Continuing WebAssembly with Effect Handlers

Daniel Hillerström

Computing Systems Laboratory Zürich Research Center Huawei Technologies Switzerland

April 26, 2023



I am but one of many



Sam Lindley



Andreas Rossberg



Daan Leijen



KC Sivaramakrishnan



Matija Pretnar



Luna Phipps-Costin



Arjun Guha

https://wasmfx.dev

WebAssembly (Wasm): neither web nor assembly (Haas et al. 2017)

What is Wasm?

- A low-level virtual machine
- Source language agnostic
- Platform independent target
- Formally specified¹ and mechanised
- A predictable performance model

Code format

- A Wasm "program" is a structured module
- Designed for streaming compilation
- The term language is statically typed and block-structured
- Control flow is structured (*i.e.* all CFGs are reducible)

https://webassembly.org

¹Wasm 1.0 has a tiny bit of nondeterminism related to floating point NaNs

Stack machine

(i32.const 2) (i32.const 5) (i32.add)

Stack machine

(i32.const 2)(i32.const 5) $(i32.add) \longrightarrow (i32.const 7)$ Stack machine

(i32.const 2)
(i32.const 5)
(i32.add) → (i32.const 7)

Syntactic sugar: (i32.add (i32.const 2) (i32.const 5))

Stack typing

(i32.const	2)	:	$[] \rightarrow [i32]$
(i32.const	5)	:	$[] \rightarrow [i32]$
(i32 .add)		:	$[i32 i32] \rightarrow [i32]$

(**block** \$1 ... (**br** \$1) ...)











continue

Wasm 1.0 & 2.0+

Wasm 1.0 is tailored for C/C++

- The instruction set is an intersection of modern CPUs
- Memory model: a flat array of bytes
- Data types: i32, i64, f32, f64
- Modules, functions, and tables

Wasm 2.0 includes high-level language support

- Tail calls
- Typed function references
- Exception handling
- Garbage collection
- SIMD v128 data type (accounts for 236 out of 437 instructions)

Wasm 1.0 & 2.0+

Wasm 1.0 is tailored for C/C++

- The instruction set is an intersection of modern CPUs
- Memory model: a flat array of bytes
- Data types: i32, i64, f32, f64
- Modules, functions, and tables

Wasm 2.0 includes high-level language support

- Tail calls
- Typed function references
- Exception handling
- Garbage collection
- SIMD v128 data type (accounts for 236 out of 437 instructions)

Beyond Wasm 2.0

- Multithreading
- Memory64
- Higher-order modules
- First-class control (this talk!)

Non-local control is pervasive in programming languages

- Async/await (e.g. C++, C#, Dart, JavaScript, Rust, Swift)
- Coroutines (e.g. C++, Kotlin, Python, Swift)
- Lightweight threads (e.g. Erlang, Go, Haskell, Java, Swift)
- Generators and iterators (e.g. C#, Dart, Haskell, JavaScript, Kotlin, Python)
- First-class continuations (e.g. Haskell, Java, OCaml, Scheme)

Non-local control is pervasive in programming languages

- Async/await (e.g. C++, C#, Dart, JavaScript, Rust, Swift)
- Coroutines (e.g. C++, Kotlin, Python, Swift)
- Lightweight threads (e.g. Erlang, Go, Haskell, Java, Swift)
- Generators and iterators (e.g. C#, Dart, Haskell, JavaScript, Kotlin, Python)
- First-class continuations (e.g. Haskell, Java, OCaml, Scheme)

The problem

How do I compile non-local control flow abstractions to Wasm?

Non-local control is pervasive in programming languages

- Async/await (e.g. C++, C#, Dart, JavaScript, Rust, Swift)
- Coroutines (e.g. C++, Kotlin, Python, Swift)
- Lightweight threads (e.g. Erlang, Go, Haskell, Java, Swift)
- Generators and iterators (e.g. C#, Dart, Haskell, JavaScript, Kotlin, Python)
- First-class continuations (e.g. Haskell, Java, OCaml, Scheme)

The problem

How do I compile non-local control flow abstractions to Wasm?

