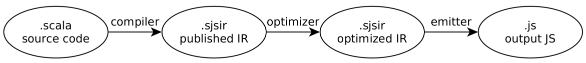
Foundations of Software
Fall 2023

Week 14

Sébastien Doeraene

Elements of the Scala.js IR type system

Scala.js compilation pipeline



Why formally study an IR

Why formally study an IR

- Optimizations may only be applicable if the type tystem is sound
- Prove that certain optimizations are correct
- Prove that the translation from source and to the target language are correct
- etc.

Mixing primitives and objects

Motivation

Featherweight Java only has objects. How do we model primitives, for example, `int` and `bool`?

Motivation

Featherweight Java only has objects. How do we model primitives, for example, `int` and `bool`?

Moreover, in Scala, primitive types are "object-like". We can use them in arbitrary type parameters, and they should behave like objects.
On the JVM, this is implemented with *boxing*. In Scala.js, however, boxing would be detrimental to *interoperability* with JavaScript.
How do we make primitives object-like without boxing?

Motivation

Featherweight Java only has objects. How do we model primitives, for example, `int` and `bool`?

Moreover, in Scala, primitive types are "object-like". We can use them in arbitrary type parameters, and they should behave like objects.
On the JVM, this is implemented with *boxing*. In Scala.js, however, boxing would be detrimental to *interoperability* with JavaScript.
How do we make primitives object-like without boxing?

Idea: make primitive types *subtypes* of their "representative classes".

Types and subtyping

`T` ::=		*types*
	`C`	*class*
	`int`	*primitive int*
	`bool`	*primitive bool*

$$\frac{CT(\texttt{C}) = \texttt{class C extends D \{...\}}}{\texttt{C} <: \texttt{D}}$$

$$\texttt{T} <: \texttt{T}$$

$$\frac{\texttt{S} <: \texttt{W} \qquad \texttt{W} <: \texttt{T}}{\texttt{S} <: \texttt{T}}$$

$$\texttt{int} <: \texttt{Integer} \qquad\qquad \texttt{bool} <: \texttt{Boolean}$$

Representative classes

$$\mathit{tpcls}(\texttt{C}) = \texttt{C}$$

$$\mathit{tpcls}(\texttt{int}) = \texttt{Integer}$$

$$\mathit{tpcls}(\texttt{bool}) = \texttt{Boolean}$$

$$\texttt{T} <: \mathit{tpcls}(\texttt{T})$$

Syntax (terms)

`t` ::=		*terms*
	`x`	*variable*
	`t.f`	*field access*
	$\texttt{t.m(}\overline{\texttt{t}}\texttt{)}$	*method invocation*
	$\texttt{new C(}\overline{\texttt{t}}\texttt{)}$	*object creation*
	`(T) t`	*cast*
	`false`	
	`true`	
	`if t then t else t`	
	`0`	
	`succ t`	
	`pred t`	
	`iszero t`	

Syntax (values)

$\texttt{v}$ ::=		*values*	
	`new C(`$\overline{\texttt{v}}$`)`		*object creation*
	`nv`		*numeric value*
	`bv`		*boolean value*
`nv` ::=		*numeric values*	
	`0`		*zero*
	`succ nv`		*non-zero*
`bv` ::=		*boolean values*	
	`false`		*false*
	`true`		*true*

Typing rules: method calls

Adapting from Featherweight Java:

$$\frac{\Gamma \vdash \texttt{t}_0 : \texttt{C}_0 \quad mtype(\texttt{m}, \texttt{C}_0) = \overline{\texttt{S}} \to \texttt{T} \quad \Gamma \vdash \overline{\texttt{t}} : \overline{\texttt{S}_1} \quad \overline{\texttt{S}_1} <: \overline{\texttt{S}}}{\Gamma \vdash \texttt{t}_0\texttt{.m(}\overline{\texttt{t}}\texttt{)} : \texttt{T}} \quad \text{(T-INVK)}$$

What if $\texttt{t}_0$ is a primitive?

