Intermediate Code Generation

Cosmin E. Oancea
cosmin.oancea@diku.dk

Modified by Marco Valtorta for CSCE 531 at UofSC
Based on Jost Berthold’s slides and Torben Mogensen’s book
Department of Computer Science
University of Copenhagen

February 2018 IPS Lecture Slides
Structure of a Compiler

Program text
↓
Lexical analysis
↓
Symbol sequence
↓
Syntax analysis
↓
Syntax tree
↓
Typecheck
↓
Syntax tree
↓
Intermediate code generation

Binary machine code
↑
Assembly and linking
↑
Ditto with named registers
↑
Register allocation
↑
Symbolic machine code
↑
Machine code generation
↑
Intermediate code
1 Why Intermediate Code?
   - Intermediate Language
   - To-Be-Translated Language

2 Syntax-Directed Translation
   - Arithmetic Expressions
   - Statements
   - Boolean Expressions, Sequential Evaluation

3 Translating More Complex Structures
   - More Control Structures
   - Arrays and Other Structured Data
   - Role of Declarations in the Translation
Why Intermediate Language (IL)?

- Compilers for different platforms and languages can share parts.

![Diagram](Diagram.png)

- **Without IL**: how many translators do I need to write to map \( n \) languages to \( m \) different hardware?
Why Intermediate Language (IL)?

- Compilers for different platforms and languages can share parts.

Without IL: how many translators do I need to write to map $n$ languages to $m$ different hardware?
Answer: $n \times m$ instead of $n + m$!

- Machine-independent optimizations are possible.
- Also enables interpretation ...
Intermediate Language (IL)

- **Machine Independent**: unlimited number of registers and memory space, no machine-specific instructions.

- **Mid-level(s)** between source and machine languages (**tradeoff**): simpler constructs, easier to generate machine code.

- What features/constructs should IL support?
  - every translation loses information $\Rightarrow$ use the information before losing it!
  - typically a chain of ILs moving from higher towards lower level.

- How complex should IL’s instruction be?
  - complex: good for interpretation (amortizes instruction-decoding overhead),
  - simple: can more easily generate optimal machine code.
Intermediate Language (IL)

Here: Low-level language, but keeping functions (procedures).
Small instructions:

- **3-address code**: one operation per expression
Intermediate Language (IL)

Here: Low-level language, but keeping functions (procedures).
Small instructions:

- 3-address code: one operation per expression
- Memory read/write (\(M\)) (address is atom).
Intermediate Language (IL)

Here: Low-level language, but keeping functions (procedures).
Small instructions:

- **3-address code**: one operation per expression
- **Memory read/write** (M) (address is atom).
- **Jump labels**, GOTO and conditional jump (IF).
Intermediate Language (IL)

Here: Low-level language, but keeping functions (procedures).
Small instructions:

- **3-address code**: one operation per expression
- **Memory read/write** ($M$) (address is atom).
- **Jump labels**, **GOTO** and conditional jump (**IF**).
- **Function calls and returns**

```
Prg    →  Fcts
Fcts   →  Fct Fcts | Fct
Fct    →  Hdr Bd
Hdr    →  functionid(Args)
Bd     →  [ Instrs ]
Instrs →  Instr , Instrs | Instr
Instr  →  id := Atom | id := unop Atom
         | id := id binop Atom
         | LABEL label | GOTO label
         | IF id relop Atom THEN label ELSE label
Atom   →  id | num
```
Intermediate Language (IL)

Here: Low-level language, but keeping functions (procedures).

Small instructions:

- **3-address code**: one operation per expression
- **Memory** read/write (M) (address is atom).
- **Jump** labels, GOTO and conditional jump (IF).
- **Function** calls and returns

\[
\begin{align*}
Prg & \rightarrow Fcts \\
Fcts & \rightarrow Fct \ Fcts \ | \ Fct \\
Fct & \rightarrow Hdr \ Bd \\
Hdr & \rightarrow functionid(Args) \\
Bd & \rightarrow [ \ Instrs \ ] \\
Instrs & \rightarrow Instr , Instrs \ | \ Instr \\
Instr & \rightarrow id := Atom \ | \ id := unop \ Atom \\
& \quad \ | \ id := id \ binop \ Atom \\
& \quad \ | \ id := M[Atom] \ | \ M[Atom] := id \\
& \quad \ | \ LABEL \ label \ | \ GOTO \ label \\
& \quad \ | \ IF \ id \ relop \ Atom \\
& \quad \quad \quad \quad \quad \ THEN \ label \ ELSE \ label \\
& \quad \ | \ id := CALL \ functionid(Args) \\
& \quad \ | \ RETURN \ id
\end{align*}
\]

\[
\begin{align*}
Atom & \rightarrow id \ | \ num \\
Args & \rightarrow id , Args \ | \ id
\end{align*}
\]
The To-Be-Translated Language

We shall translate a simple procedural language:

- Arithmetic expressions and function calls, boolean expressions,
- conditional branching (if),
- two loops constructs (while and repeat until).

Syntax-directed translation:

- In practice we work on the abstract syntax tree \texttt{AbSyn}
  (but here we use a generic grammar notation),
- Implement each syntactic category via a translation function:
  Arithmetic expressions, Boolean expressions, Statements.
- Code for subtrees is generated independent of context,
  (i.e., context is a parameter to the translation function)
1. Why Intermediate Code?
   - Intermediate Language
   - To-Be-Translated Language

2. Syntax-Directed Translation
   - Arithmetic Expressions
   - Statements
   - Boolean Expressions, Sequential Evaluation

3. Translating More Complex Structures
   - More Control Structures
   - Arrays and Other Structured Data
   - Role of Declarations in the Translation
Translating Arithmetic Expressions

Expressions in Source Language

- Variables and number literals,
- unary and binary operations,
- function calls (with argument list).

\[ \begin{align*}
\text{Exp} & \rightarrow \text{num} | \text{id} \\
& \quad | \text{unop} \text{ Exp} \\
& \quad | \text{Exp binop Exp} \\
& \quad | \text{id}(\text{Exps}) \\
\text{Exps} & \rightarrow \text{Exp} | \text{Exp} , \text{Exps}
\end{align*} \]
Translating Arithmetic Expressions

Expressions in Source Language
- Variables and number literals,
- unary and binary operations,
- function calls (with argument list).