Solution

• Ceremoniously transform my entire source programs (e.g. Asyncify, CPS)

Non-local control is pervasive in programming languages

- Async/await (e.g. C++, C#, Dart, JavaScript, Rust, Swift)
- Coroutines (e.g. C++, Kotlin, Python, Swift)
- Lightweight threads (e.g. Erlang, Go, Haskell, Java, Swift)
- Generators and iterators (e.g. C#, Dart, Haskell, JavaScript, Kotlin, Python)
- First-class continuations (e.g. Haskell, Java, OCaml, Scheme)

The problem

How do I compile non-local control flow abstractions to Wasm?

Solution

- Ceremoniously transform my entire source programs (e.g. Asyncify, CPS)
- Add each abstraction as a primitive to Wasm

Non-local control is pervasive in programming languages

- Async/await (e.g. C++, C#, Dart, JavaScript, Rust, Swift)
- Coroutines (e.g. C++, Kotlin, Python, Swift)
- Lightweight threads (e.g. Erlang, Go, Haskell, Java, Swift)
- Generators and iterators (e.g. C#, Dart, Haskell, JavaScript, Kotlin, Python)
- First-class continuations (e.g. Haskell, Java, OCaml, Scheme)

The problem

How do I compile non-local control flow abstractions to Wasm?

Solution

- Ceremoniously transform my entire source programs (e.g. Asyncify, CPS)
- Add each abstraction as a primitive to Wasm
- Use effect handlers as a unified modular basis for control in Wasm

WasmFX at glance

WasmFX is a minimal and compatible extension

- 1 new data type
- 6 new instructions (3 are core, 3 are nice-to-have)
- No new block structure

WasmFX uses effect handlers to manage stacks

- Modular and extensible basis for stack switching
 - Structured form of delimited control (intuition: first-class resumable exceptions)
 - Easy encoding of your favourite abstraction
 - Control abstractions compose (due to effect forwarding)
- Based on practical evidence
 - 100+ peer reviewed papers
 - Available in many programming languages (e.g. C++, Haskell, Pyro, OCaml, Unison)
 - Deployed in industrial technologies (e.g. GitHub's semantic, Meta's React, Uber's Pyro)
- Restriction: single-shot continuations

- ... without garbage collection
- ... without closures
- ... without the use of recursion
- ... with simply typed stacks
- ... with imperative control structure
- ... with predicable cost model
- ... with legacy code

- ... without garbage collection
- ... without closures
- ... without the use of recursion
- ... with simply typed stacks
- ... with imperative control structure
- ... with predicable cost model
- ... with legacy code

 \checkmark reference counting suffices

- ... without garbage collection
- ... without closures
- ... without the use of recursion
- ... with simply typed stacks
- ... with imperative control structure
- ... with predicable cost model
- ... with legacy code

- \checkmark reference counting suffices
- \checkmark first-order abstraction

- ... without garbage collection
- ... without closures
- \bullet \ldots without the use of recursion
- ... with simply typed stacks
- ... with imperative control structure
- ... with predicable cost model
- ... with legacy code

- \checkmark reference counting suffices
- \checkmark first-order abstraction
- $\checkmark\,$ no fundamental dependency

- ... without garbage collection
- ... without closures
- \bullet \ldots without the use of recursion
- ... with simply typed stacks
- ... with imperative control structure
- ... with predicable cost model
- ... with legacy code

- \checkmark reference counting suffices
- \checkmark first-order abstraction
- \checkmark no fundamental dependency
- \checkmark straightforward typing

- ... without garbage collection
- ... without closures
- ... without the use of recursion
- ... with simply typed stacks
- ... with imperative control structure
- ... with predicable cost model
- ... with legacy code

- \checkmark reference counting suffices
- \checkmark first-order abstraction
- \checkmark no fundamental dependency
- \checkmark straightforward typing
- $\checkmark\,$ compatible with builtin side-effects

- ... without garbage collection
- ... without closures
- \bullet \ldots without the use of recursion
- ... with simply typed stacks
- ... with imperative control structure
- ... with predicable cost model
- ... with legacy code

- \checkmark reference counting suffices
- \checkmark first-order abstraction
- \checkmark no fundamental dependency
- \checkmark straightforward typing
- $\checkmark\,$ compatible with builtin side-effects
- ✓ transparent cost of instructions

- ... without garbage collection
- ... without closures
- \bullet \ldots without the use of recursion
- ... with simply typed stacks
- ... with imperative control structure
- ... with predicable cost model
- ... with legacy code

