

# Melocoton: A Program Logic for Verified Interoperability Between OCaml and C

---

Armaël Guéneau

Johannes Hostert

Simon Spies

Michael Sammler

Lars Birkedal

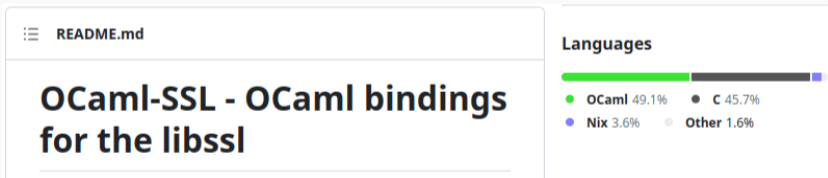
Derek Dreyer

Sept 11, 2023

# Many Real Programs Are Multi-Language

Consider the `ocaml-ssl` library:

- Exposes OpenSSL (a C library) as an OCaml library
- To do so, it is implemented using a mix of *both* OCaml *and* C code:



# Mind the gap!

OCaml

C

# Mind the gap!

## OCaml

Structured values

$\lambda_{ML}$   $V \in Val ::= (n \in \mathbb{Z}) \mid (\ell \in Loc)$   
| **true** | **false**  
|  $\langle \rangle$  |  $\langle V, V \rangle$  |  $\lambda x. e \dots$

Garbage collection

## C

Integers and pointers

$\lambda_C$   $w \in Val ::= (n \in \mathbb{Z}) \mid (a \in Addr)$

Manual memory management

# Mind the gap!

OCaml  $\longleftrightarrow$  OCaml FFI  $\longleftrightarrow$  C

Structured values

$\lambda_{ML}$   $V \in Val ::= (n \in \mathbb{Z}) \mid (\ell \in Loc)$   
| true | false  
|  $\langle \rangle$  |  $\langle V, V \rangle$  |  $\lambda x. e \dots$

Garbage collection

Integers and pointers

$\lambda_C$   $w \in Val ::= (n \in \mathbb{Z}) \mid (a \in Addr)$

Manual memory management

## Key Challenge (as an OCaml hacker)

Write “glue code” using the OCaml FFI is **tricky and unsafe**.

mistake  $\Rightarrow$  *memory corruption* (often silent and hard to debug)



Which **rules** should I follow to safely use the OCaml FFI?

distinguished  
OCaml hacker

## Key Challenge (as an expert in program logics)

We already have powerful *program logics* for OCaml and C  
but those are for programs written in a *single* language



How do we **formally reason** about such multi-language code?

program logics  
expert

## Key challenge (in this work)

Can we build a program logic for reasoning about interoperability with an FFI,  
**while preserving language-local reasoning?**

$\lambda_{ML}$  Semantics

**Iris<sub>ML</sub>**

Program Logic

given as black box

$\lambda_C$  Semantics

**Iris<sub>C</sub>**

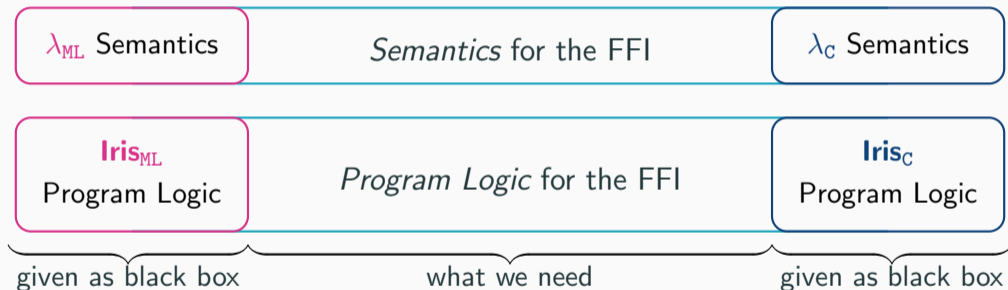
Program Logic

given as black box



## Key challenge (in this work)

Can we build a program logic for reasoning about interoperability with an FFI,  
**while preserving language-local reasoning?**



**Design choice:** reuse most of existing semantics/program logics;  
**do not** drop down to a lowest-common denominator (assembly)!

## Melocoton:

- Two instantiations of Iris for a ML-like and C-like language with *external calls*
- An *operational semantics* for the OCaml FFI, bridging between the two languages.
- A *separation logic* for the OCaml FFI, bridging between the two language logics.
- A number of interesting *case studies*

## Melocoton:

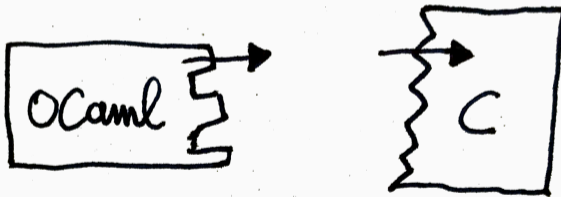
- Two instantiations of Iris for a ML-like and C-like language with *external calls*
- An *operational semantics* for the OCaml FFI, bridging between the two languages.
- A *separation logic* for the OCaml FFI, bridging between the two language logics.
- A number of interesting *case studies*

**Language-locality:** Verification of mixed OCaml/C programs can be done *almost entirely* in logics for OCaml and C!

**In Iris:** the logic is proved sound and all proofs are checked in Coq

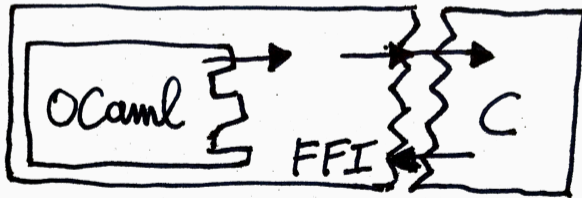
# Outline

## 1. Language-local program logics with external calls



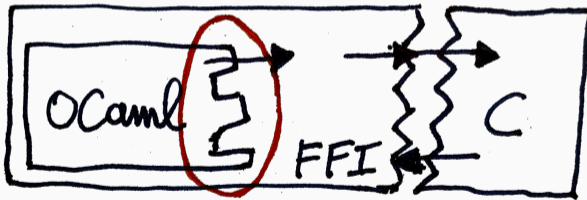
# Outline

1. Language-local program logics with external calls
2. Program logic for FFI



# Outline

1. Language-local program logics with external calls
2. Program logic for FFI
3. Focus: the language boundary



The OCaml FFI deals with two core challenges:

- mediating between the different views of the OCaml memory
- interacting with the OCaml GC

The OCaml FFI deals with two core challenges:

- **mediating between the different views of the OCaml memory**
- interacting with the OCaml GC



## Example: updating an OCaml reference from C code

OCaml code:

```
let main () =  
  let r = ref 0 in  
  update_ref r; (* TODO call C code and use rand() *)  
  print_int !r
```

C code:

```
int rand(int x) { ... }
```

## Example: updating an OCaml reference from C code

OCaml code:

```
external update_ref : int ref -> unit = "caml_update_ref"  
let main () =  
  let r = ref 0 in  
  update_ref r;  
  print_int !r
```

C code:

```
int rand(int x) { ... }
```

## Example: updating an OCaml reference from C code

OCaml code:

```
external update_ref : int ref -> unit = "caml_update_ref"  
let main () =  
  let r = ref 0 in  
  update_ref r;  
  print_int !r
```

C code:

```
int rand(int x) { ... }
```

Glue code:

```
value caml_update_ref(value r) {  
  /* TODO */  
  int y = rand(x);  
  /* TODO */  
}
```

