

Compiler Construction

Lecture 23: Code Generation IV (Implementation of Static Data Structures)

Thomas Noll

Lehrstuhl für Informatik 2
(Software Modeling and Verification)

RWTH Aachen University

`noll@cs.rwth-aachen.de`

`http://www-i2.informatik.rwth-aachen.de/i2/cc10/`

Winter semester 2010/11

- 1 Repetition: The Example Programming Language EPL
- 2 Implementation of Data Structures
- 3 Static Data Structures
- 4 Modifying the Abstract Machine
- 5 Translation of Static Data Structures into AM Programs
- 6 A Translation Example

Definition (Syntax of EPL)

The **syntax of EPL** is defined as follows:

$\mathbb{Z} :$ z (* z is an integer *)

$Ide :$ I (* I is an identifier *)

$AExp :$ $A ::= z \mid I \mid A_1 + A_2 \mid \dots$

$BExp :$ $B ::= A_1 < A_2 \mid \text{not } B \mid B_1 \text{ and } B_2 \mid B_1 \text{ or } B_2$

$Cmd :$ $C ::= I := A \mid C_1; C_2 \mid \text{if } B \text{ then } C_1 \text{ else } C_2 \mid$
 $\text{while } B \text{ do } C \mid I()$

$Dcl :$ $D ::= D_C D_V D_P$
 $D_C ::= \varepsilon \mid \text{const } I_1 := z_1, \dots, I_n := z_n;$
 $D_V ::= \varepsilon \mid \text{var } I_1, \dots, I_n;$
 $D_P ::= \varepsilon \mid \text{proc } I_1; K_1; \dots; I_n; K_n;$

$Blk :$ $K ::= D C$

$Pgm :$ $P ::= \text{in/out } I_1, \dots, I_n; K.$

- 1 Repetition: The Example Programming Language EPL
- 2 Implementation of Data Structures
- 3 Static Data Structures
- 4 Modifying the Abstract Machine
- 5 Translation of Static Data Structures into AM Programs
- 6 A Translation Example

Implementation of Data Structures

Source code: **data structures** = arrays, records, lists, trees, ...
⇒ **structured** state space, variables with components

Abstract machine: **linear** memory structure, cells for storing atomic data

Translation: **mapping** of structured state space to linear memory
(⇒ **address computation**)

- **static** data structures: memory requirements known at compile time
- **dynamic** data structures: memory requirements runtime dependent
⇒ heap, pointers, garbage collection, ...

First step:

- static data structures (arrays and records)
- inductive type definitions
- no procedures (for simplification; “orthogonal” extension)

- 1 Repetition: The Example Programming Language EPL
- 2 Implementation of Data Structures
- 3 Static Data Structures**
- 4 Modifying the Abstract Machine
- 5 Translation of Static Data Structures into AM Programs
- 6 A Translation Example

Modified Syntax of EPL

Definition 23.1 (Modified syntax of EPL)

The **modified syntax of EPL** is defined as follows (where $n \geq 1$):

$\mathbb{Z}$:	z	(* z is an integer *)
$\mathbb{B}$:	$b ::= \text{true} \mid \text{false}$	(* b is a Boolean *)
$\mathbb{R}$:	r	(* r is a real number *)
<i>Con</i> :	$c ::= z \mid b \mid r$	(* c is a constant *)
<i>Ide</i> :	I	(* I is an identifier *)
<i>Type</i> :	$T ::= \text{bool} \mid \text{int} \mid \text{real} \mid I \mid \text{array}[z_1..z_2] \text{ of } T \mid$ $\text{record } I_1:T_1; \dots; I_n:T_n \text{ end}$	
<i>Var</i> :	$V ::= I \mid V[E] \mid V.I$	
<i>Exp</i> :	$E ::= c \mid V \mid E_1 + E_2 \mid E_1 < E_2 \mid E_1 \text{ and } E_2 \mid \dots$	
<i>Cmd</i> :	$C ::= V:=E \mid C_1; C_2 \mid$ $\text{if } E \text{ then } C_1 \text{ else } C_2 \mid \text{while } E \text{ do } C$	
<i>Dcl</i> :	$D ::= D_C \text{ } D_T \text{ } D_V$ $D_C ::= \varepsilon \mid \text{const } I_1 := c_1; \dots; I_n := c_n;$ $D_T ::= \varepsilon \mid \text{type } I_1 := T_1; \dots; I_n := T_n;$ $D_V ::= \varepsilon \mid \text{var } I_1 : T_1; \dots; I_n : T_n;$	
<i>Pgm</i> :	$P ::= D \ C$	

