# **Asynchronous Effect Handling**

## Leo Poulson

Master of Science
Computer Science
School of Informatics
University of Edinburgh
2020

## **Abstract**

Features for asynchronous programming are commonplace in the programming languages of today, allowing programmers to issue tasks to run on other threads and wait for the results to come back later. This is particularly useful for programs like web programs, etc...

In this thesis we show how asynchronous programming can be very easily accomodated in a language with existing support for effect handlers. We show how, with a small change to the language implementation, truly asynchronous programming with pre-emptive concurrency is achieved.

# Acknowledgements

thanks!

# **Table of Contents**

1	Introduction					
	1.1	Related Work	1			
	1.2	Contributions	1			
2	Programming in Frank					
	2.1	Effects and Effect Handling	2			
	2.2	Concurrency	5			
		2.2.1 Simple Scheduling	5			
		2.2.2 Forking New Processes	6			
3	For	malisation of Frank	9			
4	Pre-emption					
	4.1	Motivation	15			
	4.2	Interruption with Yields	15			
	4.3	Relaxing Catches	16			
	4.4	Interrupting Arbitrary Terms	18			
	4.5	Interrupting In Practice	20			
	4.6	Soundness	21			
5	Implementation					
	5.1	In Frank	22			
		5.1.1 Handling	24			
6	Examples					
	6.1	Pre-emptive Concurrency	25			
	6.2	Async-Await	26			
Bi	bliog	raphy	29			

# **Chapter 1**

## Introduction

#### 1.1 Related Work

TFP, aeff, etc

#### 1.2 Contributions

**Asynchronous Effects Library** We present a library for programming with asynchronous effects in the style of Æff, built in Frank. We show how a complex system can be expressed concisely and elegantly when programming in a language with effect handlers, further cementing the case for effects as a foundation for concurrent programming.

**TODO:** Rewrite the end of this; slightly messy

**Pre-emptive Concurrency** We show how, by making a small change to the operational semantics of Frank, we achieve pre-emptive concurrency; that is, the suspension of running threads *without* co-operation. It is our hope that this change is simple enough to be transferrable to other languages.

**Examples** We also deliver a set of examples of the uses of asynchronous effects, and show how they have benefits to other models.

# **Chapter 2**

# **Programming in Frank**

Frank is a functional programming language, designed with the use of algebraic effects at its heart. As such, Frank has an effect type system used to track which effects a computation may use.

Frank also offers very fine-grained control over computations. It clearly distinguishes between computation and value, and offers *multihandlers* to carefully control when computations are evaluated. This combined with effect handling provides a very rich foundation for expressing complex control structures.

In this chapter, we introduce the language, and show why it is so well-suited to our task. We assume some familiarity with typed functional programming, and skip over some more traditional aspects of the language — algebraic data types, pattern matching, etc — so we can spend more time with the novel, interesting parts of the language.

## 2.1 Effects and Effect Handling

Interfaces and Operations Frank encapsulates effects through *interfaces*, which offer *commands*. For instance, the State effect (interface) offers two operations (commands), get and put. In Frank, this translates to

The type signatures of the operations mean that **get** is a 0-ary operation which is *resumed* with a value of type **x**, and **put** takes a value of type **x** and is resumed with **unit**. Programs get access to an interface's command by including them in the *ability* of the program. Commands are invoked just as normal functions;

```
xplusplus : {[State Int] Unit}
xplusplus! = put (get! + 1)
```

This familiar program increments the integer in the state by 1.

**Handling Operations** Traditioanl functions in Frank are a specialisation of Frank's handlers; that is to say, functions are handlers that handle no effects. A handler for an interface pattern matches *on the operations* that are invoked, as well as on the *values* that the computation can return. Furthermore, the handler gets access to the *continuation* of the calling function as a first-class value. Consider the handler for **State**;

```
runState : {<State S> X -> S -> X}
runState <get -> k> s = runState (k s) s
runState <put s -> k> _ = runState (k unit) s
runState x _ = x
```

The type of runState expresses that the first argument is a computation that can perform state s effects and will eventually return a value of type x, whilst the second argument is a value of type s. The state s effect is then removed from the first argument.

What happens when we run runState xplusplus! 0? When a computation is invoked, it is performed until it results in either a *value* or a *command*. Thus, runState will be paused until xplusplus! reduces; runState is resumed when xplusplus is in one of these two forms.

xplusplus instantly invokes get!. At this point, control is given to the handler runState; both in the sense that runState is now being executed by the interpreter, and that runState has control over the *continuation* of xplusplus, which is a function of type Int -> [State Int] Unit. We see that runState chooses to resume this continuation with the value of the state at that time.

**Top-Level Effects** Some effects need to be handled outside of pure Frank, as Frank is not expressive or capable enough on its own. Examples are console I/O, web requests, and ML-style state cells.

These are handled by the interpreter.

**Synchronicity and Conversations** Observe how the interaction between the effect invoking function and the handler of this effect becomes like a conversation; the caller asks the handler for a response to an operation, and the caller will then wait, blocking, for a response. This can be characterised as *synchronous* effect handling.

But what if we want to make a request for information, then do something else, then pick up the result later when we need it? This is the canonical example of asynchronous programming. It is not as simple as just invoking our e.g. <code>getRequest</code> effect; computation would block once this is invoked, meaning we are stuck waiting for the request to return.

