure 11: Usage of CQS with an Outer Atomic Variable

state of all existing cells. The possible logical states for a single cell are shown in figure 12 and there exist logical operations to change a cell's state.

The number of active cells *n* (i.e. the length of the queue) is tracked by the logical resource *CQSState n*. In normal usage of CQS, the atomic variable of the outer synchronization construct would encode the length of the queue in its value and keep this resource in an associated invariant. Changing the length of the queue is done using *enqueue* and *dequeue registration* logical operations when opening this invariant.

However, for promises the exact length of the queue is irrelevant because the *signalAll* operation will always set the length to 0. So in the adapted proof for Eio's broadcast we keep the *CQSState n* resource in the invariant of the broadcast itself. As a consequence we also move the *enqueue* and *dequeue registration* out of the public logical API because they are now done internally.

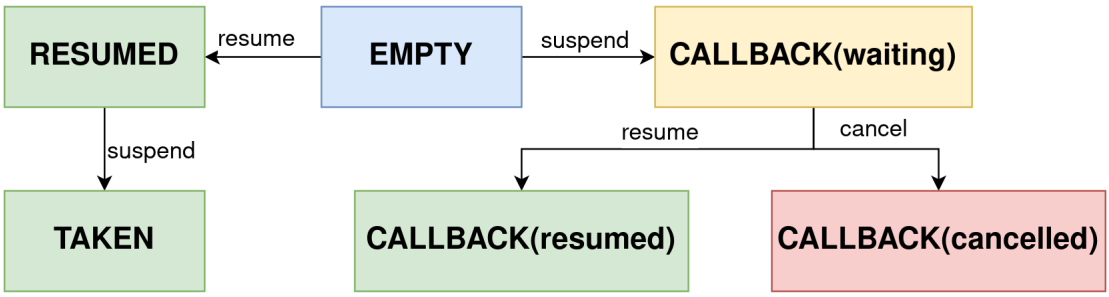

Figure 12: State Transition Diagram for a Single Cell.

3.3 Verification of Broadcast

In the following we describe the specifications we proved for the functions implemented in Eio's `Broadcast` module, and what changes we did to the internal logical state of CQS to carry out the proofs. For all three operations, the Eio implementation differs from the implementation already verified in the original CQS (e.g. some reordered instructions or a slightly different control flow) and they have different specifications. However, the specifications of the underlying operations for manipulating cell pointers are modular enough to allow us to prove the new specifications for `Broadcast.create`, `Broadcast.register`, and

`Broadcast.try_cancel`.

As for `Broadcast.signal_all`, Eio implements this function by atomically increasing the *resume pointer* by the number n of registered callbacks and then processing all n cells between the old and new pointer position. Because of technical differences between the original CQS implementation of [3] and the broadcast implementation of Eio we opted to verify a different implementation of `Broadcast.signal_all`, that increments the *resume pointer* n times in a loop. We argue this does not change the observable behaviors of the function since we ensure that it can only be called once.

3.3.1 `Broadcast.create`

Creating a broadcast requires *inv_heap_inv* which is an Iris proposition which says that we are in a garbage-collected setting. As a result we get the persistent *isBroadcast bcst* resource that shows the value *bcst* is a broadcast. We also obtain the unique resource *signalAllPermit*, which is held by the enclosing promise and allows calling the `Broadcast.signal_all` function once.

$$\textsc{Spec-BroadcastCreate}\quad \frac{\mathit{inv\_heap\_inv}}{\mathit{ewp}\ (\mathit{create}\ ())\ \langle\bot\rangle\ \{\mathit{bcst},\ \mathit{isBroadcast\ bcst} * \mathit{signalAllPermit}\}}$$

3.3.2 `Broadcast.register`

A *register* operation takes a callback *cb* and the associated resource *isCallback cb* which represents the permission to invoke the callback. We instantiate R with *promiseDone* γ so that the callback transports the knowledge that the promise has been fulfilled. *isCallback* is not persistent because the callback must be invoked only once, and it might be accessed from a different thread.

$$\begin{aligned}
\mathit{isCallback\ cb\ R} &\triangleq R \mathrel{-\!\!*} \mathit{ewp}\ (\mathit{cb}\ ())\ \langle\bot\rangle\ \{\top\} \\
\mathit{isBroadcastRegisterResult\ r\ cb\ R} &\triangleq (\ulcorner r = \mathit{Called}\urcorner) \\
&\qquad \vee (\ulcorner r = \mathit{Registered}\ h\urcorner * \mathit{isBroadcastRegisterHandle\ h\ cb\ R}) \\
\mathit{isBroadcastRegisterHandle} &: \mathit{Val} \to \mathit{Val} \to \mathit{iProp} \to \mathit{iProp}
\end{aligned}$$