- ✓ reference counting suffices
- \checkmark first-order abstraction
- \checkmark no fundamental dependency
- ✓ straightforward typing
- $\checkmark\,$ compatible with builtin side-effects
- \checkmark transparent cost of instructions
- \checkmark seamless interop

Variations on semantics of effect handlers

Deep allocation, capture, and resumption

cont.new V with $H \rightsquigarrow \operatorname{cont}_{\langle H; V \rangle}$, where $V = \lambda \langle \rangle . M$ resume V with $W \rightsquigarrow \underline{\operatorname{handle}} \mathcal{E}[W] \underline{\operatorname{with}} H$, where $V = \operatorname{cont}_{\langle H; \mathcal{E} \rangle}$ <u>handle</u> $\mathcal{E}[\operatorname{op} V] \underline{\operatorname{with}} H \rightsquigarrow N[\operatorname{cont}_{\langle H; \mathcal{E} \rangle}/r, V/x]$, where $\{\operatorname{op} p \ r \mapsto N\} \in H$

Variations on semantics of effect handlers

Deep allocation, capture, and resumption

 $\begin{array}{ll} \texttt{cont.new } V \texttt{ with } H \rightsquigarrow \texttt{cont}_{\langle H; V \rangle}, & \texttt{where } V = \lambda \langle \rangle.M \\ \texttt{resume } V \texttt{ with } W \rightsquigarrow \underline{\texttt{handle }} \mathcal{E}[W] \underline{\texttt{with }} H, \texttt{where } V = \texttt{cont}_{\langle H; \mathcal{E} \rangle} \\ \underline{\texttt{handle }} \mathcal{E}[\texttt{op } V] \underline{\texttt{with }} H \rightsquigarrow \mathcal{N}[\texttt{cont}_{\langle H; \mathcal{E} \rangle}/r, V/x], \texttt{ where } \{\texttt{op } p \ r \mapsto N\} \in H \end{array}$

Shallow allocation, capture, and resumption

 $\begin{array}{ll} & \operatorname{cont.new} V \rightsquigarrow \operatorname{cont}_{\langle V \rangle}, & \text{where } V = \lambda \langle \rangle.M \\ & \operatorname{resume} V \text{ with } W \rightsquigarrow \mathcal{E}[W], & \text{where } V = \operatorname{cont}_{\langle \mathcal{E} \rangle} \\ & \operatorname{handle} \mathcal{E}[\operatorname{op} V] \text{ with } H \rightsquigarrow N[\operatorname{cont}_{\langle \mathcal{E} \rangle}/r, V/x], \text{ where } \{\operatorname{op} p \ r \mapsto N\} \in H \end{array}$

Variations on semantics of effect handlers

Deep allocation, capture, and resumption

 $\begin{array}{ll} \texttt{cont.new } V \texttt{ with } H \rightsquigarrow \texttt{cont}_{\langle H; V \rangle}, & \texttt{where } V = \lambda \langle \rangle.M \\ \texttt{resume } V \texttt{ with } W \rightsquigarrow \underline{\texttt{handle }} \mathcal{E}[W] \underline{\texttt{with }} H, \texttt{where } V = \texttt{cont}_{\langle H; \mathcal{E} \rangle} \\ \underline{\texttt{handle }} \mathcal{E}[\texttt{op } V] \underline{\texttt{with }} H \rightsquigarrow N[\texttt{cont}_{\langle H; \mathcal{E} \rangle}/r, V/x], \texttt{ where } \{\texttt{op } p \ r \mapsto N\} \in H \end{array}$

Shallow allocation, capture, and resumption

 $\begin{array}{ll} & \operatorname{cont.new} V \rightsquigarrow \operatorname{cont}_{\langle V \rangle}, & \text{where } V = \lambda \langle \rangle.M \\ & \operatorname{resume} V \text{ with } W \rightsquigarrow \mathcal{E}[W], & \text{where } V = \operatorname{cont}_{\langle \mathcal{E} \rangle} \\ & \operatorname{handle} \mathcal{E}[\operatorname{op} V] \text{ with } H \rightsquigarrow N[\operatorname{cont}_{\langle \mathcal{E} \rangle}/r, V/x], \text{ where } \{\operatorname{op} p \ r \mapsto N\} \in H \end{array}$