## Example: updating an OCaml reference from C code

OCaml code:

```
external update_ref : int ref -> unit = "caml_update_ref"  
let main () =  
  let r = ref 0 in  
  update_ref r;  
  print_int !r
```

C code:

```
int rand(int x) { ... }
```

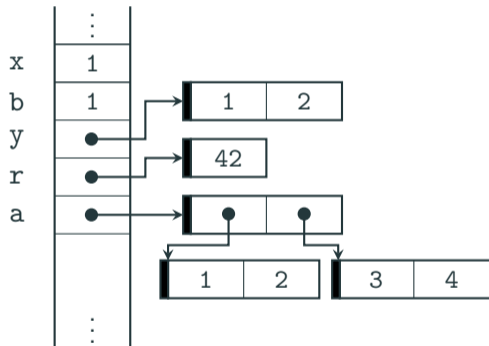
Glue code:

```
value caml_update_ref(value r) {  
  /* TODO */  
  int y = rand(x);  
  /* TODO */  
}
```

## The runtime representation of OCaml values

At runtime, an OCaml value is either an integer or a pointer to a block:

```
let x = 1
let b = true
let y = (1, 2)
let r = ref 42
let a = [| (1, 2); (3, 4) |]
```



Glue code has access to this *low-level* representation of OCaml values.

## Example: updating an OCaml reference from C code

OCaml code:

```
external update_ref : int ref -> unit = "caml_update_ref"  
let main () =  
  let r = ref 0 in  
  update_ref r;  
  print_int !r
```

C code:

```
int rand(int x) { ... }
```

Glue code:

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0));  
  int y = rand(x);  
  Store_field(r, 0, Val_int(y));  
  return Val_int(0);  
}
```

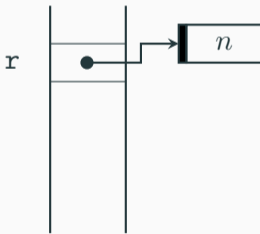
Glue code bridges between OCaml and C values by using powerful **FFI primitives**...

## Writing glue code

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0));      /* read the first field of the input block */  
  int y = rand(x);                  /* get a random integer */  
  Store_field(r, 0, Val_int(y));    /* store the value in the block */  
  return Val_int(0);                /* return () */  
}
```

## Writing glue code

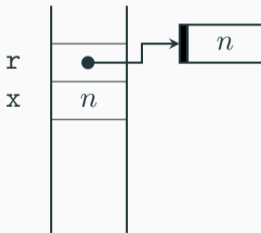
```
value caml_update_ref(value r) {<--  
  int x = Int_val(Field(r, 0));      /* read the first field of the input block */  
  int y = rand(x);                   /* get a random integer */  
  Store_field(r, 0, Val_int(y));    /* store the value in the block */  
  return Val_int(0);                 /* return () */  
}
```





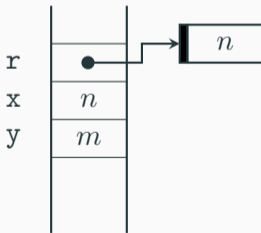
## Writing glue code

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0)); <-- /* read the first field of the input block */  
  int y = rand(x); /* get a random integer */  
  Store_field(r, 0, Val_int(y)); /* store the value in the block */  
  return Val_int(0); /* return () */  
}
```



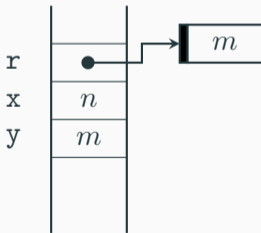
## Writing glue code

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0));      /* read the first field of the input block */  
  int y = rand(x);                   <-- /* get a random integer */  
  Store_field(r, 0, Val_int(y));     /* store the value in the block */  
  return Val_int(0);                 /* return () */  
}
```



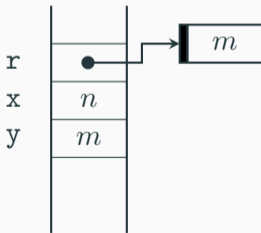
## Writing glue code

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0));      /* read the first field of the input block */  
  int y = rand(x);                   /* get a random integer */  
  Store_field(r, 0, Val_int(y));<-- /* store the value in the block */  
  return Val_int(0);                 /* return () */  
}
```



## Writing glue code

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0));      /* read the first field of the input block */  
  int y = rand(x);                   /* get a random integer */  
  Store_field(r, 0, Val_int(y));    /* store the value in the block */  
  return Val_int(0);                 <-- /* return () */  
}
```



The OCaml FFI deals with two core challenges:

- mediating between the different views of the OCaml memory
- **interacting with the OCaml GC**

## Example: swapping an OCaml pair

OCaml code:

```
external swap_pair : 'a * 'b -> 'b * 'a = "caml_swap_pair"
```

## Example: swapping an OCaml pair

OCaml code:

```
external swap_pair : 'a * 'b -> 'b * 'a = "caml_swap_pair"
```

Glue code:

(first attempt)

```
value caml_swap_pair(value p)
{
  value r = caml_alloc(0, 2); /* allocate a block for the result */
  value x = Field(p, 0);      /* read the input pair */
  value y = Field(p, 1);
  Store_field(r, 0, y);      /* initialize the output pair */
  Store_field(r, 1, x);
  return r;                  /* return it */
}
```

## Example: swapping an OCaml pair

OCaml code:

```
external swap_pair : 'a * 'b -> 'b * 'a = "caml_swap_pair"
```

Glue code:

(first attempt)

```
value caml_swap_pair(value p)
{
  value r = caml_alloc(0, 2); /* allocate a block for the result */
  value x = Field(p, 0);      /* read the input pair */
  value y = Field(p, 1);
  Store_field(r, 0, y);      /* initialize the output pair */
  Store_field(r, 1, x);
  return r;                  /* return it */
}
```

This implementation is unfortunately **incorrect** and will silently corrupt memory!

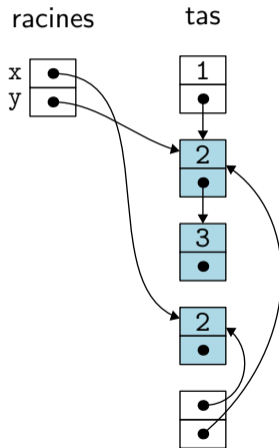
`caml_alloc` may run the GC which does not know about C variables and arguments...



OCaml has a “tracing” garbage collector.

Starts from *roots*; collects unreachable blocks; may also *move* blocks in memory.

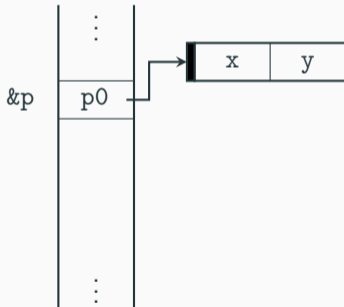
```
let x, y =
  let l = [1; 2; 3] in
  (List.filter even l, List.tl l)
...
```



## Swapping pairs in the presence of a garbage collector

This implementation is unfortunately **incorrect!**

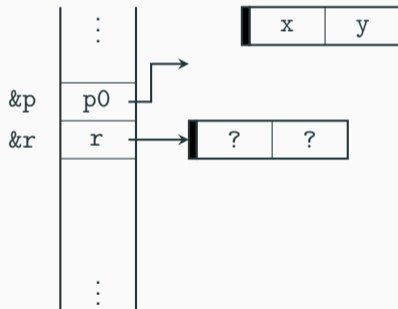
```
value caml_swap_pair(value p)
{
    value r = caml_alloc(0, 2);
    value x = Field(p, 0);
    value y = Field(p, 1);
    Store_field(r, 0, y);
    Store_field(r, 1, x);
    return r;
}
```



## Swapping pairs in the presence of a garbage collector

This implementation is unfortunately **incorrect!**

```
value caml_swap_pair(value p)
{
  value r = caml_alloc(0, 2); <--
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  return r;
}
```

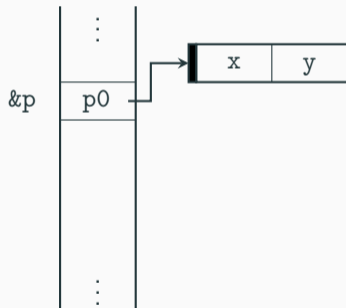


## Registering roots

`CAMLparam1(p)` registers `&p` as a GC root.