- All identifiers in a declaration D have to be **different**.
- In $T = \text{record } I_1 : T_1 ; \dots ; I_n : T_n \text{ end}$, all selectors I_j must be **different**.
- In $T = \text{array}[z_1 .. z_2] \text{ of } T$, $z_1 \leq z_2$.
- Type definitions must **not be recursive**:
if $D_T = \text{type } I_1 := T_1 ; \dots ; I_n := T_n$; and type identifier I occurs in T_j , then $I \in \{I_1, \dots, I_{j-1}\}$.
- The type identifiers used in in a variable declaration D_V must be **declared**.
- Every identifier used in a command C must be **declared** in D (as a constant or variable).
- Variables in expressions and assignments have a **base type** (**bool/int/real**; possibly via type identifiers).

Static Semantics II

- Array **indices** must have type **int**.
- **Execution conditions** (**while**) and **branching expressions** (**if**) must have type **bool**.
- The types of the left-hand side and of the right-hand side types of an assignment must be **compatible**.
- **Type compatibility**: $\mathbb{Z} \subseteq \mathbb{R}$ in mathematics, but not on computers (different representation)

⇒ **type casts**

weak typing: implicit casting by compiler ($2.5 + 1$, $1 + "42"$)

⇒ risc of undetected “real” errors;

for **programming-in-the-small** (script languages)

strong typing: explicit casting by programmer

⇒ enhanced software reliability;

for **programming-in-the-large**

- Instantiation of operators/functions/procedures/... for different parameter types: **polymorphism** or **overloading**

$+: \text{int} \times \text{int} \rightarrow \text{int}$ $+: \text{real} \times \text{real} \rightarrow \text{real}$

- 1 Repetition: The Example Programming Language EPL
- 2 Implementation of Data Structures
- 3 Static Data Structures
- 4 Modifying the Abstract Machine
- 5 Translation of Static Data Structures into AM Programs
- 6 A Translation Example

The Modified Abstract Machine AM

- Additional **main storage** for keeping data values
- **Procedure stack** not required anymore (as procedures no longer supported)

Definition 23.2 (Modified abstract machine for EPL)

The **modified abstract machine for EPL (AM)** is defined by the **state space**

$$S := PC \times DS \times MS$$

with

- the **program counter** $PC := \mathbb{N}$,
- the **data stack** $DS := \mathbb{R}^*$, and
- the **main storage** $MS := \{\sigma \mid \sigma : \mathbb{N} \rightarrow \mathbb{R}\}$.

Definition 23.3 (New AM instructions)

- **Procedure instructions** are no longer needed.
- **Transfer instructions** ($\text{LOAD}(dif, off)$, $\text{STORE}(dif, off)$) are replaced by the following instructions with the respective semantics $\llbracket O \rrbracket : S \dashrightarrow S$:

$$\begin{aligned}\llbracket \text{LOAD} \rrbracket(a, d : n, \sigma) &:= (a + 1, d : \sigma(n), \sigma) \\ &\quad \text{if } n \in \mathbb{N} \\ \llbracket \text{STORE} \rrbracket(a, d : n : r, \sigma) &:= (a + 1, d, \sigma[n \mapsto r]) \\ &\quad \text{if } n \in \mathbb{N}\end{aligned}$$

- Moreover the following **instruction for checking array bounds** is introduced:

$$\llbracket \text{CAB}(z_1, z_2) \rrbracket(a, d : z, \sigma) := \begin{cases} (a + 1, d : z, \sigma) & \text{if } z \in \{z_1, \dots, z_2\} \\ (0, d : \underbrace{\text{RTE}}_{\text{runtime error}}, \sigma) & \text{otherwise} \end{cases}$$

