



HAL
open science

Wasocaml: compiling OCaml to WebAssembly

Léo Andrès, Pierre Chambart, Jean-Christophe Filliâtre

► To cite this version:

Léo Andrès, Pierre Chambart, Jean-Christophe Filliâtre. Wasocaml: compiling OCaml to WebAssembly. IFL 2023 - The 35th Symposium on Implementation and Application of Functional Languages, João Saraiva; João Fernandes, Aug 2023, Braga, Portugal. hal-04311345v1

HAL Id: hal-04311345

<https://inria.hal.science/hal-04311345v1>

Submitted on 28 Nov 2023 (v1), last revised 13 Dec 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Wasocaml: compiling OCaml to WebAssembly

Léo Andrès*

leo@ocamlpro.com
Université Paris-Saclay, CNRS, ENS
Paris-Saclay, Inria, Laboratoire
Méthodes Formelles
Gif-sur-Yvette, France

Pierre Chambart

pierre.chambart@ocamlpro.com
OCamlPro SAS
Paris, France

Jean-Christophe Filliâtre

jean-christophe.filliatre@cnrs.fr
Université Paris-Saclay, CNRS, ENS
Paris-Saclay, Inria, Laboratoire
Méthodes Formelles
Gif-sur-Yvette, France

ABSTRACT

The limitations of JavaScript as the default language of the web led to the development of Wasm, a secure, efficient and modular language. However, compiling garbage-collected languages to Wasm presents challenges, including the need to compile or reimplement the runtime. Some Wasm extensions such as Wasm-GC are developed by the Wasm working groups to facilitate the compilation of garbage-collected languages. We present Wasocaml, an OCaml to Wasm-GC compiler. It is the first compiler for a real-world functional programming language targeting Wasm-GC. Wasocaml confirms the adequacy of the Wasm-GC proposal for a functional language and had an impact on the design of the proposal. Moreover, the compilation strategies developed within Wasocaml are applicable to other compilers and languages. Indeed, two compilers already used a design similar to our. Finally, we describe how we plan to handle the C/JavaScript FFI and effects handlers, in order to allow developers to easily deploy programs mixing C, JavaScript and OCaml code to the web, while maintaining good performances.

CCS CONCEPTS

• Software and its engineering → Compilers.

KEYWORDS

ocaml, wasm, compilation, value representation, memory, cps, ffi, effect handler

ACM Reference Format:

Léo Andrès, Pierre Chambart, and Jean-Christophe Filliâtre. 2023. Wasocaml: compiling OCaml to WebAssembly. In *Proceedings of Symposium on Implementation and Application of Functional Languages (IFL '23)*. ACM, New York, NY, USA, 12 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 CONTEXT

JavaScript is widely thought as being the de facto language of the web. However, it does show its limitations when it comes to performance, security and safety. In order to remedy this, WebAssembly

*Also with OCamlPro SAS.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

IFL '23, August 29–August 31, 2023, Braga, Minho, Portugal

© 2023 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00
<https://doi.org/XXXXXXX.XXXXXXX>

(Wasm for short) [3] has been developed as a secure modular language of predictable performance. Its usage is expanding beyond the web: finding applications in the cloud (Fastly, Cloudflare) and in the creation of portable binaries.

The first version of Wasm was meant to compile libraries in languages such as C, C++ or Rust to expose them for consumption by a JavaScript program. It was designed with the explicit aim that future versions would cater to the specific needs of other languages and uses. This version of Wasm can be seen as simplified C. Here is an example:

```
(func $fact (param $x i32) (result i32)
  (if (i32.eq (local.get $x) (i32.const 0))
    (then (i32.const 1))
    (else
      (i32.mul
        (local.get $x)
        (call $fact
          (i32.sub (local.get $x) (i32.const 1)))))))
```

A Wasm module is executed inside what is called an *embedder* or a *host*. In the context of the web, it is a browser and the host language is JavaScript. A Wasm program only manipulates scalar values (integers and floats) but not garbage collected values coming from the host language, such as JavaScript objects. This would be useful in a browser context, but requires some interaction with the GC of the embedder. This led to various extensions that we describe in the next section.

1.1 Wasm Extensions

We discuss only the most notable ones, that is to say the tail call [19], reference types [20], typed function references [21], GC [18], and exception handling [17] extensions.

1.1.1 References and GC. The aim of these proposals is to allow the program to manipulate references to different kinds of values such as opaque objects, functions, and garbage-collected values. Since Wasm has to be memory safe, those values are not exposed as raw pointers. Instead, the type system guarantees the proper manipulation of such values.

```
(type $f1 (func (param i32) (result i32)))
(type $t1 (array i31ref))
(type $t2 (struct
  ((field $a (ref $t1))
   (field $b (ref $f1))
   (field $c i32))))

(func (param $x (ref $t2)) (result i32)
  (i31.get_u
```

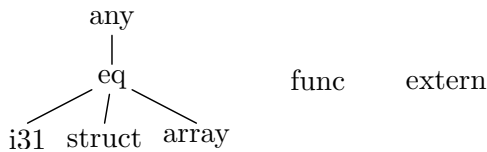
```

117 (array.get $t1
118   (i32.const 0)
119   (struct.get $t2 $a (local.get $x))))

```

However, Wasm is rather a compilation target than a programming language. It needs to be able to represent the values of any kind of source language. It was not deemed possible to have a type system powerful enough to do that. Instead, the decision was made to have a simple type system, but to rely on dynamic cast to fill the gaps, and guarantee that those casts are efficient.

There is a sub-typing relation that controls which casts are allowed. Some types are predefined by Wasm: **anyref**, **eqref**, **i31ref**, **funcref** and **funcref**. Others are user defined: array and struct. There is a structural sub-typing relation between user types. Upcasts are implicit, downcasts are explicit and incorrect ones lead to runtime errors. It is possible to dynamically test for type compatibility.



Contribution to the proposals. The general rule for the Wasm committee is to only include features with a demonstrated use case. As there are currently very few compilers targeting the GC-proposal, some features were lacking conclusive evidence of their usefulness. An example is the **i31ref** type that is not required by the Dart compiler (the only one targeting the GC-proposal at the time). Wasocaml demonstrates the usefulness of **i31ref**. It also validates the GC-proposal on a functional language. We presented Wasocaml to the Wasm-GC working group [2]. It helped in convincing the working group to keep **i31ref** in the proposal.

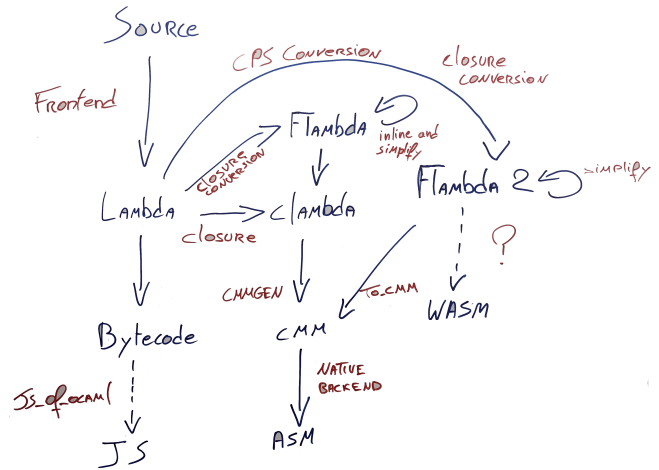
1.1.2 Exception. Another serious limitation of Wasm1 is the absence of an exception mechanism. This is not a problem for compiling either C or Rust, but C++ does require them. It is of course possible to encode exception, but this has a significant runtime cost and limits the interactions with JavaScript (that has exceptions).

1.1.3 Tail Calls. Wasm1 does not allow tail-calls to be optimized. This is a show-stopper when compiling functional languages. In the tail-call proposal, new instructions such as `return call` guarantee that a tail-call will be optimized.