This asynchrony is exactly what we search for in this project.

**Multihandlers** Recall that in Frank pure functions are just the special case of handlers that handle no effects. Naturally, this notion extends to the *n*-ary case; we can handle multiple effects from different sources are once. Handlers which handle multiple effects simultaneously are unsurprisingly called *multihandlers*. This lets us write functions such as pipe (example due to Convent et al. [2020]);

```
1 interface Send X = send : X -> Unit
2
3 interface Receive X = receive : X
4
5 pipe : {<Send X>Unit -> <Receive X>Y -> [Abort]Y}
6 pipe <send x -> s> <receive -> r> = pipe (s unit) (r x)
7 pipe <_> y = y
8 pipe unit <_> = abort!
```

Line 5 states that pipe will handle all instances of the send effect in the first argument, all instances of the Receive effect in the second, and might perform Abort commands along the way. The matching clauses are also new to the reader; line 6 implements the communication between the two functions. We reinvoke pipe, passing the payload x of send to the continuation of r. Lines 7 and 8 make use of the catchall pattern, <m>. This will match the invocation of any effect that is handled by that argument, or a value, binding this to m. In line 7, the catchall pattern matches either a send command or a value; in this case, the receiver has produced a value, so we can return that. In line 8 <\_> matches either a value or a receive; but it must be a receive command, as the value case would have been caught above. Hence we have a broken pipe, so the abort command is invoked. This can then be caught by another handler, which can implement a recovery strategy.

```
TODO: Is it worth changing the example to match request on the left earlier?
```

**Polymorphic Commands** As well as having polymorphic interfaces, such as **state x**, parametrised by e.g. the data stored in the state, Frank supports polymorphic *com*-

*mands*. These are commands which can be instantiated for any type. An example is ML-style references, realised through the **RefState** interface;

For instance, **new x** can be instantiated by supplying a value as an argument. A **Ref x** cell is then returned as answer.

### 2.2 Concurrency

Frank is a single-threaded language. It is fortunate, then, that effect handlers give us a malleable way to run multiple program-threads "simultaneously"

```
TODO: This is poorly written — fix
```

This is because the invocation of an operation not only offers up the operation's payload, but also the *continuation* of the calling computation. The handler for this operation is then free to do what it pleases with the continuation. For many effects, such as <code>getState</code>, nothing interesting happens to the continuation; in the case of <code>getState</code>, it is resumed with the value in state. But these continuations are first-class; they can resumed, sure, but also stored elsewhere or even thrown away. As such, by handling <code>Yield</code> operations, we easily pause and switch between several threads.

### 2.2.1 Simple Scheduling

We introduce some simple program threads and some scheduling multihandlers, to demonstrate how subtly different handlers generate different scheduling strategies.

```
interface Yield = yield : Unit
words : {[Console, Yield] Unit}
words! = print "one "; yield!; print "two "; yield!; print "three ";
    yield!

numbers : {[Console, Yield] Unit}
numbers! = print "1 "; yield!; print "2 "; yield!; print "3 "; yield
!
```

First note the simplicity of the Yield interface; we have one operation supported, which looks very boring; the operation yield! will just return unit — of course, it is

the way we *handle* yield that is more interesting.

```
-- Runs all of the LHS first, then the RHS.

scheduleA : {<Yield> Unit -> <Yield> Unit -> Unit}

scheduleA <yield -> m> <n> = scheduleA (m unit) n!

scheduleA <m> <yield -> n> = scheduleA m! (n unit)

scheduleA _ _ = unit

-- Lets two yields synchronise, then handles both

scheduleB : {<Yield> Unit -> <Yield> Unit -> Unit}

scheduleB <yield -> m> <yield -> n> = scheduleB (m unit) (n unit)

scheduleB <yield -> m> <n> = scheduleB (m unit) n!

scheduleB <m> <yield -> n> = scheduleB m! (n unit)

scheduleB _ _ = unit
```

**TODO:** Can maybe delete the 2nd and 3rd matches of scheduleB to make the point more clear?

We see two multihandlers above. Each take two yielding threads and schedule them, letting one run at a time. scheduleA runs the first thread to completion, and only then runs the second one; the first time that the second thread yields it is blocked, and can no longer execute. As such, the output of scheduleA words! numbers! is one 1 two three 2 3 unit.

scheduleB is fairer and more profound. We run scheduleB words! number! and receive one 1 two 2 three 3 unit; scheduleB is fair and will "match" the yields together. We step through slowly. First words! will print one, then it will yield. At this point — recalling that multihandlers pattern match left-to-right — the second thread, numbers!, is allowed to execute. In the meantime, words! is stuck as <yield -> m>; it cannot evaluate any further, it is blocked. Whilst words is blocked numbers! prints 1 and then yields. Great; now the first case matches. Both threads are resumed and the process repeats itself.

**TODO:** The second paragraph here is a more compelling explanation; maybe we can just get rid of all of the scheduleA business and /just/ have the scheduleB stuff? scheduleA is quite obvious i think whilst B is more subtle and compelling.