- 1 Repetition: The Example Programming Language EPL
- 2 Implementation of Data Structures
- 3 Static Data Structures
- 4 Modifying the Abstract Machine
- 5 Translation of Static Data Structures into AM Programs
- 6 A Translation Example

Modifying the Symbol Table

$$\begin{aligned} Tab := \{st \mid st : Ide \dashrightarrow & (\{\text{const}\} \times (\mathbb{B} \cup \mathbb{Z} \cup \mathbb{R})) \\ & \cup (\{\text{var}\} \times Ide \times \mathbb{N}) \\ & \cup (\{\text{type}\} \times \{\text{bool}, \text{int}, \text{real}\} \times \{1\}) \\ & \cup (\{\text{type}\} \times \{\text{array}\} \times \mathbb{Z}^2 \times Ide \times \mathbb{N}) \\ & \cup (\{\text{type}\} \times \{\text{record}\} \times (Ide^2 \times \mathbb{N})^* \times \mathbb{N})\} \end{aligned}$$

Remarks:

- **Variable descriptor** (var, I, n): type I , memory address n
- Last component of type entry: **memory requirement**
(base types: 1 “cell”)
- **Array descriptor** ($\text{type}, \text{array}, z_1, z_2, I, n$):
bounds z_1, z_2 , component type I
- **Record descriptor** ($\text{type}, \text{record}, I_1, J_1, o_1, \dots, I_l, J_l, o_l, n$): selector
 I_k , component type J_k , memory offset o_k
- “Indexed” table lookup: $st(I.I_k) := (J_k, o_k)$
if $st(I) = (\text{type}, \text{record}, \dots, I_k, J_k, o_k, \dots, n)$

Maintaining the Symbol Table I

The symbol table is again maintained by the function `update( $D, st$ )` which specifies the update of symbol table st according to declaration D .

For the sake of simplicity we assume that $D = D_C D_T D_V \in Dcl$ is **flattened**, i.e., that every subtype is named by an identifier:

- If $D_T = \text{type } I_1 := T_1; \dots; I_n := T_n;$, then for every $k \in [n]$
 - $T_k \in \{\text{bool}, \text{int}, \text{real}\}$ or
 - $T_k \in \{I_1, \dots, I_{k-1}\}$ or
 - $T_k = \text{array}[z_1..z_2] \text{ of } I_j$ where $j \in [k-1]$ or
 - $T_k = \text{record } J_1 : I_{j_1}; \dots; J_l : I_{j_l} \text{ end}$ where $j_1, \dots, j_l \in [k-1]$
- For D_T as above, D_V must be of the form
 $D_V = \text{var } J_1 : I_{j_1}; \dots; J_k : I_{j_k};$ where $j_1, \dots, j_k \in [n]$

Maintaining the Symbol Table II

Definition 23.1 (Modified update function)

update : $Dcl \times Tab \dashrightarrow Tab$ is defined by

$update(D_C \ D_T \ D_V, st) := update(D_V, update(D_T, update(D_C, st)))$

$update(\varepsilon, st) := st$

$update(\text{const } I_1 := c_1; \dots; I_n := c_n; , st)$

$:= st[I_1 \mapsto (\text{const}, c_1), \dots, I_n \mapsto (\text{const}, c_n)]$

$update(\text{type } I := \text{bool}; D'_T, st) := update(\text{type } D'_T, st[I \mapsto (\text{type}, \text{bool}, 1)])$

$update(\text{type } I := \text{int}; D'_T, st) := update(\text{type } D'_T, st[I \mapsto (\text{type}, \text{int}, 1)])$

$update(\text{type } I := \text{real}; D'_T, st) := update(\text{type } D'_T, st[I \mapsto (\text{type}, \text{real}, 1)])$

$update(\text{type } I := J; D'_T, st) := update(\text{type } D'_T, st[I \mapsto st(J)])$