The GC will avoid collecting the block, and will *update* `p` if the block moves.

```
value caml_swap_pair(value p)
{
    CAMLparam1(p);
    value r = caml_alloc(0, 2);
    value x = Field(p, 0);
    value y = Field(p, 1);
    Store_field(r, 0, y);
    Store_field(r, 1, x);
    return r;
}
```

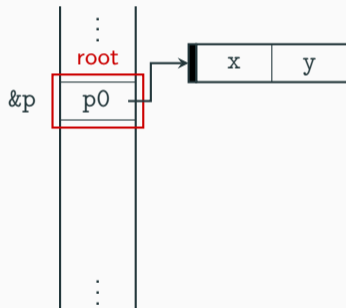


## Registering roots

`CAMLparam1(p)` registers `&p` as a GC root.

The GC will avoid collecting the block, and will *update* `p` if the block moves.

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);          <--
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  return r;
}
```

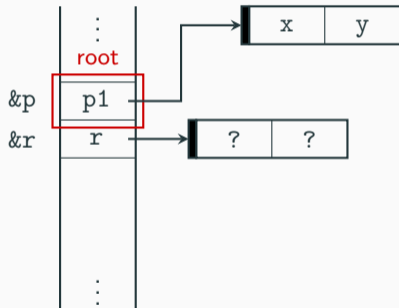


## Registering roots

`CAMLparam1(p)` registers `&p` as a GC root.

The GC will avoid collecting the block, and will *update* `p` if the block moves.

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2); <--
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  return r;
}
```

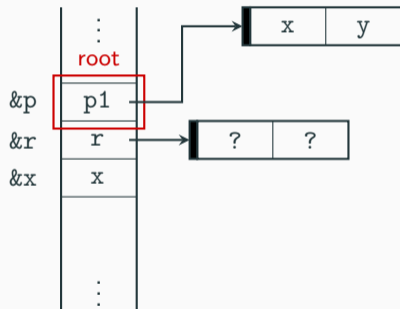


## Registering roots

`CAMLparam1(p)` registers `&p` as a GC root.

The GC will avoid collecting the block, and will *update* `p` if the block moves.

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);      <--
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  return r;
}
```

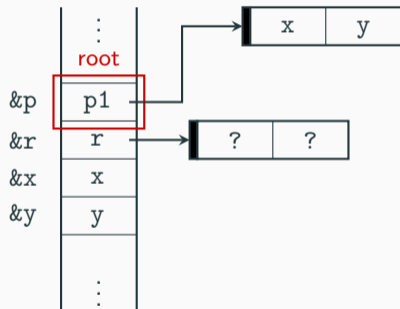


## Registering roots

`CAMLparam1(p)` registers `&p` as a GC root.

The GC will avoid collecting the block, and will *update* `p` if the block moves.

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);      <--
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  return r;
}
```



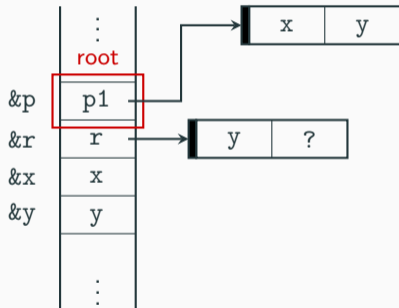


## Registering roots

`CAMLparam1(p)` registers `&p` as a GC root.

The GC will avoid collecting the block, and will *update* `p` if the block moves.

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);      <--
  Store_field(r, 1, x);
  return r;
}
```

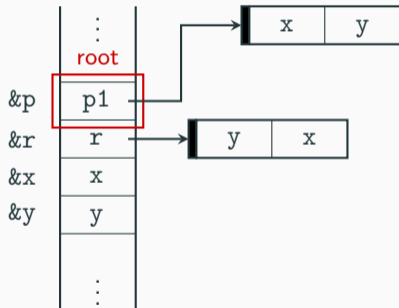


## Registering roots

`CAMLparam1(p)` registers `&p` as a GC root.

The GC will avoid collecting the block, and will *update* `p` if the block moves.

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);    <--
  return r;
}
```



## Unregistering roots

One subtle **bug** remains!

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  return r;
}
```

## Unregistering roots

One subtle **bug** remains!

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  return r;
}
```

The GC will *continue to update* `&p` after the function returns, corrupting the stack...!

We must use `CAMLreturn()` to unregister local roots when returning.

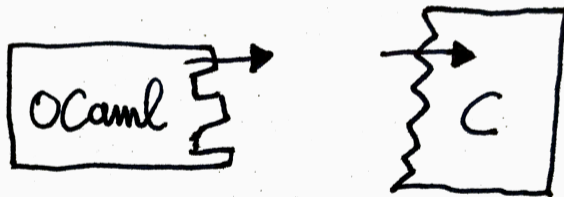
## Our final implementation for swap\_pair

```
external swap_pair : 'a * 'b -> 'b * 'a = "caml_swap_pair"
```

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  CAMLreturn(r);
}
```

## Outline: Language-local reasoning

1. Language-local program logics with external calls



# Language-local reasoning

We reuse:

$\lambda_{ML}$  Semantics

**Iris<sub>ML</sub>**  
Program Logic

$\lambda_C$  Semantics

**Iris<sub>C</sub>**  
Program Logic

The one change: a minimal extension allowing **external calls**.

## Modeling External Calls

```
external update_ref : int ref -> unit = "caml_update_ref"

let main () :=
  let r = ref 0 in
  update_ref r;
  print_int !r
```



## Modeling External Calls

```
external update_ref : int ref -> unit = "caml_update_ref"  
  
let main () :=  
  let r = ref 0 in  
  update_ref r;  
  print_int !r
```

We model external calls as a new syntactic construct (inlining the declaration):

$$e \in Expr ::= \dots \mid \text{call } fn \vec{e}$$

## Modeling External Calls

