编译原理

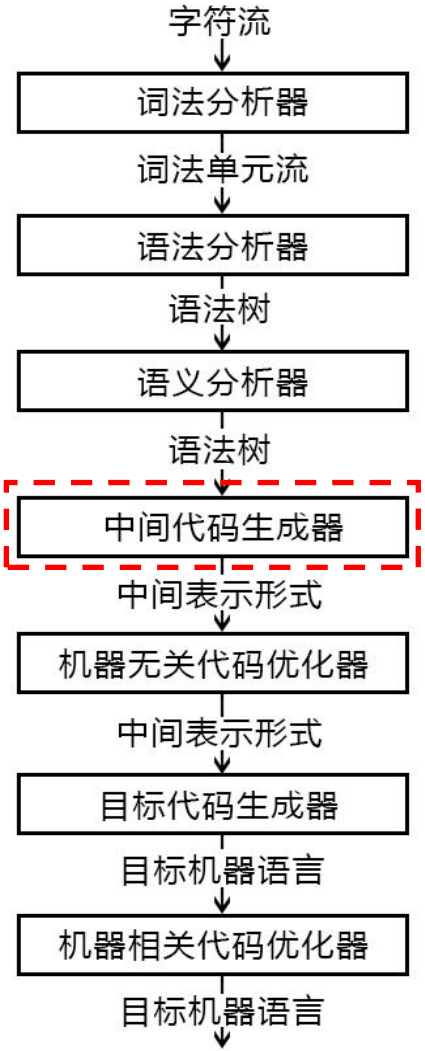
第六章

中间代码生成

哈尔滨工业大学 陈鄞 单丽莉

编译器的结构

- 中间代码的特点
  - 简单规范
  - 易于优化与转换
  - 与机器无关
  - 易于移植



常用的三地址指令

序号	指令类型	指令形式
1	赋值指令	*x* = *y* op *z* *x* = op *y*
2	复制指令	*x* = *y*
3	条件跳转	if *x* relop *y* goto *n*
4	非条件跳转	goto *n*
5	参数传递	param *x*
6	过程调用 函数调用	call *p*, *n* *y* =call *p*, *n*
7	过程返回	return *x*
8	数组引用	*x* = *y*[*i*]
9	数组赋值	*x*[*i*] = *y*
10	地址及 指针操作	*x* =& *y* *x* =* *y* **x* = *y*

三地址指令：一个运算符，三个地址；

三地址：(最多)两个运算分量地址，

一个结果分量地址。

地址可以具有如下形式之一

- 源程序中的名字 (*name*)
- 常量 (*constant*)
- 编译器生成的临时变量(*temporary*)

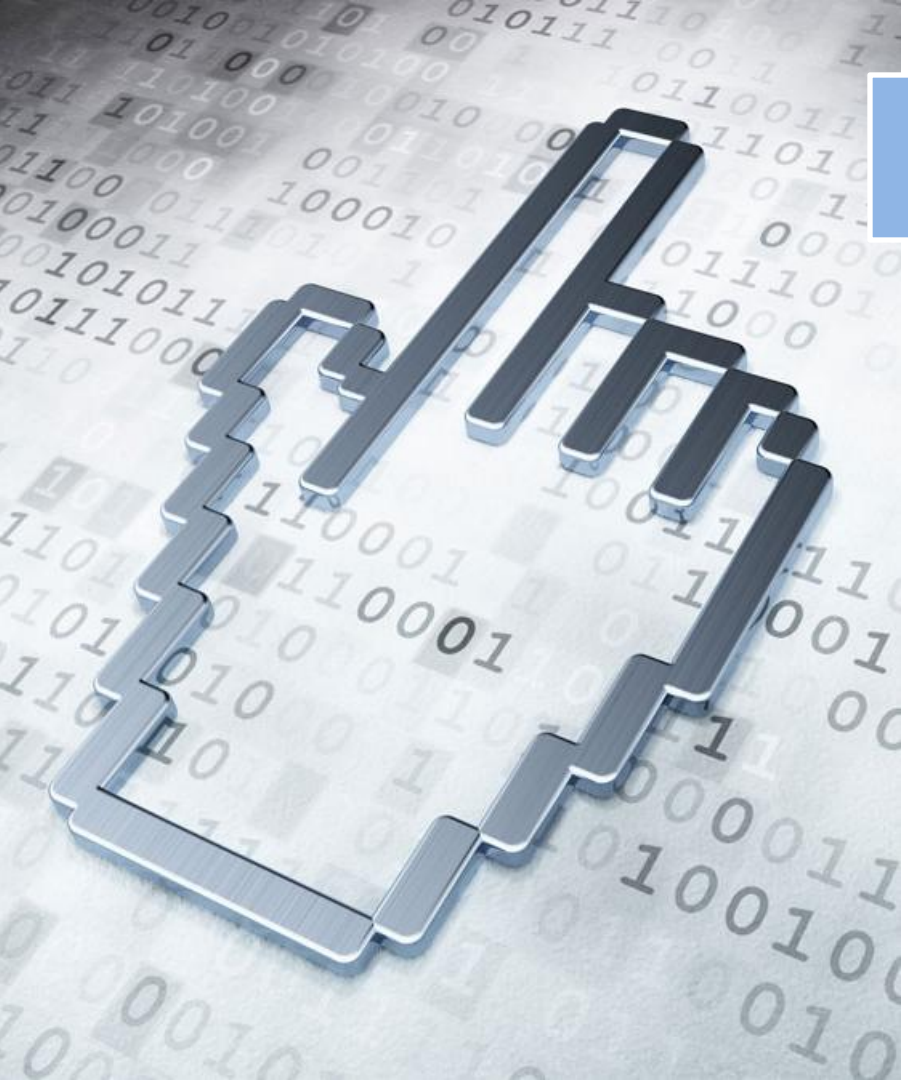
- 指令中红色的部分表示指令的操作符
- x[i]，其中的i是与首地址的偏移地址

三地址指令的四元式表示

➢ $x = y$ op z	(op , y , z , x)
➢ $x =$ op y	(op , y , _ , x)
➢ $x = y$	(= , y , _ , x)
➢ if x relop y goto n	(relop , x , y , n)
➢ goto n	(goto , _ , _ , n)
➢ param x	(param, _ , _ , x)
$y =$call p, n	(call , p , n , y)
return x	(return, _ , _ , x)
➢ $x = y[i]$	(=[] , y , i , x)
$x[i] = y$	([]= , y , x , i)
➢ $x =$ &y	(=& , y , _ , x)
$x =$ *y	(=* , y , _ , x)
*$x = y$	(*= , y , _ , x)

三地址指令序列唯一确定了运算完成的顺序

- 第一个分量是操作符
- 第二三个分量是操作数
- 第四个分量是结果地址



本章内容

6.1 声明语句的翻译

- 声明语句翻译的主要任务：收集标识符的类型等属性信息，并为每一个名字分配一个相对地址
- 类型的重要作用
  - 类型检查：验证程序运行时的行为是否遵守语言的类型的规定。语义分析任务多数都与类型检查相关。
  - 辅助翻译：计算数组引用的地址、加入显式的类型转换、选择正确版本的算术运算符都要用到类型信息。

名字的类型和相对地址信息保存在相应的符号表记录中

类型表达式 (*Type Expressions*)

- 基本类型是类型表达式
  - *integer*
  - *real*
  - *char*
  - *boolean*
  - *type_error* (出错类型)
  - *void* (无类型)

类型表达式 (*Type Expressions*)

- 基本类型是类型表达式
- 可以为类型表达式命名，类型名也是类型表达式
- 将类型构造符(*type constructor*)作用于类型表达式可以构成新的类型表达式
  - 数组构造符*array*
    - 若*T*是类型表达式，则*array* (*I*, *T*)是类型表达式(*I*是一个整数)

类型	类型表达式
int [3]	*array* (3, *int*)
int [2][3]	*array* (2, *array*(3, *int*))

类型表达式 (*Type Expressions*)

- 基本类型是类型表达式
- 可以为类型表达式命名，类型名也是类型表达式
- 将类型构造符(*type constructor*)作用于类型表达式可以构成新的类型表达式
  - 数组构造符*array*
  - 指针构造符*pointer*
    - 若*T*是类型表达式，则 *pointer* (*T*) 是类型表达式，它表示一个指针类型

类型表达式 (*Type Expressions*)

- 基本类型是类型表达式
- 可以为类型表达式命名，类型名也是类型表达式
- 将类型构造符(*type constructor*)作用于类型表达式可以构成新的类型表达式
  - 数组构造符*array*
  - 指针构造符*pointer*
  - 笛卡尔乘积构造符×
    - 若T_1和T_2是类型表达式，则笛卡尔乘积$T_1 \times T_2$是类型表达式

类型表达式 (*Type Expressions*)

- 基本类型是类型表达式
- 可以为类型表达式命名，类型名也是类型表达式
- 将类型构造符(*type constructor*)作用于类型表达式可以构成新的类型表达式
  - 数组构造符*array*
  - 指针构造符*pointer*
  - 笛卡尔乘积构造符×
  - 函数构造符→
    - 若T_1、T_2、…、T_n和R是类型表达式，则$T_1 \times T_2 \times \ldots \times T_n \rightarrow R$是类型表达式

类型表达式 (*Type Expressions*)

- 基本类型是类型表达式
- 可以为类型表达式命名，类型名也是类型表达式
- 将类型构造符(*type constructor*)作用于类型表达式可以构成新的类型表达式
  - 数组构造符*array*
  - 指针构造符*pointer*
  - 笛卡尔乘积构造符×
  - 函数构造符→
  - 记录构造符*record*
    - 若有标识符N_1、N_2、…、N_n与类型表达式T_1、T_2、…、T_n，则

      record $((N_1 \times T_1) \times (N_2 \times T_2) \times \ldots \times (N_n \times T_n))$ 是一个类型表达式

例

- 设有C程序片段：

```
struct   stype
{   char[8] name;
    int  score;
};
stype[50] table;
stype* p;
```

- 和stype绑定的类型表达式
  - record ((name×array(8, char)) × (score × integer))
- 和table绑定的类型表达式
  - array (50, stype)
- 和p绑定的类型表达式
  - pointer (stype)

类型等价

- 许多类型检查都需要判断两个类型表达式是否等价：
- 结构等价：当且仅当下列条件之一成立时，称两个类型T_1和T_2是结构等价的：
  - T_1和T_2是相同的基本类型；
  - T_1和T_2是将同一类型构造符应用于结构等价的类型上形成的；
  - T_1是表示T_2的类型名。
- 名字等价：两个类型表达式名字等价当且仅当上述前两条之一成立。关注类型名称的一致性。

类型等价

例：有一段程序声明

```
typedef  cell*  link;
link  next ,  last;
cell * p , q;
```

其中声明变量的类型表达式

变量	类型表达式
next	link
last	link
p	pointer(cell)
q	pointer(cell)

按名字等价判断：变量next和last类型相同，
变量p，q类型相同。

按结构等价判断：所有4个变量类型都相同。

局部变量的存储分配

- 对于声明语句，语义分析的主要任务就是①收集标识符的类型等属性信息，②为每一个名字分配一个相对地址
  - 从类型表达式可以知道该类型在运行时刻所需的存储单元数量称为类型的宽度(*width*)
  - 在编译时刻，可以根据类型的宽度为每一个名字分配一个相对地址（相对于数据区域开始位置的偏移量）
- 名字的类型和相对地址信息保存在符号表记录中

变量声明语句的SDT

enter(name, type, offset):在符号表中为名字*name*创建记录,将*name*的类型设置为*type*,相对地址设置为*offset*

① $P \rightarrow \{\ offset = 0\ \}\ D$

② $D \rightarrow T\ \text{id};\{\ enter(\ \text{id}.lexeme,\ T.type,\ offset\ );\ offset = offset + T.width;\ \}D$

③ $D \rightarrow \varepsilon$

④ $T \rightarrow B\ \ \{\ t = B.type;\ w = B.width;\ \}\ C\ \ \{\ T.type = C.type;\ T.width = C.width;\ \}$

⑤ $T \rightarrow \uparrow T_1\{\ T.type = pointer(\ T_1.type\ );\ T.width = 4;\ \}$

⑥ $B \rightarrow \text{int}\ \{\ B.type = int;\ B.width = 4;\ \}$

⑦ $B \rightarrow \text{real}\{\ B.type = real;\ B.width = 8;\ \}$

⑧ $C \rightarrow \varepsilon\ \ \{\ C.type = t;\ C.width = w;\ \}$

⑨ $C \rightarrow [\text{num}]C_1\ \{\ C.type = array(\ num.val,\ C_1.type\ );\ C.width = num.val * C_1.width;\ \}$

符号	综合属性
B	*type, width*
C	*type, width*
T	*type, width*

变量	作用
offset	下一个可用的相对地址
t, w	将类型和宽度信息从语法分析树中的*B*结点传递到对应于产生式$C \rightarrow \varepsilon$的结点

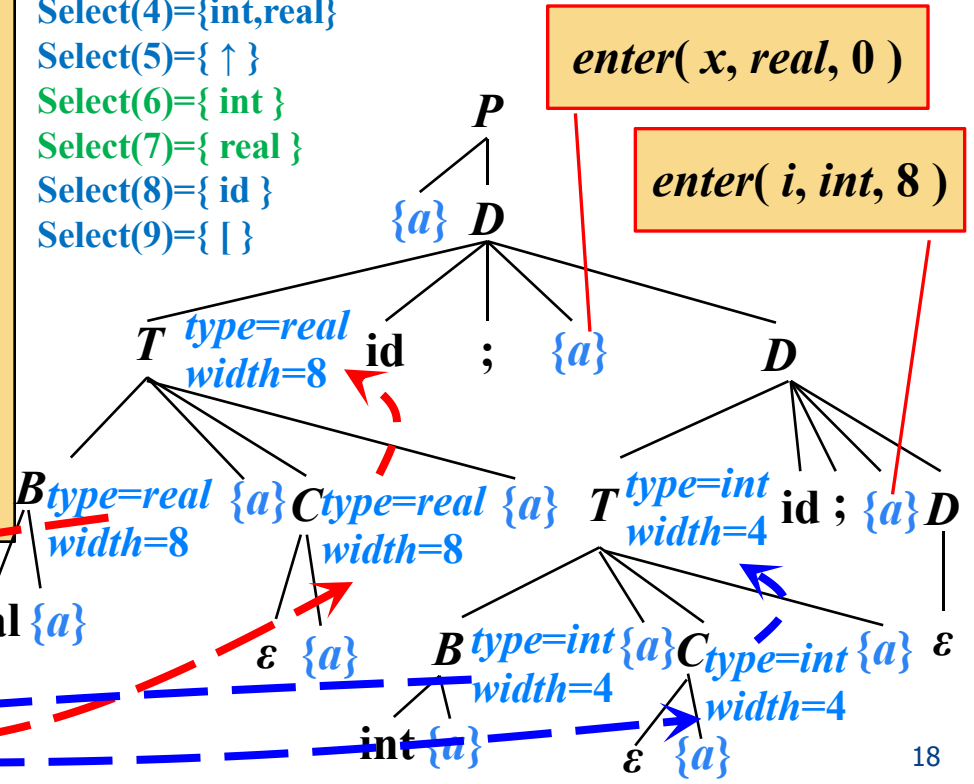
例: “*real x; int i;*”的语法制导翻译

①$P \rightarrow$ { *offset = 0* } D
②$D \rightarrow T$ id; { *enter(id.lexeme, T.type, offset);*
offset = offset + T.width; }D
③$D \rightarrow \varepsilon$
④$T \rightarrow B$ { *t =B.type; w=B.width;*}
C { *T.type = C.type; T.width = C.width;* }
⑤$T \rightarrow \uparrow T_1$ { *T.type = pointer(T_1.type); T.width = 4;* }
⑥$B \rightarrow$ int { *B.type = int; B.width = 4;* }
⑦$B \rightarrow$ real { *B.type = real; B.width = 8;* }
⑧$C \rightarrow \varepsilon$ { *C.type=t; C.width=w;* }
⑨$C \rightarrow$ [num]C_1 { *C.type = array(num.val, C_1.type);*
*C.width = num.val * C_1.width;* }

Select(1)={int,real, ↑,$}
Select(2)={int,real, ↑}
Select(3)={$}
Select(4)={int,real}
Select(5)={ ↑ }
Select(6)={ int }
Select(7)={ real }
Select(8)={ id }
Select(9)={ [}

enter(x, real, 0)

enter(i, int, 8)