2 SOURCE LANGUAGE

2.1 OCaml Intermediate Representation Choice

OCaml has many intermediate representations (IR for short). In the *Lambda* IR, closures are still implicit and the code has not been optimized. The *Bytecode* representation is not well optimized. The *Clambda* one is too low-level for Wasm because code pointers and values are mixed inside closures. *Cmm* is even more low-level: it has pointer arithmetic. The *Flambda* IR is simpler than *Flambda2*, thus it is the one we choosed for our first prototype. We will use *Flambda2* in the future, as it provides better optimisations.



2.2 Native OCaml Value Representation

The various OCaml IRs are working on a uniform representation of values, known as pointer tagging. One bit is used to indicate if the value is a small scalar or a pointer to a heap-allocated block. Granted we are working on n bits values, we have to choose one bit that will be the *tag bit*. For instance, if we choose the least significant bit, we start by testing its value:

$$b_{n-1} | b_{n-2} | \dots | b_1 | b_0$$

If $b_0 = 0$, then the whole value is a pointer:

$$b_{n-1} | b_{n-2} | \dots | b_1 | 0 \quad b_{n-1} | b_{n-2} | \dots | b_1 | 0$$

If $b_1 = 1$, then the $n - 1$ most significant bits are a small scalar and b_0 is ignored:

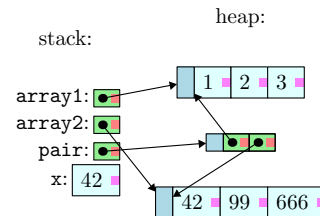
$$b_{n-1} | b_{n-2} | \dots | b_1 | 1 \quad b_{n-1} | b_{n-2} | \dots | b_1 | 1$$

In the second case, we talk about *small scalars* instead of scalars because we can only represents 2^{n-1} values instead of the 2^n that are representable when all bits are available. For pointers, we do not lose anything as they need to be *aligned* anyway and the least significant bit is always zero.

```

let array1 = [| 1; 2; 3 |]
let array2 = [| 42; 99; 666 |]
let pair = (array1, array2)
let x = 42

```



2.3 Flambda semantics

We give a formal semantics for a subset of *Flambda*. Compared to the full *Flambda* it has the following limitations: the objects related primitives have been removed, functions only have one argument,

233 exceptions have been removed but we still have static exceptions
234 that are described later.

235 2.3.1 Abstract syntax.

237 *Value.* A value v is either an integer n or an address a .

$$240 \quad v ::= n \quad (\text{integer})$$

$$241 \quad \quad \quad | a \quad (\text{address})$$

243 An integer represents a value of many types in the OCaml source
244 language such as an **int**, a **bool**, a **char** or the constant construc-
245 tors of a sum type. An address points to a heap-allocated values.

247 *Store.* A store \mathcal{M} is a map from addresses to heap-allocated val-
248 ues. It is used to represent the state of the memory. Heap-allocated
249 values can either be a closure $\mathcal{S}.x$, a set of closures \mathcal{S} or a block
250 made of a tag n and a list of values v^* . A closure is made of a set of
251 closures \mathcal{S} and of a name x which is its identifier inside the set.

$$253 \quad \mathcal{M} ::= a \mapsto n, v^* \quad (\text{block})$$

$$254 \quad \quad \quad | \mathcal{S}.x \quad (\text{closure})$$

$$255 \quad \quad \quad | \mathcal{S} \quad (\text{set of closure})$$

257 *Set of closure.* A set of closures \mathcal{S} represents a set of mutually
258 recursive functions.

$$261 \quad \mathcal{S} ::= \mathcal{V}, \mathcal{C}$$

263 It is made of two maps \mathcal{V} and \mathcal{C} . The map \mathcal{V} is the environment
264 (the values captured by the closures). An environment is a map
265 from names to values.

$$267 \quad \mathcal{V} ::= x \mapsto v$$

269 The map \mathcal{C} is the closure map (the code of each function). It is a
270 map from names to codes.

$$272 \quad \mathcal{C} ::= x \mapsto \lambda x x . t$$

274 The code is a function that takes two arguments and produces a
275 term. The first argument is the real argument of the function. The
276 second one is an address a to a closure $\mathcal{S}.x$. It is used to access the
277 environment or to call other functions of the set \mathcal{S} .

279 *Binary operator.*

$$281 \quad op ::= eq \quad (\text{equality})$$

$$282 \quad \quad \quad | add \quad (\text{addition})$$

$$283 \quad \quad \quad | sub \quad (\text{subtraction})$$

287 *Branches.*

$$288 \quad \mathcal{B} ::= n \mapsto t$$

Term.

$$t ::= v \quad (\text{value})$$

$$| x \quad (\text{identifier})$$

$$| \text{let } x = t \text{ in } t \quad (\text{let-binding})$$

$$| x x \quad (\text{application})$$

$$| x <- x \quad (\text{mutation})$$

$$| \text{if } x \text{ then } t \text{ else } t \quad (\text{conditional})$$

$$| \text{switch } x \mathcal{B} \mathcal{B} \quad (\text{switch})$$

$$| \text{sraise } x x^* \quad (\text{static raise})$$

$$| \text{scatch } t \text{ with } x x^* t \quad (\text{static catch})$$

$$| \text{while}_t t \text{ do } t \quad (\text{while loop})$$

$$| \text{get_field } n x \quad (\text{block field read})$$

$$| \text{set_field } n x x \quad (\text{block field write})$$

$$| \text{make_block } n x^* \quad (\text{block creation})$$

$$| \text{project_closure } x x \quad (\text{closure projection})$$

$$| \text{project_var } x x \quad (\text{closure environment})$$

$$| \text{move_within_closure } x x \quad (\text{closure movement})$$

$$| \text{make_set_of_closures } \mathcal{S} \quad (\text{closures allocation})$$

$$| \text{pop } t \quad (\text{pop environment})$$

$$| x \text{ op } x \quad (\text{binary operator})$$

232 2.3.2 Small-step semantics.

Reduction context.

$$E ::= \square$$

$$| \text{let } x = E \text{ in } t$$

$$| \text{while}_t E \text{ do } t$$

$$| \text{scatch } E \text{ with } x x^* t$$

$$| \text{pop } E$$

Reduction inside a context.

$$\frac{\mathcal{V}^*, \mathcal{M}, t_1 \rightarrow \mathcal{V}'^*, \mathcal{M}', t_2}{\mathcal{V}^*, \mathcal{M}, E(t_1) \rightarrow \mathcal{V}'^*, \mathcal{M}', E(t_2)}$$

Head reductions.

$$\frac{\mathcal{V}_0(x) = v}{\mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, x \rightarrow \mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, v}$$

An identifier simply reduces to its image in the current environ-
ment.

$$\text{let-binding} \frac{}{\mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, \text{let } x = v \text{ in } t \rightarrow \mathcal{V}_0[x \leftarrow v] \mathcal{V}^*, \mathcal{M}, t}$$

A let-binding $\text{let } x = v \text{ in } t$ reduces to t and adds a binding from
 x to v in the current environment.

$$\begin{array}{c} 349 \\ 350 \\ 351 \\ 352 \\ 353 \\ 354 \\ 355 \end{array} \quad \begin{array}{c} \mathcal{V}_0(x_1) = a \quad \mathcal{M}(a) = \mathcal{S}.x \\ \mathcal{S} = _ \mathcal{C} \quad \mathcal{C}(x) = \lambda x_3 x_4 . t \\ \mathcal{V}_0(x_2) = v \end{array} \quad \begin{array}{c} \text{application} \\ \hline \mathcal{V}_0 \mathcal{V}^*, \quad \mathcal{M}, \quad x_1 x_2 \\ \rightarrow [x_3 \leftarrow v; x_4 \leftarrow a] \mathcal{V}_0 \mathcal{V}^*, \quad \mathcal{M}, \quad \text{pop } t \end{array}$$

