香 港 大 學
THE UNIVERSITY OF HONG KONG
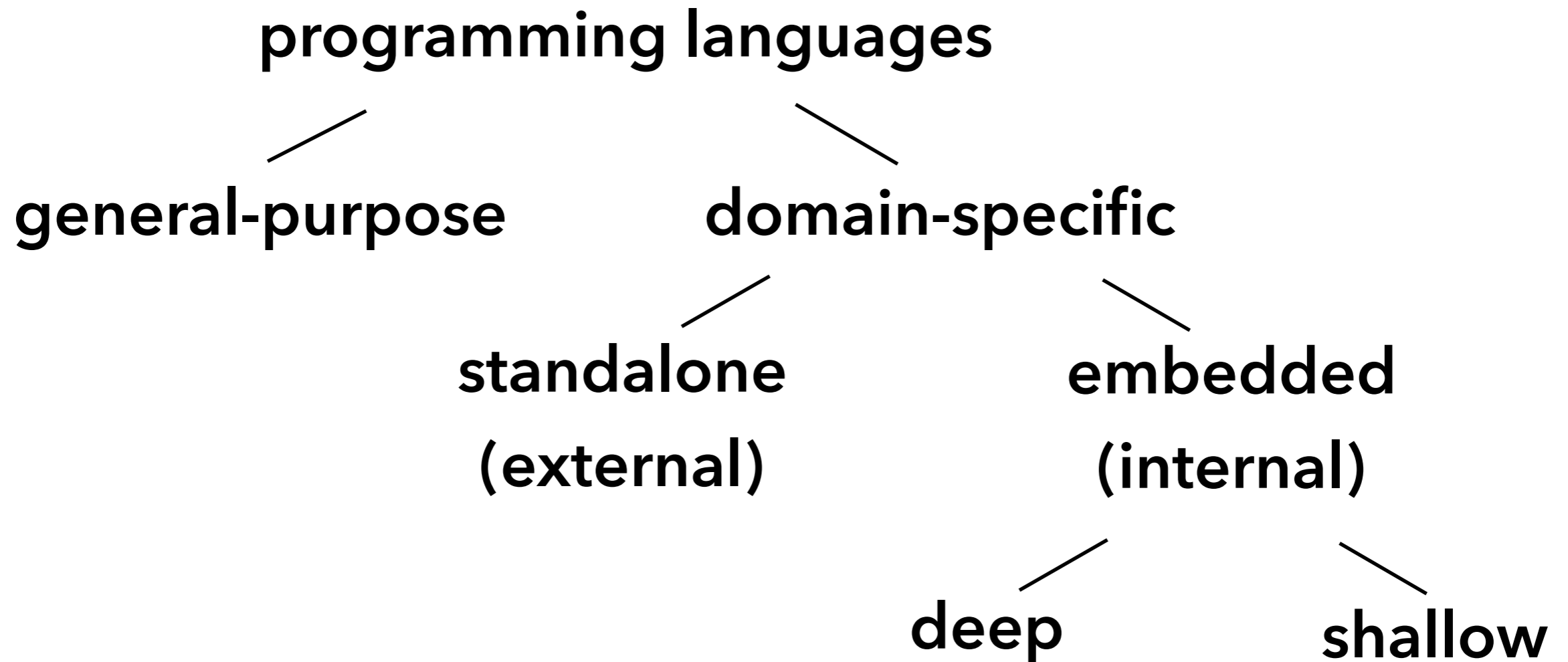
Shallow EDSLs and Object-Oriented Programming: Beyond Simple Compositionality

Weixin Zhang and Bruno C. d. S. Oliveira

<Programming> 2019

April 3, 2019

Background

programming languages

general-purpose   domain-specific

standalone   embedded

(external)   (internal)

deep   shallow

Shallow vs. deep embeddings

▶ Shallow embeddings

- ▶ Semantics first

- ▶ Compositional

- ▶ No AST

- ▶ Easy to add new language constructs

- ▶ Hard to add new interpretations

▶ Deep embeddings

- ▶ Syntax first

- ▶ Non-compositional

- ▶ Have an AST

- ▶ Easy to add new interpretations

- ▶ Hard to add new language constructs

Contribution

▸ Shallow embeddings and OOP are closely related

  ▸ Both essence is **procedural abstraction** [Reynolds,1978]