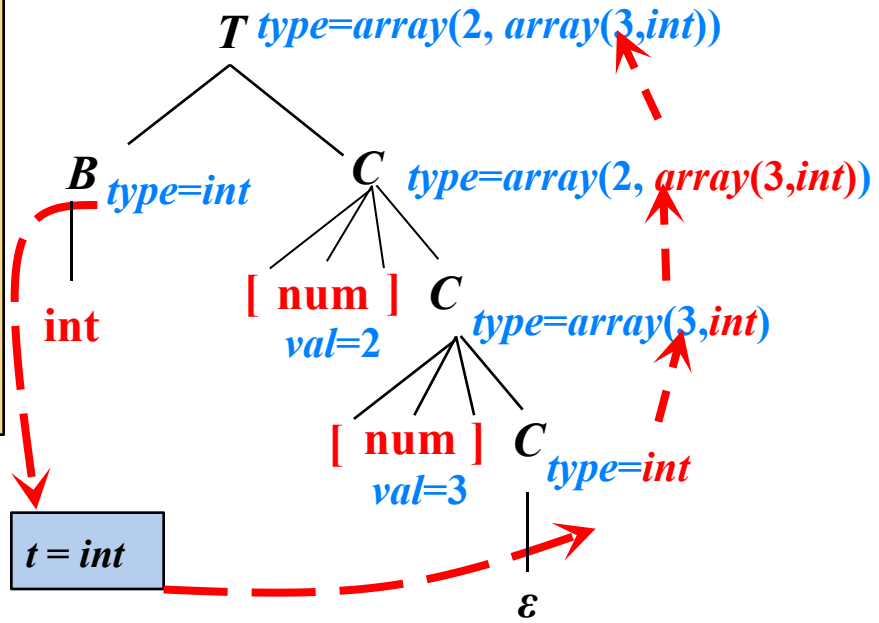
offset = 12

t = *int*
w = 4

数组类型表达式的语法制导翻译

①$P \rightarrow$ { $offset = 0$ } D
②$D \rightarrow T$ id;{ $enter($ id.$lexeme, T.type, offset$);
$offset = offset + T.width$; }$D$
③$D \rightarrow \varepsilon$
④$T \rightarrow B$ { $t = B.type$; $w = B.width$;}
C { $T.type = C.type$; $T.width = C.width$; }
⑤$T \rightarrow \uparrow T_1$ { $T.type = pointer(T_1.type)$; $T.width = 4$; }
⑥$B \rightarrow$ int { $B.type = int$; $B.width = 4$; }
⑦$B \rightarrow$ real{ $B.type = real$; $B.width = 8$; }
⑧$C \rightarrow \varepsilon$ { $C.type = t$; $C.width = w$; }
⑨$C \rightarrow$ [num]C_1 { $C.type = array(num.val, C_1.type)$;
$C.width = num.val * C_1.width$; }

例： “*int*[2][3]”



例: 数组类型表达式“*int*[2][3]”的语法制导翻译

①$P \rightarrow$ { *offset* = 0 } D

②$D \rightarrow T$ id; { *enter*(id.*lexeme*, *T.type*, *offset*);
offset = *offset* + *T.width*; }D

③$D \rightarrow \varepsilon$

④$T \rightarrow B$ { *t* =*B.type*; *w*=*B.width*;}
C { *T.type* = *C.type*; *T.width* = *C.width*; }

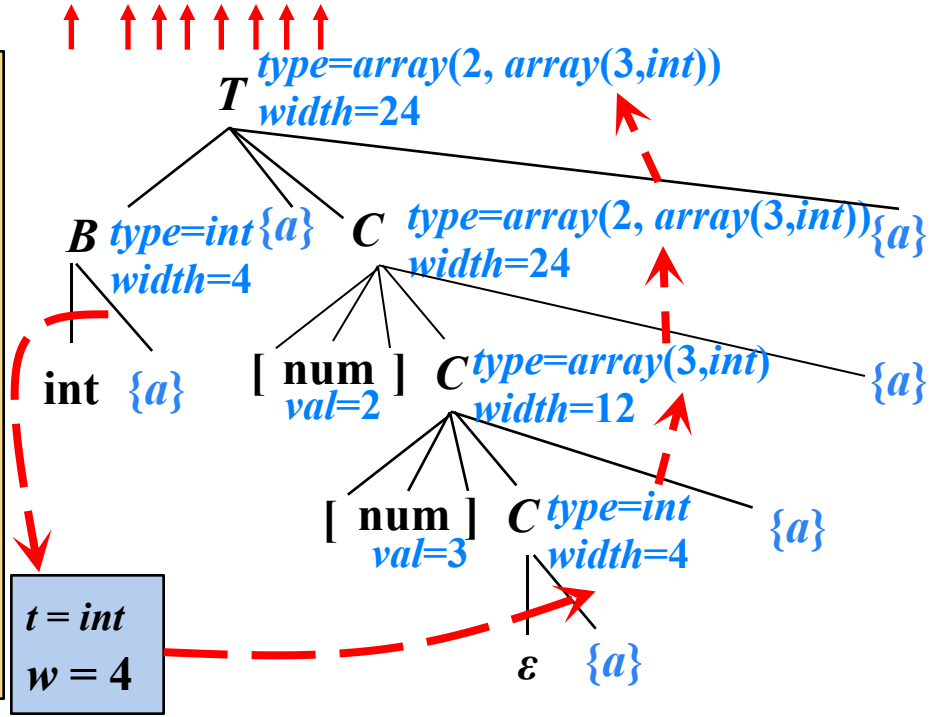
⑤$T \rightarrow \uparrow T_1$ { *T.type* = *pointer*(T_1.*type*); *T.width* = 4; }

⑥$B \rightarrow$ int { *B.type* = *int*; *B.width* = 4; }

⑦$B \rightarrow$ real { *B.type* = *real*; *B.width* = 8; }

⑧$C \rightarrow \varepsilon$ { *C.type*=*t*; *C.width*=*w*; }

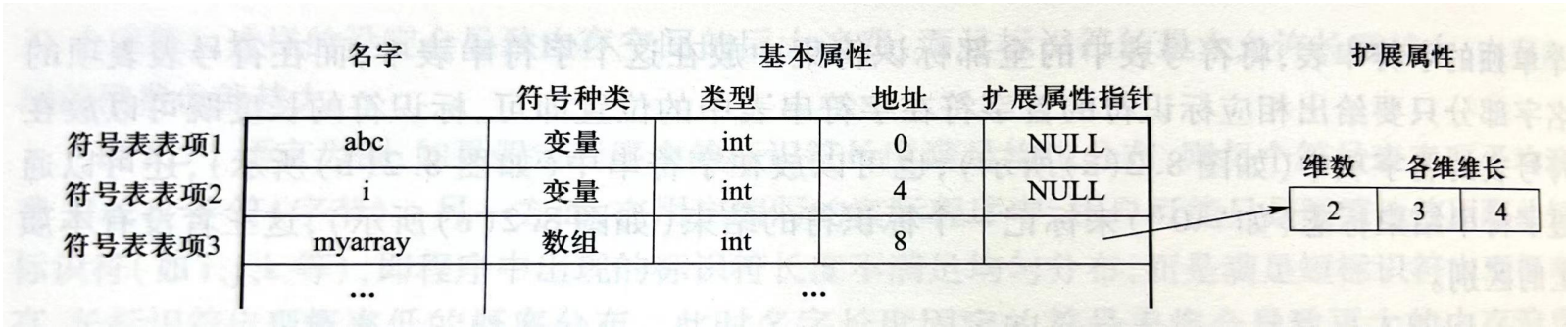
⑨$C \rightarrow$ [num]C_1 { *C.type* = *array*(*num.val*, C_1.*type*);
C.width = *num.val* * C_1.*width*; }

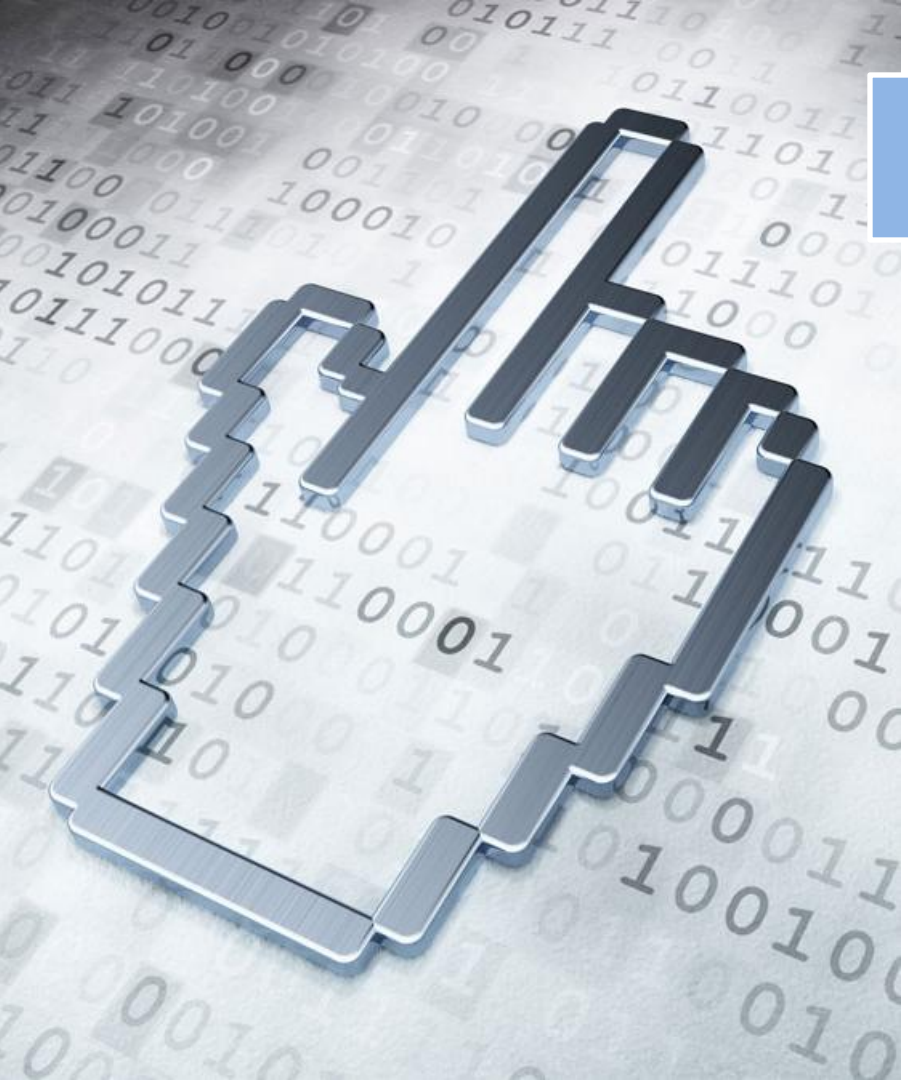


数组：只计算和保存整个数组的起始相对地址。

多维数组声明语句的翻译

- 将数组属性信息存入符号表，可以建立数组的内情向量表
  - 各维类型、各维维长、各维宽度等
  - 维数、下标上界、下标下界
  - 按行存放、按列存放——影响具体元素地址的计算
  - ……





本章内容

6.2 赋值语句的翻译

- 6.2.1简单赋值语句的翻译
- 6.2.2数组引用的翻译

6.2.1 简单赋值语句的翻译

- 赋值语句的基本文法

① $S \rightarrow \mathrm{id} = E;$

② $E \rightarrow E_1 + E_2$

③ $E \rightarrow E_1 * E_2$

④ $E \rightarrow -E_1$

⑤ $E \rightarrow (E_1)$

⑥ $E \rightarrow \mathrm{id}$

- 赋值语句翻译的主要任务
  - 生成对表达式求值的三地址码

- 例 $x = (a + b) * c;$
- 三地址码

$t_1 = a + b$

$t_2 = t_1 * c$

$x = t_2$

赋值语句的SDT

lookup(name)：查询符号表返回 *id* 对应的地址

gen(code)：生成三地址指令*code*

newtemp( )：生成一个新的临时变量*t*，返回*t*的地址

$S \rightarrow \text{id} = E;$ { *p = lookup*(id.*lexeme*); *if p == nil then error ;*
S.code = E.code ||
gen(p '=' E.addr); }

$E \rightarrow E_1 + E_2$ { *E.addr = newtemp( );*
E.code = E_1.code || E_2.code||
gen(E.addr '=' E_1.addr '+' E_2.addr); }

$E \rightarrow E_1 * E_2$ { *E.addr = newtemp( );*
E.code = E_1.code || E_2.code ||
gen(E.addr '=' E_1.addr '' E_2.addr);* }

$E \rightarrow -E_1$ { *E.addr = newtemp( );*
E.code = E_1.code||
gen(E.addr '=' 'uminus' E_1.addr); }

$E \rightarrow (E_1)$ { *E.addr = E_1.addr;*
E.code= E_1.code; }

$E \rightarrow \text{id}$ { *E.addr = lookup*(id.*lexeme*); *if E.addr == nil then error ;*
E.code = ''; }

符号	综合属性
S	*code*
E	*code* *addr*

增量翻译 (*Incremental Translation*)

$S \rightarrow$ id = E; { *p = lookup*(id.*lexeme*); *if p == nil then error* ;
S.code = E.code ||
gen(*p* '=' *E.addr*); }

$E \rightarrow E_1 + E_2$ { *E.addr = newtemp*( );
E.code = E_1*.code* || E_2*.code*||
gen(*E.addr* '=' E_1*.addr* '+' E_2*.addr*); }

$E \rightarrow E_1 * E_2$ { *E.addr = newtemp*( );
E.code = E_1*.code* || E_2*.code* ||
gen(*E.addr* '=' E_1*.addr* '*' E_2*.addr*); }

$E \rightarrow -E_1$ { *E.addr = newtemp*( );
E.code = E_1*.code*||
gen(*E.addr* '=' '*uminus*' E_1*.addr*); }

$E \rightarrow (E_1)$ { *E.addr* = E_1*.addr*;
E.code= E_1*.code*; }

$E \rightarrow$ id { *E.addr = lookup*(id.*lexeme*); *if E.addr == nil then error* ;
E.code = ''; }

不再使用*code*属性，*gen*( )负责构造出一个新的三地址指令，再将它添加到至今为止已生成的指令序列之后

- 综合属性code的值总是将产生式右部符号对应的三地址码按生成顺序连接以后，在后面追加新的三地址指令

例

①$S \rightarrow$ id $= E$; { *p = lookup*(id.*lexeme*);
if p==nil then error ;
gen(p '=' E.addr); }

②$E \rightarrow E_1 + E_2$ { *E.addr = newtemp( );*
gen(E.addr '=' E_1.addr '+' E_2.addr); }

③$E \rightarrow E_1 * E_2$ { *E.addr = newtemp( );*
gen(E.addr '=' E_1.addr '' E_2.addr);* }

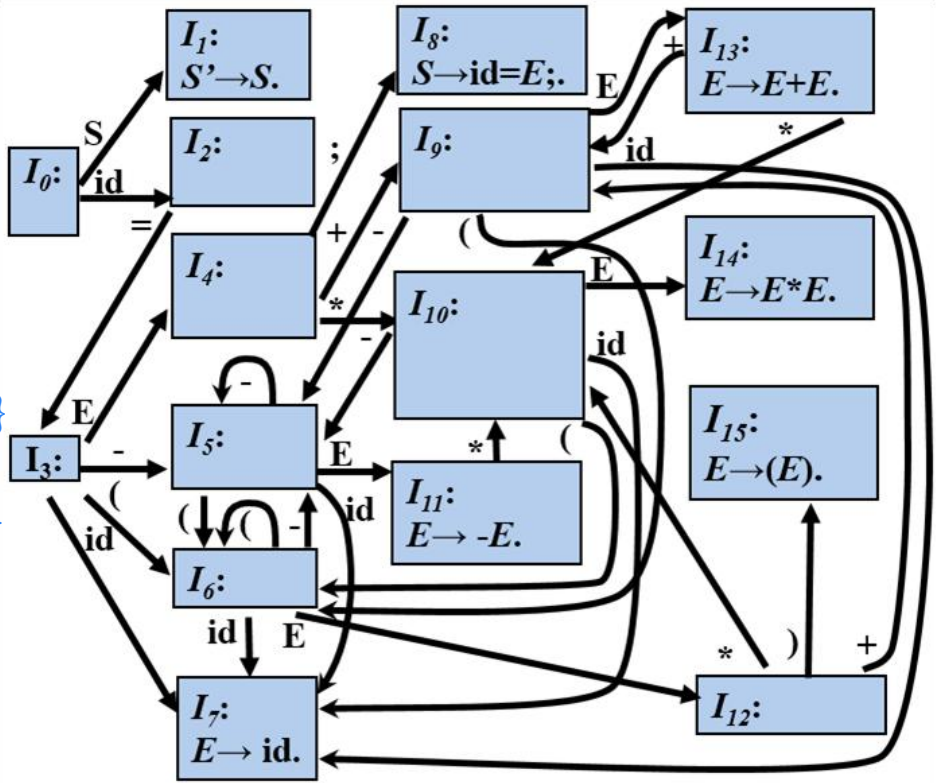
④$E \rightarrow -E_1$ { *E.addr = newtemp( );*
gen(E.addr '=' 'uminus' E_1.addr); }