An application $x_1 x_2$ looks for an address a in the current environment. The image of a in the store must be a closure $\mathcal{S}.x$ where the set of closures $\mathcal{S} = _ \mathcal{C}$. Let the value of x in the closure map \mathcal{C} be $\lambda x_3 x_4 . t$. The term reduces to $\text{pop } t$ and we create a new environment containing two bindings: x_3 is mapped to the value of x_2 in the current environment and x_4 is mapped to a .

$$\begin{array}{c} 363 \\ 364 \\ 365 \end{array} \quad \begin{array}{c} \mathcal{V}_0(x_1) = v_1 \quad \mathcal{V}_0(x_2) = v_2 \\ \text{mutate} \frac{\mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, x_1 \leftarrow x_2 \rightarrow \mathcal{V}_0[x_1 \leftarrow v_2] \mathcal{V}^*, \mathcal{M}, 0}{} \end{array}$$

A mutation $x_1 \leftarrow x_2$ simply updates the value of x_1 to be the value of x_2 in the current environment. It reduces to 0 which corresponds to the value $()$: **unit** in OCaml.

$$\begin{array}{c} 370 \\ 371 \\ 372 \\ 373 \end{array} \quad \begin{array}{c} \mathcal{V}_0(x) = n \quad n \neq 0 \\ \text{conditional 1} \frac{\mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, \text{if } x \text{ then } t_1 \text{ else } t_2}{\rightarrow \mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, t_1} \end{array}$$

A conditional $\text{if } x \text{ then } t_1 \text{ else } t_2$ reduces to t_1 when the value of x in the current environment is not 0 (i.e. when it is **true** : **bool** in OCaml).

$$\begin{array}{c} 378 \\ 379 \\ 380 \\ 381 \end{array} \quad \begin{array}{c} \mathcal{V}_0(x) = n \quad n = 0 \\ \text{conditional 2} \frac{\mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, \text{if } x \text{ then } t_1 \text{ else } t_2}{\rightarrow \mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, t_2} \end{array}$$

Similarly, it reduces to t_2 when the value of x is 0 (i.e. when it is **false** : **bool** in OCaml).

$$\begin{array}{c} 385 \\ 386 \\ 387 \end{array} \quad \begin{array}{c} \mathcal{V}_0(x) = n \quad \mathcal{B}_1(n) = t \\ \text{switch 1} \frac{\mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, \text{switch } x \mathcal{B}_1 \mathcal{B}_2 \rightarrow \mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, t}{} \end{array}$$

When the value of x in the current environment is an integer n , then $\text{switch } x \mathcal{B}_1 \mathcal{B}_2$ reduces to the image of n in \mathcal{B}_1 .

$$\begin{array}{c} 391 \\ 392 \\ 393 \end{array} \quad \begin{array}{c} \mathcal{V}_0(x) = a \quad \mathcal{M}(a) = n, v^* \quad \mathcal{B}_2(n) = t \\ \text{switch 2} \frac{\mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, \text{switch } x \mathcal{B}_1 \mathcal{B}_2 \rightarrow \mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, t}{} \end{array}$$

Similarly, when x is an address a and $\mathcal{M}(a)$ is a block with tag n , then $\text{switch } x \mathcal{B}_1 \mathcal{B}_2$ reduces to the image of n in \mathcal{B}_2 .

$$\begin{array}{c} 397 \\ 398 \\ 399 \\ 400 \\ 401 \\ 402 \end{array} \quad \begin{array}{c} x_{\text{exn}} \neq x'_{\text{exn}} \\ \text{raise 1} \frac{\mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, \text{scatch}(\text{sraise } x_{\text{exn}} x^n) \text{ with } x'_{\text{exn}} x'^n t}{\rightarrow \mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, \text{sraise } x_{\text{exn}} x^n} \end{array}$$

If a raised exception does not match the exception expected by a static catch, then the catch is discarded and we re-raise the exception.

$$\begin{array}{c} 407 \\ 408 \\ 409 \\ 410 \\ 411 \\ 412 \\ 413 \\ 414 \\ 415 \\ 416 \end{array} \quad \begin{array}{c} x_{\text{exn}} = x'_{\text{exn}} \quad x^n = x_0, \dots, x_{n-1} \\ x'^n = x'_0, \dots, x'_{n-1} \quad \forall i \in [0; n], \mathcal{V}'_0(x_i) = v_i \\ \text{raise 2} \frac{\mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, \text{scatch}(\text{sraise } x_{\text{exn}} x^n) \text{ with } x'_{\text{exn}} x'^n t}{\rightarrow \mathcal{V}'_0[x'_0 \leftarrow v_0, \dots, x'_{n-1} \leftarrow v_{n-1}] \mathcal{V}^*, \mathcal{M}, t} \end{array}$$

On the contrary, when the two exceptions match, the identifiers specified in the catch are binded to the values of the identifiers transported by the exception. The values are taken in the current environment. We have the guarantee that if the transported identifiers were in the environment when we raised, then they are also in the environment when we catch. This is because static exceptions never cross a function application. Finally, the reduced term is the one attached to the catch.

$$\begin{array}{c} 427 \\ 428 \\ 429 \\ 430 \\ 431 \end{array} \quad \begin{array}{c} \mathcal{V}^*, \mathcal{M}, \\ \text{raise 3} \frac{\text{while}_{t'}(\text{sraise } x_{\text{exn}} x^*) \text{ do } t}{\rightarrow \mathcal{V}^*, \mathcal{M}, \text{sraise } x_{\text{exn}} x^*} \end{array}$$

A while simply re-raises static exceptions.

$$\begin{array}{c} 437 \\ 438 \\ 439 \end{array} \quad \begin{array}{c} \mathcal{V}^*, \mathcal{M}, \text{let } x = (\text{sraise } x_{\text{exn}} x^*) \text{ in } t \\ \text{raise 4} \frac{}{\rightarrow \mathcal{V}^*, \mathcal{M}, \text{sraise } x_{\text{exn}} x^*} \end{array}$$

Similarly, a let-binding also re-raises static exceptions.

$$\begin{array}{c} 442 \\ 443 \\ 444 \end{array} \quad \begin{array}{c} n = 0 \\ \text{while loop 1} \frac{}{\mathcal{V}^*, \mathcal{M}, \text{while}_{t'} n \text{ do } t \rightarrow \mathcal{V}^*, \mathcal{M}, 0} \end{array}$$

A loop $\text{while}_{t'} n \text{ do } t$ reduces to 0 when $n = 0$.

$$\begin{array}{c} 447 \\ 448 \\ 449 \\ 450 \end{array} \quad \begin{array}{c} n \neq 0 \\ \text{while loop 2} \frac{}{\mathcal{V}^*, \mathcal{M}, \text{while}_{t'} n \text{ do } t \rightarrow \mathcal{V}^*, \mathcal{M}, \text{let } _ = t \text{ in while}_{t'} t' \text{ do } t} \end{array}$$

On the contrary, when $n \neq 0$, it reduces to $\text{let } _ = t \text{ in while}_{t'} t' \text{ do } t$. The term t' corresponds to the original term that was used as a condition.

$$\begin{array}{c} 456 \\ 457 \\ 458 \\ 459 \end{array} \quad \begin{array}{c} \mathcal{V}_0(x) = a \quad \mathcal{M}(a) = \text{tag}, v^* \\ v^* = v_0, \dots, v_{m-1} \quad m > n \geq 0 \\ \text{field read} \frac{\mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, \text{get_field } n x \rightarrow \mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, v_n}{} \end{array}$$