Translation function:
\[ Trans_{Exp} :: (Exp, VTable, FTable, Location) \rightarrow [ICode] \]

- Returns a list of intermediate code instructions [ICode] that . . .
- . . . upon execution, computes Exp’s result in variable Location.
- Case analysis on Exp’s abstract syntax tree ABSYN.
Symbol Tables and Helper Functions

Translation function:

\[ Trans_{Exp} :: (\text{Exp}, \text{VTable}, \text{FTable}, \text{Location}) \rightarrow \text{[ICode]} \]

Symbol Tables

- **vtable**: maps a variable name in source lang to its corresponding (translation) IL variable name.
- **ftable**: function names to function labels (for call)

Helper Functions

- **lookup**: retrieve entry from a symbol table
- **getvalue**: retrieve value of source language literal
- **getname**: retrieve name of source language variable/operation
- **newvar**: make new intermediate code variable
- **newlabel**: make new label (for jumps in intermediate code)
- **trans_op**: translates an operator name to the name in IL.
## Generating Code for an Expression

\[\text{Trans}_{\text{Exp}} : \text{(Exp, VTable, FTable, Location)} \rightarrow \text{[ICode]}\]

\[\text{Trans}_{\text{Exp}} (\text{exp, vtable, ftable, place}) = \text{case exp of}\]

<table>
<thead>
<tr>
<th>Case</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>num</code></td>
<td>( v = \text{getValue}(\text{num}) ) [( \text{place} := v )]</td>
</tr>
<tr>
<td><code>id</code></td>
<td>( x = \text{lookup}(\text{vtable, getName}(\text{id})) ) [( \text{place} := x )]</td>
</tr>
</tbody>
</table>
| `unop \text{Exp}_1` | \( \text{place}_1 = \text{newvar}() \) \(\)  \\
|               | \( \text{code}_1 = \text{Trans}_{\text{Exp}} (\text{Exp}_1, \text{vtable, ftable, place}_1) \) \(\)  \\
|               | \( \text{op} = \text{trans\_op}(\text{getName}(\text{unop})) \) \(\)  \\
|               | \( \text{code}_1 @ [\text{place} := \text{op place}_1] \) \(\)  |
| `\text{Exp}_1 \text{ binop Exp}_2` | \( \text{place}_1 = \text{newvar}() \) \(\)  \\
|               | \( \text{place}_2 = \text{newvar}() \) \(\)  \\
|               | \( \text{code}_1 = \text{Trans}_{\text{Exp}} (\text{Exp}_1, \text{vtable, ftable, place}_1) \) \(\)  \\
|               | \( \text{code}_2 = \text{Trans}_{\text{Exp}} (\text{Exp}_2, \text{vtable, ftable, place}_2) \) \(\)  \\
|               | \( \text{op} = \text{trans\_op}(\text{getName}(\text{binop})) \) \(\)  \\
|               | \( \text{code}_1 @ \text{code}_2 @ [\text{place} := \text{place}_1 \text{ op place}_2] \) \(\)  |

In this slide presentation, \( @ \) (as in SML) is used instead of ++ (as in Haskell and in the Mogensen’s book) for list concatenation.
Generating Code for a Function Call

\[ \text{Trans}_{\text{Exp}} \ (\exp, \ vtable, \ ftable, \ place) = \text{case exp of} \]

<table>
<thead>
<tr>
<th>( \text{id}(\text{Exps}) )</th>
<th>(( \text{code}<em>1, [a_1, \ldots, a_n] )) = \text{Trans}</em>{\text{Exps}}(\text{Exps, vtable, ftable})</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{fname} = \text{lookup}(\text{ftable, getname}(\text{id})) )</td>
<td></td>
</tr>
<tr>
<td>( \text{code}_1 @ [\text{place} := \text{CALL} \ \text{fname}(a_1, \ldots, a_n)] )</td>
<td></td>
</tr>
</tbody>
</table>

\text{Trans}_{\text{Exps}} \text{ returns the code that evaluates the function’s parameters,} 
\text{and the list of new-intermediate variables (that store the result).} 

\text{Trans}_{\text{Exps}} : (\text{Exps, VTable, FTable}) \rightarrow ([\text{ICode}], [\text{Location}]) 
\text{Trans}_{\text{Exps}}(\text{exp}, \ vtable, \ ftable) = \text{case exps of} 

<table>
<thead>
<tr>
<th>( \exp )</th>
<th>place = newvar()</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{Exp}</td>
<td>\text{code}<em>1 = \text{Trans}</em>{\text{Exp}}(\text{Exp, vtable, ftable, place})</td>
</tr>
<tr>
<td>\text{code}_1, [\text{place}]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \text{Exp} \ , \ \text{Exps} )</th>
<th>place = newvar()</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{Exp} \ , \ \text{Exps}</td>
<td>\text{code}<em>1 = \text{Trans}</em>{\text{Exp}}(\text{Exp, vtable, ftable, place})</td>
</tr>
<tr>
<td>\text{(code}<em>2, \text{args}) = \text{Trans}</em>{\text{Exps}}(\text{Exps, vtable, ftable})</td>
<td></td>
</tr>
<tr>
<td>\text{code}_3 = \text{code}_1 @ \text{code}_2</td>
<td></td>
</tr>
<tr>
<td>\text{args}_1 = \text{place :: args}</td>
<td></td>
</tr>
<tr>
<td>\text{(code}_3, \text{args}_1)</td>
<td></td>
</tr>
</tbody>
</table>
Translation Example

Assume the following symbol tables:

- **vtable** = \([x \mapsto v_0, \ y \mapsto v_1, \ z \mapsto v_2]\)
- **ftable** = \([f \mapsto \_F\_1, \ + \mapsto +, \ - \mapsto -]\)

Translation of \(\text{Exp}\) with place = \(t_0\):

- \(\text{Exp} = x-3\)
Translation Example

Assume the following symbol tables:

- **vtable** = \([x \mapsto v0, \ y \mapsto v1, \ z \mapsto v2]\)
- **ftable** = \([f \mapsto \_F_1, \ + \mapsto +, \ - \mapsto -]\)

Translation of **Exp** with place = \(t0\):

- **Exp=x-3**
  - \(t1 := v0\)
  - \(t2 := 3\)
  - \(t0 := t1 - t2\)

- **Exp=3+f(x-y,z)**
  - \(t1 := 3\)
  - \(t4 := v0\)
  - \(t5 := v1\)
  - \(t3 := t4 - t5\)
  - \(t6 := v2\)
  - \(t2 := \text{CALL } F_1(t3, t6)\)
  - \(t0 := t1 + t2\)
Translation Example

Assume the following symbol tables:

- \( \text{vtable} = [x \mapsto v_0, y \mapsto v_1, z \mapsto v_2] \)
- \( \text{ftable} = [f \mapsto F_1, + \mapsto +, - \mapsto -] \)

Translation of \( \text{Exp} \) with place = \( t_0 \):

- \( \text{Exp} = x - 3 \)
  - \( t_1 := v_0 \)
  - \( t_2 := 3 \)
  - \( t_0 := t_1 - t_2 \)

- \( \text{Exp} = 3 + f(x - y, z) \)
Translation Example

Assume the following symbol tables:

- vtable = \([x \mapsto v_0, \ y \mapsto v_1, \ z \mapsto v_2]\)
- ftable = \([f \mapsto _F_1, \ + \mapsto +, \ - \mapsto -]\)

Translation of Exp with place = \(t_0\):

- Exp=\(x-3\)
  
  \[
  \begin{align*}
  t_1 & := v_0 \\
  t_2 & := 3 \\
  t_0 & := t_1 - t_2 \\
  \end{align*}
  \]