⑤$E \rightarrow (E_1)$ { *E.addr = E_1.addr ;* }

⑥$E \rightarrow$ id { *E.addr = lookup*(id.*lexeme*);
if E.addr==nil then error ; }

例：$x = (a + b) * c$;

0	2	3	6	7
$	id	=	(	id
	x			*a*



例

①$S \rightarrow$ id = E; { *p = lookup*(id.*lexeme*);
if p==nil then error ;
gen(p '=' E.addr); }

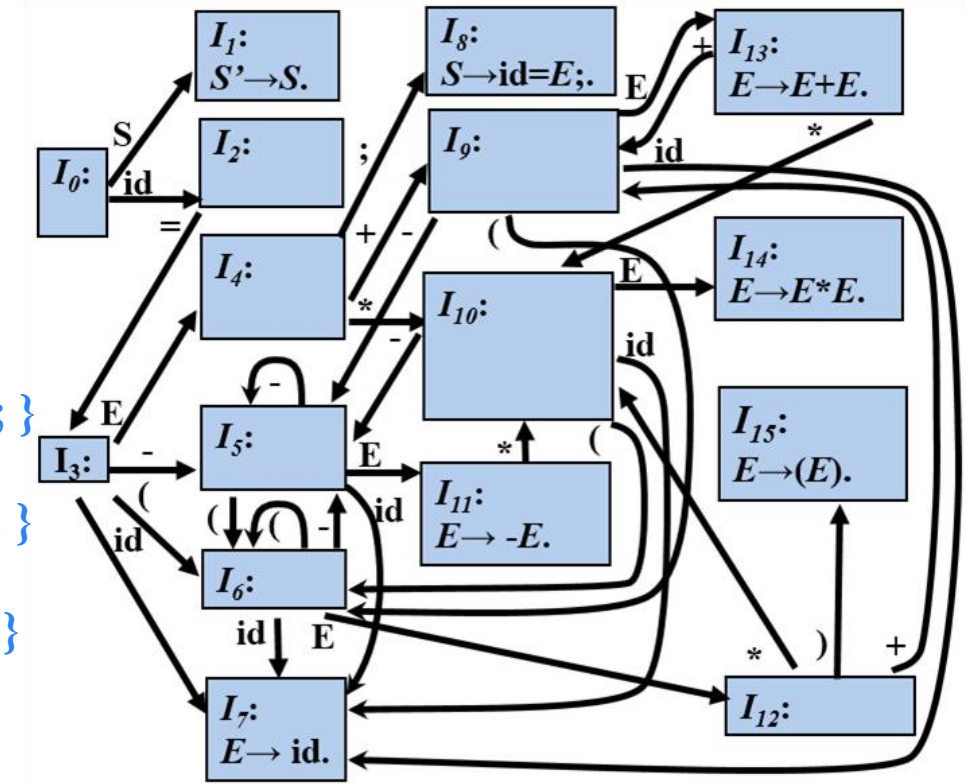
②$E \rightarrow E_1 + E_2$ { *E.addr = newtemp( );*
gen(E.addr '=' E_1.addr '+' E_2.addr); }

③$E \rightarrow E_1 * E_2$ { *E.addr = newtemp( );*
gen(E.addr '=' E_1.addr '' E_2.addr);* }

④$E \rightarrow -E_1$ { *E.addr = newtemp( );*
gen(E.addr '=' 'uminus' E_1.addr); }

⑤$E \rightarrow (E_1)$ { *E.addr = E_1.addr ;* }

⑥$E \rightarrow$ id { *E.addr = lookup*(id.*lexeme*);
if E.addr==nil then error ; }



例：$x = (a + b) * c$;

0	2	3	6	12	9	7
$	id	=	(	*E*	+	id
	x			*a*		*b*

例

①$S \rightarrow \text{id} = E$; { *p = lookup(*id.*lexeme);*
if p==nil then error ;
gen(p '=' E.addr); }

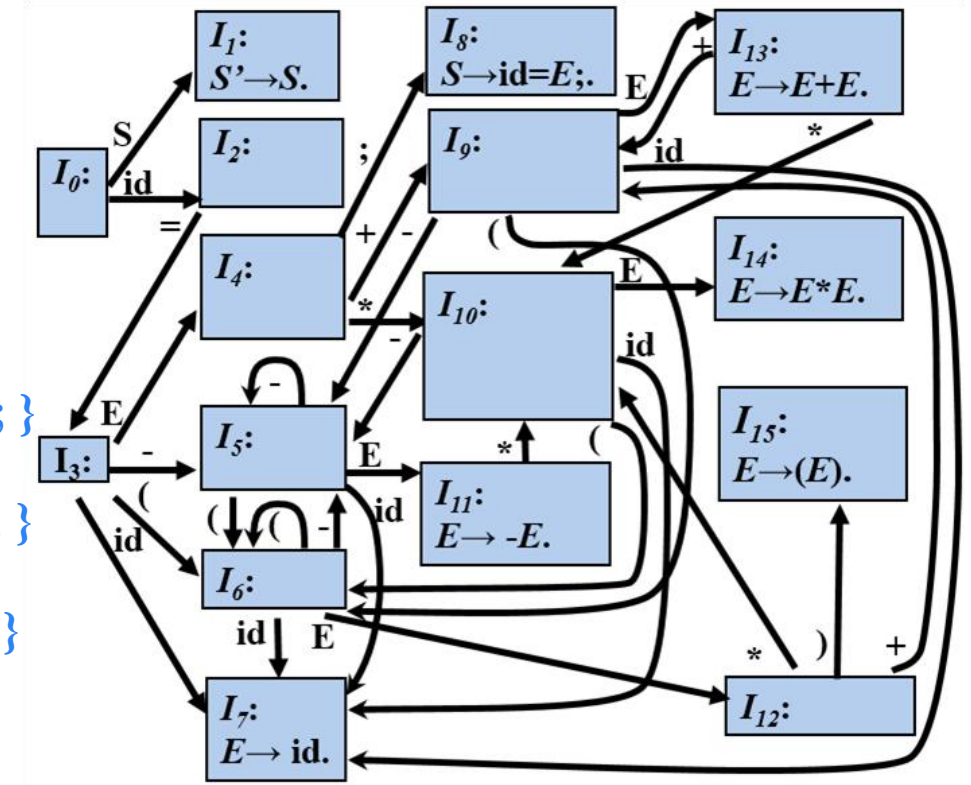
②$E \rightarrow E_1 + E_2$ { *E.addr = newtemp( );*
gen(E.addr '=' E_1.addr '+' E_2.addr); }

③$E \rightarrow E_1 * E_2$ { *E.addr = newtemp( );*
gen(E.addr '=' E_1.addr '' E_2.addr); }*

④$E \rightarrow -E_1$ { *E.addr = newtemp( );*
gen(E.addr '=' 'uminus' E_1.addr); }

⑤$E \rightarrow (E_1)$ { *E.addr = E_1.addr ; }*

⑥$E \rightarrow \text{id}$ { *E.addr = lookup(*id.*lexeme);*
if E.addr==nil then error ; }



例：$x = (a + b) * c$;

0	2	3	6	12	9	13
$	id	=	(	E	+	E
	x			a		b

$t_1 = a + b$

例

①$S \rightarrow \text{id} = E$; { *p = lookup*(id.*lexeme*);
if p==nil then error ;
gen(p '=' E.addr); }

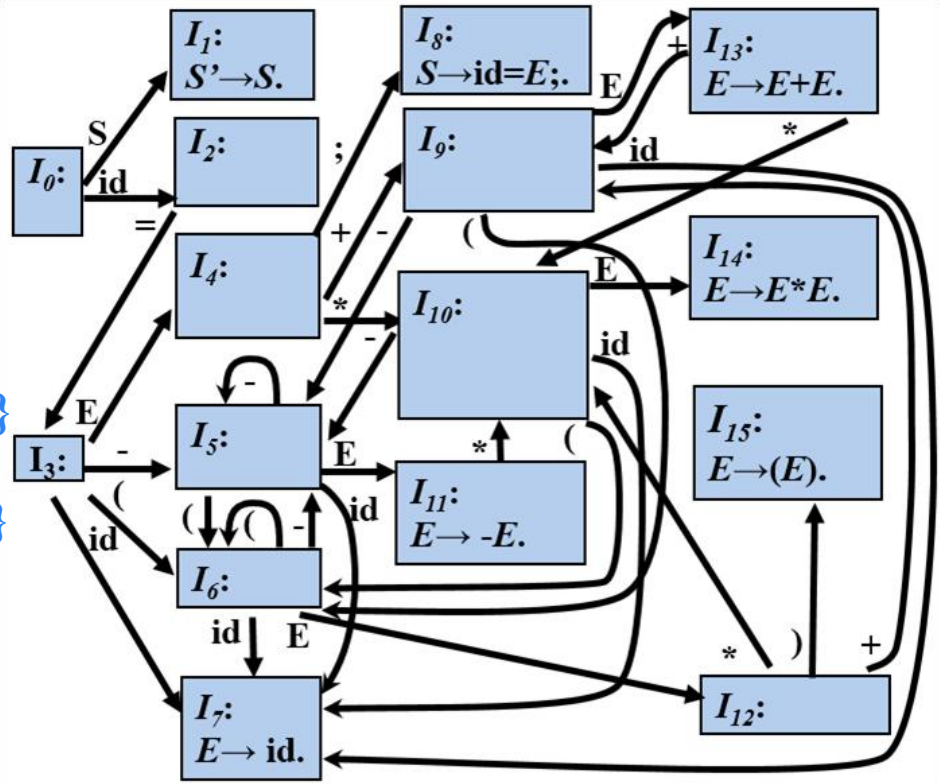
②$E \rightarrow E_1 + E_2$ { *E.addr = newtemp( );*
gen(E.addr '=' E_1.addr '+' E_2.addr); }

③$E \rightarrow E_1 * E_2$ { *E.addr = newtemp( );*
gen(E.addr '=' E_1.addr '' E_2.addr);* }

④$E \rightarrow -E_1$ { *E.addr = newtemp( );*
gen(E.addr '=' 'uminus' E_1.addr); }

⑤$E \rightarrow (E_1)$ { *E.addr = E_1.addr ;* }

⑥$E \rightarrow \text{id}$ { *E.addr = lookup*(id.*lexeme*);
if E.addr==nil then error ; }



例：$x = (a + b) * c$;

0	2	3	6	12	15
$	id	=	(	E	)
	x			t_1	

$t_1 = a + b$

例

①$S \rightarrow \text{id} = E$; { *p = lookup(id.lexeme);*
if p==nil then error ;
gen(p '=' E.addr); }

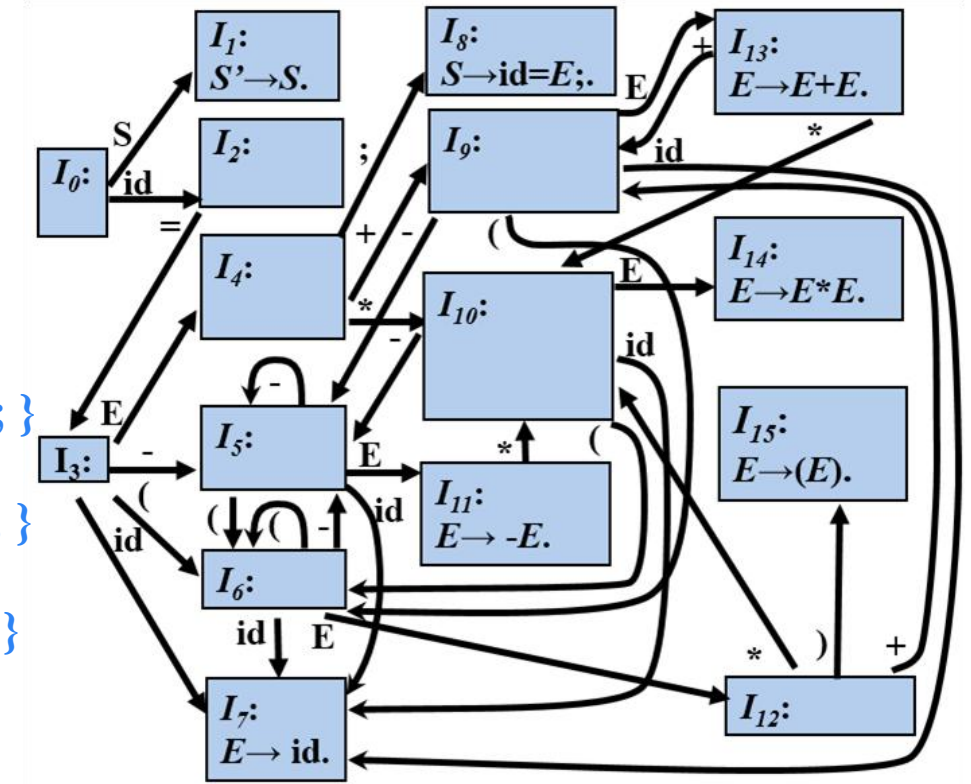
②$E \rightarrow E_1 + E_2$ { *E.addr = newtemp( );*
gen(E.addr '=' E_1.addr '+' E_2.addr); }

③$E \rightarrow E_1 * E_2$ { *E.addr = newtemp( );*
gen(E.addr '=' E_1.addr '' E_2.addr);* }

④$E \rightarrow -E_1$ { *E.addr = newtemp( );*
gen(E.addr '=' 'uminus' E_1.addr); }

⑤$E \rightarrow (E_1)$ { *E.addr = E_1.addr ;* }

⑥$E \rightarrow \text{id}$ { *E.addr = lookup(id.lexeme);*
if E.addr==nil then error ; }



例：$x = (a + b) * c$;

0	2	3	4	10	7
$	id	=	E	*	id
	x		t_1		c

$t_1 = a + b$

例

①$S \rightarrow \text{id} = E$; { *p = lookup(id.lexeme);*
if p==nil then error ;
gen(p '=' E.addr); }

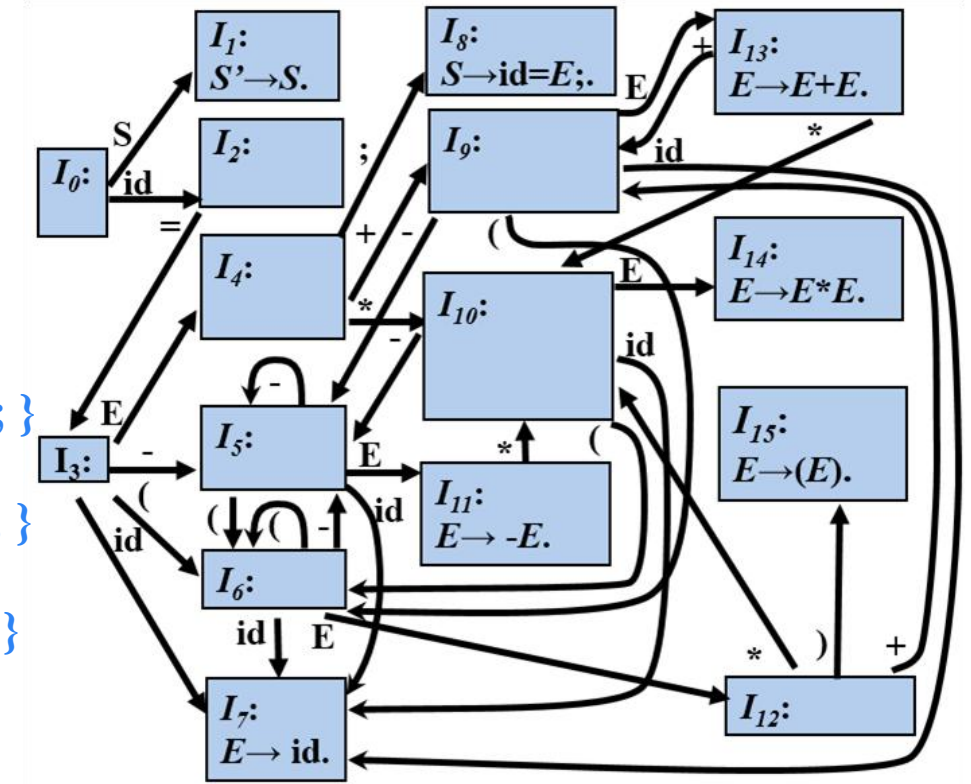
②$E \rightarrow E_1 + E_2$ { *E.addr = newtemp( );*
gen(E.addr '=' E_1.addr '+' E_2.addr); }