Typing rules: method calls

Adapting from Featherweight Java:

$$\frac{\Gamma \vdash \texttt{t}_0 : \texttt{C}_0 \quad mtype(\texttt{m}, \texttt{C}_0) = \overline{\texttt{S}} \to \texttt{T} \quad \Gamma \vdash \overline{\texttt{t}} : \overline{\texttt{S}_1} \quad \overline{\texttt{S}_1} <: \overline{\texttt{S}}}{\Gamma \vdash \texttt{t}_0\texttt{.m(}\overline{\texttt{t}}\texttt{)} : \texttt{T}} \quad \text{(T-INVK)}$$

What if $\texttt{t}_0$ is a primitive?

$$\frac{\Gamma \vdash \texttt{t}_0 : \texttt{T}_0 \quad mtype(\texttt{m}, tpcls(\texttt{T}_0)) = \overline{\texttt{S}} \to \texttt{T} \quad \Gamma \vdash \overline{\texttt{t}} : \overline{\texttt{S}_1} \quad \overline{\texttt{S}_1} <: \overline{\texttt{S}}}{\Gamma \vdash \texttt{t}_0\texttt{.m(}\overline{\texttt{t}}\texttt{)} : \texttt{T}} \quad \text{(T-INVK)}$$

If $\Gamma \vdash \texttt{x} : \texttt{int}$, the call `x.m(...)` is typed by looking up `m` in `Integer`.

Example

```
class Boolean extends Object { Boolean() { super(); } }
class Integer extends Object {
  Integer() { super(); }
  int plus(int that) {
    return if (iszero that) then ((int) this)
           else (succ this.plus(pred that)); }
}
class Pair extends Object {
  Object fst;
  Object snd;
  Pair(Object fst, Object snd) {
    super(); this.fst=fst; this.snd=snd; }
  int sum() {
    return ((int) this.fst).plus((int) this.snd); }
}

new Pair(5, 11).sum()
```

Typing rules: fields

Adapting from Featherweight Java:

$$\frac{\Gamma \vdash \mathtt{t}_0 : \mathtt{C}_0 \qquad \mathit{fields}(\mathtt{C}_0) = \overline{\mathtt{T}}\ \overline{\mathtt{f}}}{\Gamma \vdash \mathtt{t}_0.\mathtt{f}_i : \mathtt{T}_i} \quad \text{(T-Field)}$$

What if $\mathtt{t}_0$ is a primitive?

Typing rules: fields

Adapting from Featherweight Java:

$$\frac{\Gamma \vdash \mathtt{t}_0 : \mathtt{C}_0 \qquad \mathit{fields}(\mathtt{C}_0) = \overline{\mathtt{T}}\ \overline{\mathtt{f}}}{\Gamma \vdash \mathtt{t}_0.\mathtt{f}_i : \mathtt{T}_i} \quad \text{(T-Field)}$$

What if $\mathtt{t}_0$ is a primitive?

We can't have that!

Typing rules: fields

Adapting from Featherweight Java:

$$\frac{\Gamma \vdash \mathtt{t}_0 : \mathtt{C}_0 \qquad \mathit{fields}(\mathtt{C}_0) = \overline{\mathtt{T}}\ \overline{\mathtt{f}}}{\Gamma \vdash \mathtt{t}_0.\mathtt{f}_i : \mathtt{T}_i} \quad \text{(T-Field)}$$

What if $\mathtt{t}_0$ is a primitive?

We can't have that!

Add additional well-formedness conditions for representative classes:

$$\frac{\mathit{fields}(\mathtt{Integer}) = \emptyset \qquad \mathit{fields}(\mathtt{Boolean}) = \emptyset}{\mathtt{repr\ classes\ OK}}$$

Typing rules: casts

Straightforward generalization to all types.

$$\frac{\Gamma \vdash \mathtt{t}_0 : \mathtt{S} \qquad \mathtt{S} <: \mathtt{T}}{\Gamma \vdash (\mathtt{T})\mathtt{t}_0 : \mathtt{T}} \quad \text{(T-UCast)}$$

$$\frac{\Gamma \vdash \mathtt{t}_0 : \mathtt{S} \qquad \mathtt{T} <: \mathtt{S} \qquad \mathtt{T} \neq \mathtt{S}}{\Gamma \vdash (\mathtt{T})\mathtt{t}_0 : \mathtt{T}} \quad \text{(T-DCast)}$$

$$\frac{\Gamma \vdash \mathtt{t}_0 : \mathtt{S} \qquad \mathtt{T} \not<: \mathtt{S} \qquad \mathtt{S} \not<: \mathtt{T} \qquad \textit{stupid warning}}{\Gamma \vdash (\mathtt{T})\mathtt{t}_0 : \mathtt{T}} \quad \text{(T-SCast)}$$

Typing rules: casts

Since it is an Intermediate Representation, warnings are not relevant anymore. Therefore, we keep only one typing rule for casts.

$$\frac{\Gamma \vdash \mathtt{t}_0 : \mathtt{S}}{\Gamma \vdash \mathtt{(T)t}_0 : \mathtt{T}} \qquad \text{(T-CAST)}$$

Typing rules: casts

Since it is an Intermediate Representation, warnings are not relevant anymore. Therefore, we keep only one typing rule for casts.

$$\frac{\Gamma \vdash \mathtt{t}_0 : \mathtt{S}}{\Gamma \vdash \mathtt{(T)t}_0 : \mathtt{T}} \qquad \text{(T-CAST)}$$

Question: can we remove the premise of that rule?