'Sheep' allocation, capture, and resumption

 $\begin{array}{ll} & \operatorname{cont.new} V \rightsquigarrow \operatorname{cont}_{\langle V \rangle}, & \operatorname{where} V = \lambda \langle \rangle.M \\ & \operatorname{resume} V \text{ with } \langle H; W \rangle \rightsquigarrow \underline{\operatorname{handle}} \ \mathcal{E}[W] \, \underline{\operatorname{with}} \ H, \ \operatorname{where} \ V = \operatorname{cont}_{\langle \mathcal{E} \rangle} \\ & \underline{\operatorname{handle}} \ \mathcal{E}[\operatorname{op} V] \, \underline{\operatorname{with}} \ H \rightsquigarrow N[\operatorname{cont}_{\langle \mathcal{E} \rangle}/r, V/x], & \operatorname{where} \ \{\operatorname{op} \ p \ r \mapsto N\} \in H \end{array}$

'Seesaw' coroutines (Ganz, Friedman, and Wand 1999).



We will have two modules

- co2 implementing the coroutine runtime
- example interleaved streams of natural numbers

```
;; interface for running two coroutines
;; non-interleaving implementation
(module $co2
  ;; type alias task = [] \rightarrow []
  (type $task (func))
  ;; yield : [] -> []
  (func $yield (export "yield")
    (nop))
  :: run : [(ref $task) (ref $task)] -> []
  (func $run (export "run") (param $task1 (ref $task)) (param $task2 (ref $task))
    ;; run the tasks sequentially
    (call_ref $task (local.get $task1))
    (call_ref $task (local.get $task2))
```

Running example: coroutines (2)

(module \$example ;; main example: streams of odd and even naturals

```
...
;; imports yield : [] -> []
(func $yield (import "co2" "yield"))
```

. . .

Running example: coroutines (3)

(module \$example

. . .

```
. . .
;; odd : [i32] -> []
:: prints the first $niter odd natural numbers
(func $odd (param $niter i32)
                                                              :: next odd number
  (local $n i32)
  (local $i i32)
                                                              :: iterator
  (local.set $n (i32.const 1))
                                                              :: initialise locals
  (local.set $i (i32.const 1))
                                                              11 ...
  (block $b
   (loop $1
      (br_if $b (i32.gt_u (local.get $i) (local.get $niter))) ;; termination condition
      (call $print (local.get $n))
                                                              ;; print the current odd number
      (local.set $n (i32.add (local.get $n) (i32.const 2))) :: compute next odd number
      (local.set $i (i32.add (local.get $i) (i32.const 1))) ;; increment the iterator
      (call $vield)
                                                              :: vield control
      (br $1))))
                                                              :: repeat
;; even : [i32] -> []
:: prints the first $niter even natural numbers
```

```
(func $even (param $niter i32) ...)
```

Running example: coroutines (4)

(module \$example

```
...
;; odd5, even5 : [] -> []
(func $odd5 (export "odd5")
    (call $odd (i32.const 5)))
(func $even5 (export "even5")
    (call $even (i32.const 5)))
```

```
;; calling $run with $odd5 and $even5...
(call $run (ref.func $odd5) (ref.func $even5))
;; ... prints 1 3 5 7 9 2 4 6 8 10
```

Control tag declaration

```
(\texttt{tag }\texttt{\$tag} \ (\texttt{param} \ \sigma^*) \ (\texttt{result} \ \tau^*))
```

it's a mild extension of Wasm's exception tags

(known in the literature as an 'operation symbol' (Plotkin and Pretnar 2013))

```
(module $co2
 ;; type alias task = [] \rightarrow []
 (type $task (func))
 ;; yield : [] -> []
  (tag $yield)
 ;; yield : [] -> []
 (func $yield (export "yield")
   (nop))
 ;; run : [(ref $task) (ref $task)] -> []
 (func $run (export "run") (param $task1 (ref $task)) (param $task2 (ref $task))
   ...)
```

Continuation type

(cont \$ft)

cont is a new reference type constructor parameterised by a function type, $\$ft : [\sigma^*] \to [\tau^*]$