③$E \rightarrow E_1 * E_2$ { *E.addr = newtemp( );*
gen(E.addr '=' E_1.addr '' E_2.addr); }*

④$E \rightarrow -E_1$ { *E.addr = newtemp( );*
gen(E.addr '=' 'uminus' E_1.addr); }

⑤$E \rightarrow (E_1)$ { *E.addr = E_1.addr ; }*

⑥$E \rightarrow \text{id}$ { *E.addr = lookup(id.lexeme);*
if E.addr==nil then error ; }



例：$x = (a + b) * c$;

0	2	3	4	10	14
$	id	=	E	*	E
	x		t_1		c

$t_1 = a + b$

$t_2 = t_1 * c$

例

①$S \rightarrow \text{id} = E$; { *p = lookup(*id.*lexeme);*
if p==nil then error ;
gen(p '=' E.addr); }

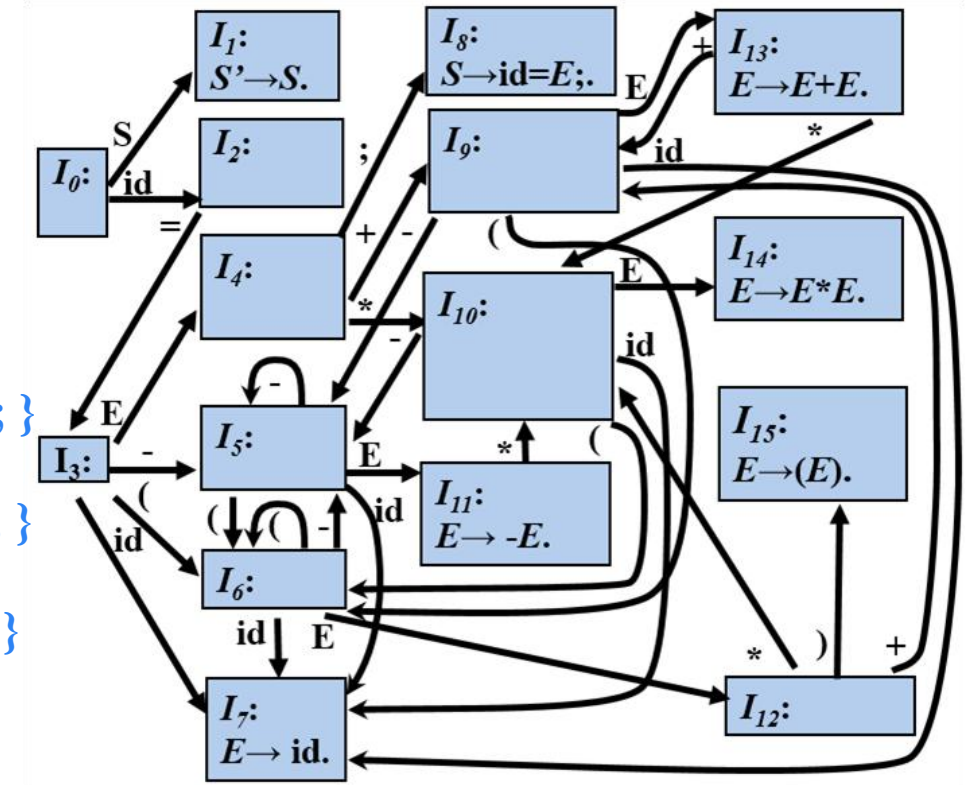
②$E \rightarrow E_1 + E_2$ { *E.addr = newtemp( );*
gen(E.addr '=' E_1.addr '+' E_2.addr); }

③$E \rightarrow E_1 * E_2$ { *E.addr = newtemp( );*
gen(E.addr '=' E_1.addr '' E_2.addr); }*

④$E \rightarrow -E_1$ { *E.addr = newtemp( );*
gen(E.addr '=' 'uminus' E_1.addr); }

⑤$E \rightarrow (E_1)$ { *E.addr = E_1.addr ; }*

⑥$E \rightarrow \text{id}$ { *E.addr = lookup(*id.*lexeme);*
if E.addr==nil then error ; }



例：$x = (a + b) * c$;

0	2	3	4	8
$	id	=	E	;
	x		t_2	

$t_1 = a + b$

$t_2 = t_1 * c$

$x = t_2$

例

①$S \rightarrow$ id = E; { *p = lookup*(id.*lexeme*);
if p==nil then error ;
gen(p '=' E.addr); }

②$E \rightarrow E_1 + E_2$ { *E.addr = newtemp( );*
gen(E.addr '=' E_1.addr '+' E_2.addr); }

③$E \rightarrow E_1 * E_2$ { *E.addr = newtemp( );*
gen(E.addr '=' E_1.addr '' E_2.addr);* }

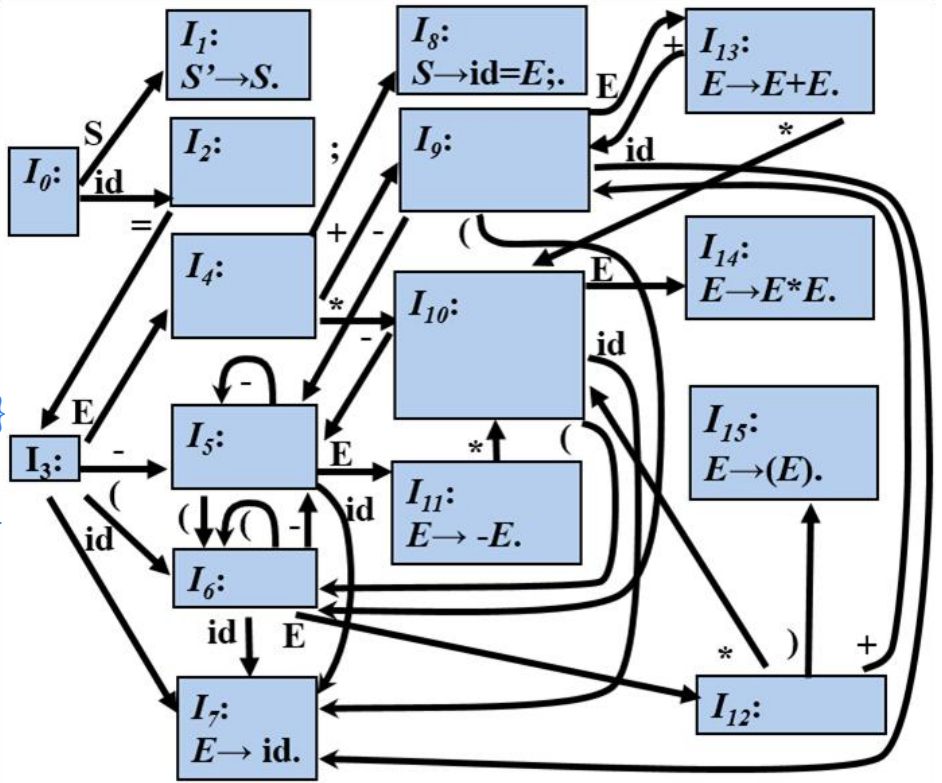
④$E \rightarrow -E_1$ { *E.addr = newtemp( );*
gen(E.addr '=' 'uminus' E_1.addr); }

⑤$E \rightarrow (E_1)$ { *E.addr = E_1.addr ;* }

⑥$E \rightarrow$ id { *E.addr = lookup*(id.*lexeme*);
if E.addr==nil then error ; }

例：$x = (a + b) * c$;

0	1
$	S



$t_1 = a + b$

$t_2 = t_1 * c$

$x = t_2$

6.2.2 数组引用的翻译

➢ 赋值语句的基本文法

$$S \rightarrow \mathrm{id} = E; \mid L = E;$$

$$E \rightarrow E_1 + E_2 \mid -E_1 \mid (E_1) \mid \mathrm{id} \mid L$$

$$L \rightarrow \mathrm{id}\ [E] \mid L_1\ [E]$$

将数组引用翻译成三地址码时要解决的主要问题是确定数组元素的存放地址，也就是数组元素的寻址

例

- *int a*[3][5][8] ;

  type(*a*) = *array*(3, *array*(5, *array*(8, *int*))),

  一个整型变量占用4个字节，

  则*addr*(*a*[2][3][4]) = *base* + 2 * w_1 + 3*w_2 + 4*w_3

  = *base* + 2*160 + 3*32 + 4*4

a[2]的宽度

a[2][3]的宽度

a[2][3][4]的宽度

数组元素寻址 (*Addressing Array Elements*)

- 一维数组：$array(n_1, type)$
  - 假设每个数组元素的宽度是w，则数组元素$a[i]$的相对地址是：

$$base+i\times w$$

其中，$base$是数组的基地址，$i\times w$是偏移地址

- 二维数组：$array(n_1, array(n_2, type))$
  - 假设一行的宽度是w_1，同一行中每个数组元素的宽度是w_2，则数组元素$a[i_1]\ [i_2]$的相对地址是：

$$\underbrace{base+i_1\times w_1+i_2\times w_2}_{\text{偏移地址}}$$

$array(n_1, array(n_2, array(n_3, \ldots array(n_k, type)\ldots)))$

- k维数组
  - 数组元素$a[i_1]\ [i_2]\ \ldots[i_k]$的相对地址是：

$$\underbrace{base+i_1\times w_1+i_2\times w_2+\ldots+i_k\times w_k}_{\text{偏移地址}}$$

$w_1 \rightarrow a[i_1]$ 的宽度
$w_2 \rightarrow a[i_1]\ [i_2]$ 的宽度
…
$w_k \rightarrow a[i_1]\ [i_2]\ \ldots[i_k]$的宽度

带有数组引用的赋值语句的翻译

- 例1

  假设 $type(a)=array(n, int)$,

  - 源程序片段
    - $c = a[i];$

  $$addr(a[i]) = base + i*4$$

  - 三地址码

    $t_1 = i * 4$

    $t_2 = a[t_1]$

    $c = t_2$

带有数组引用的赋值语句的翻译

- 例2

  假设 $type(a) = array(3, array(5, int))$，

  - 源程序片段
    - $c = a[i_1][i_2];$ $addr(a[i_1][i_2]) = base + i_1*20 + i_2*4$
  - 三地址码

    $t_1 = i_1 * 20$

    $t_2 = i_2 * 4$

    $t_3 = t_1 + t_2$

    $t_4 = a\,[\,t_3\,]$

    $c = t_4$

t_1 t_2

数组元素寻址的*SDT*

- 赋值语句的基本文法

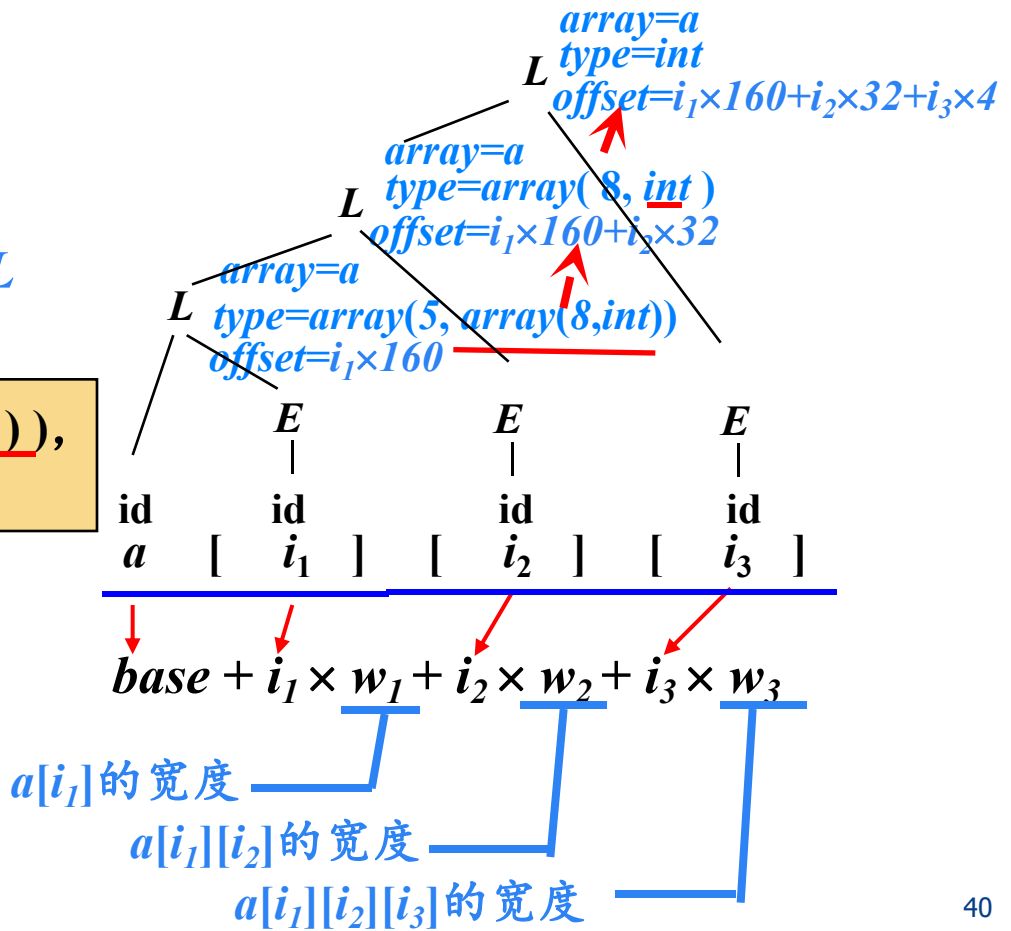
$S \rightarrow \mathrm{id} = E; \mid L = E;$

$E \rightarrow E_1 + E_2 \mid -E_1 \mid (E_1) \mid \mathrm{id} \mid L$

$L \rightarrow \mathrm{id}\ [E] \mid L_1\ [E]$

假设 *type*(*a*)= *array*(3, *array*(5, *array*(8, *int*))), 翻译语句片段 “$a[i_1][i_2][i_3]$”

- *L*的综合属性
  - *L.type*：*L*生成的数组元素的类型
  - *L.offset*：指示一个临时变量，该临时变量用于累加公式中的$i_j \times w_j$项，从而计算数组元素的偏移量
  - *L.array*：数组名在符号表的入口地址



数组元素寻址的SDT

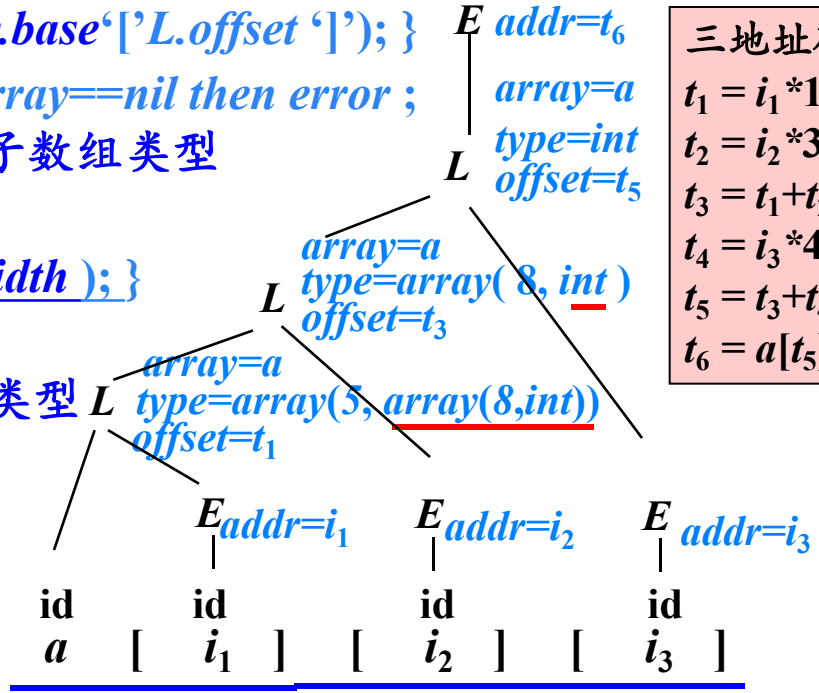
假设 type(a)= array(3, array(5, array(8, int))),
翻译语句片段 “$a[i_1][i_2][i_3]$”