Gibbons & Wu, 2015   Cook, 2009

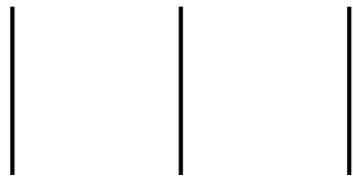| shallow embeddings | ⟷ | procedural abstraction | ⟷ | OOP |

▸ OOP mechanisms, **subtyping**, **inheritance** and **type-refinement** increase the modularity of shallow EDSLs
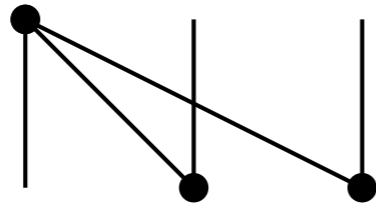
  ▸ Enable multiple (possibly dependent) interpretations

SCANS: a DSL for parallel prefix circuits

▶ Grammar: ⟨*circuit*⟩ ::= 'id' ⟨*positive-number*⟩
| 'fan' ⟨*positive-number*⟩
| ⟨*circuit*⟩ 'beside' ⟨*circuit*⟩
| ⟨*circuit*⟩ 'above' ⟨*circuit*⟩
| 'stretch' ⟨*positive-numbers*⟩ ⟨*circuit*⟩
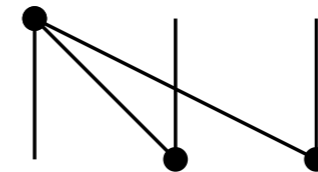| '(' ⟨*circuit*⟩ ')'

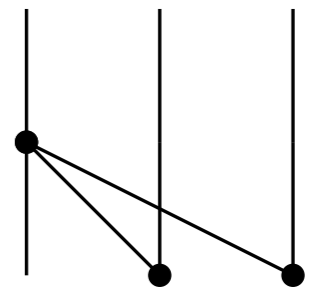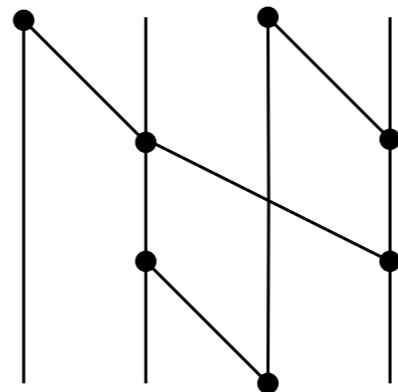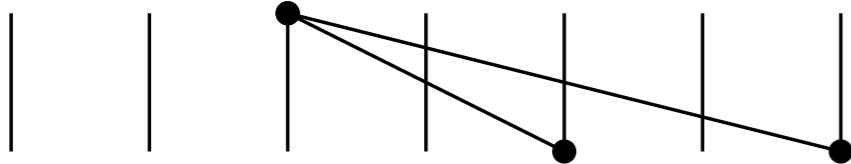id 3   fan 3   id 3 beside fan 3   id 3 above fan 3

stretch 3 2 3 fan 3

(fan 2 beside fan 2)
 above
(stretch 2 2 fan 2)
 above
(id 1 beside fan 2 beside id 1)