Evaluation rules

$$\frac{\mathit{fields}(\mathtt{C}) = \overline{\mathtt{T}}\ \overline{\mathtt{f}}}{(\mathtt{new\ C}(\overline{\mathtt{v}})).\mathtt{f}_i \longrightarrow \mathtt{v}_i} \qquad \text{(E-PROJNEW)}$$

$$\frac{\mathit{mbody}(\mathtt{m}, \mathit{tpcls}(\mathit{vtpe}(\mathtt{v}))) = (\overline{\mathtt{x}}, \mathtt{t}_0)}{\mathtt{v.m}(\overline{\mathtt{u}}) \longrightarrow [\overline{\mathtt{x}} \mapsto \overline{\mathtt{u}}, \mathtt{this} \mapsto \mathtt{v}]\mathtt{t}_0} \qquad \text{(E-INVKVAL)}$$

$$\frac{\mathit{vtpe}(\mathtt{v}) <: \mathtt{T}}{\mathtt{(T)v} \longrightarrow \mathtt{v}} \qquad \text{(E-CASTVAL)}$$

$\mathit{vtpe}(\mathtt{new\ C}(\overline{\mathtt{v}})) = \mathtt{C}$ $\qquad \mathit{vtpe}(\mathtt{nv}) = \mathtt{int}$ $\qquad \mathit{vtpe}(\mathtt{bv}) = \mathtt{bool}$

plus congruence rules and rules for `if`, `pred`, `succ` and `iszero` (omitted)

Labeled blocks

Presentation

In JavaScript, we have labeled *statements* with `break`s:

```
label: {
  ...
  if (x)
    break label;
  ...
}
```

If execution reaches `break label`, it jumps to *after* the block.

Presentation

We generalize the concept to *expressions*. A `return` to a label jumps out of the block, resulting in the specified value as the value of the block.

```
val y: T = label[T]: {
  ...
  if (x)
    return@label someT;
  ...
  someOtherT
}
```


Use cases: modeling `return`

Traditional `return`:

```
def foo(x: int): int = {
  if (x < 0) {
    return -x
  }
  x
}
```

Modeled as:

```
def foo(x: int): int = {
  ret[int]: {
    if (x < 0) {
      return@ret -x
    }
    x
  }
}
```


Use cases: modeling `break` and `continue`

Traditional `break` and `continue`:

```
def foo(x: int): unit = {
  var i: int = x
  while (i > 0) {
    if (i % 3 == 0)
      continue
    if (i % 10 == 0)
      break
    println(i)
    i = i + 1
  }
}
```

Use cases: modeling `break` and `continue`

Modeled as two nested labels:

```
def foo(x: int): unit = {
  var i: int = x
  breakLoop[unit]: {
    while (i > 0) {
      continueLoop[unit]: {
        if (i % 3 == 0)
          return@continueLoop unit
        if (i % 10 == 0)
          return@breakLoop unit
        println(i)
        i = i + 1
      }
    }
  }
}
```


Use cases: encoding of tail recursion

Tail recursive function in source code:

```
def fact(n: int, acc: int): int = {
  if (n == 0) acc
  else fact(n - 1, n * acc)
}
```


Use cases: encoding of tail recursion

Encoding with a creative use of labeled blocks:

```
def fact(var n: int, var acc: int): int = {
  ret[int]: {
    while (true) {
      tailcall[unit]: {
        return@ret {
          if (n == 0) acc
          else {
            val n' = n - 1
            val acc' = n * acc
            n = n'
            acc = acc'
            return@tailcall unit
          }
        }
      }
    }
  }
```


Use cases: encoding of tail recursion

Encoding with a creative use of labeled blocks and `loop`:

```
def fact(var n: int, var acc: int): int = {
  ret[int]: {
    loop {
      tailcall[unit]: {
        return@ret {
          if (n == 0) acc
          else {
            val n' = n - 1
            val acc' = n * acc
            n = n'
            acc = acc'
            return@tailcall unit
          }
        }
      }
    }
  }
```