```
external update_ref : int ref -> unit = "caml_update_ref"

let main () :=
  let r = ref 0 in
  update_ref r;
  print_int !r
```

We model external calls as a new syntactic construct (inlining the declaration):

$$e \in Expr ::= \dots \mid \text{call } fn \vec{e}$$

We assign **no semantics** to external calls: they are simply stuck!

## Interface Specifications

We still want to *reason* about calls to `caml_update_ref`, as if it had the specification:

$$\forall \ell n. \{ \ell \mapsto_{\text{ML}} n \} \text{ call caml\_update\_ref } [\ell] \{ V'. \exists m. V' = \langle \rangle * \ell \mapsto_{\text{ML}} m \}_{\text{ML}}$$

## Interface Specifications

We still want to *reason* about calls to `caml_update_ref`, as if it had the specification:

$$\forall \ell n. \{ \ell \mapsto_{\text{ML}} n \} \text{ call caml\_update\_ref } [\ell] \{ V'. \exists m. V' = \langle \rangle * \ell \mapsto_{\text{ML}} m \}_{\text{ML}}$$

“ $\ell \mapsto_{\text{ML}} V$ ” is a *Separation Logic assertion*

- asserts that the memory location  $\ell$  stores the value  $V$
- grants the **permission** to access the location (read/write)

## Interface Specifications

We still want to *reason* about calls to `caml_update_ref`, as if it had the specification:

$$\forall l n. \{l \mapsto_{\text{ML}} n\} \text{call caml\_update\_ref } [l] \{V'. \exists m. V' = \langle \rangle * l \mapsto_{\text{ML}} m\}_{\text{ML}}$$

To do so, we introduce **interfaces**  $\Psi$  and Hoare triples  $\{P\} e @ \Psi \{v. Q\}$  that verify programs against them. For example, for `caml_update_ref`, we assume:

$$\forall l n. \langle l \mapsto_{\text{ML}} n \rangle \text{caml\_update\_ref } [l] \langle V'. \exists m. V' = \langle \rangle * l \mapsto_{\text{ML}} m \rangle \sqsubseteq \Psi$$

## Interface Specifications

We still want to *reason* about calls to `caml_update_ref`, as if it had the specification:

$$\forall l n. \{l \mapsto_{\text{ML}} n\} \text{ call caml\_update\_ref } [l] \{V'. \exists m. V' = \langle \rangle * l \mapsto_{\text{ML}} m\}_{\text{ML}}$$

To do so, we introduce **interfaces**  $\Psi$  and Hoare triples  $\{P\} e @ \Psi \{v. Q\}$  that verify programs against them. For example, for `caml_update_ref`, we assume:

$$\forall l n. \langle l \mapsto_{\text{ML}} n \rangle \text{ caml\_update\_ref } [l] \langle V'. \exists m. V' = \langle \rangle * l \mapsto_{\text{ML}} m \rangle \sqsubseteq \Psi$$



**This is an assumption, not a Hoare triple**



## Desugaring To Predicate Transformers

Implement interface triples as a predicate transformer  $\Psi$ :

$$\Psi : \underbrace{FnName}_{\text{Name}} \rightarrow \underbrace{\vec{Val}}_{\text{Args}} \rightarrow \underbrace{(Val \rightarrow iProp)}_{\text{Postcondition}} \rightarrow \underbrace{iProp}_{\text{Precondition}}$$

## Desugaring To Predicate Transformers

Implement interface triples as a predicate transformer  $\Psi$ :

$$\Psi : \text{FnName}_e \rightarrow \vec{Val} \rightarrow (Val \rightarrow iProp) \rightarrow iProp$$

“*iProp*” is the type of *Iris propositions*, which includes:

- quantifiers  $\forall, \exists, \dots$  and pure propositions
- Separation Logic modalities
- memory assertions of both languages ( $\ell \mapsto_{ML} V, a \mapsto_C w$ )
- specifications  $\{P\} e @ \Psi \{Q\}$  of both languages



# Desugaring To Predicate Transformers

Implement interface triples as a predicate transformer  $\Psi$ :

$$\Psi : \underbrace{FnName}_{\text{Name}} \rightarrow \underbrace{\vec{Val}}_{\text{Args}} \rightarrow \underbrace{(Val \rightarrow iProp)}_{\text{Postcondition}} \rightarrow \underbrace{iProp}_{\text{Precondition}}$$

We desugar

$$\forall l n. \langle l \mapsto_{ML} n \rangle \text{caml\_update\_ref } [l] \\ \langle V'. \exists m. V' = \langle \rangle * l \mapsto_{ML} m \rangle$$

# Desugaring To Predicate Transformers

Implement interface triples as a predicate transformer  $\Psi$ :

$$\Psi : \underbrace{FnName}_{\text{Name}} \rightarrow \underbrace{\vec{Val}}_{\text{Args}} \rightarrow \underbrace{(Val \rightarrow iProp)}_{\text{Postcondition}} \rightarrow \underbrace{iProp}_{\text{Precondition}}$$

We desugar

$$\forall l n. \langle l \mapsto_{ML} n \rangle \text{caml\_update\_ref } [l] \\ \langle V'. \exists m. V' = \langle \rangle * l \mapsto_{ML} m \rangle$$

as follows:

$$\Psi_{upd} fn \vec{V} \Phi := \exists l n. l \mapsto_{ML} n * fn = \text{caml\_update\_ref} * \vec{V} = [l] \\ * (\forall V' m. V' = \langle \rangle * l \mapsto_{ML} m \multimap \Phi(V'))$$

## Implementing Interface Triples

$$\Psi : \underbrace{FnName}_{\text{Name}} \rightarrow \underbrace{\vec{Val}}_{\text{Args}} \rightarrow \underbrace{(Val \rightarrow iProp)}_{\text{Postcondition}} \rightarrow \underbrace{iProp}_{\text{Precondition}}$$

We parameterize Hoare triples by  $\Psi$  (inspired by de Vilhena and Pottier [2021]):

“ $\{P\} e @ \Psi \{Q\}$ ” means:

“Starting from a state satisfying  $P$ ,  $e$  reduces to a value arriving in a state satisfying  $Q$   
— either by normal reductions, *or by making external calls that satisfy  $\Psi$* ”

## Implementing Interface Triples

$$\Psi : \underbrace{FnName}_{\text{Name}} \rightarrow \underbrace{\vec{Val}}_{\text{Args}} \rightarrow \underbrace{(Val \rightarrow iProp)}_{\text{Postcondition}} \rightarrow \underbrace{iProp}_{\text{Precondition}}$$

We parameterize Hoare triples by  $\Psi$  (inspired by de Vilhena and Pottier [2021]):

“ $\{P\} e @ \Psi \{Q\}$ ” means:

“Starting from a state satisfying  $P$ ,  $e$  reduces to a value arriving in a state satisfying  $Q$   
— either by normal reductions, *or by making external calls that satisfy  $\Psi$* ”

**Note:** In a OCaml-and-C program (after linking), adequacy holds for  $\Psi \text{ fn } \vec{V} \Phi := \perp$

## Implementing Interface Triples

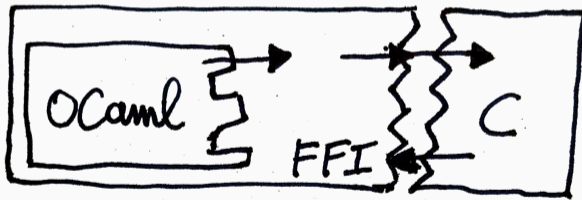
In **Iris**, we then **define** Hoare triples in terms of the operational semantics:

$$\{P\} e @ \Psi \{Q\} := \square (P \multimap \text{wp } e @ \Psi \{Q\})$$

$$\text{wp } e @ \Psi \{Q\} := \begin{cases} Q(v) & e = v \\ \forall e', (e \rightarrow e') \Rightarrow \text{wp } e' @ \Psi \{Q\} & e \text{ reducible} \\ \Psi \text{ fn } \vec{V} \underbrace{(\lambda V'. \text{wp } K[V'] @ \Psi \{Q\})}_{\text{Postcondition}} & e = K[\text{call } \text{fn } \vec{V}] \end{cases}$$

## Outline: The OCaml FFI

1. Language-local program logics with external calls
2. Glue code and program logic for FFI



## External Calls in Glue Code

In glue code we treat operations of the OCaml FFI as **external functions**.

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0));  
  int y = rand(x);  
  Store_field(r, 0, Val_int(y));  
  return Val_int(0);  
}
```

Glue code is verified using the program logic for C, but additionally **assuming an interface**  $\Psi_{\text{FFI}}$  for the OCaml FFI primitives, with resources e.g.  $\gamma \mapsto_{\text{blk}[t|m]} \vec{v}$ .