The term $\text{get_field } n x$ reduces to the n^{th} value of the block stored at address a with a being the value of x in the current environment.

$$\text{field write} \frac{\mathcal{V}_0(x_1) = a \quad \mathcal{M}(a) = \text{tag}, v^* \quad v^* = v_0, \dots, v_{m-1} \quad m > n \geq 0 \quad \mathcal{V}_0(x_2) = v}{\begin{array}{l} \mathcal{V}_0 \mathcal{V}^*, \\ \mathcal{M}, \\ \text{set_field } n \ x_1 \ x_2 \\ \rightarrow \mathcal{V}_0 \mathcal{V}^*, \\ \mathcal{M} [a \leftarrow \text{tag}, v_0, \dots, v_{n-1}, v, v_{n+1}, \dots, v_{m-1}], \\ 0 \end{array}}$$

If x_1 is the address a in the current environment, then the term `set_field n x1 x2` sets the n^{th} value of the block stored at address a to the value of x_2 in the current environment.

$$\text{block creation} \frac{\mathcal{V}_0(x_0) = v_0, \dots, \mathcal{V}_0(x_{n-1}) = v_{n-1} \quad a \notin \text{dom}(\mathcal{M})}{\begin{array}{l} \mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, \text{make_block } n_{\text{tag}} \ x_0, \dots, x_{n-1} \\ \rightarrow \mathcal{V}_0 \mathcal{V}^*, \mathcal{M} [a \leftarrow n_{\text{tag}}, v_0, \dots, v_{n-1}], 0 \end{array}}$$

This creates a new block in the store at a fresh address. The tag is given as a value while the others values of the block are read from the current environment.

$$\text{closure projection} \frac{\mathcal{V}_0(x_2) = a_1 \quad \mathcal{M}(a_1) = \mathcal{S} \quad \mathcal{M}(a_2) = \mathcal{S}.x_1}{\begin{array}{l} \mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, \text{project_closure } x_1 \ x_2 \\ \rightarrow \mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, a_2 \end{array}}$$

This reduces to the address a_2 of the closure which has the identifier x_1 in the set of closures at address a_1 , with a_1 being the value of x_2 in the current environment. In short, it allows to get a particular function from a previously allocated set of closures.

$$\text{variable projection} \frac{\mathcal{V}_0(x_2) = a \quad \mathcal{M}(a) = \mathcal{S}.x \quad \mathcal{S} = \mathcal{V}, _ \quad \mathcal{V}(x_1) = v}{\mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, \text{project_var } x_1 \ x_2 \rightarrow \mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, v}$$

This reduces to the value of x_1 in the environment of the set of closures that contains the closure at address a_a , with a_a being the value of x_2 in the current environment. In short, when inside a closure, it allows to read a variable from the set of closures the closure belongs to.

$$\text{closure movement} \frac{\mathcal{V}_0(x_2) = a_1 \quad \mathcal{M}(a_1) = \mathcal{S}.x \quad \mathcal{M}(a_2) = \mathcal{S}.x_1}{\begin{array}{l} \mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, \text{move_within_closure } x_1 \ x_2 \\ \rightarrow \mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, a_2 \end{array}}$$

This reduces to the address a_2 of the closure x_1 in the set of closures to which the closure at address a_1 belongs to. In short, when inside a closure, it allows to get another closure in the current set of closures.

$$\text{pop environment} \frac{}{\mathcal{V}_0 \mathcal{V}^*, \mathcal{M}, \text{pop } v \rightarrow \mathcal{V}^*, \mathcal{M}, v}$$

This simply throws away the current environment. This is used to model the call stack. A pop is inserted around each function application and removed once the function is fully evaluated to a value.

$$\text{set alloc} \frac{\mathcal{S} = _, \mathcal{C} \quad x_0 \in \text{dom}(\mathcal{C}), \dots, x_{n-1} \in \text{dom}(\mathcal{C}) \quad \forall a \in [a_0, \dots, a_n], a \notin \text{dom}(\mathcal{M})}{\begin{array}{l} \mathcal{V}^*, \mathcal{M}, \text{make_set_of_closures } \mathcal{S} \\ \rightarrow \mathcal{V}^*, \\ \mathcal{M} [a_n \leftarrow \mathcal{S}; a_0 \leftarrow \mathcal{S}.x_0; \dots; a_{n-1} \leftarrow \mathcal{S}.x_{n-1}], \\ a_n \end{array}}$$

This maps the set of closures \mathcal{S} and all its closures to fresh addresses in the store. The term reduces to the address of the set.

2.4 Example

The following OCaml code:

```
let f = print_int
let rec iter f l =
  match l with
  | [] -> ()
  | hd :: tl ->
    f hd;
    iter f tl
let () =
  let iter_print = iter f in
  iter_print [2; 1]
```

Would corresponds to the following Flambda code:

```
let f =
  let s =
    make_closures
    | cl_f { x } env -> print_int x
  with vars
  end
  in
  project_closure cl_f from s
in

let iter =
  let set =
    make_closures
    | cl_iter { f } env ->
      let otherset =
        make_closures
        | cl_iter_f { l } envtwo ->
          switch l
          with int
          | 0 -> const 0
          with tag
          | 0 ->
            let x = get_field 0 l in
            let f = project_var f from envtwo in
            let dummy = f x in
            let tl = get_field 1 l in
            envtwo tl
          end
        end
      with vars
      | f -> f
    end
  in
```



```

581     project_closure cl_iter_f from otherset
582     with vars
583     end
584   in
585   project_closure cl_iter from set
586   in
587
588   let lempy = const 0 in
589   let one = const 1 in
590   let lone = make_block 0 one lempy in
591   let two = const 2 in
592   let ltwo = make_block 0 two lone in
593   let iter_print = iter f in
594   iter_print ltwo
595

```

3 COMPILING FLAMBDA TO WASM

There have been attempts at targeting the early versions of Wasm from OCaml. The first one, *OCamlrun WebAssembly* [7] simply compiles the OCaml runtime (including the GC) and interpreter using emscripten [13] (a C to Wasm compiler). The second one, WASICaml [11] is a smarter and more efficient version of the same approach.

These approaches suffer from the same problem as C programs running on Wasm: they cannot natively manipulate external values, such as DOM objects in the browser. Indeed, in the first versions of Wasm, the only types are scalars (**i32**, **i64**, **f32** and **f64**). There is no way to directly manipulate values from the embedder (e.g. JavaScript objects). It could be possible to identify objects with an integer and map them to their corresponding objects on the embedder side. This approach comes with the usual limitations of having two runtimes manipulating (indirectly) garbage-collected values: it is tedious and easily leads to memory leaks (cycles between the two runtimes cannot be collected).

To properly interact with the embedder, we need to leverage the reference extension. This extension adds new types to the language, e.g. **externref** which is an opaque type representing a value from the embedder. References cannot be stored in the linear memory of Wasm thus they cannot appear inside OCaml values when using the previously described compilation scheme.

In order to use references, we require a completely different compilation strategy. We do not use the linear memory. Our strategy is close to the native OCaml one, which we describe now.

We use the Flambda IR of the OCaml compiler as input for the Wasm generation. This is a step of the compilation chain where most of the high-level OCaml-specific optimisations are already applied. Also in this IR, the closure conversion pass is already performed. Most of the constructions of this IR maps quite directly to Wasm ones.

Wasocaml is a compiler for the full Flambda IR. We describe the full compilation scheme but for the sake of simplicity we only formalize a subset of Flambda.

3.1 OCaml Value Representation in Wasm

As in native OCaml, we use a uniform representation. We cannot use integers, as they cannot be used as pointers. We need to use a reference type. The most general one is **anyref**. We can be more

precise and use **eqref**. This is the type of all values that can be tested for physical equality. It is a super type of OCaml values. This also allows us to test for equality without requiring an additional cast.