Formalization

On the board

Typing rules

$$\frac{\Gamma \mid \Delta \mid \Sigma \vdash \mathtt{t}_1 : \mathtt{T}_1}{\Gamma \mid \Delta \mid \Sigma \vdash \mathtt{loop}\ \mathtt{t}_1 : \mathtt{nothing}} \quad \text{(T-Loop)}$$

$$\frac{\Gamma \mid \Delta, \alpha : \mathtt{T} \mid \Sigma \vdash \mathtt{t} : \mathtt{T}}{\Gamma \mid \Delta \mid \Sigma \vdash \alpha\,\mathtt{[T]}\ \mathtt{\{t\}} : \mathtt{T}} \quad \text{(T-Labeled)}$$

$$\frac{\alpha : \mathtt{T}_1 \in \Delta \qquad \Gamma \mid \Delta \mid \Sigma \vdash \mathtt{t}_1 : \mathtt{T}_1}{\Gamma \mid \Delta \mid \Sigma \vdash \mathtt{return@}\alpha\ \mathtt{t}_1 : \mathtt{nothing}} \quad \text{(T-Return)}$$

Evaluation rules

$$\mathtt{loop}\ \mathtt{t}_1 \mid \mu \longrightarrow \mathtt{t}_1\mathtt{;}\ \mathtt{loop}\ \mathtt{t}_1 \mid \mu \quad \text{(E-Loop)}$$

$$\alpha\,\mathtt{[T]}\ \mathtt{\{v_1\}} \mid \mu \longrightarrow \mathtt{v}_1 \mid \mu \quad \text{(E-LabeledValue)}$$

$$\alpha\,\mathtt{[T]}\ \mathtt{\{return@}\alpha\ \mathtt{v_1\}} \mid \mu \longrightarrow \mathtt{v}_1 \mid \mu \quad \text{(E-LabeledRetMatch)}$$

$$\frac{\beta \neq \alpha}{\alpha\,\mathtt{[T]}\ \mathtt{\{return@}\beta\ \mathtt{v_1\}} \mid \mu \longrightarrow \mathtt{return@}\beta\ \mathtt{v}_1 \mid \mu} \quad \text{(E-LabeledRetDiff)}$$

Plus congruence rules and propagation rules for `return`, for example:

$$\mathtt{(return@}\alpha\ \mathtt{v_1)}\ \mathtt{t}_2 \mid \mu \longrightarrow \mathtt{return@}\alpha\ \mathtt{v}_1 \mid \mu \quad \text{(E-AppRet1)}$$

$$\mathtt{(return@}\alpha\ \mathtt{v_1);}\ \mathtt{t}_2 \mid \mu \longrightarrow \mathtt{return@}\alpha\ \mathtt{v}_1 \mid \mu \quad \text{(E-SeqRet)}$$

Proofs

On the board

Typing rules, fixed

$$\frac{\Gamma \mid \Delta \mid \Sigma \vdash \mathtt{t_1} : \mathtt{T_1}}{\Gamma \mid \Delta \mid \Sigma \vdash \mathtt{loop\ t_1} : \mathtt{nothing}} \quad \text{(T-Loop)}$$

$$\frac{\Gamma \mid \Delta, \alpha : \mathtt{T} \mid \Sigma \vdash \mathtt{t} : \mathtt{T}}{\Gamma \mid \Delta \mid \Sigma \vdash \alpha\mathtt{[T]\ \{t\}} : \mathtt{T}} \quad \text{(T-Labeled)}$$

$$\frac{\alpha{:}\mathtt{T_1} \in \Delta \qquad \Gamma \mid \Delta \mid \Sigma \vdash \mathtt{t_1} : \mathtt{T_1}}{\Gamma \mid \Delta \mid \Sigma \vdash \mathtt{return@}\alpha\ \mathtt{t_1} : \mathtt{nothing}} \quad \text{(T-Return)}$$

$$\frac{\Gamma, \mathtt{x}{:}\mathtt{T_1} \mid \emptyset \mid \Sigma \vdash \mathtt{t_2} : \mathtt{T_2}}{\Gamma \mid \Delta \mid \Sigma \vdash \lambda\mathtt{x}{:}\mathtt{T_1}.\mathtt{t_2} : \mathtt{T_1}{\to}\mathtt{T_2}} \quad \text{(T-Abs)}$$