## External Calls in Glue Code

In glue code we treat operations of the OCaml FFI as **external functions**.

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0));  
  int y = rand(x);  
  Store_field(r, 0, Val_int(y));  
  return Val_int(0);  
}
```

Glue code is verified using the program logic for C, but additionally **assuming an interface**  $\Psi_{\text{FFI}}$  for the OCaml FFI primitives, with resources e.g.  $\gamma \mapsto_{\text{blk}[t|m]} \vec{v}$ .

$$\begin{aligned} & \langle \text{GC}(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} \vec{v} * \gamma \sim_{\text{C}}^{\theta} w * v' \sim_{\text{C}}^{\theta} w' \rangle \\ & \quad \text{Store\_field}(w, i, w') \quad \sqsubseteq \Psi_{\text{FFI}} \\ & \langle \text{GC}(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} \vec{v}[i := v'] \rangle \end{aligned}$$



## External Calls in Glue Code

In glue code we treat operations of the OCaml FFI as **external functions**.

```
value caml_update_ref(value r) {  
  int x = Int_val(Field(r, 0));  
  int y = rand(x);  
  Store_field(r, 0, Val_int(y));  
  return Val_int(0);  
}
```

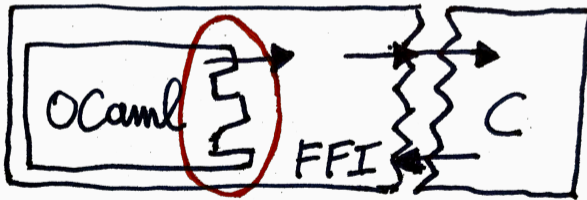
$$\{GC(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} [n] * \gamma \sim_{\text{C}}^{\theta} w\}$$
$$\text{call caml\_update\_ref } [w] @ \Psi_{\text{FFI}}$$
$$\{w'. \exists m. GC(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} [m] * w' \sim_{\text{C}}^{\theta} 0\}$$

Glue code is verified using the program logic for C, but additionally **assuming an interface**  $\Psi_{\text{FFI}}$  for the OCaml FFI primitives, with resources e.g.  $\gamma \mapsto_{\text{blk}[t|m]} \vec{v}$ .

$$\langle GC(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} \vec{v} * \gamma \sim_{\text{C}}^{\theta} w * v' \sim_{\text{C}}^{\theta} w' \rangle$$
$$\text{Store\_field}(w, i, w') \quad \sqsubseteq \Psi_{\text{FFI}}$$
$$\langle GC(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} \vec{v}[i := v'] \rangle$$

## Outline: The OCaml-FFI boundary

1. Language-local program logics with external calls
2. Glue code and program logic for FFI
3. Focus: the OCaml-FFI boundary



## View Reconciliation

We assumed an interface for `caml_update_ref` that uses ML points-tos:

$$\forall \ell n. \langle \ell \mapsto_{\text{ML}} n \rangle \text{ caml\_update\_ref } [\ell] \langle V'. \exists m. V' = \langle \rangle * \ell \mapsto_{\text{ML}} m \rangle$$

Meanwhile, we proved the following specification for `caml_update_ref` using  $\Psi_{\text{FFI}}$ :

$$\begin{aligned} & \{ \text{GC}(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} [n] * \gamma \sim_{\text{C}}^{\theta} w \} \\ & \quad \text{call caml\_update\_ref } [w] @ \Psi_{\text{FFI}} \\ & \{ w'. \exists m. \text{GC}(\theta) * w' \sim_{\text{C}}^{\theta} 0 * \gamma \mapsto_{\text{blk}[0|\text{mut}]} [m] \} \end{aligned}$$

These express two different views about the **same piece of state!**

## View Reconciliation: Update Rules

Idea:

- make  $l \mapsto_{\text{ML}} \vec{V}$  and  $\gamma \mapsto_{\text{blk}[0|\text{mut}]} \vec{v}$  mutually exclusive (for related  $l$  and  $\gamma$ )
- have *view reconciliation* rules to switch between the two representations

$$\text{GC}(\theta) * l \mapsto_{\text{ML}} \vec{V} \equiv * \exists \vec{v}, \gamma. \text{GC}(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} \vec{v} * l \sim_{\text{ML}} \gamma * \vec{V} \sim_{\text{ML}} \vec{v} \quad (\text{ML-TO-FFI})$$

$$\text{GC}(\theta) * \gamma \mapsto_{\text{blk}[0|\text{mut}]} \vec{v} * \vec{V} \sim_{\text{ML}} \vec{v} \equiv * \exists l. \text{GC}(\theta) * l \mapsto_{\text{ML}} \vec{V} * l \sim_{\text{ML}} \gamma \quad (\text{FFI-TO-ML})$$

## View Reconciliation: Challenge

Challenge: **proving** that the view reconciliation rules are sound is hard!

Challenge: **proving** that the view reconciliation rules are sound is hard!

The standard workflow in Iris:

- have Separation Logic memory assertions ( $\ell \mapsto_{\text{ML}} V$ )
- have the state of the operational semantics (finite map: Location  $\rightarrow$  Value)
- *relate the two* (“state interpretation”). *Often* straightforward...

## View Reconciliation: Challenge

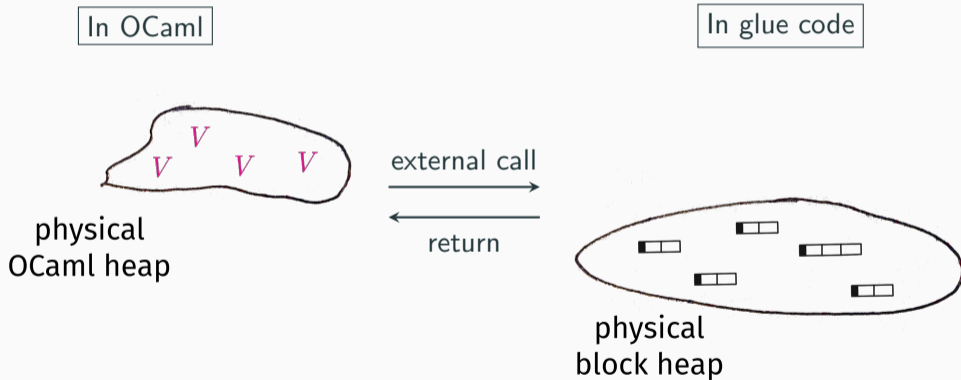
Challenge: **proving** that the view reconciliation rules are sound is hard!

- in the **program logic**, we can hold a mix of  $\ell \mapsto_{\text{ML}} \vec{V}$  and  $\gamma \mapsto_{\text{blk}[0|\text{mut}]} \vec{v}$
- the **operational semantics** has *only one simultaneous view* of the OCaml state

How can we relate the assertions and the operational semantics state?

## View Reconciliation: Challenge (2)

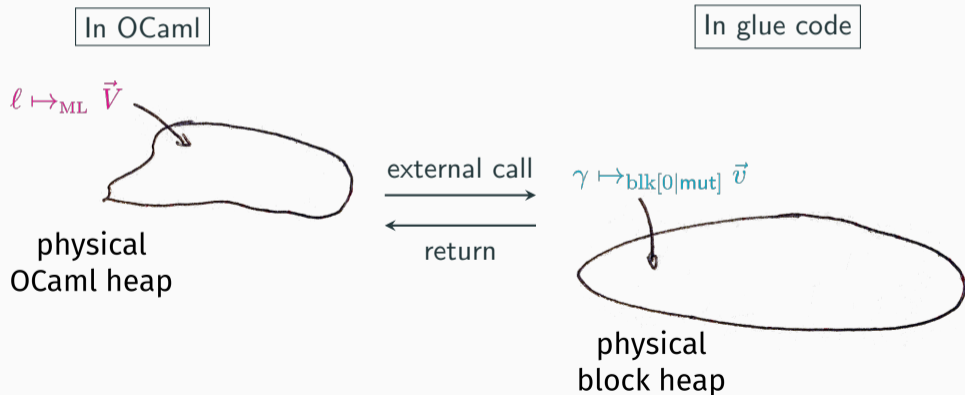
In the **operational semantics**, there is *only one simultaneous view* of the OCaml state.





## View Reconciliation: Challenge (2) and Solution

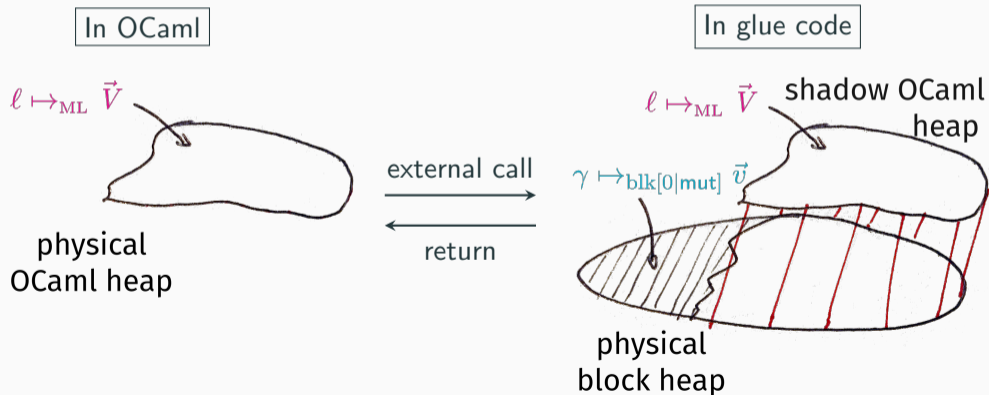
In the **program logic**: what happens to OCaml points-to?



## View Reconciliation: Challenge (2) and Solution

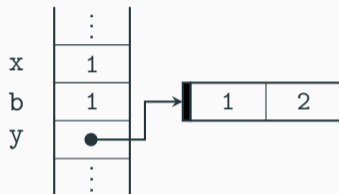
In the **program logic**: what happens to OCaml points-to?

**Solution:** track *both* views of the state in the program logic



## Changing The Representation: Making Difficult Choices

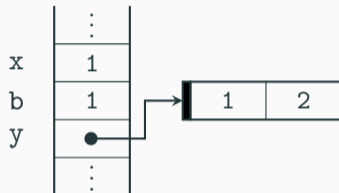
```
let x = ?  
let b = ?  
let y = (?, ?)
```



Quiz Time: What are the OCaml values of `x`, `b`, and `y`?

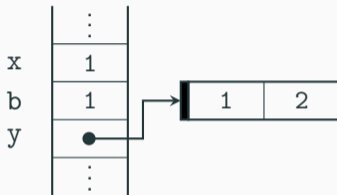
## Changing The Representation: Making Difficult Choices