- Exp=\(3+f(x-y,z)\)
  
  \[
  \begin{align*}
  t_1 & := 3 \\
  t_4 & := v_0 \\
  t_5 & := v_1 \\
  t_3 & := t_4 - t_5 \\
  t_6 & := v_2 \\
  t_2 & := CALL _F_1(t_3, t_6) \\
  t_0 & := t_1 + t_2 \\
  \end{align*}
  \]
1 Why Intermediate Code?
   - Intermediate Language
   - To-Be-Translated Language

2 Syntax-Directed Translation
   - Arithmetic Expressions
   - Statements
     - Boolean Expressions, Sequential Evaluation

3 Translating More Complex Structures
   - More Control Structures
   - Arrays and Other Structured Data
   - Role of Declarations in the Translation
Translating Statements

Statements in Source Language

- Sequence of statements
  \[ Stat \rightarrow Stat ; Stat \]
- Assignment
  \[ id := Exp \]
- Conditional Branching
  \[ if Cond then \{ Stat \} \]
  \[ if Cond then \{ Stat \} else \{ Stat \} \]
- Loops: while and repeat
  (simple conditions for now)
  \[ Cond \rightarrow Exp \text{ relop } Exp \]
  \[ while Cond do \{ Stat \} \]
  \[ repeat \{ Stat \} until Cond \]

We assume relational operators translate directly (using \text{trans}_\text{op}).
Translating Statements

Statements in Source Language

- Sequence of statements
  \[ Stat \rightarrow Stat ; Stat \]
- Assignment
  \[ id := Exp \]
- Conditional Branching
  \[ if \ Cond \ then \{ Stat \} \]
  \[ if \ Cond \ then \{ Stat \} \ else \{ Stat \} \]
- Loops: \texttt{while} and \texttt{repeat}
  \[ while \ Cond \ do \{ Stat \} \]
  \[ repeat \{ Stat \} \ until \ Cond \]

\[ Cond \rightarrow Exp \ relop \ Exp \]

We assume relational operators translate directly (using \texttt{trans\_op}).

Translation function:

\[ Trans_{Stat} : (Stat, VTable, FTable) \rightarrow [ICode] \]

- As before: syntax-directed, \texttt{case analysis} on \texttt{Stat}
- Intermediate code instructions for statements
 Generating Code for Sequences, Assignments, . . .

\[ Trans_{Stat} : (Stat, Vtable, Ftable) \rightarrow [ICode] \]
\[ Trans_{Stat}(stat, vtable, ftable) = \text{case } stat \text{ of} \]
\[ Stat_1 ; Stat_2 \quad code_1 = Trans_{Stat}(Stat_1, vtable, ftable) \]
\[ \quad code_2 = Trans_{Stat}(Stat_2, vtable, ftable) \]
\[ \quad code_1 \@ code_2 \]
\[ \text{id := Exp} \quad place = \text{lookup}(vtable, \text{getname(id)}) \]
\[ Trans_{Exp}(Exp, vtable, ftable, place) \]

. . . (rest coming soon)

- Sequence of statements, sequence of code.
- Symbol tables are inherited attributes.
Generating Code for Conditional Jumps: Helper

- Helper function for loops and branches
- Evaluates Cond, i.e., a boolean expression, then jumps to one of two labels, depending on result

\[
\text{Trans}_\text{Cond} \colon (\text{Cond}, \text{Label}, \text{Label}, \text{Vtable}, \text{Ftable}) \rightarrow [\text{ICode}]
\]

\[
\text{Trans}_\text{Cond}(\text{cond}, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable}) = \text{case} \ \text{cond} \ \text{of}
\]

\[
\begin{align*}
\text{Exp}_1 \ \text{relop} \ \text{Exp}_2 & \quad t_1 = \text{newvar}() \\
& \quad t_2 = \text{newvar}() \\
& \quad \text{code}_1 = \text{Trans}_\text{Exp}(\text{Exp}_1, \text{vtable}, \text{ftable}, t_1) \\
& \quad \text{code}_2 = \text{Trans}_\text{Exp}(\text{Exp}_2, \text{vtable}, \text{ftable}, t_2) \\
& \quad \text{op} = \text{trans\_op(getname(relop))} \\
& \quad \text{code}_1 \ @ \ \text{code}_2 \ @ \ [\text{IF} \ t_1 \ \text{op} \ t_2 \ \text{THEN} \ \text{label}_t \ \text{ELSE} \ \text{label}_f]
\end{align*}
\]

- Uses the IF of the intermediate language
- Expressions need to be evaluated before
  (restricted IF: only variables and atoms can be used)
Generating Code for If-Statements

- Generate **new labels** for branches and following code
- Translate **If** statement to a **conditional jump**
### Generating Code for If-Statements

- Generate **new labels** for branches and following code
- Translate `if` statement to a **conditional jump**

\[
Trans_{\text{Stat}}(\text{stat}, \text{vtable}, \text{ftable}) = \text{case stat of}
\]

- \(\text{if } \text{Cond} \text{ then } \text{Stat}_1\)
  - \(\text{label}_t = \text{newlabel()}\)
  - \(\text{label}_f = \text{newlabel()}\)
  - \(\text{code}_c = Trans_{\text{Cond}}(\text{Cond}, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable})\)
  - \(\text{code}_s = Trans_{\text{Stat}}(\text{Stat}_1, \text{vtable}, \text{ftable})\)
  - \(\text{code}_c @ [\text{LABEL } \text{label}_t] @ \text{code}_s @ [\text{LABEL } \text{label}_f]\)

- \(\text{if } \text{Cond} \text{ then } \text{Stat}_1 \text{ else } \text{Stat}_2\)
  - \(\text{label}_t = \text{newlabel()}\)
  - \(\text{label}_f = \text{newlabel()}\)
  - \(\text{label}_e = \text{newlabel()}\)
  - \(\text{code}_c = Trans_{\text{Cond}}(\text{Cond}, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable})\)
  - \(\text{code}_1 = Trans_{\text{Stat}}(\text{Stat}_1, \text{vtable}, \text{ftable})\)
  - \(\text{code}_2 = Trans_{\text{Stat}}(\text{Stat}_2, \text{vtable}, \text{ftable})\)
  - \(\text{code}_c @ [\text{LABEL } \text{label}_t] @ \text{code}_1 @ [\text{GOTO } \text{label}_e]\)
  - \(\@ [\text{LABEL } \text{label}_f] @ \text{code}_2 @ [\text{LABEL } \text{label}_e]\)
Generating Code for Loops

- repeat-until loop is the easy case:
  Execute body, check condition, jump back if false.
- while loop needs check before body, one extra label needed.
Generating Code for Loops

- repeat-until loop is the easy case:
  Execute body, check condition, jump back if false.
- while loop needs check before body, one extra label needed.

\[
\text{Trans}_{\text{Stat}}(\text{stat}, \text{vtable}, \text{ftable}) = \text{case stat of}
\]