The `Broadcast.register` function will advance the *suspend pointer* to allocate a fresh cell in the **EMPTY** logical state. If there is a concurrent call to `Broadcast.signal_all` which changed the cell to the **RESUMED** logical state before this function can save the callback into the fresh cell, the callback is invoked immediately and the value *Called* is returned. In this case, the state of the cell will be set to **TAKEN**. Otherwise, the callback is saved in the cell, which is advanced to the **CALLBACK(waiting)** logical state, and a *Registered handle* value is returned along with a *isBroadcastRegisterHandle* resource as the cancellation permit.

$$\textsc{Spec-BroadcastRegister}\quad \frac{\mathit{isBroadcast\ bcst} * \mathit{isCallback\ callback\ R}}{\mathit{ewp}\ (\mathit{register\ bcst\ callback})\ \langle\bot\rangle\ \{r,\ \mathit{isBroadcastRegisterResult\ r\ callback\ R}\}}$$

3.3.3 `Broadcast.try_cancel`

Given a *textitisBroadcastRegisterHandle h cb R*, `Broadcast.try_cancel` will try to cancel the registration of the callback.

If the callback had already been invoked by a call to `Broadcast.signal_all` (i.e. the logical state is **CALLBACK(resumed)**) the function returns `false` and no resources are returned to the caller. Otherwise, the permission to invoke the callback *isCallback cb* is returned, and the cell is advanced to the **CALLBACK(cancelled)** logical state.

$$\textsc{Spec-BroadcastTryCancel}\quad \frac{\mathit{isBroadcastRegisterHandle\ h\ cb\ R}}{\mathit{ewp}\ (\mathit{try\_cancel}\ h)\ \langle\bot\rangle\ \{b,\ \mathit{if}\ b\ \mathit{then}\ \mathit{isCallback\ cb\ R}\ \mathit{else}\ \top\}}$$

3.3.4 `Broadcast.signal_all`

To call `Broadcast.signal_all` the unique *signalAllPermit* resource is needed. The `R` resource must be duplicable because it will be used to invoke multiple callbacks, which have `R` as their precondition. The function does not return any resources because its only effect is making an unknown number of fibers resume execution, which is not something we can easily formalize in Iris.

$$\frac{\textsc{Spec-BroadcastSignalAll} \quad \mathit{isBroadcast}\ \mathit{bcst} * \square\ R * \mathit{signalAllPermit}}{\mathit{ewp}\ (\mathit{signalAll}\ \mathit{bcst})\ \langle \bot \rangle\ \{\top\}}$$

3.4 Features Removed from Original CQS

The original CQS supports multiple additional features like a synchronous mode for suspend and resume, and also a smart cancellation mode. These features enlarge the state space of CQS and complicate the verification but are not used in Eio so when we ported the verification of CQS to our Eio development we removed support for these features. This reduced the state space of a cell shown in figure 13 to something more manageable for us when adapting the proofs.

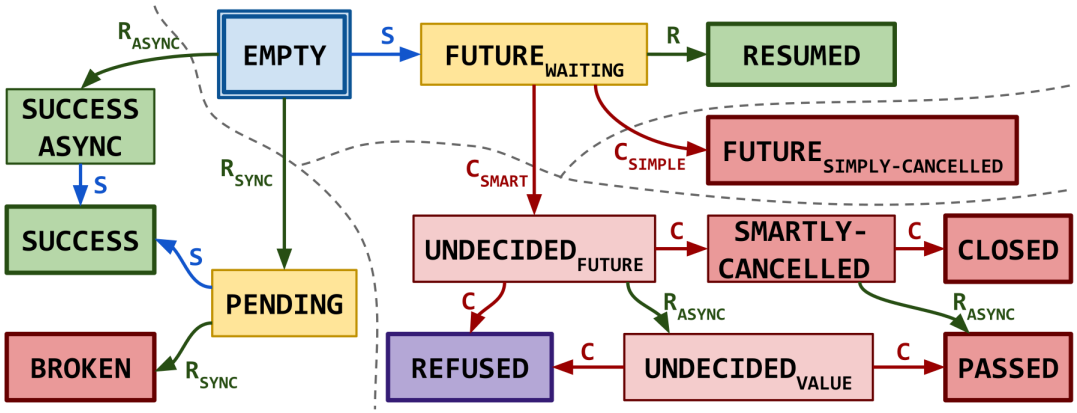

Figure 13: Cell States in the Original CQS from [3] (page 42).