Small scalars. All OCaml values that are represented as integers in the native OCaml are represented as **i31ref**. This includes the **int** and **char** types, but also all constant constructors of sum types.

Arrays. The OCaml array type is directly represented as a Wasm array.

Blocks. For other kinds of blocks, there are two possible choices: we can either use a struct with a field for the tag and one field per OCaml value field, or use arrays of **eqref** with the tag stored at position 0.

3.1.1 Blocks as Structs. A block of size one is represented using the type **\$block1** while for size two it is **\$block2**:

```

(type $block1 (struct
  (field $tag i8)
  (field $field0 eqref)))

```

```

(type $block2 (sub $block1) (struct
  (field $tag i8)
  (field $field0 eqref)
  (field $field1 eqref)))

```

The type **\$block2** is declared as a subtype of **\$block1** because in the OCaml IRs, the primitive for accessing a block field is untyped: when accessing the n -th field of the block the only available information is that the block has size at least $n + 1$. It could be possible to propagate size metadata in the IRs but that wouldn't be sufficient because of untyped primitives such as **Obj**.field that we need to support in order to be compatible with existing code.

```

(func $snd (param $x eqref) (result eqref)
  (struct.get $block2 1
    (ref.cast $block2 (local.get $x))))

```

3.1.2 Blocks as Arrays. We represent blocks with an array of **eqref**:

```

(type $block (array eqref))

```

The tag is stored in the cell at index 0 of the array. Reading its value is implemented by getting the cell and casting it to an integer:

```

(func $tag (param $x eqref) (result i32)
  (i31.get_u (ref.cast i31ref
    (array.get $block
      (i32.const 0)
      (ref.cast $block (local.get $x))))))

```

Thus accessing the field n of the OCaml block amounts to accessing the field $n + 1$ of the array:

```

(func $snd (param $x eqref) (result eqref)
  (array.get $block
    (i32.const 2)
    (ref.cast $block (local.get $x))))

```

3.1.3 Block Representation Tradeoffs. The array representation is simpler but requires (implicit) bound checks at each field access and a cast to read the tag.

On the other hand, the struct representation requires a more complex cast for the Wasm runtime (a subtyping test). A compiler propagating more types could use finer Wasm type information, providing a precise type for each struct field. This would allow fewer casts.

The V8 runtime allows to consider casts as no-ops (only for test purposes). The speedup we measured is around 10%. That gives us an upper bound on the actual runtime cost of casts.

OCaml blocks can be arbitrarily large, and very large ones do occur in practice. In particular, modules are compiled as blocks and tend to be on the larger side. We have seen in the wild some examples containing thousands of fields. Wasm only allows a subtyping chain of length 64, which is far from sufficient for the struct encoding.

In the formalization we use the *blocks as arrays* representation.

3.1.4 Boxed Numbers. Raw Wasm scalars are not subtypes of `eqref` thus they cannot be used directly to represent OCaml boxed numbers. We need to box them inside a struct in order to make them compatible with our representation:

```
(type $boxed_float (struct (field $v f64)))
(type $boxed_int64 (struct (field $v i64)))
```

3.1.5 Closures. Wasm `funcref` are functions, not closures, hence we need to produce values containing both the function and its environment. The only Wasm type construction that can contain both `funcref` and other values are structs. Thus, a closure is a struct containing a `funcref` and the captured variables. As an example, here is the type of a closure with two captured variables:

```
(type $closure1 (struct
  (field funcref)
  (field $v1 eqref)
  (field $v2 eqref)))
```

In order to reduce casts and to handle mutually recursive functions, the actual representation is a bit more complex. This is the only place where we need recursive Wasm types.

3.2 Control flow

Control flow and continuations have a direct equivalent with Wasm `block`, `loop`, `br_table`, and `if` instructions. Low level OCaml primitives to handle exceptions are quite similar to Wasm ones. In OCaml, it is possible to generate new exceptions at runtime by using *e.g.* the `let exception` syntax or functors and first-class modules. This is not possible in the Wasm exception proposal. Thus, we use the same Wasm exception everywhere and manage the rest on the side by ourselves, using an identifier to discriminate between different exceptions.

3.3 Currification

The main difference revolves around functions. In OCaml, functions take only one argument. However, in practice, functions look like they have more than one. Without any special management this would mean that most of the code would be functions producing closures that would be immediately applied. To handle that, internally, OCaml does have functions taking multiple arguments, with some special handling for times when they are partially applied. This information is explicit at the Flambda level. In the native

OCaml compiler, the transformation handling that occurs in a later step called `cmmgen`. Hence, we have to duplicate this in Wasocaml. Compiling this requires some kind of structural subtyping on closures such that, closures for functions of arity one are supertypes of all the other closures. Thankfully there are easy encodings for that in Wasm.

3.4 Stack Representation

Most of the remaining work revolves around translating from let bound expressions to a stack based language without producing overly naive code. Also, we do not need to care too much about low level Wasm specific optimisations as we rely on Binaryen [16] (a quite efficient Wasm to Wasm optimizer) for those.

3.5 Unboxing

The main optimisation available in native OCaml that we are missing is number unboxing. As OCaml values have a uniform representation, only small scalars can fit in a true OCaml value. This means that the types `nativeint`, `int32`, `int64` and `float` have to be boxed. In numerical code, lots of intermediate values of type float are created, and in that case, the allocation time to box numbers completely dominates the actual computation time. To limit that problem, there is an optimisation called unboxing performed during the `cmmgen` pass that tends to eliminate most of the useless allocations. As this pass is performed after Flambda, and was not required to produce a complete working compiler, this was left for future work. Note that the end plan is to use the next version of Flambda, which does a much better job at unboxing. For the time being, we can expect Binaryen to perform local unboxing in some cases.

3.6 Expected Performances

We are able to run the classical functional microbenchmarks using the experimental branch of V8 and the results are quite convincing. While we do not expect to be competitive with the native code compiler, the performance degradation seems to be almost constant (around twice slower).

We are currently working on producing benchmarks on real sized programs, it is not easy as Wasocaml is still a prototype and is not yet integrated into existing build systems. Nonetheless, we are able to compile an OCaml implementation of the Knuth-Bendix algorithm [5]. On V8 the Wasm exception handling proposal is really slow compared to native OCaml: a raise is around a hundred times slower. In SpiderMonkey (the Firefox engine) they are much faster but other proposals are missing and we can't run code we generate. We have to discuss with the V8 team if this is expected and if this can be improved.

Compared to a JavaScript VM, a Wasm compiler is a much simpler beast that can compile ahead of time. For this reason, various Wasm engines are expected to behave quite similarly. They do not show any of the wild unpredictability that browsers tend to demonstrate with JavaScript. Indeed, compiling OCaml to JS using `jsoo` leads to results that are usually also twice as slow as native code in the best cases, but can sometimes be much slower in an unpredictable fashion.

Currently there is no other Wasm runtime supporting all the extensions we require. SpiderMonkey does not have tail-call. The reference interpreter implementation of the various extensions are split in separate repository and merging them requires some work. Nevertheless, we expect performances of Wasm-GC to vary across implementations in a not too different way than they do for C-compiled Wasm programs. Although the implementation choices space is larger when it comes to a full GC than when implementing what is needed in Wasm1 (e.g. register allocation).