$update(\text{type } I := \text{array}[z_1 \dots z_2] \text{ of } J; D'_T, st)$

$:= update(\text{type } D'_T,$

$st[I \mapsto (\text{type}, \text{array}, z_1, z_2, J, k \cdot n)])$

if $st(J) = (\text{type}, \dots, n)$ and $k = z_2 - z_1 + 1$

$update(\text{type } I := \text{record } I_1 : J_1; \dots; I_l : J_l \text{ end}; D'_T, st)$

$:= update(\text{type } D'_T, st[I \mapsto$

$(\text{type}, \text{record}, I_1, J_1, 0, I_2, J_2, n_1, \dots,$

$I_l, J_l, \sum_{i=1}^{l-1} n_i, \sum_{i=1}^l n_i)])$

if $st(J_i) = (\text{type}, \dots, n_i)$ for $i \in [l]$

$update(\text{var } I_1 : J_1; \dots; I_n : J_n; , st) := st[I_1 \mapsto (\text{var}, J_1, 0), I_2 \mapsto (\text{var}, J_2, n_1), \dots,$

$I_n \mapsto (\text{var}, J_n, \sum_{i=1}^{n-1} n_i)]$

if $st(J_i) = (\text{type}, \dots, n_i)$ for $i \in [l]$

Example 23.2 (Modified update function)

```
Let  $D :=$  type Bool=bool; Int=int;  
      Array=array[1..20] of Bool;  
      Record=record S:Array; T:Int end;  
      var x:Int; y:Array; z:Record;
```

Then

$$\text{update}(D, \text{st}) = \text{st}[\begin{array}{l} \text{Bool} \mapsto (\text{type}, \text{bool}, 1), \\ \text{Int} \mapsto (\text{type}, \text{int}, 1), \\ \text{Array} \mapsto (\text{type}, \text{array}, 1, 20, \text{Bool}, 20), \\ \text{Record} \mapsto (\text{type}, \text{record}, \text{S}, \text{Array}, 0, \text{T}, \text{Int}, 20, 21), \\ \text{x} \mapsto (\text{var}, \text{Int}, 0), \\ \text{y} \mapsto (\text{var}, \text{Array}, 1), \\ \text{z} \mapsto (\text{var}, \text{Record}, 21) \end{array}]$$

Translation of Variables I

The translation employs the following auxiliary function to determine the type identifier of a given variable:

Definition 23.3 (vtype function)

The mapping

$$\text{vtype} : \text{Var} \times \text{Tab} \dashrightarrow \text{Ide}$$

is given by

$$\begin{aligned} \text{vtype}(I, \text{st}) &:= J \\ &\text{if } \text{st}(I) = (\text{var}, J, n) \end{aligned}$$

$$\begin{aligned} \text{vtype}(V[E], \text{st}) &:= J \\ &\text{if } \text{vtype}(V, \text{st}) = I \\ &\text{and } \text{st}(I) = (\text{type}, \text{array}, z_1, z_2, J, n) \end{aligned}$$

$$\begin{aligned} \text{vtype}(V.I, \text{st}) &:= J \\ &\text{if } \text{vtype}(V, \text{st}) = I' \text{ and } \text{st}(I'.I) = (J, o) \end{aligned}$$

Translation of Variables II

The function `vt` generates code for computing the memory address of a variable (and storing it on the data stack):

Definition 23.4 (Translation of variables)

The mapping

$$vt : Var \times Tab \dashrightarrow AM$$

is given by

```
vt(I, st) := LIT(n);
           if st(I) = (var, J, n)
vt(V[E], st) := vt(V, st)           % address of V
               et(E, st)             % array index
               CAB(z1, z2);         % bounds checking
               LIT(z1); SUB;        % index difference
               LIT(n); MULT;         % relative address
               ADD;                  % address of V[E]
           if vtype(V, st) = I and st(I) = (type, array, z1, z2, J, m)
           and st(J) = (type, ..., n)
vt(V.I, st) := vt(V, st)           % address of V
               LIT(o);              % offset
               ADD;                  % address of V.I
           if vtype(V, st) = I' and st(I'.I) = (J, o)
```

Definition 23.5 (Translation of expressions)