$addr(a[i_1][i_2][i_3]) = base + i_1 \times w_1 + i_2 \times w_2 + i_3 \times w_3$
(t_1, t_2, t_4)

$S \rightarrow$ id = E;
| L = E; { gen($L.array.base$ ‘[’ $L.offset$ ‘]’ ‘=’ $E.addr$); }

$E \rightarrow E_1 + E_2 \mid -E_1 \mid (E_1) \mid$ id
| L { $E.addr$=newtemp(); gen($E.addr$ ‘=’ $L.array.base$ ‘[’ $L.offset$ ‘]’); }

$L \rightarrow$ id [E] { $L.array$ = lookup(id.lexeme); if $L.array$==nil then error ;
$L.type = L.array.type.elem$; //获得子数组类型
$L.offset$ = newtemp();
gen($L.offset$ ‘=’ $E.addr$ ‘*’ $L.type.width$); }
| L_1 [E] { $L.array = L_1.array$;
$L.type = L_1.type.elem$; //获得子数组类型
t = newtemp();
gen(t ‘=’ $E.addr$ ‘*’ $L.type.width$);
$L.offset$ = newtemp();
gen($L.offset$ ‘=’ $L_1.offset$ ‘+’ t); }

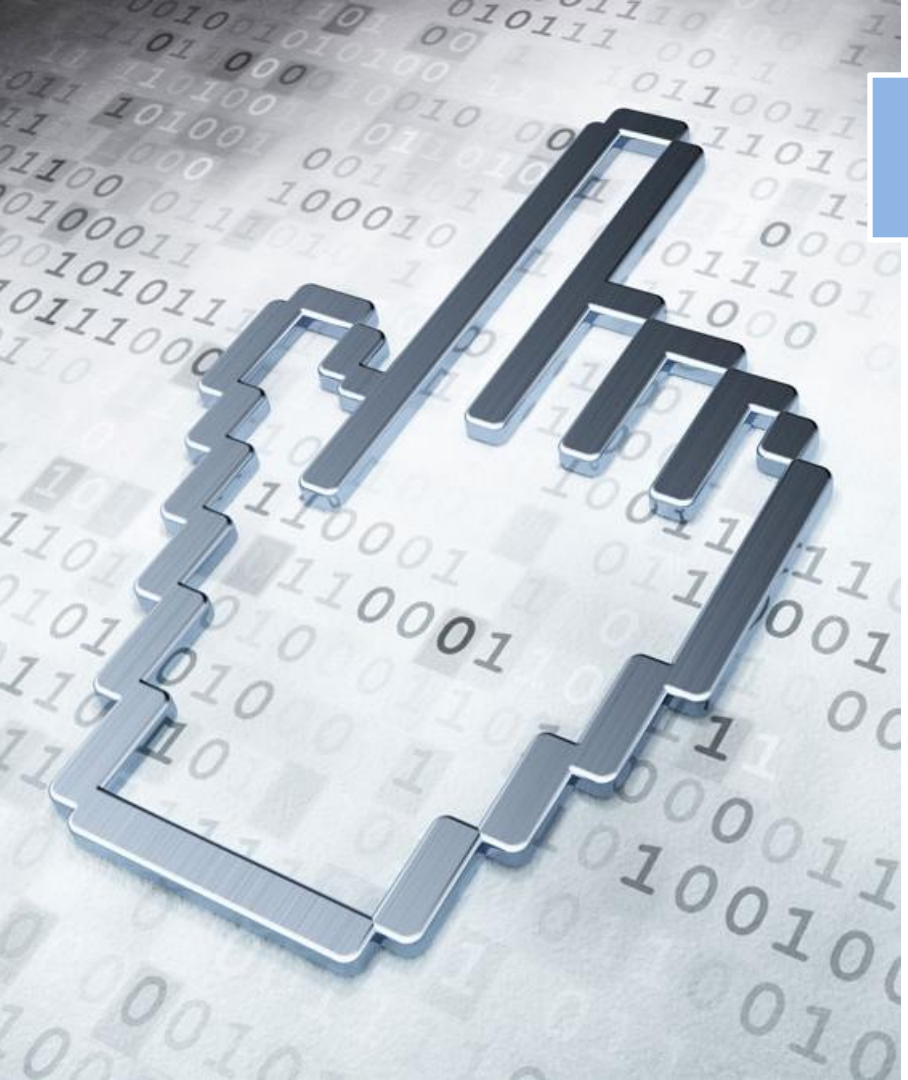


三地址码
$t_1 = i_1 * 160$
$t_2 = i_2 * 32$
$t_3 = t_1 + t_2$
$t_4 = i_3 * 4$
$t_5 = t_3 + t_4$
$t_6 = a[t_5]$

练习1

- 使用讲义中的翻译方案翻译下面赋值语句生成三地址码序列。假设*a*和*b*的类型表达式都是 array(3, array(5, real)), real类型占8个存储单元。

$$x = a[i][j] + b[i][j]$$



本章内容

6.3 类型检查

- 为每个语法成分赋予一个类型表达式
  - 名字的类型表达式保存在符号表中
  - 其他语法成分(如表达式E或语句S)的类型表达式则作为属性保存在语义栈中。
- 类型检查的任务就是确定这些类型表达式是否符合一定的规则，这些规则的集合通常称为源程序的类型系统。
- 类型检查具有发现程序错误的能力。

6.3.1 类型检查的规则

- 类型综合
  - 从子表达式的类型确定表达式的类型
  - 要求名字在引用之前必须先进行声明
  - if *f*的类型为$s \to t$ and *x*的类型为*s* then 表达式*f*(*x*)的类型为*t*

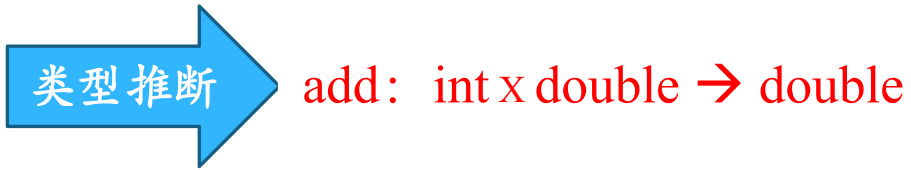
```
double a = 5;
double b = 3.14;
int c = a + b;
```

a+b 类型是double,
可能需要类型转换

6.3.1 类型检查的规则

- 类型综合
- 类型推断
  - 根据语言结构的使用方式来确定其类型
  - 经常被用于从函数体推断函数类型
  - if $f(x)$是一个表达式 then 对某两个类型变量α和β，f具有类型 $\alpha \rightarrow \beta$ and x具有类型α

```
auto add( int t, double u) {
    return t + u;
}
```



6.3.2 类型转换

- 不同类型机内表示不同，且机器运算指令不同
- 编译器需要将运算分量转换为相同类型

```
int a = 5,  b = 3.14;
float c = 2;
int z = a * b + c;
```

- 中间代码

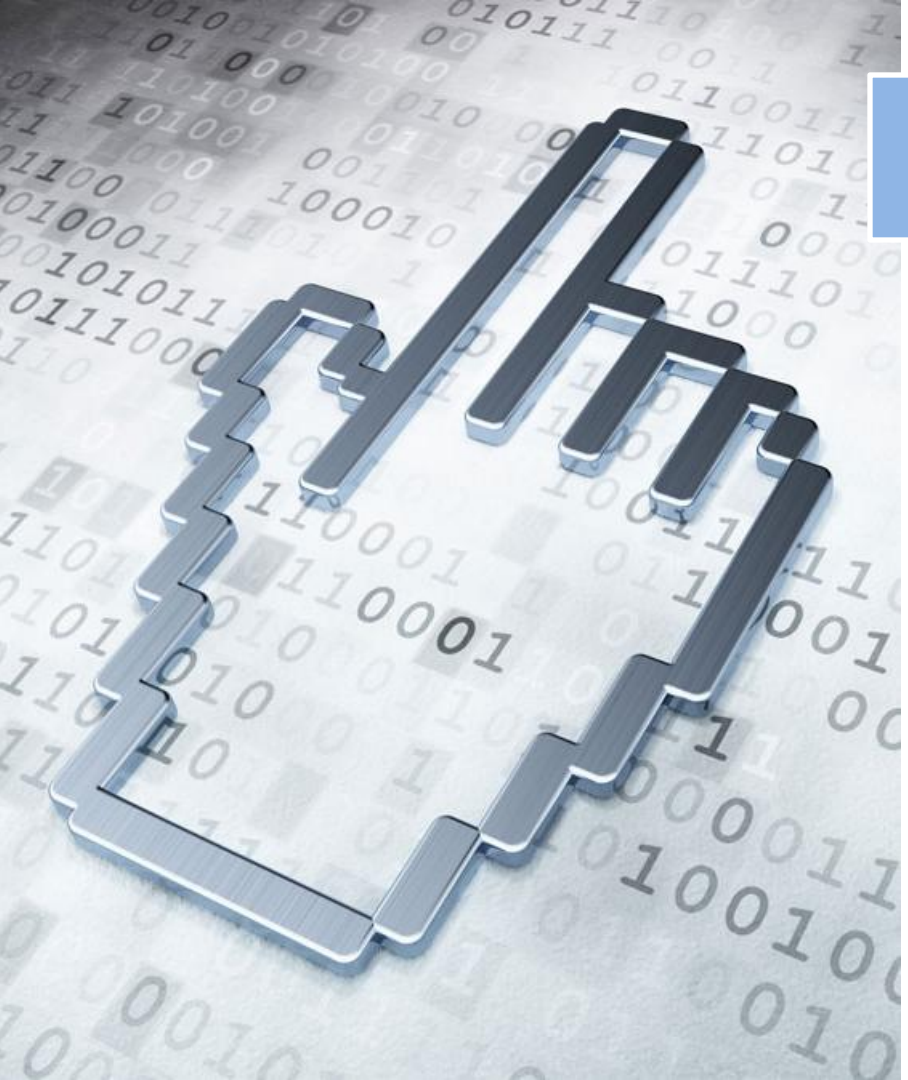
t_1 = a **int*** b
t_2 = (float) t_1
t_3 = c **float+** t_2
z = (int) t_3

6.3.2 类型转换

➢ 加入类型检查的$E \rightarrow E_1 + E_2$的SDT

为语法成分设置type属性

```
{ E.addr := newtemp()
if E1.type = integer and E2.type = integer then begin
    gencode( E.addr, ':=', E1.addr, 'int+', E2.addr ) ;
    E.type = integer
end
else if E1.type = integer and E2.type = float then begin
    u := newtemp();
    gencode( u, ':=', '(float)', E1.addr ) ;
    gencode( E.addr, ':=', u, 'float+', E2.addr ) ;
    E.type := float
end        . . . }
```



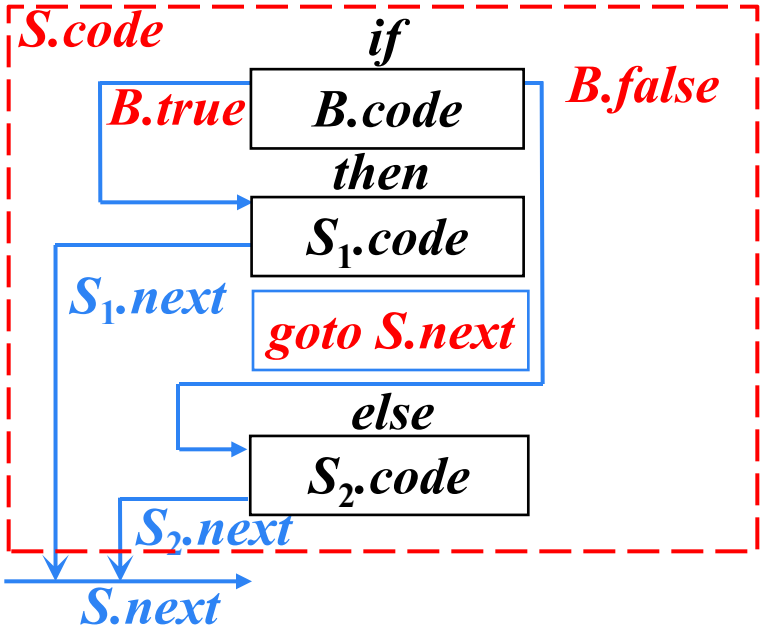
本章内容

6.4 控制语句的翻译

- 控制流语句的基本文法
  - $P \rightarrow S$
  - $S \rightarrow S_1 S_2$
  - $S \rightarrow \text{id} = E\ ; \mid L = E\ ;$
  - $S \rightarrow \text{if}\ B\ \text{then}\ S_1$
    $\mid \text{if}\ B\ \text{then}\ S_1\ \text{else}\ S_2$
    $\mid \text{while}\ B\ \text{do}\ S_1$

控制流语句的代码结构

➢ 例 $S \rightarrow if\ B\ then\ S_1\ else\ S_2$



布尔表达式B被翻译成由跳转指令构成的跳转代码

难点：确定跳转指令中目标地址。

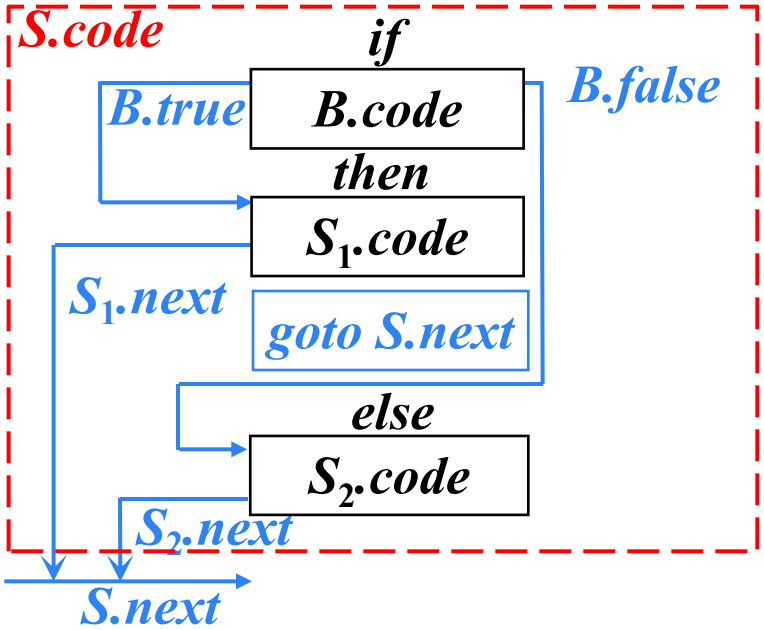
- 第一种方法：为跳转目标预分配标号并通过继承属性传递到标注位置
  - 需第二遍扫描绑定标号与地址
- 第二种方法：回填技术
  - 一遍扫描完成

例：if a<b then y=0 else y=1

```
1:      if a < b goto L1      } B
2:      goto L2
3:  L1: y=0 ——— S1
4:      goto L3
5:  L2: y=1 ——— S2
6:  L3:
```

控制流语句的代码结构

➢ 例 $S \rightarrow if\ B\ then\ S_1\ else\ S_2$



布尔表达式B被翻译成由跳转指令构成的跳转代码

➢ 继承属性

- ➢ *B.true*：用来存放当*B*为真时控制流转向的指令的标号
- ➢ *B.false*：用来存放当*B*为假时控制流转向的指令的标号
- ➢ *S.next*：用来存放紧跟在*S*代码之后的指令(*S*的后继指令)的标号