**TODO:** It's not true that it matches L-R as much as runs all computations L - R until they are all a command / value - fix this

#### 2.2.2 Forking New Processes

We can make use of Frank's higher-order effects to dynamically create new threads at runtime. We strengthen the Yield interface by adding a new operation fork;

The type of **fork** expresses that **fork** takes a suspended computation that can perform further **co** effects, and returns unit when handled. We can now run programs that allocate new threads at runtime, such as the below

We can now choose a strategy for handling fork operations; we can either lazily run them, by finishing executing the current thread, or eagerly run them, suspending the currently executing thread and running the forked process. The handler for the former, breadth-first style of scheduling, is

```
scheduleBF : {<Co> Unit -> [Queue Proc] Unit}
scheduleBF <fork p -> k> =
    enqueue (proc {scheduleBF (<Queue> p!)});
    scheduleBF (k unit)
scheduleBF <yield -> k> =
    enqueue (proc {scheduleBF (k unit)});
    runNext!
scheduleBF <exit -> _> =
    runNext!
scheduleBF unit =
    runNext!
```

Notice the use of the Queue effect; we have to handle the computation scheduleBF forker! with a handler for Queue effects afterwards. Moreover, notice how concisely we can express the scheduler; this is due to the handler having access to te continuation of the caller, and treating it as a first-class object that can be stored elsewhere. We can see a diagram of how scheduleBF treats continuations in Figure 2.1, and a similar diagram of how the depth-first handling differs in Figure 2.2.

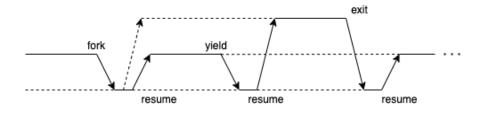


Figure 2.1: Breadth-First scheduling

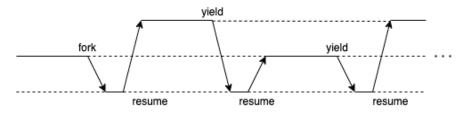


Figure 2.2: Depth-First scheduling

# Chapter 3 Formalisation of Frank

(1, , , , )	D	(interfaces)	I
(data types)	D	(term variables)	x, y, z, f
(value type variables)	X	•	
(effect type variables)	E	(instance variables)	s,a,b,c
,		(seeds)	$\sigma ::= \emptyset \mid E$
(value types)	$A,B ::= D \overline{R}$	(abilities)	$\Sigma ::= \sigma \!\mid\! \Xi$
	$  \{C\}   X$	(extensions)	$\Xi ::= \iota \mid \Xi, I \ \overline{R}$
(computation types)	$C ::= \overline{T  o G}$	,	
(argument types)	$T::=\langle\Delta angle A$	(adaptors)	$\Theta ::= \iota \mid \Theta, I(S \to S')$
( )	,	(adjustments)	$\Delta ::= \Theta   \Xi$
(return types)	$G ::= [\Sigma]A$	(instance patterns)	$S ::= s \mid S a$
(type binders)	$Z ::= X \mid [E]$	,	ı
(type arguments)	$R ::= A \mid [\Sigma]$	(kind environments)	$\Phi,\Psi ::= \cdot \mid \Phi, Z$
(***		(type environments)	$\Gamma ::= \cdot \mid \Gamma, x : A \mid \Gamma, f : P$
(polytypes)	$P ::= \forall \overline{Z}.A$	(instance environments	$\Omega ::= s : \Sigma \mid \Omega, a : I \overline{R}$
			•

Figure 3.1: Types

```
\begin{array}{ll} \text{(constructors)} & k \\ \text{(commands)} & c \\ \text{(uses)} & m ::= x \mid f \; \overline{R} \mid m \; \overline{n} \mid \uparrow (n : A) \\ \text{(constructions)} & n ::= \downarrow m \mid k \; \overline{n} \mid c \; \overline{R} \; \overline{n} \mid \{e\} \\ & \mid \; \text{let} \; f : P = n \; \text{in} \; n' \mid \text{letrec} \; \overline{f : P = e} \; \text{in} \; n \\ & \mid \; \langle \Theta \rangle \; n \\ \text{(computations)} & e ::= \overline{r} \mapsto n \\ \text{(computation patterns)} & r ::= p \mid \langle c \; \overline{p} \to z \rangle \mid \langle x \rangle \\ \text{(value patterns)} & p ::= k \; \overline{p} \mid x \\ \end{array}
```

Figure 3.2: Terms

$$\begin{array}{c} \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ T - VAR \\ \underline{x : A \in \Gamma} \\ \overline{\Phi}; \Gamma[\underline{\varepsilon}] - x \Rightarrow A \\ \hline \\ T - APP \\ \hline \\ \Sigma' = \underline{\Sigma} \\ (\Sigma \vdash \Delta_i \dashv \Sigma'_i)_i \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow \{\overline{\langle \Delta \rangle} A \to [\Sigma'] B\} \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow B \\ \hline \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ \Phi; \Gamma[\underline{\varepsilon}] - m \Rightarrow A \\ \hline \\ A = B \\ \hline \\ A = B \\ \hline \\ A = D \overline{R} \\$$

$$\frac{(\Phi \vdash r_{i,j} : T_j \vdash [\Sigma] \exists \Psi_{i,j} . \Gamma'_{i,j})_{i,j}}{(\Phi, (\Psi_{i,j})_j ; \Gamma, (\Gamma'_{i,j})_j [\Sigma \vdash n_i : B)_i \qquad ((r_{i,j})_i \text{ covers } T_j)_j}{\Phi; \Gamma \vdash ((r_{i,j})_j \mapsto n_i)_i : (T_j \to)_j [\Sigma] B}$$