Continuation allocation

```
cont.new ct : [(ref null <math>ft)] \rightarrow [(ref ct)]
```

where $\$ft : [\sigma^*] \rightarrow [\tau^*]$ and \$ct : cont \$ft

Refactoring the co2 module (2)

```
(module $co2
  ;; type alias $task = [] -> []
  (type $task (func))
```

;; type alias \$ct = \$task
(type \$ct (cont \$task))

. . .

```
;; run : [(ref $task) (ref $task)] -> []
;; implements a 'seesaw' (c.f. Ganz et al. (ICFP@99))
(func $run (export "run") (param $task1 (ref $task)) (param $task2 (ref $task))
;; locals to manage continuations
(local $up (ref null $ct))
(local $down (ref null $ct))
(local $down (ref null $ct))
(local $isOtherDone i32)
;; initialise locals
(local.set $up (cont.new $ct (local.get $task1)))
(local.set $down (cont.new $ct (local.get $task2)))
...)
```



cont.new allocates a new stack segment New segments are initially suspended



 $\verb|cont.new||$ allocates a new stack segment

New segments are initially suspended



cont.new allocates a new stack segment

New segments are initially suspended

Continuation resumption

```
\begin{aligned} & \textbf{resume } \$ct \ (\textbf{tag } \$tag \ \$h)^* : [\sigma^* \ (\textbf{ref null } \$ct)] \rightarrow [\tau^*] \\ & \textbf{where} \ \{\$tag_i : [\sigma_i^*] \rightarrow [\tau_i^*] \ \text{and} \ \$h_i : [\sigma_i^* \ (\textbf{ref null } \$ct_i)] \text{ and} \\ & \$ct_i \ : \textbf{cont} \ \$ft_i \ & \textbf{and} \ \$ft_i : [\tau_i^*] \rightarrow [\tau^*] \}_i \\ & \textbf{and} \ \$ct : \textbf{cont} \ \$ft \\ & \textbf{and} \ \$ft \ : [\sigma^*] \rightarrow [\tau^*] \end{aligned}
```

The instruction fully consume the continuation argument

Refactoring the co2 module (3)

```
(module $co2
                                                              ;; declarations of $task, $yield, etc
  . . .
 ;; run : [(ref $task) (ref $task)] -> []
 (func $run (export "run") (param $task1 (ref $task)) (param $task2 (ref $task))
                                                              ;; initialisation of $up and $down
    . . .
   :: run $up
   (loop $h
                                                              :: handling loop
      (block $on_yield (result (ref $ct))
       (resume $ct (tag $yield $on_yield) (local.get $up)) ;; resume $up; handle $yield using $on_yield
       (if (i32.eq (local.get $is0therDone) (i32.const 1)) ;; $up finished; $down is already done?
         (then (return)))
                                                              :: ... then exit
       (local.get $down)
                                                              :: ... otherwise prepare to run $down
       (local.set $up)
                                                              :: $up := $down
       (local.set $isOtherDone (i32.const 1))
                                                              ;; mark other as done
       (br $h)
                                                              :: repeat
                                                              :: vield-case definition: stack: [(cont $ct)]
      (local.set $up)
                                                              :: set $up to the current continuation
      (if (i32.eqz (local.get $is0therDone))
                                                             ;; is $down already done?
       (then (local.get $down)
                                                              :: ... then swap $up and $down
              (local.set $down (local.get $up))
             (local.set $up)))
      (br $h)))
                                                              :: repeat
```



 $\ensuremath{\textit{resume}}$ transfers control from the parent to the child stack

The pointer between parent and child is inverted



 $\ensuremath{\textit{resume}}$ transfers control from the parent to the child stack

The pointer between parent and child is inverted

Continuation suspension

suspend
$$tag: [\sigma^*] \rightarrow [\tau^*]$$

where $tag : [\sigma^*] \rightarrow [\tau^*]$

Refactoring the co2 module (4)

```
(module $co2
;; type alias task = [] -> []
(type $task (func))
;; type alias ct = $task
(type $ct (cont $task))
;; yield : [] -> []
(tag $yield (param) (result))
;; yield : [] -> []
(func $yield (export "yield")
(suspend $yield))
```

```
;; run : [(ref $task) (ref $task)] -> []
;; implements a 'seesaw' (c.f. Ganz et al. (ICFP@99))
(func $run (export "run") (param $task1 (ref $task)) (param $task2 (ref $task))
... )
```