4 FORMALIZED COMPILATION SCHEME

When compiling a program, we start with the following template:

```
(module $wasocaml_generated_modul
  (type $mlfun (sub final
    (func (param eqref) (param eqref) (result eqref))))
  (type $data (sub final
    (array (mut (ref null eq))))))
  (type $block (sub final (struct
    (field i32)
    (field (ref null $data)))))
  (type $vars (sub final (array (mut eqref))))
  (rec
    (type $set_of_closures (sub final (struct
      (field (mut (ref null $closures)))
      (field (ref null $vars)))))
    (type $closures (sub final
      (array (mut (ref $closure)))))
    (type $closure (sub final (struct
      (field (ref $mlfun))
      (field (mut (ref null $set_of_closures)))))))
  (global $tmp_set_of_closures
    (mut (ref null $set_of_closures))
    ref.null $set_of_closures)
  (global $tmp_switch_value
    (mut i32) i32.const 0)
  (elem declare (ref $mlfun)
    ;; ...
  )
  (func $start
    ;; ...
    drop
  )
  (start $start)
)
```

To compile a term we define a function \mathcal{T} from Flambda terms to Wasm expressions. We define it case by case. Here we do not only use the S-expression notation for Wasm but also use the stack notation which makes these rules clearer.

$$\mathcal{T}(n) = \text{i32.const } n$$

$$\text{i31.new}$$

An integer literal is turned into an **i31ref** in order to match with the uniform representation.

$$\mathcal{T}(x) = \text{local.get } \$x$$

To access the value of an identifier, we simply access the value stored in the local variable of the current function.

$$\mathcal{T}(\text{let } x = t_1 \text{ in } t_2) = \mathcal{T}(t_1)$$

$$\text{local.set } \$x$$

$$\mathcal{T}(t_2)$$

A let-binding is compiled by evaluating t_1 , storing its result in the local x and evaluating t_2 .

$$\mathcal{T}(x_1 \ x_2) = \text{local.get } \$x_1$$

$$\text{ref.cast (ref $closure)}$$

$$\text{local.get } \$x_2$$

$$\text{ref.as_non_null}$$

$$\text{local.get } \$x_1$$

$$\text{ref.cast (ref $closure)}$$

$$\text{struct.get } \$closure \ 0$$

$$\text{call_ref } \$mlfun$$

To apply a function t_1 to a value t_2 we first put the two parameters of the function on the stack. The first one is the closure itself, which is needed to access the environment and other functions in the set of closures. The second one can be any value of type **eqref**. Then we need to put the function of the stack. This is done by accessing the field 0 of the closure, which is a **funcref**. Then we use **call_ref** to call the function with its two arguments. The type of the function reference is always **\$mlfun**.

$$\mathcal{T}(x_1 \leftarrow x_2) = \text{local.get } \$x_2$$

$$\text{local.set } \$x_1$$

$$\text{i32.const } 0$$

$$\text{i31.new}$$

A mutation simply puts the content of x_2 into x_1 and leaves 0 on the stack (this is **()** : **unit** in OCaml).

```

929
930
931  $\mathcal{T}(\text{if } x_1 \text{ then } t_1 \text{ else } t_2) = \text{local.get } \$x_1$ 
932
933     ref.cast i31ref
934     i31.get_s
935     (if (result (ref null eq))
936         (then  $\mathcal{T}(t_1)$ )
937         (else  $\mathcal{T}(t_2)$ ))
938     ref.as_non_null
939
940
941

```

A conditional converts the value of x_1 to an `i32` and then uses a Wasm `if`. The two branches are simply the result of compiling t_1 and t_2 .

$\mathcal{T}(\text{switch } x \ B_1 \ B_2) = \text{TODO}$

```

942
943
944  $\mathcal{T}(\text{sraise } x_{exn} \ x^n) = \text{local.get } \$x_{n-1}$ 
945
946     ...
947     local.get $x_0
948     br $exn_xexn
949
950
951
952
953
954
955
956
957
958
959

```

A static raise puts all the value transported by the exception on the stack and branches to the block corresponding to the raise.

```

960
961
962
963
964  $\mathcal{T}(\text{scatch } t_1 \text{ with } x_{exn} \ x^n \ t_2) =$ 
965     (block $after_try_catch) (result (ref null eq))
966     (block $exn_xexn
967         (result (ref null eq)) ;; 0
968         (result (ref null eq)) ;; ...
969         (result (ref null eq)) ;; n - 1
970          $\mathcal{T}(t_1)$ 
971         br $after_try_catch)
972     local.set $x_0
973     ...
974     local.set $x_{n-1}
975      $\mathcal{T}(t_2)$ 
976
977
978
979
980

```

A catch has two nested blocks. The first one is used in case no exception is raised and directly returns the value of t_1 . The second one corresponds to the case where an exception is raised. It needs to have many results corresponding to all the values transported by the exception. It will put all these values in the right locals.

```

981
982
983  $\mathcal{T}(\text{while}_{t_1} \ t_1 \ \text{do } t_2) = (\text{block } \$\text{exit\_while\_loop}$ 
984     (loop $while_loop
985          $\mathcal{T}(t_1)$ 
986         i31.get_s
987         i32.eqz
988         br_if $exit_while_loop
989          $\mathcal{T}(t_2)$ 
990         drop
991         br $while_loop ))
992
993     i32.const 0
994     i31.new
995
996
997
998
999
1000
1001
1002

```

A while loop maps quite directly to a Wasm `loop`. We also need an other block in order to exit the loop when the condition is false.

```

1003
1004
1005  $\mathcal{T}(\text{get\_field } n \ x) = \text{local.get } \$x$ 
1006     ref.cast (ref $block)
1007     struct.get $block 1
1008     ref.as_non_null
1009     i32.const n
1010     array.get $data
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030

```

To read a field of a block, we need to cast it to a block, access its second fields which contains the values of the block (the first field being the tag). Then we simply read the element at the index n .

```

1031
1032
1033  $\mathcal{T}(\text{set\_field } n \ x_1 \ x_2) = \text{local.get } \$x_1$ 
1034     ref.cast (ref $block)
1035     struct.get $block 1
1036     ref.cast (ref $data)
1037     i32.const x
1038     local.get $x_2
1039     array.set $data
1040     i32.const 0
1041     i31.new
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100

```

Similarly, to set a field of a block, we cast it, access its data field and then set the element at index n to the value of x_2 . There array is mutable and there is no need to update the structure after the element has been set.

```

1101
1102
1103  $\mathcal{T}(\text{make\_block } n \ x_0, \dots, x_{m-1}) = \text{i32.const } n$ 
1104     local.get $x_0
1105     ...
1106     local.get $x_{m-1}
1107     array.new_fixed $data m
1108     struct.new $block
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144

```

To create a new block, we put its tag on the stack. Then we put all its data and make a new array from it. With the tag and the array we then make a new structure of type `$block`.

```

1049  $\mathcal{T}(\text{project\_closure } x_1 \ x_2) =$ 
1050   local.get $x_2
1051   ref.cast (ref $set_of_closures)
1052   struct.get $set_of_closures 0
1053   global.get $closure_offset_within_set_x1
1054   array.get $closures

```

To read a closure from a set of closures, we cast x_2 to a set. Then we access its first field, which is an array of closures. Then we look for the offset of x_1 and access it.

```

1061  $\mathcal{T}(\text{project\_var } x_1 \ x_2) =$ 
1062   local.get $x_2
1063   ref.cast (ref $closure)
1064   struct.get $closure 1
1065   ref.as_non_null
1066   struct.get $set_of_closures 1
1067   global.get $var_offset_within_set_x1
1068   array.get $vars

```

This is similar to the closure case but the difference here is that x_2 is a closure and not a set. We need to access its parent first. Then we access the second field, which is the array of the captured variables. Then we need again to look for the offset of the variable before reading the array.

```

1078  $\mathcal{T}(\text{move\_within\_closure } x_1 \ x_2) =$ 
1079   local.get $x_2
1080   ref.cast (ref $closure)
1081   struct.get $closure 1
1082   ref.as_non_null
1083   struct.get $set_of_closures 0
1084   global.get $closure_offset_within_set_x1
1085   array.get $closures

```

This is the same as the previous case: x_2 is a closure and not a set. The only difference being that we look in the first field instead of the second one.

```

1093  $\mathcal{T}(\text{make\_set\_of\_closures } \mathcal{S})$ 
1094   =TODO

```

The code of the compiled Flambda program is inserted as the code of the `$start` function. But we also need to analyse the term in order to do a few more things. Each function defined in a set of closures is also added as a new Wasm function and marked as a declarative element (in order to be referenced by the `ref.func` instruction). Moreover, each function and each variable in a set

of closures has its own offset stored as a global variable (for instance `global $closure_offset_within_set_myfunc`). Finally, each function is analysed in order to find which locals it requires to be declared.