Figure 3.3: Term Typing Rules

(uses) 
$$m ::= \cdots \mid \lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil$$
  
(constructions)  $n ::= \cdots \mid \lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil$   
(use values)  $u ::= x \mid f \ \overline{R} \mid \uparrow(v : A)$   
(non-use values)  $v ::= k \ \overline{w} \mid \{e\}$   
(construction values)  $w ::= \downarrow u \mid v$   
(normal forms)  $t ::= w \mid \lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil$   
(evaluation frames)  $\mathcal{F} ::= [\ ] \ \overline{n} \mid u \ (\overline{t}, [\ ], \overline{n}) \mid \uparrow([\ ] : A)$   
 $| \ \downarrow[\ ] \mid k \ (\overline{w}, [\ ], \overline{n}) \mid c \ \overline{R} \ (\overline{w}, [\ ], \overline{n})$   
 $| \ \text{let} \ f : P = [\ ] \ \text{in} \ n \mid \langle \Theta \rangle \ [\ ]$   
(evaluation contexts)  $\mathcal{E} ::= [\ ] \mid \mathcal{F}[\mathcal{E}]$ 

Figure 3.4: Runtime Syntax

$$\begin{array}{c|c} \Phi; \Gamma[\Sigma \vdash m \Rightarrow A] & \Phi; \Gamma[\Sigma \vdash n : A] \\ \hline \\ T\text{-Freeze-Use} \\ \neg (\mathcal{E} \text{ handles } c) & \Phi; \Gamma[\Sigma \vdash \mathcal{E}[c \ \overline{R} \ \overline{w}] \Rightarrow A \\ \hline \\ \Phi; \Gamma[\Sigma \vdash \lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil \Rightarrow A \\ \hline \\ T\text{-Freeze-Cons} \\ \neg (\mathcal{E} \text{ handles } c) & \Phi; \Gamma[\Sigma \vdash \mathcal{E}[c \ \overline{R} \ \overline{w}] : A \\ \hline \\ \Phi; \Gamma[\Sigma \vdash \lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil : A \\ \hline \end{array}$$

Figure 3.5: Frozen Commands

Figure 3.6: Operational Semantics

$$r: T \leftarrow t - [\Sigma] \theta$$

B-VALUE  

$$\Sigma \vdash \Delta \dashv \Sigma'$$

$$p: A \leftarrow w \dashv \theta$$

$$p: \langle \Delta \rangle A \leftarrow w \dashv \Sigma \mid \theta$$

**B-REQUEST** 

$$\begin{array}{c|c}
\Sigma \vdash \Delta \dashv \Sigma' & \mathcal{E} \text{ poised for } c \\
\Delta = \Theta \mid \Xi & c : \forall \overline{Z}.\overline{B \to} B' \in \Xi & (p_i : B_i \leftarrow w_i \dashv \theta_i)_i \\
\hline
\langle c \ \overline{p} \to z \rangle : \langle \Delta \rangle A \leftarrow \lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil \dashv [\Sigma] \ \overline{\theta}[\uparrow(\{x \mapsto \mathcal{E}[x]\} : \{B' \to [\Sigma']A\})/z] \\
B-CATCHALL-VALUE \\
\underline{\Sigma \vdash \Delta \dashv \Sigma'} \\
\hline
\langle x \rangle : \langle \Delta \rangle A \leftarrow w \dashv [\Sigma] \ [\uparrow(\{w\} : \{[\Sigma']A\})/x]
\end{array}$$

B-CATCHALL-REQUEST

$$\begin{split} \Sigma \vdash \Delta \dashv \Sigma' & \mathcal{E} \text{ poisedfor } c \\ \Delta &= \Theta \mid \Xi & c : \forall \overline{Z}. \overline{B \to} B' \in \Xi \\ \hline \langle x \rangle : \langle \Delta \rangle A \leftarrow \lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil \dashv_{\Sigma} \lceil \uparrow (\{\lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil\} : \{\lceil \Sigma' | A \}) / x \rceil \end{split}$$

 $p: A \leftarrow w \dashv \theta$ 

B-VAR 
$$\frac{k \, \overline{A} \in D \, \overline{R}}{x : A \leftarrow w \, \exists \, [\uparrow(w : A)/x]} \qquad \frac{k \, \overline{A} \in D \, \overline{R}}{k \, \overline{p} : D \, \overline{R} \leftarrow k \, \overline{w} \, \exists \, \overline{\theta}}$$

Figure 3.7: Pattern Binding

# **Chapter 4**

## **Pre-emption**

#### 4.1 Motivation

One important part of our asynchronous effect handling system is the ability to interrupt arbitrary computations. This is essential for pre-emptive concurrency, which relies on being able to suspend a computation *non-cooperatively*; the computation gets suspended without being aware of its suspension.

```
TODO: Rewrite this - clumsy
```

Consider the two programs below;

```
controller : {[Stop, Go, Console] Unit}
controller! =
    stop!; print "stop "; sleep 200000; go!; controller!

runner : {[Console] Unit}
runner! = print ''1 ''; print ''2 ''; print ''3 '';
```

We want a multihandler that uses the stop and go commands from controller to control the execution of runner. The console output of this multihandler should be then 1 stop 2 stop 3 stop.