```
let x = 1
let b = true
let y = (1, 2)
```



## Changing The Representation: Making Difficult Choices

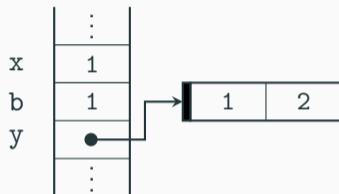
```
let x = 1
let b = true
let y = (1, 2)
```



High-level representation is **not unique!**

## Changing The Representation: Making Difficult Choices

```
let x = 1
let b = true
let y = (1, 2)
```



High-level representation is **not unique!**

How does Operational Semantics choose the right value when switching to ML values?

## Angelic Non-Determinism And The Weakest Pre

We use angelic nondeterminism, based on multi-relations (see DimSum, CCR)!

## Angelic Non-Determinism And The Weakest Pre

We use angelic nondeterminism, based on multi-relations (see DimSum, CCR)!

$\text{wp } e \{ \Phi \} : \hat{=} \dots \vee (e \text{ reducible} * \forall e'. e \rightarrow e' \multimap \text{wp } e' \{ \Phi \})$  usual Iris

$\text{wp } e \{ \Phi \} : \hat{=} \dots \vee (\exists X. e \twoheadrightarrow X * \forall e'. e' \in X \multimap \text{wp } e' \{ \Phi \})$  multi-relations

Regular C and ML, not having angelic non-determinism, retain usual SOS



## Angelic Non-Determinism And The Weakest Pre

We use angelic nondeterminism, based on multi-relations (see DimSum, CCR)!

$\text{wp } e \{ \Phi \} : \hat{=} \dots \vee (e \text{ reducible} * \forall e'. e \rightarrow e' \multimap \text{wp } e' \{ \Phi \})$  usual Iris

$\text{wp } e \{ \Phi \} : \hat{=} \dots \vee (\exists X. e \rightarrow X * \forall e'. e' \in X \multimap \text{wp } e' \{ \Phi \})$  multi-relations

Regular C and ML, not having angelic non-determinism, retain usual SOS

For adequacy, **existential** needs to be extracted  $\Rightarrow$  *transfinite Iris*

Contribution: An Iris for toy C+ML+FFI, emphasizing **language-local reasoning**.

Contribution: An Iris for toy C+ML+FFI, emphasizing **language-local reasoning**.

We give a **general recipe** for merging two languages:

1. Abstract over “the other side” using **interfaces and external calls**
2. Formalize the **semantics of the FFI** (memory model and primitives)
3. Bridge between memory models using **view reconciliation**

# Conclusion

Contribution: An Iris for toy C+ML+FFI, emphasizing **language-local reasoning**.

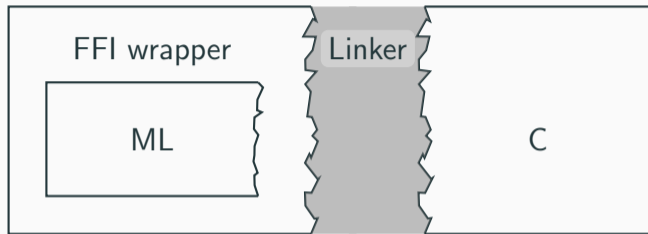
We give a **general recipe** for merging two languages:

1. Abstract over “the other side” using **interfaces and external calls**
2. Formalize the **semantics of the FFI** (memory model and primitives)
3. Bridge between memory models using **view reconciliation**

More in the paper: <https://melocoton-project.github.io>

- **more detailed FFI**: callbacks, custom blocks, GC interaction
- **logical relation** for semantic typing of external functions

bonus slides



## The FFI wrapper

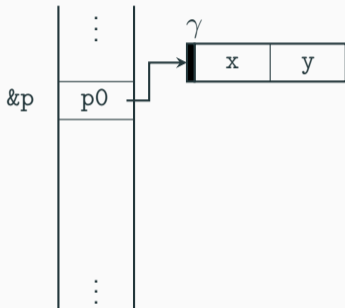
- Convert ML values to block-level
- Provide FFI: a C calling convention for ML

## The Linker

- Link programs using the same calling convention
- Resolve external calls

# Checking swap\_pair

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  CAMLreturn(r);
}
```



## Permissions:

$GC(\alpha)$

$\gamma \mapsto_{\text{blk}[0|\text{imm}]} [x; y]$

$\&p \mapsto_c p0$

## Facts:

$\text{blkaddr}(\alpha, \gamma) = p0$

A *permission* describes the right to access some resources or memory:

$\boxed{\text{GC}(\alpha)}$ : permission to use C functions of the FFI

→  $\alpha$ : an abstract name that identifies a **specific layout** of the GC memory.