Embedding SCANS in Haskell

▸ A shallow implementation should conform to the following signatures

$$\textbf{type } Circuit = \dots$$

semantic domain

$$
\begin{aligned}
id &:: Int \rightarrow Circuit \\
fan &:: Int \rightarrow Circuit \\
beside &:: Circuit \rightarrow Circuit \rightarrow Circuit \\
above &:: Circuit \rightarrow Circuit \rightarrow Circuit \\
stretch &:: [Int] \rightarrow Circuit \rightarrow Circuit
\end{aligned}
$$

procedural abstraction

▸ E.g. an interpretation calculating the **width**

$$
\begin{aligned}
\textbf{type } Circuit &= Int \\
id\ n &= n \\
fan\ n &= n \\
beside\ c_1\ c_2 &= c_1 + c_2 \\
above\ c_1\ c_2 &= c_1 \\
stretch\ ns\ c &= sum\ ns
\end{aligned}
$$

$>$ ((fan 2 'beside' fan 2) 'above'
| stretch [2, 2] (fan 2) 'above'
| (id 1 'beside' fan 2 'beside' id 1))
4

Towards OOP

▸ An *isomorphic* encoding of **width**

$$\textbf{type } Circuit = Int$$

$$id\ n \qquad\qquad = n$$

$$fan\ n \qquad\qquad = n$$

$$beside\ c_1\ c_2 \quad = c_1 + c_2$$

$$above\ c_1\ c_2 \quad = c_1$$

$$stretch\ ns\ c \quad = sum\ ns$$

$$\textbf{newtype } Circuit_1 = Circuit_1\ \{width_1 :: Int\}$$

$$id_1\ n \qquad\qquad\qquad = Circuit_1\ \{width_1 = n\}$$

$$fan_1\ n \qquad\qquad\qquad = Circuit_1\ \{width_1 = n\}$$

$$beside_1\ c_1\ c_2 \qquad\quad = Circuit_1\ \{width_1 = width_1\ c_1 + width_1\ c_2\}$$

$$above_1\ c_1\ c_2 \qquad\quad = Circuit_1\ \{width_1 = width_1\ c_1\}$$

$$stretch_1\ ns\ c \qquad\quad = Circuit_1\ \{width_1 = sum\ ns\}$$

Embedding SCANS in OOP

▸ It is easy to port the definition into an OOP language like Scala

```
// object interface
trait Circuit₁ {def width : Int}
// concrete implementations
```

```
class Id₁ (n : Int) extends Circuit₁ {
    def width = n
}
```

```
trait Fan₁ extends Circuit₁ {
    val n : Int
    def width = n
}
```

```
trait Beside₁ extends Circuit₁ {
    val c₁, c₂ : Circuit₁
    def width = c₁.width + c₂.width
}
```

```
trait Above₁ extends Circuit₁ {
    val c₁, c₂ : Circuit₁
    def width = c₁.width
}
```

```
trait Stretch₁ extends Circuit₁ {
    val ns : List[Int]; val c : Circuit₁
    def width = ns.sum
}
```

Smart constructors

▸ Smart constructors are needed for building a circuit object conveniently

$$
\begin{aligned}
&\textbf{def } id(x:Int) &&= \textbf{new } Id_1 &&\{\textbf{val } n = x\} \\
&\textbf{def } fan(x:Int) &&= \textbf{new } Fan_1 &&\{\textbf{val } n = x\} \\
&\textbf{def } beside(x:Circuit_1, y:Circuit_1) &&= \textbf{new } Beside_1 &&\{\textbf{val } c_1 = x; \textbf{val } c_2 = y\} \\
&\textbf{def } above(x:Circuit_1, y:Circuit_1) &&= \textbf{new } Above_1 &&\{\textbf{val } c_1 = x; \textbf{val } c_2 = y\} \\
&\textbf{def } stretch(x:Circuit_1, xs:Int*) &&= \textbf{new } Stretch_1 &&\{\textbf{val } ns = xs.toList; \textbf{val } c = x\}
\end{aligned}
$$