The part of the verification of the original CQS that we had to customize for Eio was originally 3600 lines of Coq code but – due to these simplifications – we could reduce it by approximately 1300 lines of Coq code. Additionally, there are 4000 lines of Coq code about lower-level functionality that we did not need to adapt when porting them to our development.

```
let fiber1 () =
  let ctx = perform (GetContext ()) in
  let ctx2 = perform (GetContext ()) in
  assert (ctx.tlv == ctx2.tlv)

let fiber2 () =
  let ctx = perform (GetContext ()) in
  let v = !ctx.tlv in
  (* some computation that does not perform Fork/Suspend *)
  ...
  assert (!ctx.tlv == v)
```

Figure 14: Constructed example of safety for thread-local variables.

4 Extending the Scheduler with Thread-Local Variables

So far we have looked at a protocol *Coop* that has two effects which suffice to model fibers that can suspend and fork off new fibers. But in Eio fibers can use an additional effect called *GetContext* that we discuss in this section. For each fiber the scheduler keeps track of context metadata, one part of which are *thread-local variables*. Thread-local variables are state that is shared between all fibers of one scheduler (hence thread-local) and a fiber gets access to them via the *GetContext* effect.

Since all fibers of one scheduler execute concurrently on one system-level thread, they have exclusive access to the thread-local variables while they are running. This allows a practical form of shared state without the overhead of synchronization primitives of multithreaded data structures. Two example use-cases are per-scheduler tracing of events, where all fibers of one scheduler write to a common log, and inter-fiber message passing, where fibers use a simple queue to exchange messages. Of course, this comes with the restriction that it is only usable for fibers running in the same thread.

In Eio thread-local variables are represented by a dictionary from variable names to arbitrary values and expose an intended API that only allows adding new entries. However, it is still possible for fibers to arbitrarily modify the whole dictionary, so for demonstration purposes we model thread-local variables as a single mutable reference that is part of the context record: `ctx.tlv`. Properties we want to prove about thread-local variables are:

1. Each time a fiber performs a *GetContext* effect it will receive the same reference.
2. As long as a fiber does not perform other effects like *Fork* or *Suspend*, it holds exclusive ownership of the reference.

Code examples illustrating the properties are shown in figure 14. Note that these are only the most basic properties showing that `ctx.tlv` acts like a normal reference, but one that can be accessed via an effect. To enable modular proofs of concrete fibers using thread-local variables, we include in our logical state a predicate T on the stored value that can be instantiated by fibers as needed.

4.1 Changes to Logical State

To handle thread-local variables in our development we must change both the implementation and logical state definitions. The necessary changes to the implementation are trivial, so we just refer to the mechanization[5]. For the logical state we define *fiberResources* that a fiber receives when it starts running and relinquishes when it stops. The new definitions are described in figure 15. *tlvAg* δ *tlv* is used to show the uniqueness of the location *tlv*. *isFiberContext* δ *tlv* represents the context that is tracked for each fiber, where δ is a shorthand for multiple ghost names. It expresses that the location *tlv* is a thread-local variable which maps to some value v satisfying T. The predicate T is hidden behind a *savedPred* indirection to make the mechanization easier. *fiberResources* δ is then used to abstract away the concrete location *tlv*. Finally, we must change the definition of *Ready* to require *fiberResources* as a precondition because it is needed to invoke the continuations saved in the scheduler's run-queue.

The effect protocols of *Fork* and *Suspend* are amended so that they pass *fiberResources* from a fiber to the scheduler and from there to the next running fiber via the protocol pre- and postconditions as shown