## 4.2 Interruption with Yields

We can simulate this behaviour by using the familiar Yield interface from Section 2.2.1.

```
runner : {[Console, Yield] Unit}
runner! = print "1 "; yield!; print "2 "; yield!; print "3 "; yield!
```

B-CATCHALL-REQUEST 
$$\Sigma \vdash \Delta \dashv \Sigma' \qquad \mathcal{E} \text{ poised for } c$$
 
$$\Delta = \Theta \mid \Xi \qquad c : \forall \overline{Z}.\overline{B \to} B' \in \Xi$$
 
$$\overline{\langle x \rangle} : \langle \Delta \rangle A \leftarrow \lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil - [\Sigma] \left[ \uparrow (\{\lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil\} : \{[\Sigma']A\})/x \right]$$
 B-CATCHALL-REQUEST-LOOSE 
$$\Sigma \vdash \Delta \dashv \Sigma'$$
 
$$\overline{\langle x \rangle} : \langle \Delta \rangle A \leftarrow \lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil - [\Sigma] \left[ \uparrow (\{\lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil\} : \{[\Sigma']A\})/x \right]$$

Figure 4.1: Updated B-CATCHALL-REQUEST

Running suspend runner! controller! nothing then prints out 1 stop 2 stop 3 as desired. This is due to the same synchronisation behaviour that we saw in Section 2.2.1; runner is evaluated until it becomes a command or a value, and then controller is given the same treatment. Once both are a command or a value, pattern matching is done.

So far so good; this works as we hoped. However, observe that we had to change the code of the runner to yield every time it prints. This is not in the spirit of preemptive concurrency; we are still operating co-operatively. Threads should be unaware of the fact they are even being pre-empted. Furthermore, see that the yield operation adds no more information; it is just used as a placeholder operation; any operation would work. As such, we keep searching for a better solution.

```
TODO: Penultimate sentence could maybe go, a bit clumsy / weird
```

## 4.3 Relaxing Catches

The key to this lies in the catchall pattern,  $\langle x \rangle$ , and the pattern binding rules of Figure 3.7; specifically B-CATCHALL-REQUEST. We quickly go into detail on this rule

B-CATCHALL-REQUEST 
$$\Sigma \vdash \Delta \dashv \Sigma' \qquad \mathcal{E} \text{ poised for } c$$
 
$$\Delta = \Theta \mid \Xi \qquad c : \forall \overline{Z}.\overline{B \to} B' \in \Xi$$
 
$$\overline{\langle x \rangle} : \langle \Delta \rangle A \leftarrow \lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil \dashv_{\Sigma} [\uparrow (\{\lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil\} : \{[\Sigma']A\})/x]$$
 B-CATCHALL-REQUEST-LOOSE 
$$\Sigma \vdash \Delta \dashv \Sigma'$$
 
$$\overline{\langle x \rangle} : \langle \Delta \rangle A \leftarrow \lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil \dashv_{\Sigma} [\uparrow (\{\lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil\} : \{[\Sigma']A\})/x]$$

Figure 4.2: Updated B-CATCHALL-REQUEST

now.  $\langle x \rangle$ :  $\langle \Delta \rangle A$  states that  $\langle x \rangle$  is a term with value type A and *adjustment*  $\Delta = \Theta \mid \Xi$ , made up of an adaptor  $\Theta$  and an extension  $\Xi$ . This extension is made up of a list of interface instantiations  $I \overline{R}$ .

The crux is that the command c that is invoked in the frozen term  $\lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil$  must be an element of this extension  $\Xi$ ; that is, it must be handled by the current use of R-HANDLE. Refer back to the example of Section 4.2. This rules means that the catch-all pattern <\_> in the final pattern matching case of suspend can match against stop or go, as they are present in the extension of the second argument, but not print commands; although the Console interface is present in the ability of controller, it is not in the extension in suspend.

In the interests of pre-emption, we propose to remove this constraint from B-CATCHALL-REQUEST. The resulting rule B-CATCHALL-REQUEST-LOOSE can be seen in Figure 4.2. This lets us update the previous **suspend** code to the following, which yields the same results as last time;

**TODO:** Check that the above still works...

```
(uses) m ::= ... | \lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil | \boxed{m}

(constructions) n ::= ... | \lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil | \boxed{m}

(use values) u ::= x | f \ \overline{R} | \uparrow (v : A)

(non-use values) v ::= k \ \overline{w} | \{e\}

(construction values) w ::= \downarrow u | v

(normal forms) t ::= w | \lceil \mathcal{E}[c \ \overline{R} \ \overline{w}] \rceil | \boxed{m}

(evaluation frames) \mathcal{F} ::= [\ ] \ \overline{n} | u \ (\overline{t}, [\ ], \overline{n}) | \uparrow ([\ ] : A)

| \downarrow [\ ] | k \ (\overline{w}, [\ ], \overline{n}) | c \ \overline{R} \ (\overline{w}, [\ ], \overline{n})

| \text{let } f : P = [\ ] \text{ in } n | \langle \Theta \rangle [\ ]

(evaluation contexts) \mathcal{E} ::= [\ ] | \mathcal{F}[\mathcal{E}]
```

Figure 4.3: Runtime Syntax, Updated with Freezing of Uses

Now when we run suspend runner! controller! nothing, the suspend handler is able to

**TODO:** Talk about how this still maintains the "no-snooping" policy; we can't inspect or access the effects that are invoked, but we know they happen.