<table>
<thead>
<tr>
<th>repeat Stat until Cond</th>
<th>( label_f = \text{newlabel}() )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( label_t = \text{newlabel}() )</td>
</tr>
<tr>
<td></td>
<td>( \text{code}<em>1 = \text{Trans}</em>{\text{Stat}}(\text{Stat}, \text{vtable}, \text{ftable}) )</td>
</tr>
<tr>
<td></td>
<td>( \text{code}<em>2 = \text{Trans}</em>{\text{Cond}}(\text{Cond}, label_t, label_f, \text{vtable}, \text{ftable}) )</td>
</tr>
<tr>
<td></td>
<td>([\text{LABEL } label_f] @ \text{code}_1 @ \text{code}_2 @ [\text{LABEL } label_t] )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>while Cond do Stat</th>
<th>( label_s = \text{newlabel}() )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( label_t = \text{newlabel}() )</td>
</tr>
<tr>
<td></td>
<td>( label_f = \text{newlabel}() )</td>
</tr>
<tr>
<td></td>
<td>( \text{code}<em>1 = \text{Trans}</em>{\text{Cond}}(\text{Cond}, label_t, label_f, \text{vtable}, \text{ftable}) )</td>
</tr>
<tr>
<td></td>
<td>( \text{code}<em>2 = \text{Trans}</em>{\text{Stat}}(\text{Stat}, \text{vtable}, \text{ftable}) )</td>
</tr>
<tr>
<td></td>
<td>([\text{LABEL } label_s] @ \text{code}_1 @ [\text{LABEL } label_t] @ \text{code}_2 @ [\text{GOTO } label_s] @ [\text{LABEL } label_f] )</td>
</tr>
</tbody>
</table>
Translation Example

- Symbol table vtable: \[x \mapsto v_0, \ y \mapsto v_1, \ z \mapsto v_2\]
- Symbol table ftable: \[\text{getInt} \mapsto \text{libIO}\_\text{getInt}\]

\[
x := 3;
\]
\[
y := \text{getInt}();
\]
\[
z := 1;
\]
\[
\text{while } y > 0
\]
\[
\quad y := y - 1;
\]
\[
\quad z := z \times x
\]
Translation Example

- Symbol table vtable: $[x \mapsto v_0, y \mapsto v_1, z \mapsto v_2]$
- Symbol table ftable: $[\text{getInt} \mapsto \text{libIO\_getInt}]$

```plaintext
x := 3;
y := getInt();
z := 1;
while y > 0
    y := y - 1;
    z := z * x
```

```plaintext
v_0 := 3
v_1 := CALL libIO\_getInt()
v_2 := 1
```
Translation Example

- Symbol table vtable: \([x \mapsto v_0, y \mapsto v_1, z \mapsto v_2]\)
- Symbol table ftable: \([\text{getInt} \mapsto \text{libIO_getInt}]\)

\[\begin{align*}
\text{x} & := 3; \\
\text{y} & := \text{getInt}(); \\
\text{z} & := 1; \\
\text{while} & \ \text{y} > 0 \\
    \quad & \text{y} := \text{y} - 1; \\
    \quad & \text{z} := \text{z} \ast \text{x}
\end{align*}\]

\vspace{1em}

\[\begin{align*}
v_0 & := 3 \\
v_1 & := \text{CALL libIO_getInt}() \\
v_2 & := 1 \\
\text{LABEL l_s} \\
    & \text{t}_1 := v_1 \\
    & \text{t}_2 := 0 \\
    \text{IF} & \ \text{t}_1 > \text{t}_2 \ \text{THEN} \ l_t \ \text{else} \ l_f \\
\text{LABEL l_t}
\end{align*}\]

\[\begin{align*}
\text{GOTO l_s} \\
\text{LABEL l_f}
\end{align*}\]
Translation Example

- Symbol table vtable: \([x \mapsto v_0, \ y \mapsto v_1, \ z \mapsto v_2]\)
- Symbol table ftable: \([\text{getInt} \mapsto \text{libIO_getInt}]\)

\[
x := 3;
y := \text{getInt}();
z := 1;
while y > 0 
  \begin{align*}
  y &:= y - 1; \\
  z &:= z * x
  \end{align*}
\]

\[
v_0 := 3 \\
v_1 := \text{CALL libIO_getInt}() \\
v_2 := 1 \\
\text{LABEL } l_s \\
t_1 := v_1 \\
t_2 := 0 \\
\text{IF } t_1 > t_2 \text{ THEN } l_t \text{ else } l_f \\
\text{LABEL } l_t \\
t_3 := v_1 \\
t_4 := 1 \\
v_1 := t_3 - t_4
\]

GOTO \ l_s
\text{LABEL } l_f
Translation Example

- Symbol table \textbf{vtable}: \([x \mapsto v_0, \ y \mapsto v_1, \ z \mapsto v_2]\)
- Symbol table \textbf{ftable}: \([\text{getInt} \mapsto \text{libIO\_getInt}]\)

\begin{verbatim}
x := 3;
y := \text{getInt}();
z := 1;
while y > 0 
    y := y - 1;
    z := z \ast x
\end{verbatim}

\begin{verbatim}
v_0 := 3
v_1 := \text{CALL libIO\_getInt()}
v_2 := 1
LABEL l_s
    t_1 := v_1
t_2 := 0
    IF t_1 > t_2 THEN l_t else l_f
LABEL l_t
    t_3 := v_1
t_4 := 1
    v_1 := t_3 - t_4
t_5 := v_2
t_6 := v_0
    v_2 := t_5 \ast t_6
GOTO l_s
LABEL l_f
\end{verbatim}
1. Why Intermediate Code?
   - Intermediate Language
   - To-Be-Translated Language

2. Syntax-Directed Translation
   - Arithmetic Expressions
   - Statements
   - Boolean Expressions, Sequential Evaluation

3. Translating More Complex Structures
   - More Control Structures
   - Arrays and Other Structured Data
   - Role of Declarations in the Translation
More Complex Conditions, Boolean Expressions

Boolean Expressions as Conditions

- Arithmetic expressions used as Boolean
- Logical operators (not, and, or)
- Boolean expressions used in arithmetics

\[
\begin{align*}
\text{Cond} & \rightarrow \text{Exp relop Exp} \\
& | \text{Exp} \\
& | \text{not Cond} \\
& | \text{Cond and Cond} \\
& | \text{Cond or Cond} \\
\text{Exp} & \rightarrow \ldots | \text{Cond}
\end{align*}
\]
More Complex Conditions, Boolean Expressions

Boolean Expressions as Conditions

- Arithmetic expressions used as Boolean
- Logical operators (not, and, or)
- Boolean expressions used in arithmetics

\[ Cond \rightarrow Exp \text{ relop } Exp \]
\[ \text{ or } \]
\[ \text{not } Cond \]
\[ Cond \text{ and } Cond \]
\[ Cond \text{ or } Cond \]

\[ Exp \rightarrow \ldots \mid Cond \]