( $\alpha$  changes when the GC moves or deallocates block)

$\boxed{\gamma \mapsto_{\text{blk}[0|\text{imm}]} [x; y; \dots]}$ : permission to access a block in the GC memory

→  $\gamma$ : abstract **label** of the block

→  $[x; y; \dots]$ : contents of the block

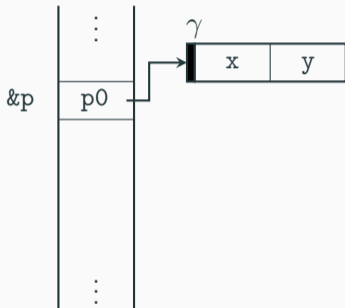
$\boxed{\&p \mapsto_{\text{C}} p0}$ : permission to access the C variable  $p$

→  $p0$ : current value of the variable



# Checking swap\_pair

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  CAMLreturn(r);
}
```



## Permissions:

$GC(\alpha)$

$\gamma \mapsto_{\text{blk}[0|\text{imm}]} [x; y]$

$\&p \mapsto_c p0$

## Facts:

$\text{blkaddr}(\alpha, \gamma) = p0$

## Checking swap\_pair

```
value caml_swap_pair(value p)
{
  CAMLpa
  value
  value
  value
  Store_
  Store_
  CAMLre
}
```

Permissions:

GC( $\alpha$ )

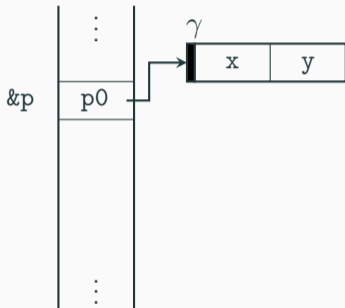
We also collect “facts” (mathematical equalities) of the form:

$$\text{blkaddr}(\alpha, \gamma) = p_0$$

means: the block with label  $\gamma$  has concrete address  $p_0$ ,  
*when the GC memory is in layout  $\alpha$*

# Checking swap\_pair

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  CAMLreturn(r);
}
```



## Permissions:

$GC(\alpha)$

$\gamma \mapsto_{\text{blk}[0|\text{imm}]} [x; y]$

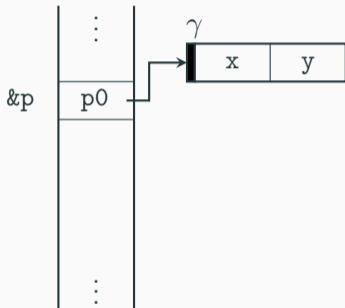
$\&p \mapsto_c p0$

## Facts:

$\text{blkaddr}(\alpha, \gamma) = p0$

# Checking swap\_pair

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  CAMLreturn(r);
}
```



## Permissions:

$GC(\alpha)$

$\gamma \mapsto_{\text{blk}[0|\text{imm}]} [x; y]$

$\&p \mapsto_c p0$

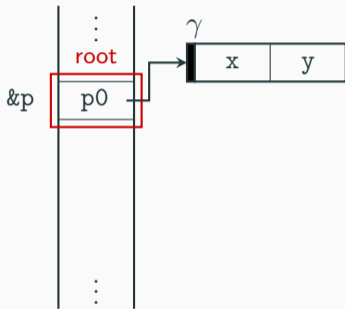
## Facts:

$\text{blkaddr}(\alpha, \gamma) = p0$

$\langle GC(\alpha) * \&p \mapsto_c p0 * \text{blkaddr}(\alpha, \gamma) = p0 \rangle$   
 $\text{CAMLparam1}(p)$   
 $\langle GC(\alpha) * \&p \mapsto_{\text{root}} \gamma \rangle$

# Checking swap\_pair

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  CAMLreturn(r);
}
```



## Permissions:

$GC(\alpha)$

$\gamma \mapsto_{\text{blk}[0|\text{imm}]} [x; y]$

$\&p \mapsto_{\text{root}} \gamma$

## Facts:

$\text{blkaddr}(\alpha, \gamma) = p0$

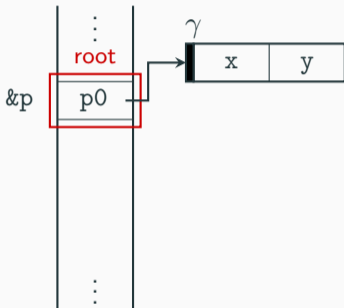
$\langle GC(\alpha) * \&p \mapsto_C p0 * \text{blkaddr}(\alpha, \gamma) = p0 \rangle$

$\text{CAMLparam1}(p)$

$\langle GC(\alpha) * \&p \mapsto_{\text{root}} \gamma \rangle$

# Checking swap\_pair

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  CAMLreturn(r);
}
```



Permissions:

$GC(\alpha)$

$\gamma \mapsto_{\text{blk}[0|\text{imm}]} [x; y]$

$\&p \mapsto_{\text{root}} \gamma$

Facts:

$\text{blkaddr}(\alpha, \gamma) = p0$

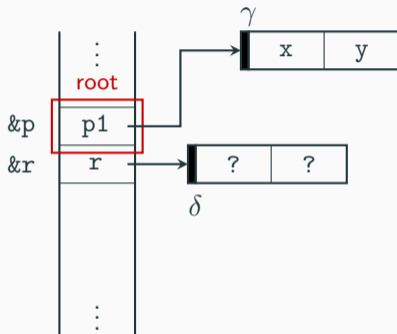
$\langle GC(\alpha) \rangle$

$\text{caml\_alloc}(0, n)$

$\langle r. \exists \beta. \text{blkaddr}(\beta, \delta) = r * GC(\beta) * \delta \mapsto_{\text{blk}[0|\text{imm}]} [?; \dots; ?] \rangle$

# Checking swap\_pair

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  CAMLreturn(r);
}
```



## Permissions:

$GC(\beta)$

$\gamma \mapsto_{\text{blk}[0|\text{imm}]} [x; y]$

$\&p \mapsto_{\text{root}} \gamma$

$\delta \mapsto_{\text{blk}[0|\text{imm}]} [?; ?]$

## Facts:

$\text{blkaddr}(\alpha, \gamma) = p_0$

$\text{blkaddr}(\beta, \delta) = r$

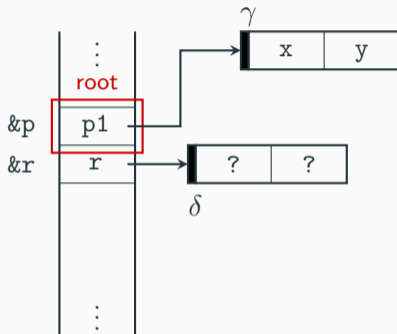
$\langle GC(\alpha) \rangle$

$\text{caml\_alloc}(0, n)$

$\langle r. \exists \beta. \text{blkaddr}(\beta, \delta) = r * GC(\beta) * \delta \mapsto_{\text{blk}[0|\text{imm}]} [?; \dots; ?] \rangle$