Even with this extension,

## 4.4 Interrupting Arbitrary Terms

The approach of Section 4.3 can only interrupt command invocations. If **runner** were instead a sequence of pure computations<sup>1</sup>, we would be unable to interrupt it; it does not invoke commands.

As such, we need to further change the pattern binding rules of Figure 3.7. This is to let us interrupt arbitrary computation terms. In Figure ??, we see an updated version of the runtime syntax; this allows for the suspension of arbitrary *uses*, essentially just function applications.

```
TODO: verify that uses are "just" apps and constructions
```

Note that frozen terms here behave in a similar way to frozen commands, by freezing the rest of the term around it as well. This continues up until a handler is reached, at which point the term is unfrozen and resumed.

With this in mind, we now give the updated rule for the catchall pattern matching on frozen terms. This can be seen in Figure 4.5. It expresses that an arbitrary frozen

<sup>&</sup>lt;sup>1</sup>I.e. runner! = 1 + 1; 1 + 1; 1 + 1; ...

Figure 4.4: Updated Freezing

TODO: Maybe need to add that the eval ctx is NOT a handler?

B-CATCHALL-INTERRUPT 
$$\frac{\Sigma \vdash \Delta \dashv \Sigma'}{\langle x \rangle : \langle \Delta \rangle A \leftarrow \lceil m \rceil \dashv_{\Sigma} \rceil [\uparrow (\{m\} : \{ [\Sigma']A \}) / x]}$$

Figure 4.5: Catching Interrupts rule.

use can be matched against the computation pattern  $\langle x \rangle$ . The suspended, unfrozen computation  $\{m\}$  is then bound to x, in a similar way to other B-CATCHALL rules.

Figure 4.6 shows how uses m become interrupted. This rule supplements the operational semantics of Figure 3.6. It states that at any point, a use term can reduce to the same term but suspended. At this point, the term cannot reduce any further; observe that  $\lceil \rceil$  is not an evaluation context. The term then blocks until unfrozen. Note how similar this is to regular command invocations, which block until their continuation is invoked.

TODO: Check that just uses is enough

TODO: Talk about how this achieves our goal.

The addition of this rule introduces non-determinism into the language; at any point, a use can either step as normal (e.g. through the R-HANDLE rule), or it can be interrupted. An interrupted term !(m) can no longer reduce; it is blocked until it is resumed by the R-HANDLE rule.

TODO: Talk about non-determinism as a result of this

TODO: Maybe move non-determinism to the next section?

$$\frac{R\text{-Interrupt}}{m \leadsto_{\mathbf{u}} ! (m)}$$

Figure 4.6: Use interruption rule

## 4.5 Interrupting In Practice

Due to the non-determinism introduced by R-INTERRUPT, this system is difficult to implement; how do we choose whether to apply a handler to its arguments or to just interrupt it?

In our implementation of this system, we instead maintain a counter which is incremented every time a handler is evaluated. When the counter reaches a certain threshold value t, we interrupt the current term m, applying the R-Interrupt rule. This converts the non-deterministic system to a deterministic one; we never have any question of what to do.

Observe that the process of interrupting a computation is a familiar one; we stop computing and offer up the continuation to the programmer. Where is a similar control flow already around? That's right — invocation of an effect.

**TODO:** Rewrite above paragraph to be less camp

As such, we can just use a normal effect to perform the interruption. Sticking with previous themes, we choose to invoke a **yield** effect when interrupting. This lets the programmer choose to handle the effect as they wish. Note that now interrupts are not restricted to the catch-all pattern  $\langle m \rangle$  but are normal computation patterns.

These yield effects are only inserted in a computation when the Yield interface is present in the ability of this computation. This is important, as it lets the programmer get fine-grained control over which computations can be interrupted and which cannot. Consider the example in Section 4.1; we want the runner computation to be controller by the controller. As such, we want the runner to be interruptible, whilst the controller is not. We can reflect this by adding the Yield interface to the ability of the former computation.

Thus our example from before becomes;

**TODO:** updated example

## 4.6 Soundness

We now state the soundness property for our extended system, as well as the subject reduction theorem needed for this proof.

**Theorem 1 (Subject Reduction)** • *If*  $\Phi$ ;  $\Gamma[\Sigma]$ -  $m \Rightarrow A$  and  $m \leadsto_{\mathrm{u}} m'$  then  $\Phi$ ;  $\Gamma[\Sigma]$ -  $m' \Rightarrow A$ .

• If  $\Phi$ ;  $\Gamma[\Sigma]$ -n:A and  $n \leadsto_{\mathbb{C}} n'$  then  $\Phi$ ;  $\Gamma[\Sigma]$ -n':A.

By induction on the transitions....