用指令的标号标识一条三地址指令

控制流语句的SDT

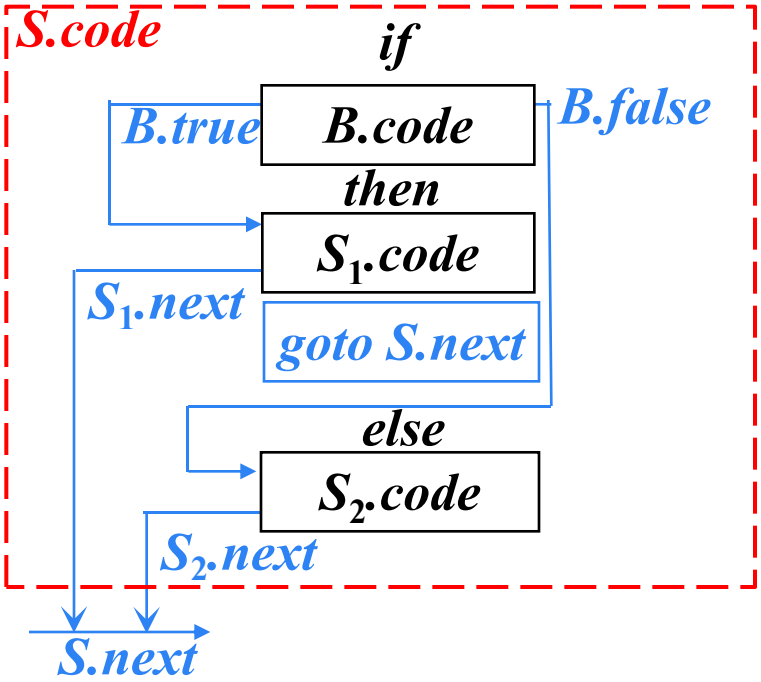
newlabel( ): 生成一个新指令标号并返回

- $P \rightarrow \{ S.next = newlabel(); \} S \{ label(S.next); \}$
- $S \rightarrow \{ S_1.next = newlabel(); \} S_1$
  $\{ label(S_1.next); S_2.next = S.next ; \} S_2$
- $S \rightarrow \mathbf{id} = E ; \mid L = E ;$
- $S \rightarrow if\ B\ then\ S_1$
  $\mid if\ B\ then\ S_1\ else\ S_2$
  $\mid while\ B\ do\ S_1$

label(L): 将标号*L*附加到下一条三地址指令前

*if-then-else*语句的*SDT*

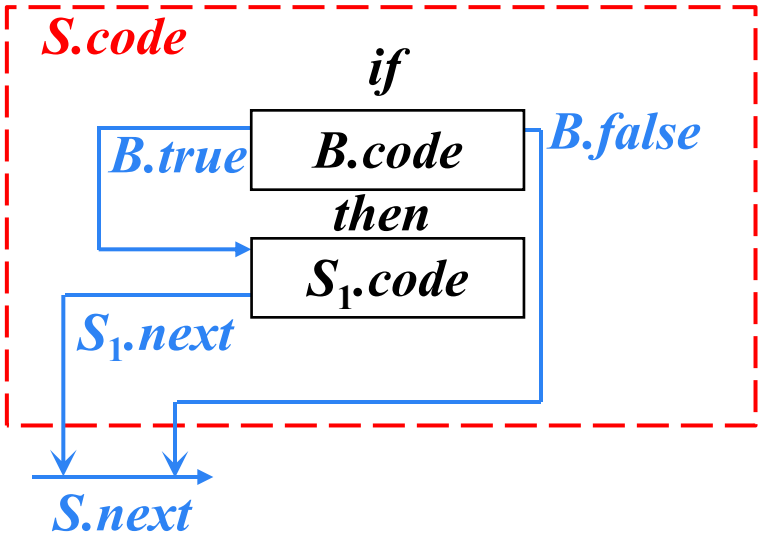
$S \rightarrow if\ B\ then\ S_1\ else\ S_2$



$S \rightarrow if$ { $B.true = newlabel(); B.false = newlabel();$ } B
$then$ { $label(B.true); S_1.next = S.next;$ } S_1 { $gen(\text{'goto'}\ S.next)$ }
$else$ { $label(B.false); S_2.next = S.next;$ } S_2

*if-then*语句的*SDT*

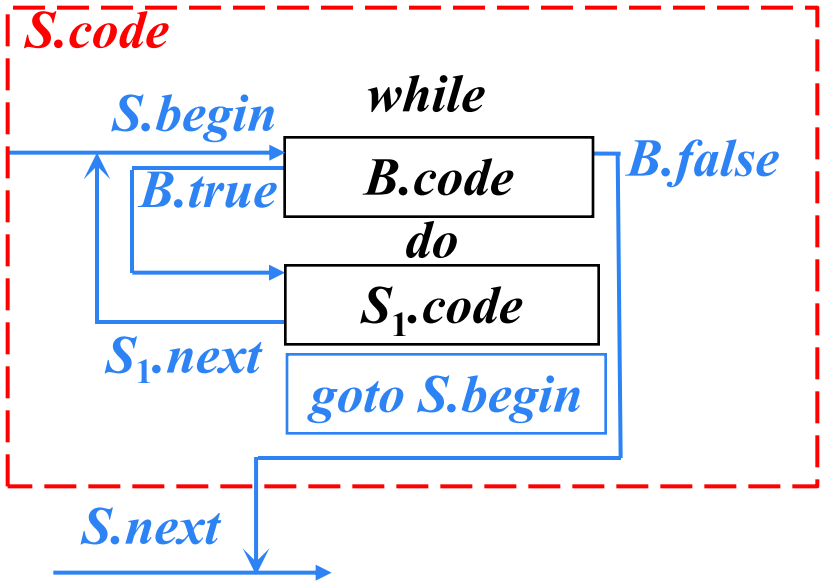
$S \rightarrow if\ B\ then\ S_1$



$S \rightarrow if\ \{\ B.true = newlabel();\ B.false = S.next;\ \}\ B$
$then\ \{\ label(B.true);\ S_1.next = S.next;\ \}\ S_1$

*while-do*语句的*SDT*

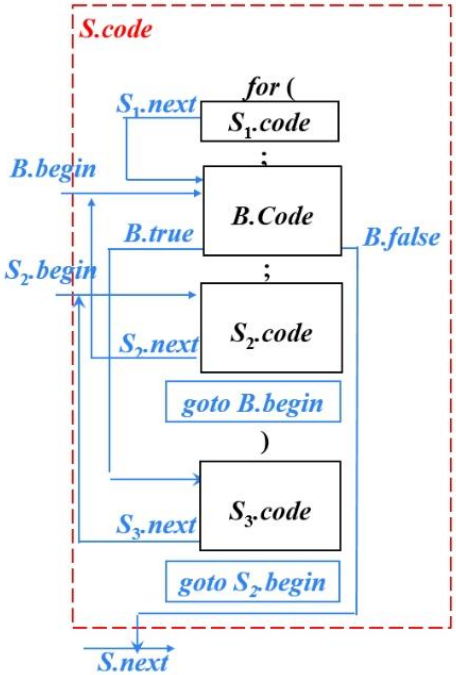
$S \rightarrow$ *while* B *do* S_1



$S \rightarrow$ *while* {*S.begin* = *newlabel*();
label(*S.begin*);
B.true = *newlabel*();
B.false = *S.next*;} B
do { *label*(*B.true*); S_1*.next* = *S.begin*;} S_1
{ *gen*('*goto*' *S.begin*); }

练习2

- 根据给定的代码结构图，参考6.4控制流语法的翻译方案，构造预标号的SDT，实现for控制流的翻译。



布尔表达式的翻译

➢布尔表达式的基本文法

$B \rightarrow B$ or B
| B and B
| not B
| (B)
| E relop E
| true
| false

优先级：not > and > or

关系表达式

relop（关系运算符）：
<, <=, >, >=, ==, !=

布尔表达式的翻译

- 逻辑运算符&&、||和！被翻译成跳转指令。运算符本身不出现在代码中，布尔表达式的值是通过代码序列中的跳转位置来表示的
- 例

```
if ( x<100 || x>200 && x!=y )
        x=0;
```

- 语句
- 三地址代码

```
        if x<100 goto L2
        goto L3
L3 :    if x>200 goto L4
        goto L1
L4 :    if x!=y goto L2
        goto L1
L2 :    x=0
L1 :
```

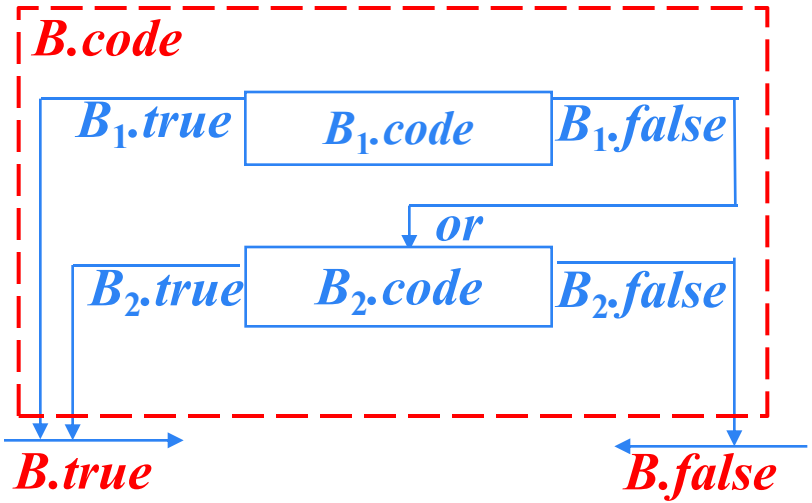
布尔表达式的*SDT*

- $B \rightarrow E_1$ relop E_2 { *gen*(‘*if* ’ E_1.*addr relop* E_2.*addr* ‘*goto*’ *B.true*);
  gen(‘*goto*’ *B.false*); }
- $B \rightarrow$ true { *gen*(‘*goto*’ *B.true*); }
- $B \rightarrow$ false { *gen*(‘*goto*’ *B.false*); }
- $B \rightarrow$ ({B_1.*true* =*B. true*; B_1.*false* =*B.false*; } B_1)
- $B \rightarrow$ not { B_1.*true* = *B.false*; B_1.*false* = *B.true*; } B_1

$B \rightarrow B_1$ or B_2 的SDT

- $B \rightarrow B_1$ *or* B_2
  - $B \rightarrow \{ B_1.true = B.true;\ B_1.false = newlabel(); \} B_1$
    or $\{ label(B_1.false);\ B_2.true = B.true;\ B_2.false = B.false; \} B_2$

生成短路代码

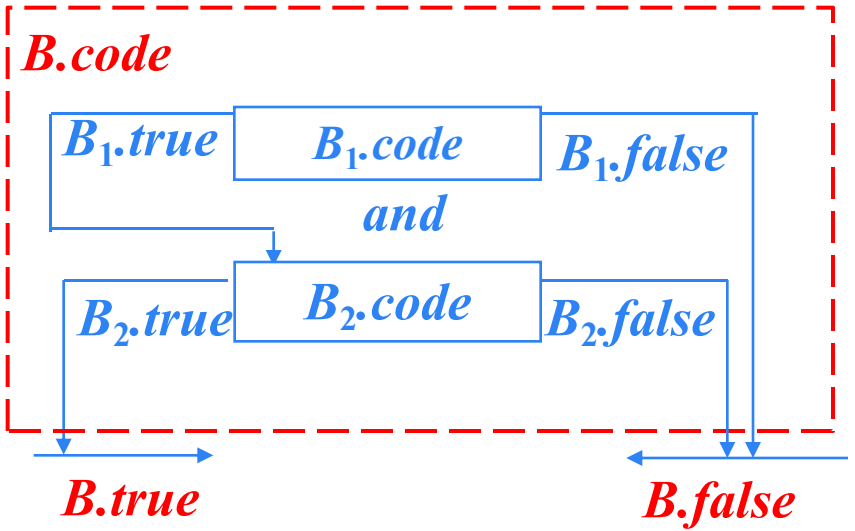


$B \rightarrow B_1$ and B_2 的SDT

- $B \rightarrow B_1$ *and* B_2
  - $B \rightarrow \{ B_1.true = newlabel(); B_1.false = B.false; \} B_1$

    and $\{ label(B_1.true); B_2.true = B.true; B_2.false = B.false; \} B_2$

生成短路代码



控制流语句的SDT

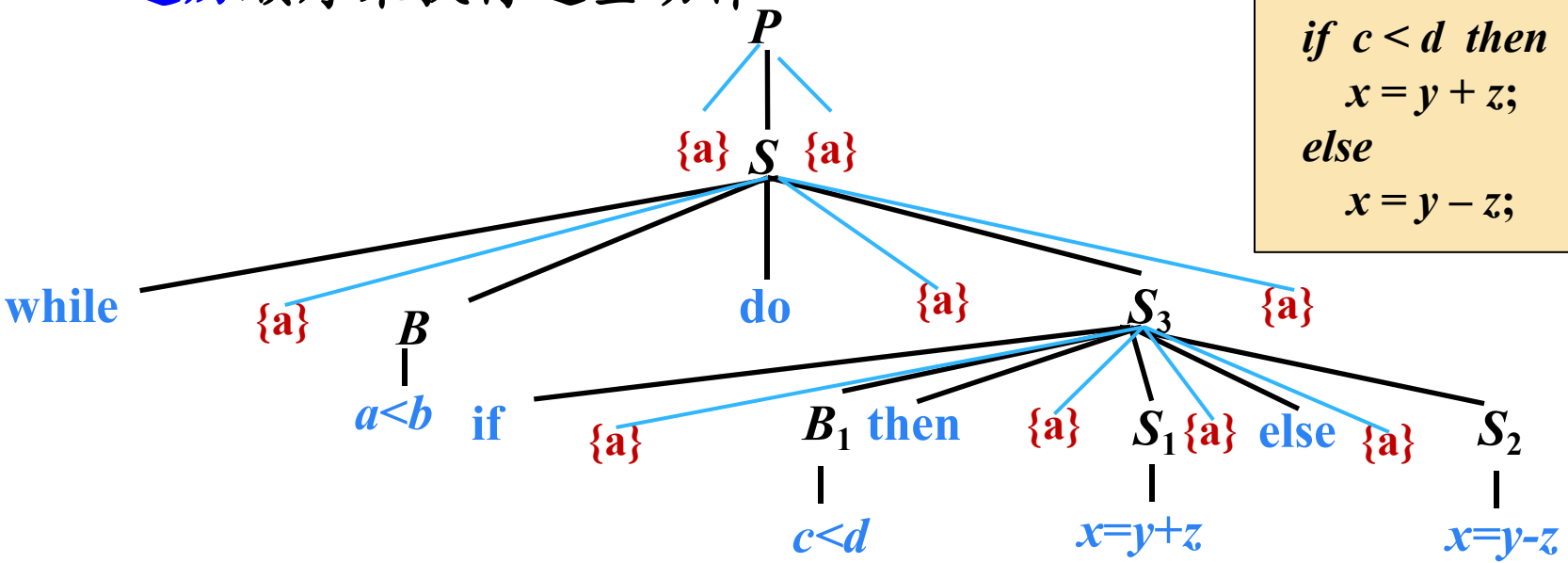
- $P \rightarrow$ {a}S{a}
- $S \rightarrow$ {a}S_1{a}S_2
- $S \rightarrow$ id=E;{a} | L=E;{a}
- $E \rightarrow E_1+E_2${a} | $-E_1${a} | (E_1){a}| id{a}| L{a}
- $L \rightarrow$ id[E]{a} | L_1[E]{a}
- $S \rightarrow$ if {a}B then {a}S_1
  | if {a}B then {a}S_1 {a} else {a}S_2
  | while {a}B do {a}S_1{a}
- $B \rightarrow$ {a}B or {a}B | {a}B and {a}B | not {a}B | ({a}B)
  | E relop E{a} | true{a} | false{a}

```
while a < b do
  if c < d then
    x = y + z;
  else
    x = y – z;
```