We extend the translation functions \( Trans_{Exp} \) and \( Trans_{Cond} \):

- Interpret numeric values as Boolean expressions: 0 is \textbf{false}, all other values \textbf{true}.
- Likewise: truth values as arithmetic expressions
Numbers and Boolean Values, Negation

Expressions as Boolean values, negation:

\( \text{Trans}_{\text{Cond}} : (\text{Cond}, \text{Label}, \text{Label}, \text{Vtable}, \text{Ftable}) \rightarrow [\text{ICode}] \)
\( \text{Trans}_{\text{Cond}}(\text{cond}, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable}) = \text{case } \text{cond } \text{of} \)

\[ \begin{align*}
\text{Exp} & \quad t = \text{newvar}() \\
& \quad \text{code} = \text{Trans}_{\text{Exp}}(\text{Exp}, \text{vtable}, \text{ftable}, t) \\
& \quad \text{code } @ \text{[IF } t \neq 0 \text{ THEN } \text{label}_t \text{ ELSE } \text{label}_f] \\
\text{notCond} & \quad \text{Trans}_{\text{Cond}}(\text{Cond}, \text{label}_f, \text{label}_t, \text{vtable}, \text{ftable})
\end{align*} \]

\[ \ldots \]
Numbers and Boolean Values, Negation

Expressions as Boolean values, negation:

\[
\text{Trans}_{\text{Cond}} : (\text{Cond}, \text{Label}, \text{Label}, \text{Vtable}, \text{Ftable}) \rightarrow [\text{ICode}]
\]

\[
\text{Trans}_{\text{Cond}}(\text{cond}, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable}) = \text{case } \text{cond} \text{ of}
\]

\[
\]

\[
\text{Exp}
\]
\[
t = \text{newvar}()
\]
\[
\text{code} = \text{Trans}_{\text{Exp}}(\text{Exp}, \text{vtable}, \text{ftable}, t)
\]
\[
\text{code} @ [\text{IF } t \neq 0 \text{ THEN } \text{label}_t \text{ ELSE } \text{label}_f]
\]

\[
\text{notCond} \quad \text{Trans}_{\text{Cond}}(\text{Cond}, \text{label}_f, \text{label}_t, \text{vtable}, \text{ftable})
\]

\[
\]

Conversion of Boolean values to numbers (by jumps):

\[
\text{Trans}_{\text{Exp}} : (\text{Exp}, \text{Vtable}, \text{Ftable}) \rightarrow [\text{ICode}]
\]

\[
\text{Trans}_{\text{Exp}}(\text{exp}, \text{vtable}, \text{ftable}, \text{place}) = \text{case } \text{exp} \text{ of}
\]

\[
\]

\[
\text{Cond}
\]
\[
\text{label}_1 = \text{newlabel}()
\]
\[
\text{label}_2 = \text{newlabel}()
\]
\[
t = \text{newvar}()
\]
\[
\text{code} = \text{Trans}_{\text{Cond}}(\text{Cond}, \text{label}_1, \text{label}_2, \text{vtable}, \text{ftable})
\]
\[
[t := 0] @ \text{code} @ [\text{LABEL label}_1, t := 1] @ [\text{LABEL label}_2, \text{place} := t]
\]
Sequential Evaluation of Conditions

Moscow ML version 2.01 (January 2004)
Enter ‘quit();’ to quit.
- fun f l = if (hd l = 1) then "one" else "not one";
> val f = fn : int list -> string
- f [];
! Uncaught exception:
! Empty
Sequential Evaluation of Conditions

Moscow ML version 2.01 (January 2004)
Enter ‘quit();’ to quit.
- fun f l = if (hd l = 1) then "one" else "not one";
> val f = fn : int list -> string
- f [];
  ! Uncaught exception:
  ! Empty

In most languages, logical operators are evaluated sequentially.
  • If $B_1 = false$, do not evaluate $B_2$ in $B_1 \&\& B_2$ (anyway false).
  • If $B_1 = true$, do not evaluate $B_2$ in $B_1 || B_2$ (anyway true).
Sequential Evaluation of Conditions

Moscow ML version 2.01 (January 2004)
Enter ‘quit();’ to quit.
- fun f l = if (hd l = 1) then "one" else "not one";
> val f = fn : int list -> string
- f [];
! Uncaught exception:
! Empty

In most languages, logical operators are **evaluated sequentially**.

- If $B_1 = \text{false}$, do not evaluate $B_2$ in $B_1 \&\& B_2$ (anyway $false$).
- If $B_1 = \text{true}$, do not evaluate $B_2$ in $B_1 || B_2$ (anyway $true$).

- fun g l = if not (null l) andalso (hd l = 1) then "one" else "not one";
> val g = fn : int list -> string
- g [];
> val it = "not one" : string
Sequential Evaluation by “Jumping Code”

\[ \text{Trans}_{\text{Cond}} : (\text{Cond}, \text{Label}, \text{Label}, \text{Vtable}, \text{Ftable}) \rightarrow [\text{ICode}] \]

\[ \text{Trans}_{\text{Cond}}(\text{cond}, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable}) = \text{case cond of} \]

\[ \begin{aligned}
\text{Cond}_1 & \quad \text{label}_{\text{next}} = \text{newlabel}() \\
\text{and} & \quad \text{code}_1 = \text{Trans}_{\text{Cond}}(\text{Cond}_1, \text{label}_{\text{next}}, \text{label}_f, \text{vtable}, \text{ftable}) \\
\text{Cond}_2 & \quad \text{code}_2 = \text{Trans}_{\text{Cond}}(\text{Cond}_2, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable}) \\
\text{code}_1 \ & \odot [\text{LABEL} \ \text{label}_{\text{next}}] \ & \odot \text{code}_2 \\
\end{aligned} \]

\[ \begin{aligned}
\text{Cond}_1 & \quad \text{label}_{\text{next}} = \text{newlabel}() \\
\text{or} & \quad \text{code}_1 = \text{Trans}_{\text{Cond}}(\text{Cond}_1, \text{label}_t, \text{label}_{\text{next}}, \text{vtable}, \text{ftable}) \\
\text{Cond}_2 & \quad \text{code}_2 = \text{Trans}_{\text{Cond}}(\text{Cond}_2, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable}) \\
\text{code}_1 \ & \odot [\text{LABEL} \ \text{label}_{\text{next}}] \ & \odot \text{code}_2 \\
\end{aligned} \]

Note: No logical operations in intermediate language!

Logics of \text{and} and \text{or} encoded by jumps.

Alternative: Logical operators in intermediate language

\[ \text{Cond} \Rightarrow \text{Exp} \Rightarrow \text{Exp} \text{ binop} \text{Exp} \]

Translated as an arithmetic operation.