▸ Constructing the example circuit again

$$
\begin{aligned}
\textbf{val } circuit = above(&beside(fan(2), fan(2)), \\
&above(stretch(fan(2), 2, 2), \\
&\quad beside(beside(id(1), fan(2)), id(1))))
\end{aligned}
$$

$$> circuit.width$$
4

Multiple interpretations in Haskell

▸ Often claimed as a limitation of shallow embedding

▸ Typical workaround is to use tuples

    ▸ e.g. additionally supporting **depth** for SCANS

```
type Circuit₂ = (Int, Int)
id₂ n          = (n, 0)
fan₂ n         = (n, 1)
above₂ c₁ c₂   = (width c₁, depth c₁ + depth c₂)
beside₂ c₁ c₂  = (width c₁ + width c₂, depth c₁ `max` depth c₂)
stretch₂ ns c  = (sum ns, depth c)

width = fst
depth = snd
```

▸ However, this implementation is *non-modular*

Multiple interpretations in Scala

▸ Multiple interpretations can be modular with Scala

Subtyping

trait $Circuit_2$ **extends** $Circuit_1$ {**def** $depth : Int$} // extended semantic domain
trait Id_2 **extends** Id_1 **with** $Circuit_2$ {**def** $depth = 0$}
trait Fan_2 **extends** Fan_1 **with** $Circuit_2$ {**def** $depth = 1$}
trait $Above_2$ **extends** $Above_1$ **with** $Circuit_2$ {
    override val $c_1, c_2 : Circuit_2$ // type-refinement that allows depth invocations
    def $depth = c_1.depth + c_2.depth$
}

Type-refinement

Inheritance

trait $Beside_2$ **extends** $Beside_1$ **with** $Circuit_2$ {
    override val $c_1, c_2 : Circuit_2$ // type-refinement that allows depth invocations
    def $depth = Math.max\,(c_1.depth, c_2.depth)$
}
trait $Stretch_2$ **extends** $Stretch_1$ **with** $Circuit_2$ {
    override val $c : Circuit_2$ // type-refinement that allows depth invocations
    def $depth = c.depth$
}

Dependent interpretations in Haskell

▸ An interpretation depends not *only* on itself but also on *other* interpretations

  ▸ E.g. **wellSized**, which depends on **width**

$$\textbf{type } Circuit_3 = (Int, Bool)$$
$$id_3\ n \qquad\qquad = (n, True)$$
$$fan_3\ n \qquad\qquad = (n, True)$$
$$above_3\ c_1\ c_2 \quad = (width\ c_1, wellSized\ c_1 \wedge wellSized\ c_2 \wedge \boxed{width\ c_1 \equiv width\ c_2})$$
$$beside_3\ c_1\ c_2 \quad = (width\ c_1 + width\ c_2, wellSized\ c_1 \wedge wellSized\ c_2)$$
$$stretch_3\ ns\ c \quad = (sum\ ns, wellSized\ c \wedge length\ ns \equiv \boxed{width\ c})$$
$$wellSized = snd$$

Dependent interpretations in Scala

▸ Again, modular dependent interpretations are unproblematic in Scala

trait $Circuit_3$ **extends** $Circuit_1$ {**def** $wellSized : Boolean$} // extended semantic domain
trait Id_3 **extends** Id_1 **with** $Circuit_3$ {**def** $wellSized = true$}
trait Fan_3 **extends** Fan_1 **with** $Circuit_3$ {**def** $wellSized = true$}
trait $Above_3$ **extends** $Above_1$ **with** $Circuit_3$ {
    override val $c_1, c_2 : Circuit_3$
    def $wellSized =$
       $c_1.wellSized \wedge c_2.wellSized \wedge \boxed{c_1.width \equiv c_2.width}$ // width dependency
}
trait $Beside_3$ **extends** $Beside_1$ **with** $Circuit_3$ {
    override val $c_1, c_2 : Circuit_3$
    def $wellSized = c_1.wellSized \wedge c_2.wellSized$
}
trait $Stretch_3$ **extends** $Stretch_1$ **with** $Circuit_3$ {
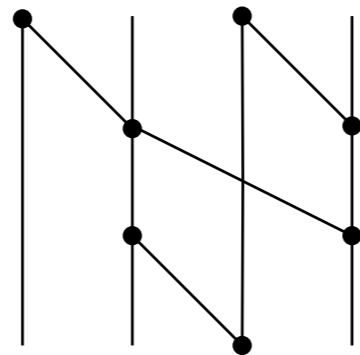   **override val** $c : Circuit_3$
    def $wellSized = c.wellSized \wedge \boxed{ns.length \equiv c.width}$ // width dependency
}