[5]TODO insert link

$$
\begin{aligned}
\mathit{tlvAg}\ \delta\ \mathit{tlv} &\triangleq \boxed{\mathit{agree}(\mathit{tlv})}^{\delta} \qquad \text{Persistent}(\mathit{tlvAg}\ \delta\ \mathit{tlv}) \\
\mathit{isFiberContext}\ \delta\ \mathit{tlv} &\triangleq \mathit{tlvAg}\ \delta\ \mathit{tlv} * \exists T\ v.\ \mathit{tlv} \mapsto v * \mathit{savedPred}\ \delta\ T * T\ v \\
\mathit{fiberResources}\ \delta &\triangleq \exists\ \mathit{tlv}.\ \mathit{isFiberContext}\ \delta\ \mathit{tlv} \\
\mathit{Ready}\ \delta\ f &\triangleq \mathit{fiberResources}\ \delta \twoheadrightarrow \mathit{ewp}\ (f\ ())\ \langle \bot \rangle\ \{\top\}
\end{aligned}
$$

Figure 15: Logical State Definitions for the Verification of Scheduler & Promise Modules

in figure 16. The *Fork* effect now also passes the concrete reference that should be used as the thread-local variable of the new fiber. A fiber uses the *GetContext* effect to receive the fiber context value and a copy of *tlvAg*. This is used to show that the reference *ctx.tlv* is equal to the one from *fiberResources* that the fiber already owns so that the contained points-to predicate can be used.

The crux is that now the protocol $\mathit{Coop}\ \delta$ is parameterized by the ghost name δ that identifies the thread-local variable. This so that both the fiber and the scheduler agree on this ghost name.

$$
\begin{aligned}
\mathit{Coop}\ \delta \triangleq \quad & \mathit{Fork}\ \#\ !\ \mathit{tlv}\ e\ ((\mathit{tlv}, e))\ \big\{\mathit{fiberResources}\ \delta\ T * \mathit{tlvAg}\ \delta\ \mathit{tlv} * \\
& \qquad \triangleright\ (\mathit{fiberResources}\ \delta\ T \twoheadrightarrow \mathit{ewp}\ (e)\ \langle \mathit{Coop} \rangle\ \{\mathit{fiberResources}\ \delta\ T\})\big\}. \\
& \quad ?\ ()\ \{\mathit{fiberResources}\ \delta\ T\} \\
& \mathit{Suspend}\ \#\ !\ \mathit{reg}\ P\ (\mathit{reg})\ \{\mathit{fiberResources}\ \delta\ T * \mathit{isRegister}\ \mathit{reg}\ P\}. \\
& \quad ?\ y\ (y)\ \{\mathit{fiberResources}\ \delta\ T * P\ y\} \\
& \mathit{GetContext}\ \#\ !\ ()\ \{\top\}.\ ?\ \mathit{ctx}\ (\mathit{ctx})\ \{\mathit{tlvAg}\ \delta\ \mathit{ctx.tlv}\}
\end{aligned}
$$

Figure 16: Definition of extended $\mathit{Coop}\ \delta$ protocol with *Fork*, *Suspend*, and *GetContext* effects.

These changes suffice to prove the safety of the two examples in figure 14.

5 Evaluation

6 Conclusion

Appendix

A Translation Table

Eio	Thesis	Mechanization
`enqueue`	`waker` function	`waker`
`f`	`register` function	`register`
`Fiber.fork_promise`	`Fiber.fork_promise`	`fork_promise`
`Promise.await`	`Promise.await`	`await`
`Sched.run`	`Scheduler.run`	`run`

B Towards A multithreaded Scheduler

OCaml 5 added not only effect handlers but also the ability to use multiple threads of execution, which are called *domains* (in the following we use the terms interchangeably). Each domain in OCaml 5 corresponds to one system-level thread and the usual rules of multithreaded execution apply, i.e. domains are preemtively scheduled and can share memory. Eio defines an operation to make use of multi-threading by forking off a new thread and running a separate scheduler in it. So while each Eio scheduler is only responsible for fibers in a single thread, fibers can await and communicate with fibers running in other threads.

In order for a fiber to be able to await fibers in another thread, the `wakers_queue` [note it will be in the Simple Scheduler section] from above is actually a thread-safe queue based on something called CQS, which we will discuss in detail in a later section.

Heaplang supports reasoning about multithreaded programs by implementing fork and join operations for threads and defining atomic steps in the operational semantics, which enables the use of Iris *invariants*. In contrast, Hazel did not define any multithreaded operational semantics but it contained most of the building blocks for using invariants. In the following we explain how we added a multithreaded operational semantics and enabled the use of invariants.

Adding Invariants to Hazel

Invariants in Iris are used to share resources between threads. They encapsulate a resource to be shared and can be opened for a single atomic step of execution. During this step the resource can be taken out of the invariant and used in the proof but at the end of the step the invariant must be restored.