**TODO:** Finish this

# **Chapter 5**

## **Implementation**

We now introduce the Frank library used for asynchronous effects. Our design closely follows the design of Æff (Ahman and Pretnar [2020]). Æff is a language designed around writing multithreaded programs that communicate by sending *interrupts*. A thread dictates how it will respond to an interrupt by installing an *interrupt handler*. Interrupts and interrupt handlers can be seen as a less expressive version of effects and effect handlers; an interrupt handler describes how to behave on receipt of an interrupt in the same way to an effect handler, but it does not get access to the continuation of the calling code.

Interrupts and interrupt handlers have one particularly compelling feature; when we invoke an interrupt (in the case of synchronous effects, this is just invoking a command), we can carry on computing the rest of the code whilst we wait for a response. This is a stark difference to traditional effects, where the rest of the computation is blocked whilst we wait for an answer. The programmer can then choose to await the response from interrupt, which does block computation if an answer is not already received.

Our system untyped in that there is no tracking of which asynchronous effects are issued in which functions, however it is typed in that the traditional Frank effects promises can perform are tracked.

Æff has an effect tracking system for asynchronous effects; our system does not. Ours does however track the effects that can be performed by interrupt handlers.

#### 5.1 In Frank

Æff's interrupt handlers are manifested in Frank through the Promise interface.

We have a lot to unpack here, so we start slowly. The promise R command is a polymorphic command, which takes a function of type Sig -> Maybe {[E|] R}. This function is an interrupt handler; it dictates what to do on receipt of an interrupt, which is a thing of type Sig. The return type of the interrupt handler is Maybe {[E|] R}; this is because the programmer has the chance to return nothing, which will mean the promise goes unfulfilled and waits for another message. The programmer would want to do this on receipt of other types of message, or if a certain condition regarding the interrupt is not fulfilled<sup>1</sup>. The promise operation returns a Pid R (promise id), which is used to check if the installed promise has completed or not. The R type parameter is determined by the return type of the interrupt handler. This is later used in await.

**signal** is a more simple operation. The **sig** data type is the type of signals that the thread can invoke. These can have extra bits of information — also called *payloads* — such as the parameters for a remote request. The handler for **Promise** will then send the signals to each other thread, possibly executing the interrupt handler if needs be.

await takes a Pid R and returns a value of type R. This R is the returned value of the promise. If the promise has been fulfilled then the promise status stored in Pid will be done res; hence we just return this value. If it is incomplete — i.e. the promise status is empty — we add the resumption to resume.

Effect Typing We can track and control the effects that promises can perform using Frank's effect type system. For instance, a Frank program of type [Promise [Console 11 x can install promises that print to the console, a program of type [Promise [Console, Web]] x can install promises that perform web requests and print to the console, etc.

<sup>&</sup>lt;sup>1</sup>Interrupt handlers which put conditions on the incoming interrupts are called *guarded* interrupt handlers — we come back to these later.

Note however that the effect typing is slightly complicated in the definition of the interface; a type of Promise [Console] means that the installed promise can really perform the effects [Promise[Console], Console, RefState, Yield]. This is expressed by the [E | Promise[E|], RefState, Yield] part of the promise R definition. A recursive type is needed as the promises can themselves invoke other promises.

**TODO:** Flesh this out, rewrite it

#### 5.1.1 Handling

In order to run threads in parallel, we need to maintain a collection of thread states; when we stop executing a thread we store its continuation so far and start executing a new one, in the same style as Section 2.2.

We also need to store in this collection the promises that each thread has installed. These are stored as a stack so as to maintain the order of installation. Installing a promise is as straightforward as pushing it onto the corresponding thread's promise stack.

# **Chapter 6**

# **Examples**

## 6.1 Pre-emptive Concurrency

An essential feature of our asynchronous effects system is that it supports pre-emptive concurrency; that is, the suspension and resumption of threads non-cooperatively. Naturally, this relies on the insertion of yields as discussed in Chapter ??.

We supplement the signals supported in our program with two more;

```
data Sig = ... | stop Int | go Int
```

The integer payload can act as a counter, or as a way to tell specific threads to stop or go. The blocking or non-blocking behaviour then depends on the promises for these signals.

```
onStop : {Int -> [Promise[Console], Console, Yield, RefState] Unit}
onStop id =
   let gp = promise (prom {s -> goPromise id s}) in
   await gp;
   promise (prom {s -> stopPromise id s});
   unit

stopPromise : {Int -> Sig -> Maybe {[Promise[Console], Console,
        Yield, RefState]Unit}}
stopPromise id (stop n) =
   if (n == id)
        { just { onStop id } }
        { nothing }
stopPromise id _ = nothing
```

**stopPromise** is another guarded interrupt handler; it will only fire its body if the payload to **stop** is the id of the thread. The body is then fairly simple; it installs a

promise waiting for **go** and immediately starts blocking. The rest of the computation can not proceed until the corresponding go message is received. Once the go promise is fulfilled, the non-blocking stop promise is reinstalled.

```
goPromise : {Int -> Sig -> Maybe {[Promise[Console], Console, Yield,
    RefState]Unit}}
goPromise id (go n) =
    if (n == id)
        { just {unit} }
        { nothing }
goPromise id _ = nothing
```

goPromise is simple in comparison; if it receives the correct go signal it just returns unit.