Evaluates both sides!
Sequential Evaluation by “Jumping Code”

Trans\textsubscript{Cond} : (Cond, Label, Label, Vtable, Ftable) → [ICode]

\[\text{Trans}_{\text{Cond}}(\text{cond}, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable}) = \text{case } \text{cond} \text{ of}\]

\[
\begin{array}{ll}
\text{Cond}_1 & \text{label}_{next} = \text{newlabel}() \\
\text{and} & \text{code}_1 = \text{Trans}_{\text{Cond}}(\text{Cond}_1, \text{label}_{next}, \text{label}_f, \text{vtable}, \text{ftable}) \\
\text{Cond}_2 & \text{code}_2 = \text{Trans}_{\text{Cond}}(\text{Cond}_2, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable}) \\
& \text{code}_1 \ @ [\text{LABEL } \text{label}_{next}] \ @ \text{code}_2 \\
\end{array}
\]

\[
\begin{array}{ll}
\text{Cond}_1 & \text{label}_{next} = \text{newlabel}() \\
\text{or} & \text{code}_1 = \text{Trans}_{\text{Cond}}(\text{Cond}_1, \text{label}_t, \text{label}_{next}, \text{vtable}, \text{ftable}) \\
\text{Cond}_2 & \text{code}_2 = \text{Trans}_{\text{Cond}}(\text{Cond}_2, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable}) \\
& \text{code}_1 \ @ [\text{LABEL } \text{label}_{next}] \ @ \text{code}_2 \\
\end{array}
\]

- Note: No logical operations in intermediate language!
  Logics of \textbf{and} and \textbf{or} encoded by jumps.
Sequential Evaluation by “Jumping Code”

\[
\text{Trans}_{\text{Cond}} : (\text{Cond}, \text{Label}, \text{Label}, \text{Vtable}, \text{Ftable}) \rightarrow \text{[ICode]}
\]

\[
\text{Trans}_{\text{Cond}}(\text{cond}, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable}) = \text{case } \text{cond} \text{ of}
\]

\[
\begin{align*}
\text{Cond}_1 & \quad \text{label}_{next} = \text{newlabel()}
\text{and} & \quad \text{code}_1 = \text{Trans}_{\text{Cond}}(\text{Cond}_1, \text{label}_{next}, \text{label}_f, \text{vtable}, \text{ftable}) \\
\text{Cond}_2 & \quad \text{code}_2 = \text{Trans}_{\text{Cond}}(\text{Cond}_2, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable}) \\
& \quad \text{code}_1 \ @ \ [\text{LABEL } \text{label}_{next}] \ @ \ \text{code}_2
\end{align*}
\]

\[
\begin{align*}
\text{Cond}_1 & \quad \text{label}_{next} = \text{newlabel()}
\text{or} & \quad \text{code}_1 = \text{Trans}_{\text{Cond}}(\text{Cond}_1, \text{label}_t, \text{label}_{next}, \text{vtable}, \text{ftable}) \\
\text{Cond}_2 & \quad \text{code}_2 = \text{Trans}_{\text{Cond}}(\text{Cond}_2, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable}) \\
& \quad \text{code}_1 \ @ \ [\text{LABEL } \text{label}_{next}] \ @ \ \text{code}_2
\end{align*}
\]

- **Note:** No logical operations in intermediate language!
  Logics of **and** and **or** encoded by jumps.

- **Alternative:** Logical operators in intermediate language
  \[
  \text{Cond} \Rightarrow \text{Exp} \Rightarrow \text{Exp binop Exp}
  \]
  Translated as an arithmetic operation.
# Sequential Evaluation by “Jumping Code”

\[ \text{Trans}_\text{Cond} : (\text{Cond}, \text{Label}, \text{Label}, \text{Vtable}, \text{Ftable}) \rightarrow [\text{ICode}] \]

\[ \text{Trans}_\text{Cond}(\text{cond}, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable}) = \text{case cond of} \]

\[ \ldots \]

\begin{align*}
\text{Cond}_1 & \quad \text{label}_{\text{next}} = \text{newlabel}() \\
\text{and} & \\
\text{Cond}_2 & \quad \text{code}_1 = \text{Trans}_\text{Cond}(\text{Cond}_1, \text{label}_{\text{next}}, \text{label}_f, \text{vtable}, \text{ftable}) \\
& \quad \text{code}_2 = \text{Trans}_\text{Cond}(\text{Cond}_2, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable}) \\
& \quad \text{code}_1 \ @ \ [\text{LABEL label}_{\text{next}}] \ @ \ \text{code}_2
\end{align*}

\begin{align*}
\text{Cond}_1 & \quad \text{label}_{\text{next}} = \text{newlabel}() \\
\text{or} & \\
\text{Cond}_2 & \quad \text{code}_1 = \text{Trans}_\text{Cond}(\text{Cond}_1, \text{label}_t, \text{label}_{\text{next}}, \text{vtable}, \text{ftable}) \\
& \quad \text{code}_2 = \text{Trans}_\text{Cond}(\text{Cond}_2, \text{label}_t, \text{label}_f, \text{vtable}, \text{ftable}) \\
& \quad \text{code}_1 \ @ \ [\text{LABEL label}_{\text{next}}] \ @ \ \text{code}_2
\end{align*}

- **Note:** No logical operations in intermediate language!
  Logics of **and** and **or** encoded by jumps.

- **Alternative:** Logical operators in intermediate language
  \[ \text{Cond} \Rightarrow \text{Exp} \Rightarrow \text{Exp binop Exp} \]
  Translated as an arithmetic operation. **Evaluates both sides!**
1 Why Intermediate Code?
   - Intermediate Language
   - To-Be-Translated Language

2 Syntax-Directed Translation
   - Arithmetic Expressions
   - Statements
   - Boolean Expressions, Sequential Evaluation

3 Translating More Complex Structures
   - More Control Structures
   - Arrays and Other Structured Data
   - Role of Declarations in the Translation
More Control Structures

- Control structures determine control flow: which instruction to execute next
- A while-loop is enough
More Control Structures

- Control structures determine control flow: which instruction to execute next
- A **while**-loop is enough ... but ... languages usually offer more.
- **Explicit jumps**:

  $$\text{Stat} \rightarrow \text{label : goto label}$$

  Necessary instructions are in the intermediate language.
  Needs to build symbol table of labels.
More Control Structures

- Control structures determine control flow: which instruction to execute next
- A while-loop is enough . . . but . . . languages usually offer more.
- Explicit jumps:  
  \[
  Stat \rightarrow \text{label} : \\
  \mid \text{goto label}
  \]
  Necessary instructions are in the intermediate language. Needs to build symbol table of labels.
- Case/Switch:  
  \[
  Stat \rightarrow \text{case Exp of [ Alts ]} \\
  Alts \rightarrow \text{num : Stat} \mid \text{num : Stat, Alts}
  \]
  When exited after each case: chain of if-then-else
  When “falling through” (e.g., in C): if-then-else and goto.
More Control Structures

- **Control structures** determine control flow: which instruction to execute next

- A **while**-loop is enough . . . but . . . languages usually offer more.