Hazel did already have the basic elements necessary to support using invariants. It defined a ghost cell to hold invariants and proved an invariant access lemma which allows opening an invariant if the current expression is atomic. In order to use invariant we only had to provide proofs for which evaluation steps are atomic. We provided proofs for all primitive evaluation steps. The proofs are the same for all steps so we just explain the one for `Load`.

```
Lemma ectx_language_atomic a e :
head_atomic a e → sub_exprs_are_values e → Atomic a e.

Instance load_atomic v : Atomic StronglyAtomic (Load (Val v)).
Instance store_atomic v1 v2 : Atomic StronglyAtomic (Store (Val v1) (Val v2)).
...
```

An expression is atomic if it takes one step to a value, and if all subexpressions are already values. The first condition follows by definition of the step relation and the second follows by case analysis of the expression.

Since performing an effect starts a chain of evaluation steps to capture the current continuation, it is not atomic. For the same reason an effect handler and invoking a continuation are not atomic except in degenerate cases. Therefore, invariants and effects do not interact in any interesting way.

How we add support for the iInv tactic to use invariants more easily.

Adding Multi-Threading to Hazel

To allow reasoning in Hazel about multithreaded programs we need a multithreaded operational semantics as well as specifications for the new primitive operations *Fork*, `Cmpxcgh` and `FAA`.

The language interface of Iris provides a multithreaded operational semantics that is based on a thread-pool. The thread-pool is a list of expressions that represents threads running in parallel. At each step, one expressions is picked out of the pool at random and executed for one thread-local step. Each thread-local step additionally returns a list of forked-off threads, which are then added to the pool. This is only relevant for the *Fork* operation as all other operations naturally don't fork off threads.

Heaplang implements multi-threading like this and for Hazel we do the same thing. We adapt Hazel's thread-local operational semantics to include *Fork*, `Cmpxchg` and `FAA` operations and to track forked-off threads and get a multithreaded operational semantics "for free" from Iris' language interface.

Additionally, we need to prove specifications for these three operations. `Cmpxchg` and `FAA` are standard so we will not discuss them here. The only interesting design decision in the case of Hazel is how effects and *Fork* interact. This decision is guided by the fact that in OCaml 5 effects never cross thread-boundaries. An unhandled effect just terminates the current thread. As such we must impose the empty protocol on the argument of *Fork*.

Using these primitive operations we can then build the standard `CAS`, `Spawn`, and `Join` operations on top and prove their specifications. For `Spawn` & `Join` we already need invariants as the point-to assertion for the done flag must be shared between the two threads.

Note that for `Spawn` we must also impose the empty protocol on `f` as this expression will be forked-off.

This allows us to implement standard multithreaded programs which also use effect handlers. For example, we can prove the specification of the function below that is based on an analogous function in Eio which forks a thread and runs a new scheduler inside it. Note that same as in Eio the function blocks until the thread has finished executing, so it should be called in separate fiber.

The scheduler `run` and therefore also the `spawn_scheduler` function don't have interesting return values, so this part of the specification is uninteresting. What is more interesting is that they encapsulate the possible effects the given function `f` performs.

C A Note on Cancellation

- That we tried to model cancellation but the feature is too permissive to give it a specification.
- There is still an interesting question of safety (fibers cannot be added to a cancelled Switch).
- But including switches & cancellation in our model would entail too much work so we leave it for future work.

References

[1] Paulo De Vilhena. “Proof of Programs with Effect Handlers”. PhD thesis. Université Paris Cité, 2022.

[2] Stephen Dolan et al. “Concurrent system programming with effect handlers”. In: *Trends in Functional Programming: 18th International Symposium, TFP 2017, Canterbury, UK, June 19-21, 2017, Revised Selected Papers 18*. Springer. 2018, pp. 98–117.

[3] Nikita Koval, Dmitry Khalanskiy, and Dan Alistarh. “CQS: A Formally-Verified Framework for Fair and Abortable Synchronization”. In: *Proceedings of the ACM on Programming Languages* 7.PLDI (2023), pp. 244–266.

[4] Daan Leijen. “Structured asynchrony with algebraic effects”. In: *Proceedings of the 2nd ACM SIGPLAN International Workshop on Type-Driven Development*. 2017, pp. 16–29.

[5] Paulo Emílio de Vilhena and François Pottier. “A separation logic for effect handlers”. In: *Proceedings of the ACM on Programming Languages* 5.POPL (2021), pp. 1–28.