For now, we don't have a proof of the correctness of our compilation scheme. The formal semantics of the various Wasm extensions we are needing is still an ongoing work by the various working groups. We would like to use these semantics in the future.

4.1 Example of generated code

The example given in 2.4 would be compiled to the following Wasm code:

```

1103 (module $wasocaml_generated_modul
1104   (func $start (local $f (ref null eq)));; ...
1105   ref.null $closures
1106   array.new_fixed $vars 0
1107   struct.new $set_of_closures
1108   global.set $tmp_set_of_closures
1109   global.get $tmp_set_of_closures
1110   ref.func $cl_f
1111   ;; ...
1112   local.get $iter_print
1113   ref.cast (ref $closure)
1114   struct.get $closure 0
1115   call_ref $mlfun
1116   drop)
1117 (func $cl_f (param $env eqref) (param $x eqref)
1118   (result eqref)
1119   local.get $x
1120   ref.cast (ref i31)
1121   i31.get_u
1122   call $print_i32
1123   i32.const 0
1124   i31.new)
1125 (func $cl_iter_f (param $envtwo eqref) (param $l eqref)
1126   (result eqref) (local $x eqref) (local $f eqref)
1127   (local $dummy eqref) (local $tl eqref)
1128   (block $switch_a (result (ref null eq))
1129     (block $switch_b (result (ref null eq))
1130       local.get $l
1131       br_on_cast $switch_b (ref null eq) i31ref
1132       ref.cast (ref $block)
1133       struct.get $block 0
1134       global.set $tmp_switch_value
1135       global.get $tmp_switch_value
1136       i32.const 0
1137       i32.eq
1138       (if (result (ref null eq))
1139         (then
1140           local.get $l
1141           ref.cast (ref $block)
1142           struct.get $block 1
1143           ;; ...
1144           local.get $envtwo
1145           ref.cast (ref $closure)
1146           struct.get $closure 0

```

```

1161         call_ref $mlfun)
1162         (else unreachable))
1163     br $switch_a)
1164     ref.cast (ref i31)
1165     i31.get_s
1166     global.set $tmp_switch_value
1167     global.get $tmp_switch_value
1168     i32.const 0
1169     i32.eq
1170     (if (result (ref null eq))
1171         (then
1172             i32.const 0
1173             i31.new)
1174         (else
1175             unreachable))))
1176     (func $cl_iter ;; ...
1177     ))

```

5 PERSPECTIVES

As the first version of the implementation was intended as a demonstration, it remains a bit rough around the edges. In particular only a fraction of the externals from the standard library externals are provided as handwritten Wasm. The only unsupported part of the language is the objects fragment (by lack of time rather than due to any specific complexity). The source code of Wasocaml is publicly available [1].

5.1 FFI

A lot of OCaml code in the wild interacts with C or JavaScript code through bindings written using dedicated FFI mechanisms. We plan to allow re-use of existing bindings when compiling to Wasm. Currently, using C bindings when targeting the browser is not possible. People have to rewrite their code in pure OCaml or using JavaScript bindings. Compiling C bindings directly to Wasm allows to re-use them as-is.

C bindings. Some extensions recently added to Clang [6] allow to compile C code to Wasm in a way that makes reusing existing OCaml bindings to C code possible with almost no change. We would only have to provide a modified version of the OCaml native FFI headers files, replacing the usual macros with hand written Wasm functions. The only limitation that we foresee is that the `Field` macro will not be an l-value anymore; a new `Set_field` macro will be needed instead (as it was originally proposed for OCaml 5).

JavaScript bindings. To re-use existing bindings of OCaml to JavaScript code, we need one more extension: the reference-typed strings proposal [24]. Almost all external calls in the JavaScript FFI goes through the `Js.Unsafe.meth_call` function which has the type `'a -> string -> any array -> 'b`. It can be exposed to the Wasm module by the embedder as a function of type:

```

1213 (func $meth_call
1214     (param $obj externref)
1215     (param $method stringref)
1216     (param $args $anyarray)
1217     (result externref))

```

This calls a method of name `$method` on the object `$obj` with the arguments `$args`. The JavaScript side manages all the dynamic typing.

5.2 Effect Handlers Support

Our compiler is based on OCaml 4.14. So effect handlers were out of the scope. There are three strategies to handle them.

CPS Compilation. It is possible to represent effect handlers as continuations, using whole-program CPS transformation [4]. However it is not ideal for two reasons. It requires the program to contain OCaml code only. Moreover, it prevents the use of the stack of the target language and instead stores that stack in the heap, with a non-negligible cost. It can be mitigated by only applying the transformation to code that cannot be proved to not use effect handlers but then it is no longer compatible with separate compilation and the optimisation needs to be done as a whole program transformation. This is the choice made by `Js_of_ocaml`.

Wasm Stack Switching. There is an ongoing proposal, called stack switching [23], that exactly matches the needs for OCaml effect handlers. With it the translation would be quite straightforward.

JavaScript Promise Integration. Another strategy is available for runtimes providing both Wasm and JavaScript. There is another ongoing proposal called JavaScript promise integration [22]. With this proposal, effects handlers can be implemented with a JavaScript device. This proposal is likely to land before the proper stack switching one. It might be a temporary solution.

6 RELATED WORK

6.1 Garbage Collected Languages to Wasm

6.1.1 Targeting Wasm1.

Go. Go compiles to Wasm1 [9]. In order to be able to re-use the Go GC, it must be able to inspect the stack. This is not possible within Wasm. Thus it has to maintain the stack by itself using the memory, which is much slower than using the Wasm stack.

Haskell. Haskell also compiles to Wasm [10]. It has constraints similar to Go and uses similar solutions. In particular, it also uses the memory to manage the stack.

6.1.2 Targeting Wasm-GC.

Dart. Dart [12] compiles to Wasm-GC. It does not use `i31ref` and boxes scalars in a `struct`.

Scheme. The last addition to the small family of compiler targeting Wasm-GC is the Guile Scheme compiler. Scheme has many similar constraints as OCaml and Guile uses many similar solutions. The compiler was presented to the Wasm-GC working group [26]. A more detailed explanation is published [25]. The implementation and its in-depth technical description are available [27].

6.2 OCaml Web Compilers

The history of OCaml compilers targeting web languages is quite crowded. Surely thanks to the great pleasure that is writing compilers in OCaml.

1277 **6.2.1 Targeting JavaScript.** There are multiple OCaml compilers
 1278 targeting JavaScript. Many approaches were experimented, the
 1279 main two live ones are `jsoo` and `melange`. Naive compilation of
 1280 OCaml to JavaScript is quite simple as it can almost be summa-
 1281 rized as “type erasure”. There are some limitations in JavaScript
 1282 that prevent that to be a complete compilation strategy, and, of
 1283 course, a proper compiler producing efficient and small JavaScript
 1284 code is quite more complex.

1285 `Jsoo`. `Jsoo` [15] tries to be as close as possible from the native
 1286 semantics. It compiles OCaml bytecode programs to JavaScript by
 1287 decompiling it back to a simple untyped lambda calculus then per-
 1288 forms many optimisation and minimisation passes.