- **Explicit jumps:** 
  
  \[ \text{Stat} \rightarrow \text{label} : \text{goto label} \]

  Necessary instructions are in the intermediate language.

  Needs to build symbol table of labels.

- **Case/Switch:** 
  
  \[ \text{Stat} \rightarrow \text{case Exp of [ Alts ]} \]

  \[ \text{Alts} \rightarrow \text{num : Stat} | \text{num : Stat, Alts} \]

  When exited after each case: chain of if-then-else

  When “falling through” (e.g., in C): if-then-else and goto.

- **Break and Continue:**
  
  \[ \text{Stat} \rightarrow \text{break} | \text{continue} \]

  (break: jump behind loop, continue: jump to end of loop body).

  Needs two jump target labels used only inside loop bodies

  (parameters to translation function \( Trans_{\text{Stat}} \))
More Control Structures

- **Control structures** determine control flow: which instruction to execute next
- A **while**-loop is enough . . . but . . . languages usually offer more.
- **Explicit jumps:** $Stat \rightarrow \text{label: } | \text{goto label}$
  
  Necessary instructions are in the intermediate language. 
  Needs to build symbol table of labels.
- **Case/Switch:** $Stat \rightarrow \text{case Exp of [ Alts ]}$
  $Alts \rightarrow \text{num : Stat} | \text{num : Stat, Alts}$
  
  When exited after each case: chain of if-then-else
  When “falling through” (e.g., in C): if-then-else and goto.
- **Break and Continue:** $Stat \rightarrow \text{break} | \text{continue}$
  (break: jump behind loop, continue: jump to end of loop body).
  Needs two jump target labels used only inside loop bodies
  (parameters to translation function $Trans_{Stat}$)
1 Why Intermediate Code?
   • Intermediate Language
   • To-Be-Translated Language

2 Syntax-Directed Translation
   • Arithmetic Expressions
   • Statements
   • Boolean Expressions, Sequential Evaluation

3 Translating More Complex Structures
   • More Control Structures
   • Arrays and Other Structured Data
   • Role of Declarations in the Translation
Translating Arrays (of int elements)

Extending the Source Language

- Array elements used as an expression
- Assignment to an array element
- Array elements accessed by an index (expression)

\[
\begin{align*}
\text{Exp} & \rightarrow \ldots | \text{Idx} \\
\text{Stat} & \rightarrow \ldots | \text{Idx} := \text{Exp} \\
\text{Idx} & \rightarrow \text{id[ Exp ]}
\end{align*}
\]
Translating Arrays (of int elements)

Extending the Source Language

- Array elements used as an expression
- Assignment to an array element
- Array elements accessed by an index (expression)

\[
\begin{align*}
\text{Exp} & \rightarrow \ldots | \text{Idx} \\
\text{Stat} & \rightarrow \ldots | \text{Idx} := \text{Exp} \\
\text{Idx} & \rightarrow \text{id}[\text{Exp}]
\end{align*}
\]

Again we extend \(\text{Trans}_{\text{Exp}}\) and \(\text{Trans}_{\text{Stat}}\).

- Arrays stored in pre-allocated memory area, generated code will use memory access instructions.
- Static (compile-time) or dynamic (run-time) allocation possible.
Generating Code for Address Calculation

- \textit{vtable} contains the base address of the array.
- Elements are \texttt{int} here, so 4 bytes per element for address.

\[ Trans_{\text{Idx}}(index, \text{vtable}, \text{ftable}) = \text{case } index \text{ of} \]
\begin{align*}
\text{id}[\text{Exp}] & \quad \text{base} = \text{lookup(} \text{vtable}, \text{getname} (\text{id}) \text{)} \\
& \quad \text{addr} = \text{newvar} () \\
& \quad \text{code}_1 = Trans_{\text{Exp}} (\text{Exp}, \text{vtable}, \text{ftable}, \text{addr}) \\
& \quad \text{code}_2 = \text{code}_1 \odot [\text{addr} := \text{addr} \times 4, \text{addr} := \text{addr} + \text{base}] \\
& \quad (\text{code}_2, \text{addr})
\end{align*}

Returns:

- Code to calculate the absolute address . . .
- of the array element in memory (corresponding to \texttt{index}), . . .
- . . . and a new variable (\texttt{addr}) where it will be stored.
Generating Code for Array Access

Address-calculation code: in expression and statement translation.

- **Read access** inside expressions:

  \[
  \text{Trans}_{\text{Exp}}(\text{exp}, \text{vtable}, \text{ftable}, \text{place}) = \text{case } \text{exp} \text{ of}
  \]

  \[
  \cdots
  \]

  \[
  \text{Idx} (\text{code}_1, \text{address}) = \text{Trans}_{\text{Idx}}(\text{Idx}, \text{vtable}, \text{ftable})
  \]

  \[
  \text{code}_1 \oplus [\text{place} := M[\text{address}]]
  \]

- **Write access** in assignments:

  \[
  \text{Trans}_{\text{Stat}}(\text{stat}, \text{vtable}, \text{ftable}) = \text{case } \text{stat} \text{ of}
  \]

  \[
  \cdots
  \]

  \[
  \text{Idx} := \text{Exp} (\text{code}_1, \text{address}) = \text{Trans}_{\text{Idx}}(\text{Index}, \text{vtable}, \text{ftable})
  \]

  \[
  t = \text{newvar}()
  \]

  \[
  \text{code}_2 = \text{Trans}_{\text{Exp}}(\text{Exp}, \text{vtable}, \text{ftable}, t)
  \]

  \[
  \text{code}_1 \oplus \text{code}_2 \oplus [M[\text{address}] := t]
  \]
Multi-Dimensional Arrays

Arrays in Multiple Dimensions

- Only a small change to previous grammar: $Idx$ can now be recursive.
- Needs to be mapped to an address in one dimension.

$$
\begin{align*}
\text{Exp} & \rightarrow \ldots | \text{Idx} \\
\text{Stat} & \rightarrow \ldots | \text{Idx} \ := \ \text{Exp} \\
\text{Idx} & \rightarrow \ \text{id}[\text{Exp}] \ | \ \text{Idx}[\text{Exp}]
\end{align*}
$$

Arrays stored in row-major or column-major order. Standard: row-major, index of $a[k][l]$ is $k \cdot \text{dim}_1 + l$.

Address calculation need to know sizes in each dimension. Symbol table: base address and list of array-dimension sizes. Need to change $\text{Trans}$ $\text{Idx}$, i.e., add recursive index calculation.
Multi-Dimensional Arrays

Arrays in Multiple Dimensions

- Only a small change to previous grammar: \( Idx \) can now be recursive.
- Needs to be mapped to an address in one dimension.

Arrays stored in row-major or column-major order.

**Standard: row-major**, index of \( a[k][l] \) is \( k \cdot dim_1 + l \)

(\( \text{Index of } b[k][l][m] \) is \( k \cdot dim_1 \cdot dim_2 + l \cdot dim_2 + m \))
Multi-Dimensional Arrays

Arrays in Multiple Dimensions

- Only a small change to previous grammar: $ldx$ can now be recursive.
- Needs to be mapped to an address in one dimension.