The mapping

$$et : Exp \times Tab \dashrightarrow AM$$

is given by

$$\begin{aligned} et(c, st) &:= \text{LIT}(c); \\ et(V, st) &:= \begin{cases} \text{LIT}(c); & \text{if } V \in Ide \text{ and } st(V) = (\text{const}, c) \\ vt(V, st) & \text{otherwise} \\ \text{LOAD}; \end{cases} \\ et(E_1 + E_2, st) &:= \begin{array}{l} et(E_1, st) \\ et(E_2, st) \\ \text{ADD}; \end{array} \\ &\vdots \end{aligned}$$

Translation of Commands and Programs

Definition 23.6 (Translation of commands)

For the mapping

$$ct : Cmd \times Tab \dashrightarrow AM$$

only the handling of assignments needs to be adapted:

$$\begin{aligned} ct(V := E, st) &:= vt(V, st) \quad \% \text{ address of left-hand side} \\ &\quad et(E, st) \quad \% \text{ value of right-hand side} \\ &\quad \text{STORE;} \end{aligned}$$

Definition 23.7 (Translation of programs)

The mapping

$$trans : Pgm \dashrightarrow AM$$

is defined by

$$trans(D \ C) := ct(C, \text{update}(D, st_{\emptyset}))$$

- 1 Repetition: The Example Programming Language EPL
- 2 Implementation of Data Structures
- 3 Static Data Structures
- 4 Modifying the Abstract Machine
- 5 Translation of Static Data Structures into AM Programs
- 6 A Translation Example

Example 23.8

$$P = \left. \begin{array}{l} \text{type Int=int; Array=array[1..10] of Int;} \\ \text{var a:Array; i:Int;} \\ \text{i:=1;} \\ \text{while i<=10 do} \\ \quad \text{a[i]:=i; i:=i+1;} \end{array} \right\}^D \quad \left. \begin{array}{l} \\ \\ \\ \end{array} \right\}^C$$
$$\begin{aligned} \text{trans}(P) &= \text{ct}(C, \text{update}(D, \text{st}_\emptyset)) \\ \text{st} &:= \text{update}(D, \text{st}_\emptyset) \\ &= \text{st}_\emptyset[\quad \text{Int} \mapsto (\text{type}, \text{int}, 1), \\ &\quad \text{Array} \mapsto (\text{type}, \text{array}, 1, 10, \text{Int}, 10), \\ &\quad \text{a} \mapsto (\text{var}, \text{Array}, 0), \\ &\quad \text{i} \mapsto (\text{var}, \text{Int}, 10)] \end{aligned}$$
$$\begin{aligned} \text{ct}(C, \text{st}) &= \text{ct}(\text{i:=1}, \text{st}) \\ &\quad \text{ct}(\text{while i<=10 do a[i]:=i; i:=i+1}, \text{st}) \\ \text{ct}(\text{i:=1}, \text{st}) &= \text{vt}(\text{i}, \text{st}) \quad \% \text{ adr}(\text{i}) \\ &\quad \text{et}(1, \text{st}) \quad \% \text{ val}(1) \\ &\quad \text{STORE;} \\ &= \text{LIT}(10); \text{LIT}(1); \text{STORE;} \end{aligned}$$

Example 23.8 (continued)

```
ct(while i<=10 do a[i]:=i; i:=i+1, st)
    = a : et(i<=10, st)
          JFALSE(a');
          ct(a[i]:=i; i:=i+1, st)
          JMP(a);
          a' :
et(i<=10, st) = LIT(10); LOAD; LIT(10); LE;
ct(a[i]:=i; i:=i+1, st) = ct(a[i]:=i, st) ct(i:=i+1, st)
ct(a[i]:=i, st) = vt(a[i], st)    % adr(a[i])
                  et(i, st)      % val(i)
                  STORE;
vt(a[i], st) = vt(a, st)          % adr(a)
                  et(i, st)      % val(i)
                  CAB(1,10);      % bounds checking
                  LIT(1); SUB;    % index diff.
                  LIT(1); MULT;   % rel. address
                  ADD;            % adr(a[i])
vt(a, st) = LIT(0);
et(i, st) = LIT(10); LOAD;
ct(i:=i+1, st) = LIT(10); LIT(10); LOAD; LIT(1); ADD; STORE;
```