Context–sensitive interpretations in Haskell

▸ An interpretation relies on some context

▸ e.g. **layout**

$$[[(0,1),(2,3)],[(1,3)],[(1,2)]]$$

type $Circuit_4 = (Int, \boxed{(Int \to Int)} \to [[(Int,Int)]])$

$id_4\ n\qquad = (n, \lambda f \to [\,])$

$fan_4\ n\qquad = (n, \lambda f \to [[(f\ 0, f\ j)\,|\,j \leftarrow [1..n-1]]])$

$above_4\ c_1\ c_2\ = (width\ c_1, \lambda f \to layout\ c_1\ f \,\text{++}\, layout\ c_2\ f)$

$beside_4\ c_1\ c_2\ = (width\ c_1 + width\ c_2,$
$\qquad\qquad\qquad \lambda f \to lzw\ (\text{++})\ (layout\ c_1\ f)\ (layout\ c_2\ (f \circ (width\ c_1+))))$

$stretch_4\ ns\ c\ = (sum\ ns, \lambda f \to layout\ c\ (f \circ pred \circ (scanl1\ (+)\ ns!!)))$

accumulating parameter

$layout = snd$

Context-sensitive interpretations in Scala

trait $Circuit_4$ **extends** $Circuit_1$ { **def** $layout$ $\boxed{(f : Int \Rightarrow Int)}$: $List[List[(Int, Int)]]$ }
trait Id_4 **extends** Id_1 **with** $Circuit_4$ { **def** $layout(f : Int \Rightarrow Int) = List()$ }
trait Fan_4 **extends** Fan_1 **with** $Circuit_4$ {
   def $layout(f : Int \Rightarrow Int) = List(\textbf{for}(i \leftarrow List.range(1, n)) \textbf{ yield }(f(0), f(i)))$
}
trait $Above_4$ **extends** $Above_1$ **with** $Circuit_4$ {
   override val $c_1, c_2 : Circuit_4$
   def $layout(f : Int \Rightarrow Int) = c_1.layout(f) \mathbin{+\!\!+} c_2.layout(f)$
}
trait $Beside_4$ **extends** $Beside_1$ **with** $Circuit_4$ {
   override val $c_1, c_2 : Circuit_4$
   def $layout(f : Int \Rightarrow Int) =$
     $lzw(c_1.layout(f), c_2.layout(f.compose(c_1.width + \_)))(\_ \mathbin{+\!\!+} \_)$
}
trait $Stretch_4$ **extends** $Stretch_1$ **with** $Circuit_4$ {
   override val $c : Circuit_4$
   def $layout(f : Int \Rightarrow Int) = \{$
     val $vs = ns.scanLeft(0)(\_ + \_).tail$
     $c.layout(f.compose(vs(\_) - 1))\}$
}

An alternative encoding of modular interpretations

▸ Allow non-linear extensions and loose dependencies

  ▸ e.g. **wellSized**

$$\textbf{trait } Circuit_3 \textbf{ extends } Circuit_1 \, \{ \textbf{def } wellSized : Boolean \}$$
$$\textbf{trait } Id_3 \textbf{ extends } Circuit_3 \, \{ \textbf{def } wellSized = true \}$$
  ...
$$\textbf{trait } Stretch_3 \textbf{ extends } Circuit_3 \, \{$$
$$\quad \textbf{val } c : Circuit_3 ; \textbf{val } ns : List[Int]$$
$$\quad \textbf{def } wellSized = c.wellSized \wedge ns.length \equiv c.width$$
$$\}$$