We can then make a function pre-emptible by just installing a stop-waiting promise in front of the function code;

```
counter : {Int -> [Console, Yield]Unit}
counter x = ouint x; print " "; sleep 200000; counter (x + 1)
thread1 : {[Promise[Console], Console, RefState, Yield] Unit}
thread1! = promise (prom {s -> stopPromise 0 s}); counter 0
```

Observe that all we have to do is precompose with the promise installer; the rest of the code goes on unaware that it is being pre-empted. Threads can also then communicate etc on top of this.

```
TODO: Is this example even interesting now that we've already got it baked into the language?
```

## 6.2 Async-Await

Here we show how our asynchronous effects system can express the familiar asyncawait abstraction.

Consider that we want to asynchronously run web requests using the built-in getRequest operation. These return a value of type String, being hte result of the request. First we add two more signals to our set of available signals;

```
data Sig = ... | call {String} Int | result String Int
```

```
TODO: Observe that this is a higher-order effect - something aeff lacks!
```

The signal call is used to start an asynchronous operation; the thunked argument is the computation we want to run. result is the signal used by the running thread to

indicate that it has completed the computation and is returning the **String** result. The **Int** arguments are for call IDs, so that the wrong results are not re-read.

Unlike other implementations, the Frank realisation of async-await does not dynamically create new threads to run asynchronous tasks. Instead, we have a dedicated thread that only performs these asynced processes. This may seem inefficient, however see that when not executing a process the thread will be instantly skipped in the scheduler, so we have no overhead costs.

We how show the **async** function that a caller would use to issue a new asynchronous task;

So async takes the process to be run and a reference to the callcounter. It then installs another promise, resultWaiter, which waits for the corresponding result signal to be received. resultWaiter is an example of a guarded interrupt handler; it only fires if a certain condition regarding to the signal's payload holds (i.e. that callNo == callNo'). After installing resultWaiter, async sends a call signal with the process and callNo as argument, increments the call counter and returns the resultwaiting promise.

```
onRun : {{[Console, Web] String} -> Int -> [WebThreads] Unit}
onRun proc callId =
   let res = <Promise, RefState, Yield> proc! in
   signal (result res callId);
   <Console, RefState, Web, Yield> runner!;
   unit
```

runner is the process that runs on the worker thread. This simply installs a promise that responds to call signals. On receipt of a call it runs the delivered process synchronously; once it is finished it sends a result signal and then finally reinvokes the runner. Note that proc can still have yield calls inserted into it, so that this doesn't cause the whole program to block.

**TODO:** Example of how it gets used.

# **Bibliography**

Danel Ahman and Matija Pretnar. Asynchronous effects. *arXiv preprint* arXiv:2003.02110, 2020.

Lukas Convent, Sam Lindley, Conor McBride, and Craig McLaughlin. Doo bee doo bee doo. *J. Funct. Program.*, 30:e9, 2020.

# Appendix A Remaining Formalisms

 $\Omega \vdash s : I \dashv \iota$ 

Figure A.1: Action of an Adjustment on an Ability and Auxiliary Judgements

 $\Omega \vdash S \ a : I \dashv \Xi, I \ \overline{R}$ 

$$X ::= A \mid C \mid T \mid G \mid Z \mid R \mid P \mid \sigma \mid \Sigma \mid \Xi \mid \Theta \mid \Delta \mid \Gamma \mid \exists \Psi.\Gamma \mid \Omega$$

$$\frac{\Phi \vdash X}{\Phi \vdash X}$$

$$\frac{WF\text{-Val}}{\Phi, X \vdash X} \qquad \frac{WF\text{-Eff}}{\Phi, [E] \vdash E} \qquad \frac{WF\text{-Poly}}{\Phi \vdash \sqrt{Z}.A}$$

$$\frac{WF\text{-Data}}{\Phi \vdash DR} \qquad \frac{WF\text{-Thunk}}{\Phi \vdash C} \qquad \frac{WF\text{-Comp}}{\Phi \vdash T \to G} \qquad \frac{WF\text{-Arg}}{\Phi \vdash \Delta} \qquad \frac{WF\text{-Arg}}{\Phi \vdash \Delta \land \Phi \vdash A}$$

$$\frac{WF\text{-Ret}}{\Phi \vdash \Sigma} \qquad \frac{WF\text{-Abillity}}{\Phi \vdash (E)} \qquad \frac{WF\text{-Pure}}{\Phi \vdash \emptyset} \qquad \frac{WF\text{-Id}}{\Phi \vdash 1} \qquad \frac{WF\text{-Ext}}{\Phi \vdash \Xi, I R}$$

$$\frac{WF\text{-Adapt}}{\Phi \vdash \Theta} \qquad \frac{WF\text{-Adapt}}{\Phi \vdash (E)} \qquad \frac{WF\text{-Mono}}{\Phi \vdash (E)} \qquad \frac{WF\text{-Poly}}{\Phi \vdash \Gamma, x : A} \qquad \frac{WF\text{-Poly}}{\Phi \vdash \Gamma, f : P}$$

$$\frac{WF\text{-Existential}}{\Phi \vdash \Theta \vdash \Gamma} \qquad \frac{WF\text{-Interface}}{\Phi \vdash \Omega, x : I R}$$

Figure A.2: Well-Formedness Rules

Figure A.3: Pattern Matching Typing Rules