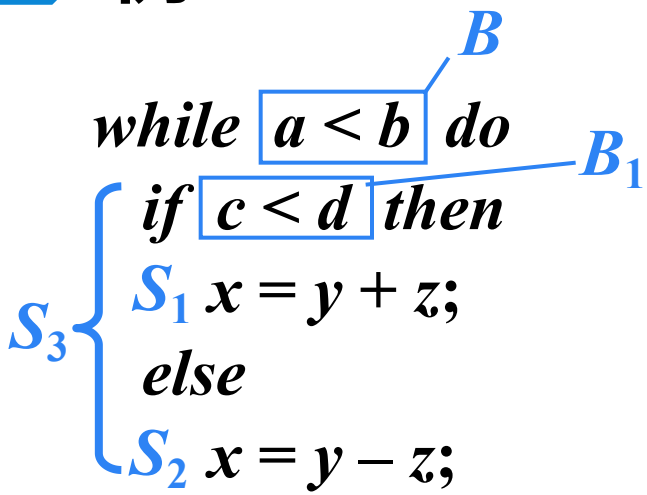
SDT的通用实现方法

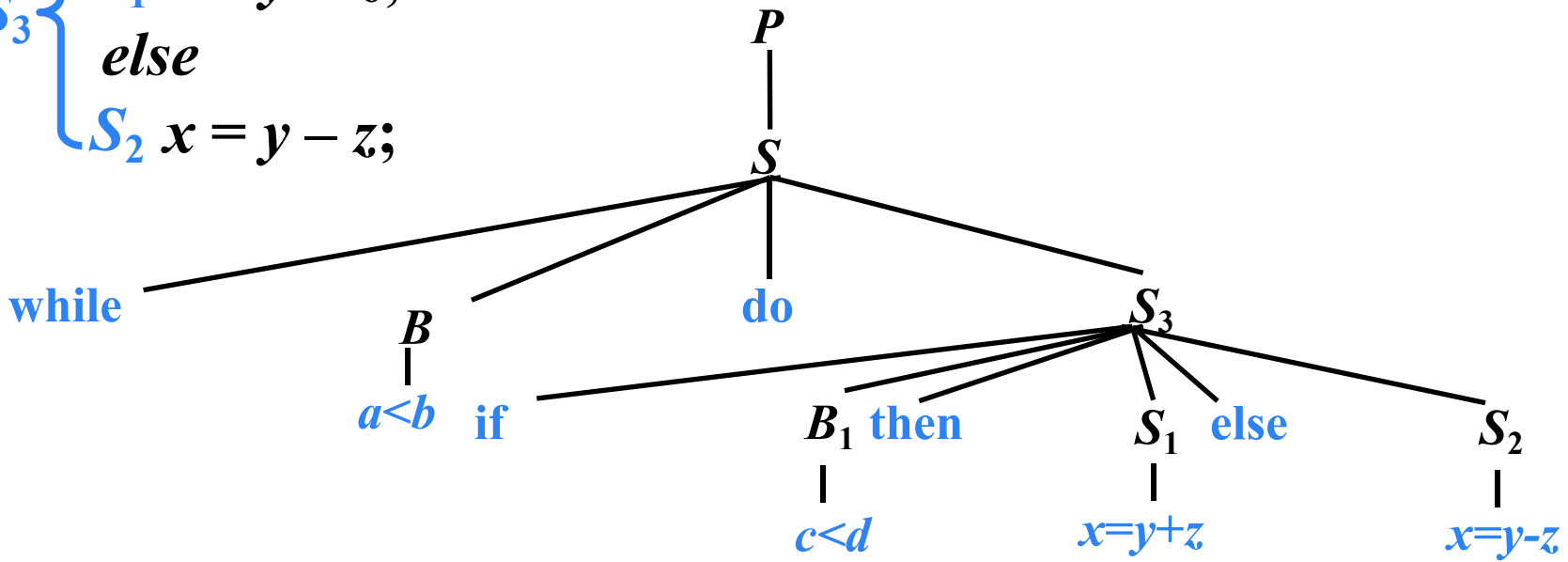
- 任何SDT都可以通过下面的方法实现
  - 首先建立一棵语法分析树，然后按照从左到右的深度优先遍历顺序来执行这些动作

```
while a < b do
  if c < d then
    x = y + z;
  else
    x = y – z;
```



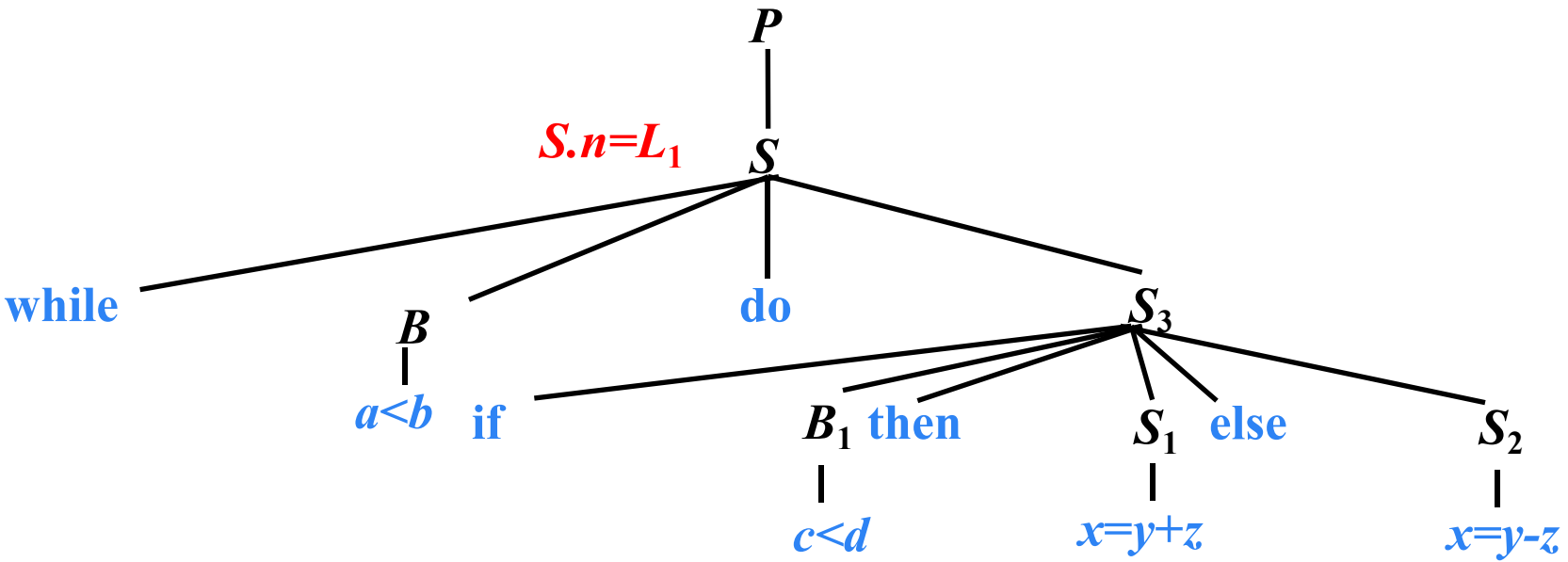
例


B
while a < b do
B1
if c < d then
S3
S1 x = y + z;
else
S2 x = y – z;


P
S
while
B
a<b
do
S3
if
B1
c<d
then
S1
x=y+z
else
S2
x=y-z

例

$P \rightarrow \{\ S.next = newlabel(\ );\ \} S \{\ label(S.next);\}$



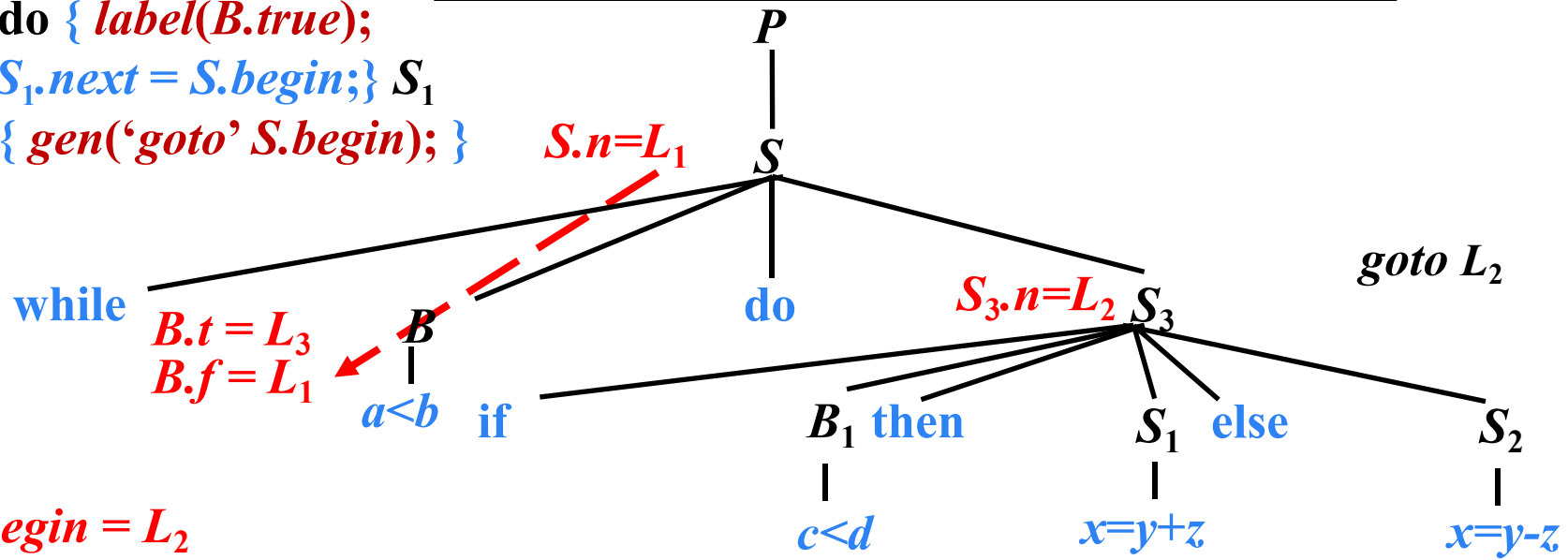
例

L_2 : *if* $a < b$ *goto* L_3
goto L_1

$S \rightarrow$ while {*S.begin = newlabel( );*
label(S.begin);
B.true = newlabel( );
B.false = S.next;} B
do { *label(B.true);*
S_1*.next = S.begin;*} S_1
{ *gen('goto' S.begin);* }

$B \rightarrow E_1$ relop E_2
{ *gen('if' E_1.addr relop E_2.addr 'goto' B.true);*
gen('goto' B.false); }

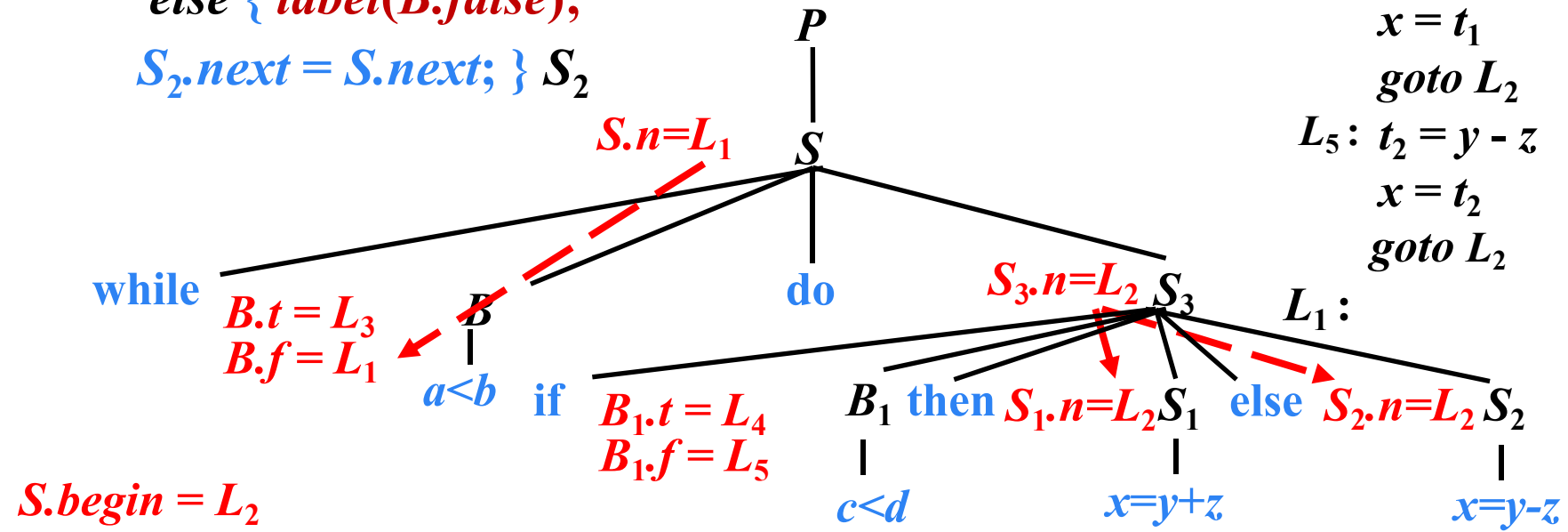
L_3 :



S.begin = L_2

例

$S \rightarrow$ *if* { *B.true* = *newlabel*( ); *B.false* = *newlabel*( ); } *B*
then { *label*(*B.true*); S_1.*next* = *S.next*; } S_1
{ *gen*('goto' *S.next*); }
else { *label*(*B.false*);
S_2.*next* = *S.next*; } S_2



L_2 : *if a* < *b* *goto* L_3
goto L_1
L_3 : *if c* < *d* *goto* L_4
goto L_5
L_4 : $t_1 = y + z$
$x = t_1$
goto L_2
L_5 : $t_2 = y - z$
$x = t_2$
goto L_2

语句“*while a<b do if c<d then x=y+z else x=y-z*”的三地址代码

1: L_2 : *if* $a < b$ *goto* L_3
2: *goto* L_1
3: L_3 : *if* $c < d$ *goto* L_4
4: *goto* L_5
5: L_4 : $t_1 = y + z$
6: $x = t_1$
7: *goto* L_2
8: L_5 : $t_2 = y - z$
9: $x = t_2$
10: *goto* L_2
11: L_1 :

需要第二趟扫描
绑定标号与
指令序号

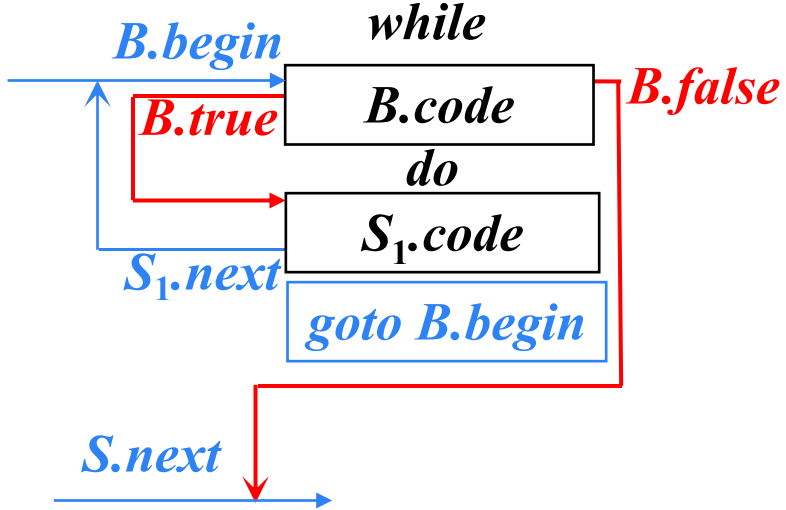
1: (*j*<, *a*, *b* , 3)
2: (*j* , - , - , 11)
3: (*j*<, *c*, *d* , 5)
4: (*j* , - , - , 8)
5: (+ , *y* , *z* , t_1)
6: (= , t_1, - , *x*)
7: (*j* , - , - , 1)
8: (- , *y*, *z* , t_2)
9: (= , t_2, - , *x*)
10: (*j* , - , - , 1)
11:

语句“*while a<b do if c<d then x=y+z else x=y-z*”的三地址代码

B.first →
1: *if a < b goto* 3
2: *goto* 11
S$_1$*.first* →
3: *if c < d goto* 5
4: *goto* 8
5: $t_1 = y + z$
6: $x = t_1$
7: *goto* 1
8: $t_2 = y - z$
9: $x = t_2$
10: *goto* 1
11:

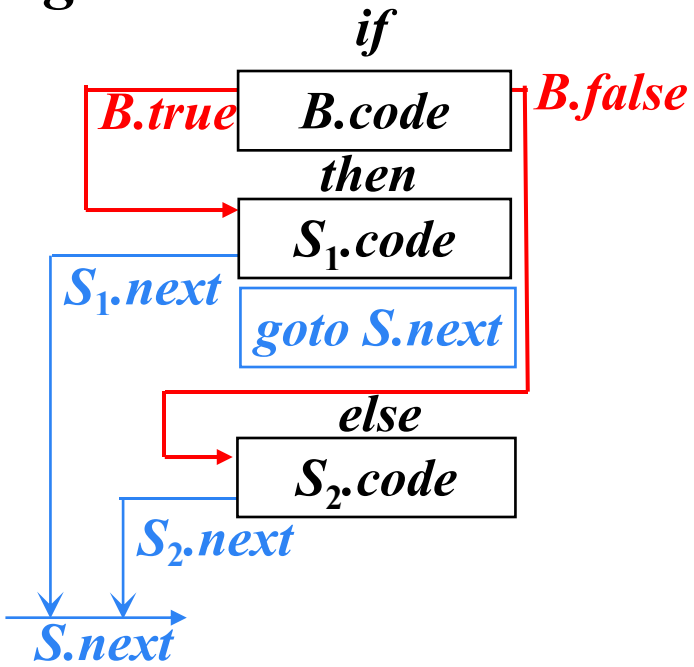
1: *ifFalse a < b goto* 11
2:
3: *ifFalse c < d goto* 8
4:
5: $t_1 = y + z$
6: $x = t_1$
7: *goto* 1
8: $t_2 = y - z$
9: $x = t_2$
10: *goto* 1
11:

如何避免生成冗余的goto指令?