▸ Require an extra step for combining **wellSized** and **width**

$$\textbf{trait } Id_{13} \textbf{ extends } Id_1 \textbf{ with } Id_3$$
  ...
$$\textbf{trait } Stretch_{13} \textbf{ extends } Stretch_1 \textbf{ with } Stretch_3$$

Adding language constructs

▸ Extend SCANS with right stretches

$$rstretch \quad :: [Int] \rightarrow Circuit_4 \rightarrow Circuit_4$$
$$rstretch \; ns \; c = stretch_4 \; (1 : init \; ns) \; c \; `beside_4` \; id_4 \; (last \; ns - 1)$$

$$\textbf{def} \; rstretch \, (ns : List[Int], c : Circuit_4) =$$
$$stretch \, (1 :: ns.init, beside \, (c, id \, (ns.last - 1)))$$

```
trait RStretch extends Stretch₄ {
    override def layout(f : Int ⇒ Int) = {
        val vs = ns.scanLeft(ns.last − 1)(_ + _).init
        c.layout(f.compose(vs(_)))}
}
```

Modular terms

▸ Object Algebras [Oliveira & Cook, 2012] come to the rescue

```
trait Circuit[C] {                    def circuit[C](f : Circuit[C]) =
    def id(x : Int) : C                   f.above(f.beside(f.fan(2), f.fan(2)),
    def fan(x : Int) : C                  f.above(f.stretch(f.fan(2), 2, 2),
    def above(x : C, y : C) : C               f.beside(f.beside(f.id(1), f.fan(2)), f.id(1))))
    def beside(x : C, y : C) : C
    def stretch(x : C, xs : Int*) : C
}

trait Factory₁ extends Circuit[Circuit₁] {        trait Factory₄ extends Circuit[Circuit₄] {...}
    def id(x : Int)                         = new Id₁       {val n = x}
    def fan(x : Int)                        = new Fan₁      {val n = x}
    def beside(x : Circuit₁, y : Circuit₁) = new Beside₁   {val c₁ = x; val c₂ = y}
    def above(x : Circuit₁, y : Circuit₁)  = new Above₁    {val c₁ = x; val c₂ = y}
    def stretch(x : Circuit₁, xs : Int*)    = new Stretch₁ {val ns = xs.toList; val c = x}
}

circuit(new Factory₁ {}).width                // 4
circuit(new Factory₄ {}).layout {x ⇒ x} // List(List((0,1),(2,3)),List((1,3)),List((1,2)))
```

Modular terms, extended

trait $ExtendedCircuit[C]$ **extends** $Circuit[C]$ {
   def $rstretch(x : C, xs : Int*) : C$
}

trait $ExtendedFactory_4$ **extends** $ExtendedCircuit[Circuit_4]$ **with** $Factory_4$ {
   def $rstretch(x : Circuit_4, xs : Int*) =$ **new** $RStretch$ {**val** $c = x;$ **val** $ns = xs.toList$}
}

def $circuit_2[C](f : ExtendedCircuit[C]) = f.rstretch(circuit(f), 2, 2, 2, 2)$

Case study

▸ We refactored an external SQL query processor [Rompf & Amin, 2015] to make it more *modular*, *shallow*, and *embedded*

tid, time,   *title,*   *room*
1,   09 : 30 *AM, Tuning IoT Devices into Robust and Safe Computers, Paganini*
2,   11 : 00 *AM, Separating Use and Reuse to Improve Both,*   *Paganini*
…

talks.csv

select ∗ **from** *talks.csv*

select *room, title* **from** *talks.csv*
where *time* = '09:00 AM'

select ∗
from (**select** *time, room, title* **as** $title_1$ **from** *talks.csv*)
join (**select** *time, room, title* **as** $title_2$ **from** *talks.csv*)
where $title_1$ <> $title_2$

def $q_0 = FROM$ ("talks.csv")

def $q_1 = q_0$ *WHERE* 'time === "09:00 AM"
          SELECT ('room, 'title)

def $q_2 =$
  q_0 *SELECT* ('time, 'room, 'title AS 'title$_1$)   *JOIN*
  (q_0 *SELECT* ('time, 'room, 'title AS 'title$_2$)) *WHERE*
  'title$_1$ <> 'title$_2$