- Arrays stored in row-major or column-major order.
  **Standard:** row-major, index of $a[k][l]$ is $k \cdot dim_1 + l$
  (Index of $b[k][l][m]$ is $k \cdot dim_1 \cdot dim_2 + l \cdot dim_2 + m$)

- Address calculation need to know sizes in each dimension.
  Symbol table: base address and list of array-dimension sizes.

- Need to change $Trans_{ldx}$, i.e., add recursive index calculation.
Address Calculation in Multiple Dimensions

\[
Trans_{idx}(\text{index}, \text{vtable}, \text{ftable}) = \]

\[
(code_1, t, base, []) = Calc_{idx}(\text{index}, \text{vtable}, \text{ftable})
\]

\[
\begin{align*}
\text{code}_2 &= \text{code}_1 @ [t := t \ast 4, t := t + base] \\
(code_2, t)
\end{align*}
\]
Address Calculation in Multiple Dimensions

\[
Trans_{Idx}(index, vtable, ftable) =
\]
\[
(code_1, t, base, []) = Calc_{Idx}(index, vtable, ftable)
\]
\[
\text{code}_2 = code_1 \; \&\; [t := t \times 4, t := t + base]
\]
\[
(code_2, t)
\]

Recursive index calculation, multiplies with dimension at each step.

\[
Calc_{Idx}(index, vtable, ftable) = \text{case } index \; \text{of}
\]
\[
\text{id[Exp]} \quad (base, \text{dims}) = \text{lookup}(vtable, \text{getname}(\text{id}))
\]
\[
\quad addr = \text{newvar}()
\]
\[
\quad code = Trans_{Exp}(Exp, vtable, ftable, addr)
\]
\[
\quad (code, addr, base, tail(dims))
\]
\[
\text{Index[Exp]} \quad (\text{code}_1, addr, base, \text{dims}) = Calc_{Idx}(\text{Index}, vtable, ftable)
\]
\[
\quad d = \text{head}(\text{dims})
\]
\[
\quad t = \text{newvar}()
\]
\[
\quad code_2 = Trans_{Exp}(Exp, vtable, ftable, t)
\]
\[
\quad code_3 = code_1 \; \&\; code_2 \; \&\; [addr := addr \times d, addr := addr + t]
\]
\[
\quad (code_3, addr, base, tail(dims))
\]
1 Why Intermediate Code?
   - Intermediate Language
   - To-Be-Translated Language

2 Syntax-Directed Translation
   - Arithmetic Expressions
   - Statements
   - Boolean Expressions, Sequential Evaluation

3 Translating More Complex Structures
   - More Control Structures
   - Arrays and Other Structured Data
   - Role of Declarations in the Translation
Declarations in the Translation

Declarations are necessary

- to allocate space for arrays,
- to compute addresses for multi-dimensional arrays,
- ...and when the language allows local declarations (scope).
Declarations in the Translation

Declarations are necessary

- to allocate space for arrays,
- to compute addresses for multi-dimensional arrays,
- ... and when the language allows local declarations (scope).

Declarations and scope

- Statements following a declarations can see declared data.
- Declaration of variables and arrays
- Here: Constant size, one dimension

\[
\begin{align*}
Stat & \rightarrow \text{Decl}; Stat \\
Decl & \rightarrow \text{int id} \\
& | \text{int id}[\text{num}] \\
\end{align*}
\]

Function \( Trans_{\text{Decl}} \) : (Decl, VTable) \( \rightarrow \) ([ICode], VTable)

- translates declarations to code and new symbol table.
Translating Declarations to Scope and Allocation

Code with local scope (extended symbol table):

\[ Trans_{\text{Stat}}(\text{stat}, \text{vtable}, \text{ftable}) = \text{case } \text{stat} \text{ of} \]

\[ Decl ; \text{Stat}_1 \]

\[ (\text{code}_1, \text{vtable}_1) = Trans_{\text{Decl}}(\text{Decl}, \text{vtable}) \]

\[ \text{code}_2 = Trans_{\text{Stat}}(\text{Stat}_1, \text{vtable}_1, \text{ftable}) \]

\[ \text{code}_1 \otimes \text{code}_2 \]
Translating Declarations to Scope and Allocation

Code with local scope (extended symbol table):

\[ Trans_{Stat}(stat, vtable, ftable) = \text{case } stat \text{ of} \]
\[ Decl ; Stat_1 \quad (code_1, vtable_1) = Trans_{Decl}(Decl, vtable) \]
\[ code_2 = Trans_{Stat}(Stat_1, vtable_1, ftable) \]
\[ code_1 \circ code_2 \]

Building the symbol table and allocating:

\[ Trans_{Decl} : (Decl, VTable) \to ([ICode], VTable) \]
\[ Trans_{Decl}(decl, vtable) = \text{case } decl \text{ of} \]
\[ \text{int } id \quad t_1 = \text{newvar()} \]
\[ vtable_1 = \text{bind}(vtable, \text{getname}(id), t_1) \]
\[ ([], vtable_1) \]
\[ \text{int } id[num] \quad t_1 = \text{newvar()} \]
\[ vtable_1 = \text{bind}(vtable, \text{getname}(id), t_1) \]
\[ ([t_1 := HP, HP := HP + (4 \ast \text{getvalue}(num))], vtable_1) \]

... where HP is the heap pointer, indicating the first free space in a managed heap at runtime; used for dynamic allocation.
Other Structures that Require Special Treatment

- **Floating-Point values:**
  - Often stored in different registers
  - Always require different machine operations
  - Symbol table needs type information when creating variables in intermediate code.
Other Structures that Require Special Treatment

- **Floating-Point values:**
  Often stored in different registers
  Always require different machine operations
  Symbol table needs type information when creating variables in intermediate code.

- **Strings**
  Sometimes just arrays of (1-byte) char type, but variable length.
  In modern languages/implementations, elements can be char or unicode (UTF-8 and UTF-16 variable size!)
  Usually handled by library functions.
Other Structures that Require Special Treatment

- **Floating-Point values:**
  Often stored in different registers
  Always require different machine operations
  Symbol table needs type information when creating variables in intermediate code.

- **Strings**
  Sometimes just arrays of (1-byte) char type, but variable length.
  In modern languages/implementations, elements can be char or unicode (UTF-8 and UTF-16 variable size!)
  Usually handled by library functions.

- **Records and Unions**
  Linear in memory. Field types and sizes can be different.
  Field selector known at compile time: compute offset from base.
Structure of a Compiler

Program text
↓
Lexical analysis
↓
Symbol sequence
↓
Syntax analysis
↓
Syntax tree
↓
Typecheck
↓
Syntax tree
↓
Intermediate code generation

Binary machine code
↑
Assembly and linking
↑
Ditto with named registers
↑
Register allocation
↑
Symbolic machine code
↑
Machine code generation
↑
Intermediate code
→