1290 `Melange`. `Melange` [8] tries to produce readable JavaScript, with
 1291 a closer integration in the JavaScript module system. `Melange` starts
 1292 from a modified version of the lambda IR which provides more type
 1293 information. This allows the use of JavaScript features that match
 1294 the uses of source features at the cost of some small semantical
 1295 differences.

1297 6.2.2 Targeting Wasm.

1298 `OCamlrun WebAssembly`. `OCamlrun WebAssembly` [7] compiles
 1299 the OCaml runtime and interpreter to Wasm. They are both writ-
 1300 ten in C so it is possible without any Wasm extension. In some
 1301 cases, it allows to start executing code faster than when compilin
 1302 OCaml to Wasm. This is because there is less WebAssembly code
 1303 to compile for the embedder. On the other hand, the execution is
 1304 slower as it is interpreting OCaml bytecode inside Wasm instead
 1305 of running a compiled version.

1307 `WASICaml`. `WASICaml` [11] is quite similar to `OCamlrun We-`
 1308 `bAssembly`. The main difference being that `WASICaml` does not
 1309 interpret the bytecode directly but translates it partially to Wasm.
 1310 It leads to a faster execution.

1312 `Wasm_of_ocaml`. `Wasm_of_ocaml` [14] is a fork of `Js_of_ocaml`.
 1313 It compiles the bytecode to Wasm-GC and re-uses many of the
 1314 techniques developed by `Wasocaml`. It does not benefit from the
 1315 optimisations provided by `Flambda`. It is also quite restricted in its
 1316 value representation choice and must use a representation based
 1317 on arrays.

1319 7 CONCLUSION

1320 The original goal of this implementation was to validate the Wasm-
 1321 GC proposal and discuss its limits. This part was successful. No ma-
 1322 jor changes were required for OCaml (mainly keeping the `i31ref`
 1323 type and changes to the maximal length of subtyping chains). And
 1324 we are quite happy to claim that Wasm-GC is, to our opinion a
 1325 very good design, that is a perfect compilation target for a garbage
 1326 collected functional language like OCaml. And we think that the
 1327 experience for most other garbage-collected languages will proba-
 1328 bly be similar.

1329 As a side-effect, we now have an OCaml to Wasm-GC compiler.
 1330 It is not yet usable because there are no user-side Wasm-GC run-
 1331 time available. In order to test our compiler we are using V8 with
 1332 various flags to enable experimental features such as the GC sup-
 1333 port. The design of the various extensions required by `Wasocaml`

1335 is stable and quite close to completion, but some details are still in
 1336 flux. For this reason, we cannot expect it them be widely available
 1337 soon. On the other hand, it means that our compiler will be ready
 1338 when browsers start deploying new Wasm extensions.

1339 Our future plans are to complete the Wasm implementations of
 1340 OCaml externals, to implement the various FFI mechanisms, sup-
 1341 port effect handlers and to move `Wasocaml` to `Flambda2`. All of
 1342 these would allow to easily deploy multi-language software on the
 1343 browser while having good and predictable performances.

1345 REFERENCES

- 1346 [1] Léo Andrés and Pierre Chambart. 2022. *Wasocaml*. <https://github.com/ocaml-wasm/wasocaml>
- 1347 [2] Léo Andrés and Pierre Chambart. 2023. OCaml on WasmGC. <https://github.com/WebAssembly/meetings/blob/main/gc/2023/GC-01-10.md>.
- 1348 [3] Andreas Haas, Andreas Rossberg, Derek L Schuff, Ben L Titzer, Michael Holman, Dan Gohman, Luke Wagner, Alon Zakai, and JF Bastien. 2017. Bringing the web up to speed with WebAssembly. In *Proceedings of the 38th ACM SIGPLAN Conference on Programming Language Design and Implementation*. 185–200.
- 1349 [4] Daniel Hillerström, Sam Lindley, Robert Atkey, and KC Sivaramakrishnan. 2017. Continuation passing style for effect handlers. (2017).
- 1350 [5] Donald E Knuth and Peter B Bendix. 1983. Simple word problems in universal algebras. *Automation of Reasoning: 2: Classical Papers on Computational Logic 1967–1970* (1983), 342–376.
- 1351 [6] Chris Lattner. 2008. LLVM and Clang: Next generation compiler technology. In *The BSD conference*, Vol. 5, 1–20.
- 1352 [7] Sebastian Markbåge. 2017. *OCamlrun WebAssembly*. <https://github.com/sebmarkbage/ocamlrun-wasm>
- 1353 [8] Antonio Nuno Monteiro. 2022. *Melange*. <https://github.com/melange-re/melange>
- 1354 [9] Richard Musiol. 2018. WebAssembly architecture for Go. https://docs.google.com/document/d/131vjr4DH6JFnB-blm_uRdaC0_Nv3OUwjEY5qVCxCup4.
- 1355 [10] Cheng Shao. 2022. WebAssembly backend merged into GHC. <https://www.tweag.io/blog/2022-11-22-wasm-backend-merged-in-ghc>.
- 1356 [11] Gerd Stolpmann. 2021. *WASICaml*. <https://github.com/remixlabs/wasicaml>
- 1357 [12] The Dart Project Authors. 2022. *dart2wasm*. <https://github.com/dart-lang/sdk/tree/main/pkg/dart2wasm>
- 1358 [13] The Emscripten Authors. 2014. *Emscripten*. <https://github.com/emscripten-core/emscripten>
- 1359 [14] Jérôme Vouillon. 2023. *Wasm_of_ocaml*. https://github.com/ocaml-wasm/wasm_of_ocaml
- 1360 [15] Jérôme Vouillon and Vincent Balat. 2014. From bytecode to JavaScript: the `Js_of_ocaml` compiler. *Software: Practice and Experience* 44, 8 (2014), 951–972.
- 1361 [16] WebAssembly Community Group participants. 2015. *Binaryen*. <https://github.com/WebAssembly/binaryen>
- 1362 [17] WebAssembly Community Group participants. 2017. *Exception Handling Proposal for WebAssembly*. <https://github.com/WebAssembly/exception-handling>
- 1363 [18] WebAssembly Community Group participants. 2017. *GC Proposal for WebAssembly*. <https://github.com/WebAssembly/gc>
- 1364 [19] WebAssembly Community Group participants. 2017. *Tail Call Proposal for WebAssembly*. <https://github.com/WebAssembly/tail-call>
- 1365 [20] WebAssembly Community Group participants. 2018. *Reference Types Proposal for WebAssembly*. <https://github.com/WebAssembly/reference-types>
- 1366 [21] WebAssembly Community Group participants. 2020. *Typed Function References Proposal for WebAssembly*. <https://github.com/WebAssembly/function-references>
- 1367 [22] WebAssembly Community Group participants. 2021. *JavaScript-Promise Integration Proposal for WebAssembly*. <https://github.com/WebAssembly/js-promise-integration>
- 1368 [23] WebAssembly Community Group participants. 2021. *Stack Switching Proposal for WebAssembly*. <https://github.com/WebAssembly/stack-switching>
- 1369 [24] WebAssembly Community Group participants. 2022. *Reference-Typed Strings Proposal for WebAssembly*. <https://github.com/WebAssembly/stringref>
- 1370 [25] Andy Wingo. 2023. a world to win: webassembly for the rest of us. <https://www.wingolog.org/archives/2023/03/20/a-world-to-win-webassembly-for-the-rest-of-us>.
- 1371 [26] Andy Wingo. 2023. Compiling Scheme to WasmGC. <https://github.com/WebAssembly/meetings/blob/main/gc/2023/GC-04-18.md>.
- 1372 [27] Andy Wingo. 2023. *Hoot*. <https://gitlab.com/spritely/guile-hoot/-/blob/main/design/ABI.md>