A relational algebra interpreter

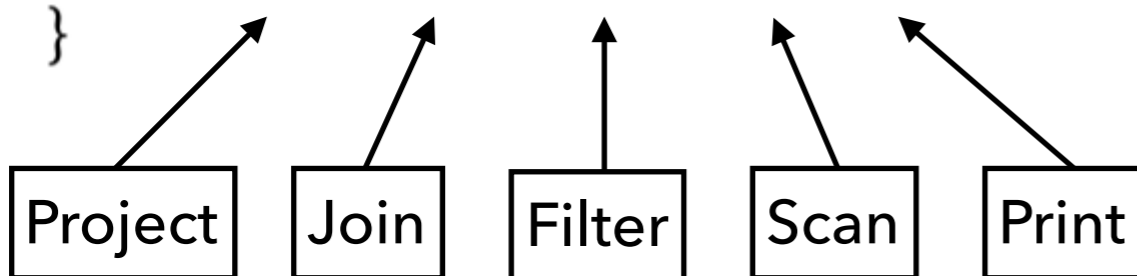▸ Under the surface syntax, a relational algebra expression is constructed

$FROM\,("talks.csv")$
$WHERE\,'time === "o9:oo\ AM"$
$SELECT\,('room, 'title)$

$Project\,(Schema\,("room", "title"),$
$\qquad Filter\,(Eq\,(Field\,("time"), Value\,("o9:oo\ AM")),$
$\qquad\qquad Scan\,("talks.csv")))$

▸ Each relational algebra operator implements the following interface

trait $Operator$ {
   def $resultSchema : Schema$
   def $execOp\,(yld : Record \Rightarrow Unit) : Unit$
}

Project	Join	Filter	Scan	Print

trait $Join$ **extends** $Operator$ {
   val $op_1, op_2 : Operator$
   def $resultSchema =$
     $op_1.resultSchema + op_2.resultSchema$
   def $execOp\,(yld : Record \Rightarrow Unit) =$
     $op_1.execOp\ \{rec_1 \Rightarrow$
       $op_2.execOp\ \{rec_2 \Rightarrow$
         val $keys = rec_1.schema\ intersect\ rec_2.schema$
         if$(rec_1\,(keys) \equiv rec_2\,(keys))$
           $yld\,(Record\,(rec_1.fields + rec_2.fields,$
             $rec_1.schema + rec_2.schema))$
     }}
 }

From interpreter to compiler

▸ The interpreter is simple but slow

▸ Turning a slow interpreter into a fast compiler while keeping the simplicity – **staging** (LMS [Rompf & Odersky, 2010])

   ▸ Actions on records are delayed to the generated code

$$\textbf{def } execOp\,(yld : Record \Rightarrow Unit) : Unit$$

$$\textbf{def } execOp\,(yld : Record \Rightarrow \boxed{Rep\,[Unit\,]}) : \boxed{Rep\,[Unit\,]}$$

▸ Two backends are supported (Scala and C), modularly

Syntax extensions

▸ Add aggregations (**group by**) and hash joins

```
trait Group extends Operator {
    val keys, agg : Schema; val op : Operator
    def resultSchema = keys ++ agg
    def execOp (yld : Record ⇒ Unit) {…}
}
trait HashJoin extends Join {
    override def execOp (yld : Record ⇒ Unit) = {
        val keys = op₁.resultSchema intersect op₂.resultSchema
        val hm = new HashMapBuffer (keys, op₁.resultSchema)
        op₁.execOp {rec₁ ⇒
            hm (rec₁ (keys)) += rec₁.fields }
        op₂.execOp {rec₂ ⇒
            hm (rec₂ (keys)) foreach {rec₁ ⇒
                yld (Record (rec₁.fields ++ rec₂.fields, rec₁.schema ++ rec₂.schema))}}}
}
```