Now (call \$run (ref.func \$odd5) (ref.func \$even5)) prints 1 2 3 4 5 6 7 8 9 10

Thinking of **suspend** in terms of stacks



suspend transfers control a child to a (grand)parent The pointer between child and parent is inverted

Thinking of **suspend** in terms of stacks



suspend transfers control a child to a (grand)parent The pointer between child and parent is inverted

Current status of the proposal

What has already been done

- Formal specification
- Informal explainer documentation
- Reference implementation
- A proof-of-concept implementation in Wasmtime

What is happening now

• Fine-tune the implementation

What is going to happen next

• Gathering performance evidence

Context switching microbenchmark

	Relative speed	Binary size	Memory usage
Asyncify	-	36 kb	66.0 mb
Bespoke	0.5 imes	27 kb	15.7 mb
WasmFX	0.25 imes	24 kb	63.9 mb

Table: Performance characteristics for webserver microbenchmark

Binary file size microbenchmarks

	main-kjp.go	coroutines.go
Asyncify	597 kb	40.0 kb
WasmFX	156 kb	7.2 kb

Table: Binary size comparison for TinyGo Programs

Summary

- Effect handlers provide a modular and extensible basis for stack switching in Wasm
- Effect handlers are a proven technology
- WasmFX is a minimal and compatible extension to Wasm
- Proof-of-concept implementation in Wasmtime

The work is actively being turned into a proposal; for more details see

https://wasmfx.dev

Comments and feedback are welcome!

References

- Sitaram, Dorai (1993). "Handling Control". In: PLDI. ACM, pp. 147–155.
- Ganz, Steven E., Daniel P. Friedman, and Mitchell Wand (1999). "Trampolined Style". In: *ICFP*. ACM, pp. 18–27.
- Plotkin, Gordon D. and Matija Pretnar (2013). "Handling Algebraic Effects". In: Logical Methods in Computer Science 9.4.
- Haas, Andreas et al. (2017). "Bringing the web up to speed with WebAssembly". In: *PLDI*. ACM, pp. 185–200.
- Forster, Yannick et al. (2019). "On the expressive power of user-defined effects: Effect handlers, monadic reflection, delimited control". In: *J. Funct. Program.* 29, e15.
- Hillerström, Daniel (2021). "Foundations for Programming and Implementing Effect Handlers". PhD thesis. The University of Edinburgh, Scotland, UK.
- Sivaramakrishnan, K. C. et al. (2021). "Retrofitting effect handlers onto OCaml". In: *PLDI*. ACM, pp. 206–221.
- Ghica, Dan et al. (2022). "High-Level Type-Safe Effect Handlers in C++". In: *Proc. ACM Program. Lang.* 6.00PSLA, pp. 1–30.
- Thomson, Patrick et al. (2022). "Fusing industry and academia at GitHub (experience report)". In: *Proc. ACM Program. Lang.* 6.ICFP, pp. 496–511.

Continuation binding, cancellation, and trapping

Partial continuation application

```
cont.bind (type ct) : [\sigma_0^* (ref null ct)] \rightarrow [(ref ct')]
```

where $ct : cont ft and ft : [\sigma_0^* \sigma_1^*] \rightarrow [\tau^*]$ and $ct' : cont ft' and ft' : [\sigma_1^*] \rightarrow [\tau^*]$

Continuation cancellation

resume_throw (tag exn) (tag $h)^* : [\sigma_0^* (ref null <math>ct)] \rightarrow [\tau^*]$

where $\$exn: [\sigma_0^*] \rightarrow [], \{\$tag_i: [\sigma_i^*] \rightarrow [\tau_i^*] \text{ and } \$h_i: [\sigma_i^* (ref null \$ct_i)] \text{ and } \$ct_i : cont \$ft_i \text{ and } \$ft_i: [\tau_i^*] \rightarrow [\tau^*]\}_i$ and $\$ct: cont ([\sigma^*] \rightarrow [\tau^*]$

Control barriers

barrier bl (type bt) $instr^* : [\sigma^*] \to [\tau^*]$

where $\$bt = [\sigma^*] \rightarrow [\tau^*]$ and $\textit{instr}^*: [\sigma^*] \rightarrow [\tau^*]$