# Checking swap\_pair

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  CAMLreturn(r);
}
```



## Permissions:

$GC(\beta)$

$\gamma \mapsto_{\text{blk}[0|\text{imm}]} [x; y]$

$\&p \mapsto_{\text{root}} \gamma$

$\delta \mapsto_{\text{blk}[0|\text{imm}]} [?; ?]$

## Facts:

$\text{blkaddr}(\alpha, \gamma) = p0$

$\text{blkaddr}(\beta, \delta) = r$

$\text{blkaddr}(\beta, \gamma) = p1$

## Rule for reading a root $\&p$

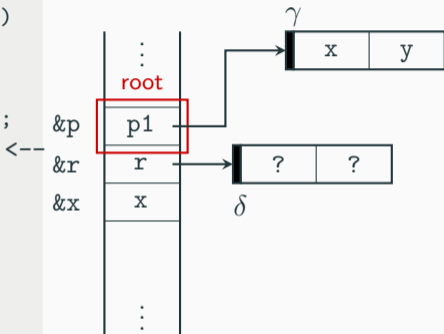
$\langle GC(\beta) * \&p \mapsto_{\text{root}} \gamma \rangle \quad *(&p)$

$\langle p1. \text{blkaddr}(\beta, \gamma) = p1 * GC(\beta), \&p \mapsto_{\text{root}} \gamma \rangle$



# Checking swap\_pair

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  CAMLreturn(r);
}
```



## Permissions:

$GC(\beta)$

$\gamma \mapsto_{\text{blk}[0|\text{imm}]} [x; y]$

$\&p \mapsto_{\text{root}} \gamma$

$\delta \mapsto_{\text{blk}[0|\text{imm}]} [?; ?]$

## Facts:

$\text{blkaddr}(\alpha, \gamma) = p0$

$\text{blkaddr}(\beta, \delta) = r$

$\text{blkaddr}(\beta, \gamma) = p1$

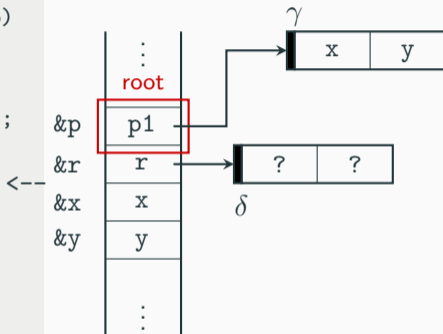
$\langle GC(\beta) * \gamma \mapsto_{\text{blk}[0|\text{imm}]} [..; v_i; ..] * \text{blkaddr}(\beta, \gamma) = p \rangle$

$\text{Field}(p, i)$

$\langle v_i. GC(\beta), \gamma \mapsto_{\text{blk}[0|\text{imm}]} [..; v_i; ..] \rangle$

# Checking swap\_pair

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  CAMLreturn(r);
}
```



## Permissions:

$GC(\beta)$

$\gamma \mapsto_{\text{blk}[0|\text{imm}]} [x; y]$

$\&p \mapsto_{\text{root}} \gamma$

$\delta \mapsto_{\text{blk}[0|\text{imm}]} [?; ?]$

## Facts:

$\text{blkaddr}(\alpha, \gamma) = p0$

$\text{blkaddr}(\beta, \delta) = r$

$\text{blkaddr}(\beta, \gamma) = p1$

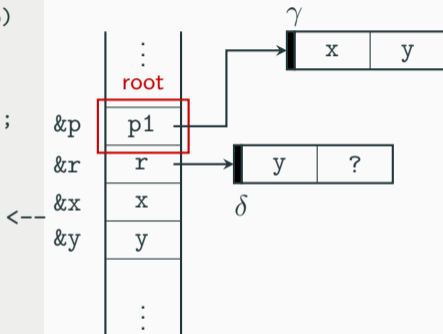
$\langle GC(\beta) * \gamma \mapsto_{\text{blk}[0|\text{imm}]} [..; v_i; ..] * \text{blkaddr}(\beta, \gamma) = p \rangle$

$\text{Field}(p, i)$

$\langle v_i. GC(\beta), \gamma \mapsto_{\text{blk}[0|\text{imm}]} [..; v_i; ..] \rangle$

# Checking swap\_pair

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  CAMLreturn(r);
}
```



## Permissions:

$GC(\beta)$

$\gamma \mapsto_{\text{blk}[0|\text{imm}]} [x; y]$

$\&p \mapsto_{\text{root}} \gamma$

$\delta \mapsto_{\text{blk}[0|\text{imm}]} [y; ?]$

## Facts:

$\text{blkaddr}(\alpha, \gamma) = p0$

$\text{blkaddr}(\beta, \delta) = r$

$\text{blkaddr}(\beta, \gamma) = p1$

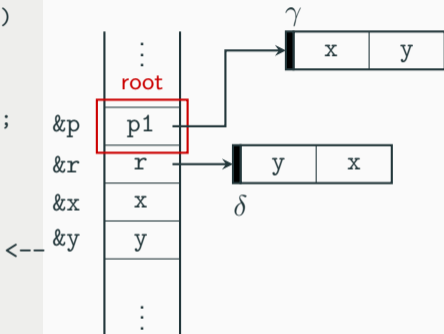
$\langle GC(\beta) * \delta \mapsto_{\text{blk}[0|\text{imm}]} [..; v_i; ..] * \text{blkaddr}(\beta, \delta) = r \rangle$

$\text{Store\_field}(r, i, v)$

$\langle GC(\beta) * \delta \mapsto_{\text{blk}[0|\text{imm}]} [..; v; ..] \rangle$

# Checking swap\_pair

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  CAMLreturn(r);
}
```



## Permissions:

$GC(\beta)$

$\gamma \mapsto_{\text{blk}[0|\text{imm}]} [x; y]$

$\&p \mapsto_{\text{root}} \gamma$

$\delta \mapsto_{\text{blk}[0|\text{imm}]} [y; x]$

## Facts:

$\text{blkaddr}(\alpha, \gamma) = p0$

$\text{blkaddr}(\beta, \delta) = r$

$\text{blkaddr}(\beta, \gamma) = p1$

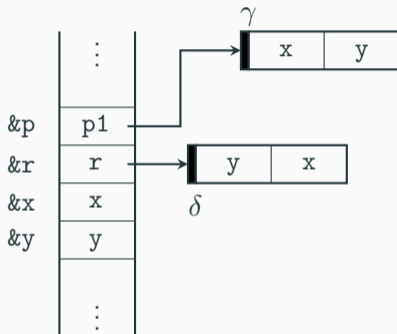
$\langle GC(\beta) * \delta \mapsto_{\text{blk}[0|\text{imm}]} [..; v_i; ..] * \text{blkaddr}(\beta, \delta) = r \rangle$

$\text{Store\_field}(r, i, v)$

$\langle GC(\beta) * \delta \mapsto_{\text{blk}[0|\text{imm}]} [..; v; ..] \rangle$

# Checking swap\_pair

```
value caml_swap_pair(value p)
{
  CAMLparam1(p);
  value r = caml_alloc(0, 2);
  value x = Field(p, 0);
  value y = Field(p, 1);
  Store_field(r, 0, y);
  Store_field(r, 1, x);
  CAMLreturn(r);
}
```



## Permissions:

$GC(\beta)$

$\gamma \mapsto_{\text{blk}[0|\text{imm}]} [x; y]$

$\&p \mapsto_C p1$

$\delta \mapsto_{\text{blk}[0|\text{imm}]} [y; x]$

## Facts:

$\text{blkaddr}(\alpha, \gamma) = p0$

$\text{blkaddr}(\beta, \delta) = r$

$\text{blkaddr}(\beta, \gamma) = p1$

$\langle GC(\beta) * \&p \mapsto_{\text{root}} \gamma \rangle$

$\text{CAMLreturn}(r)$

$\langle GC(\beta) * \&p \mapsto_C \text{blkaddr}(\beta, \gamma) \rangle$