Evaluation

▸ The **same** code is generated, thus performance is similar

▸ The modularity comes with a few more lines of code

Source	Functionality	Deep	Shallow
query_unstaged	SQL interpreter	83	98
query_staged	SQL to Scala compiler	179	194
query_optc	SQL to C compiler	245	262

More in the paper

Shallow EDSLs and Object-Oriented Programming
Beyond Simple Compositionality

Weixin Zhang[a] and Bruno C. d. S. Oliveira[a]
a   The University of Hong Kong, Hong Kong, China

Abstract
   Context. Embedded Domain-Specific Languages (EDSLs) are a common and widely used approach to DSLs in various languages, including Haskell and Scala. There are two main implementation techniques for EDSLs: *shallow embeddings* and *deep embeddings*.
   Inquiry. Shallow embeddings are quite simple, but they have been criticized in the past for being quite limited in terms of modularity and reuse. In particular, it is often argued that supporting multiple DSL interpretations in shallow embeddings is difficult.
   Approach. This paper argues that shallow EDSLs and Object-Oriented Programming (OOP) are closely related. Gibbons and Wu already discussed the relationship between shallow EDSLs and procedural abstraction, while Cook discussed the connection between procedural abstraction and OOP. We make the transitive step in this paper by connecting shallow EDSLs directly to OOP via procedural abstraction. The knowledge about this relationship enables us to improve on implementation techniques for EDSLs.
   Knowledge. This paper argues that common OOP mechanisms (including *inheritance, subtyping*, and *type-refinement*) increase the modularity and reuse of shallow EDSLs when compared to classical procedural abstraction by enabling a simple way to express *multiple, possibly dependent, interpretations*.
   Grounding. We make our arguments by using Gibbons and Wu's examples, where procedural abstraction is used in Haskell to model a simple shallow EDSL. We recode that EDSL in Scala and with an improved OO-inspired Haskell encoding. We further illustrate our approach with a case study on refactoring a deep external SQL DSL implementation to make it more modular, shallow, and embedded.
   Importance. This work is important for two reasons. Firstly, from an intellectual point of view, this work establishes the connection between shallow embeddings and OOP, which enables a better understanding of both concepts. Secondly, this work illustrates programming techniques that can be used to improve the modularity and reuse of shallow EDSLs.

ACM CCS 2012
   ▪ Software and its engineering → Language features; Domain specific languages;
Keywords   embedded domain-specific languages, shallow embedding, object-oriented programming

▶ **An OOP inspired Haskell encoding of modular (dependent) interpretations**

$$\textbf{class } Circuit\ c\ \textbf{where}$$

$$id \quad :: Int \to c$$
$$fan \quad :: Int \to c$$
$$above \ :: c \to c \to c$$
$$beside \ :: c \to c \to c$$
$$stretch :: [Int] \to c \to c$$

$$\textbf{class } a \prec b\ \textbf{where}$$
$$prj :: a \to b$$

$$\textbf{instance } a \prec a\ \textbf{where}$$
$$prj\ x = x$$

$$\textbf{instance } (a, b) \prec a\ \textbf{where}$$
$$prj = fst$$

$$\textbf{instance } (b \prec c) \Rightarrow (a, b) \prec c\ \textbf{where}$$
$$prj = prj \circ snd$$

Conclusion

Thank you!

▸ OOP and shallow embeddings are closely related

  ▸ The essence of both is *procedural abstraction*

▸ OOP abstractions bring extra modularity to shallow embeddings

  ▸ Subtyping, inheritance and type-refinement

▸ Combine extensible interpreters with Object Algebras for greater good

  ▸ Modular multiple (possibly dependent) interpretations and terms

▸ Shallow embeddings can be performant with *staging*

▸ The motivation to employ deep embeddings becomes weaker

  ▸ Mostly reduced to the need for AST transformations

https://github.com/wxzh/shallow-dsl