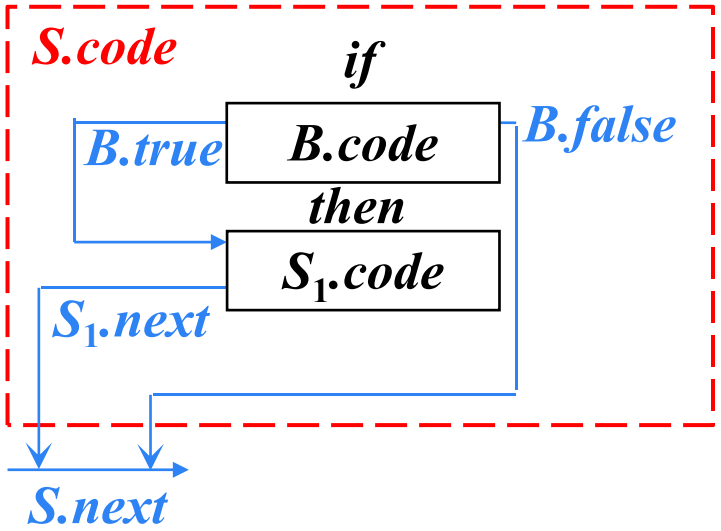
语句“*while a<b do if c<d then x=y+z else x=y-z*”的三地址代码

	1: *if a < b goto 3*	1: *ifFalse a < b goto 11*
	2: *goto 11*	2:
B.first →	3: *if c < d goto 5*	3: *ifFalse c < d goto 8*
	4: *goto 8*	4:
S_1.first →	5: $t_1 = y + z$	5: $t_1 = y + z$
	6: $x = t_1$	6: $x = t_1$
	7: *goto 1*	7: *goto 1*
S_2.first →	8: $t_2 = y - z$	8: $t_2 = y - z$
	9: $x = t_2$	9: $x = t_2$
	10: *goto 1*	10: *goto 1*
	11:	11:



避免生成冗余的*goto*指令

修改*if-then*语句的*SDT*



$S \rightarrow if\ \{\ B.true = newlabel();\ B.false = S.next;\ \}\ B$
$\quad then\ \{\ label(B.true);\ S_1.next = S.next;\ \}\ S_1$

不要生成任何跳转指令

$S \rightarrow if\ \{\ B.true = fall;\ B.false = S.next;\ \}\ B$
$\quad then\ \{\ S_1.next = S.next;\ \}\ S_1$

修改$B \rightarrow E_1$ relop E_2的*SDT*

- $B \rightarrow E_1$ relop E_2 { *gen(‘if’ E_1.addr relop E_2.addr ‘goto’ B.true); gen(‘goto’ B.false); }*

- $B \rightarrow E_1$ relop E_2

{ if *B.true* ≠*fall and B.false* ≠*fall then*

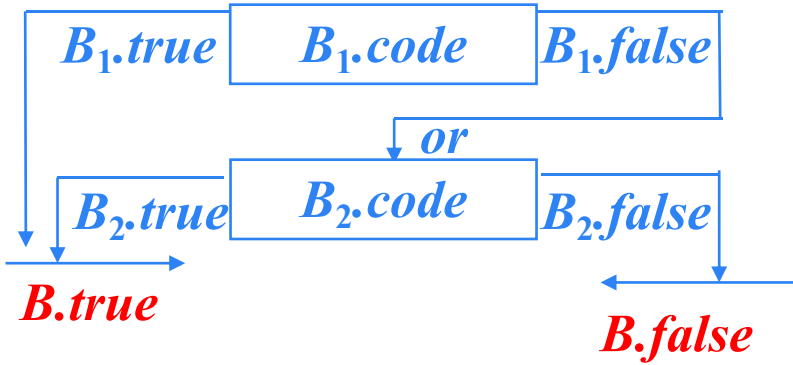
gen(‘if’ E_1.addr relop E_2.addr ‘goto’ B.true); gen(‘goto’ B.false); }

else if *B.true* ≠ *fall then gen(‘if’ E_1.addr relop E_2.addr ‘goto’ B.true);*

else if *B.false* ≠ *fall then gen(‘ifFalse’ E_1.addr relop E_2.addr‘goto’ B.false);*

else ′ ′ }

修改$B \rightarrow B_1$ or B_2 的SDT



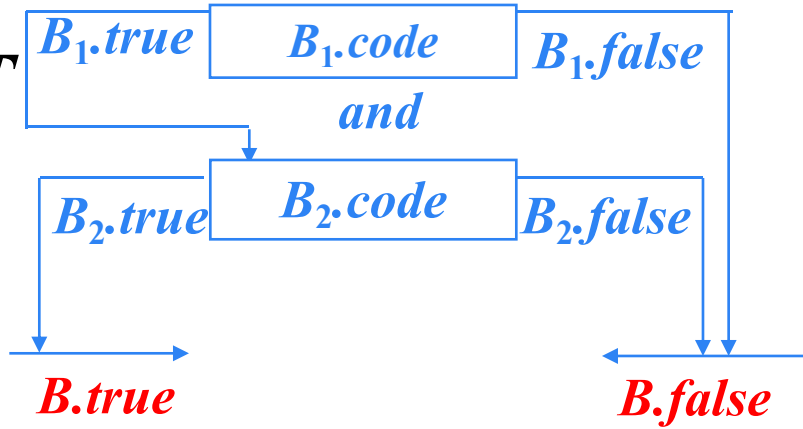
$B \rightarrow \{ B_1.true = B.true;\ B_1.false = newlabel(); \} B_1$

or $\{ label(B_1.false);\ B_2.true = B.true;\ B_2.false = B.false; \} B_2$

$B \rightarrow \{ B_1.true =$ if $B.true \neq fall$ then $B.true$ else $newlabel();\ B_1.false = fall; \} B_1$

or $\{ B_2.true = B.true;\ B_2.false = B.false; \} B_2 \{ label(B_1.true); \}$

修改$B \rightarrow B_1$ and B_2 的SDT



$B \rightarrow \{ B_1.true = newlabel(); B_1.false = B.false; \} B_1$

and $\{ label(B_1.true); B_2.true = B.true; B_2.false = B.false; \} B_2$

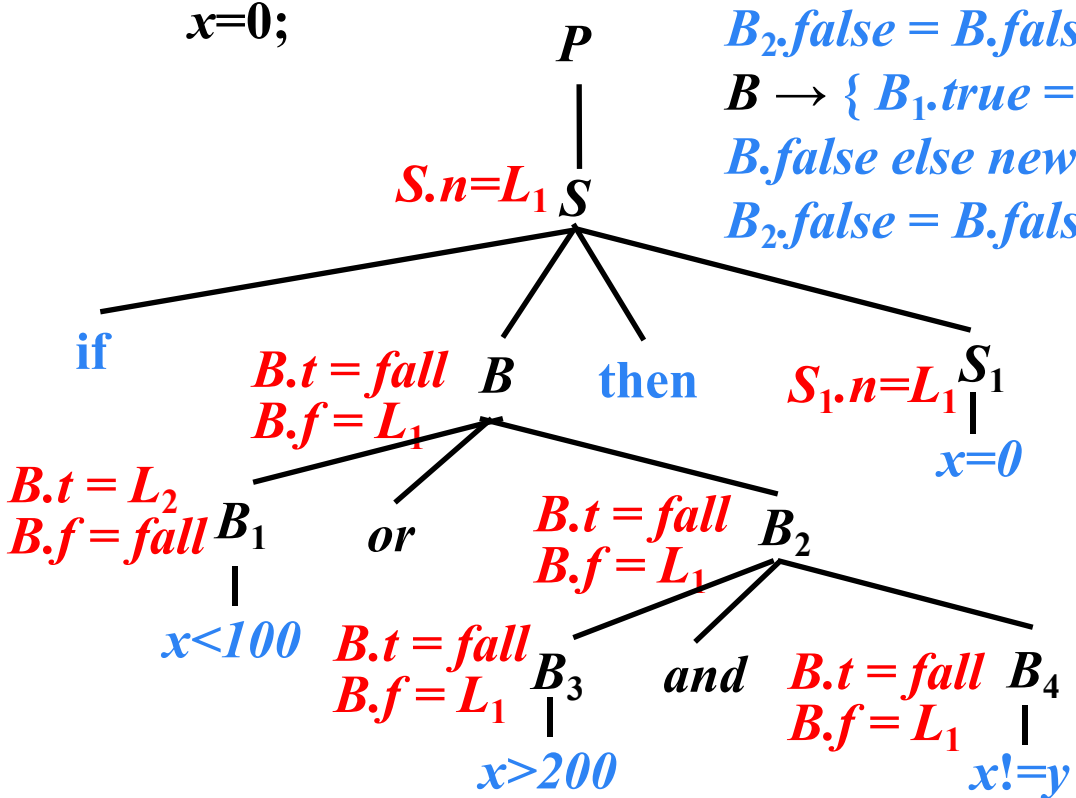
$B \rightarrow \{ B_1.true = fall; B_1.false =$ if $B.false \neq fall$ then $B.false$ else $newlabel(); \} B_1$

and $\{ B_2.true = B.true; B_2.false = B.false; \} B_2 \{ label(B_1.false); \}$

例

if (*x*<100 || *x*>200 && *x*!=*y*)

x=0;



$S \rightarrow$ *if* { *B.true* = *fall*; *B.false* = *S.next*; } *B*
then { S_1*.next* = *S.next*; } S_1

$B \rightarrow$ { B_1*.true* = if *B.true* ≠*fall then B.true else newlabel*(); B_1*.false* = *fall*; } B_1 **or** {B_2*.true* = *B.true*; B_2*.false* = *B.false*; } B_2 { *label*(B_1*.true*); }

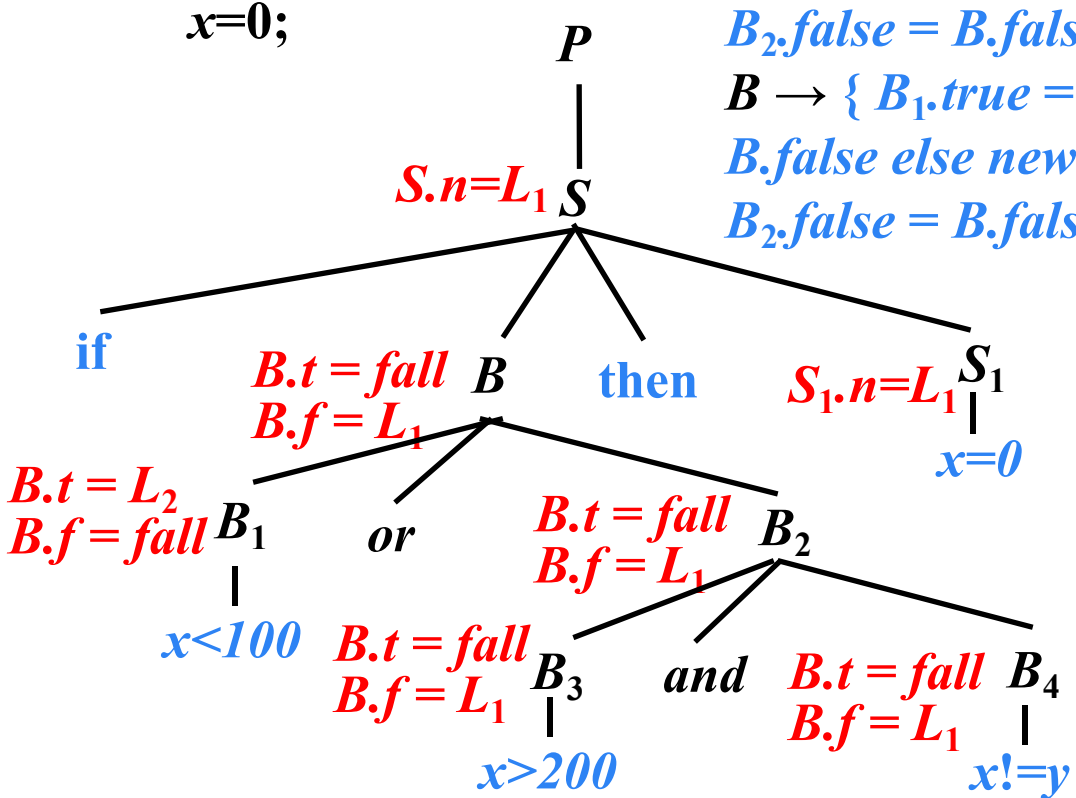
$B \rightarrow$ { B_1*.true* = *fall*; B_1*.false* = if *B.false* ≠ *fall then B.false else newlabel*();} B_1 **and** { B_2*.true* = *B.true*; B_2*.false* = *B.false*; } B_2 { *label*(B_1*.false*); }

if x<100 *goto* L_2
ifFalse x>200 *goto* L_1
ifFalse x!=*y goto* L_1
L_1 L_2 : *x*=0

例

if (*x*<100 || *x*>200 && *x*!=*y*)

x=0;



$S \rightarrow$ *if* { *B.true* = *fall*; *B.false* = *S.next*; } *B*
then { S_1*.next* = *S.next*; } S_1

$B \rightarrow$ { B_1*.true* = if *B.true* ≠*fall then B.true else newlabel*(); B_1*.false* = *fall*; } B_1 **or** {B_2*.true* = *B.true*; B_2*.false* = *B.false*; } B_2 { *label*(B_1*.true*); }

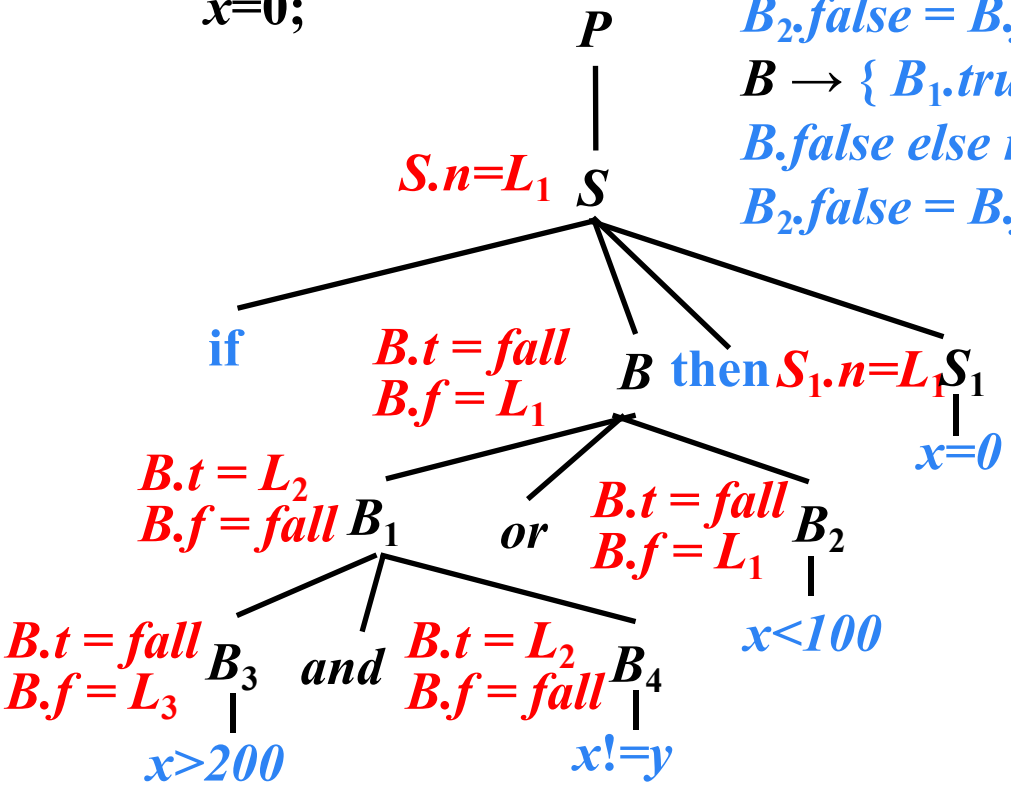
$B \rightarrow$ { B_1*.true* = *fall*; B_1*.false* = if *B.false* ≠ *fall then B.false else newlabel*();} B_1 **and** { B_2*.true* = *B.true*; B_2*.false* = *B.false*; } B_2 { *label*(B_1*.false*); }

if x<100 *goto* L_2
ifFalse x>200 *goto* L_1
ifFalse x!=*y goto* L_1
L_2 : *x*=0
L_1 :

例

if (*x*>200 && *x*!=*y* || *x*<100)

x=0;

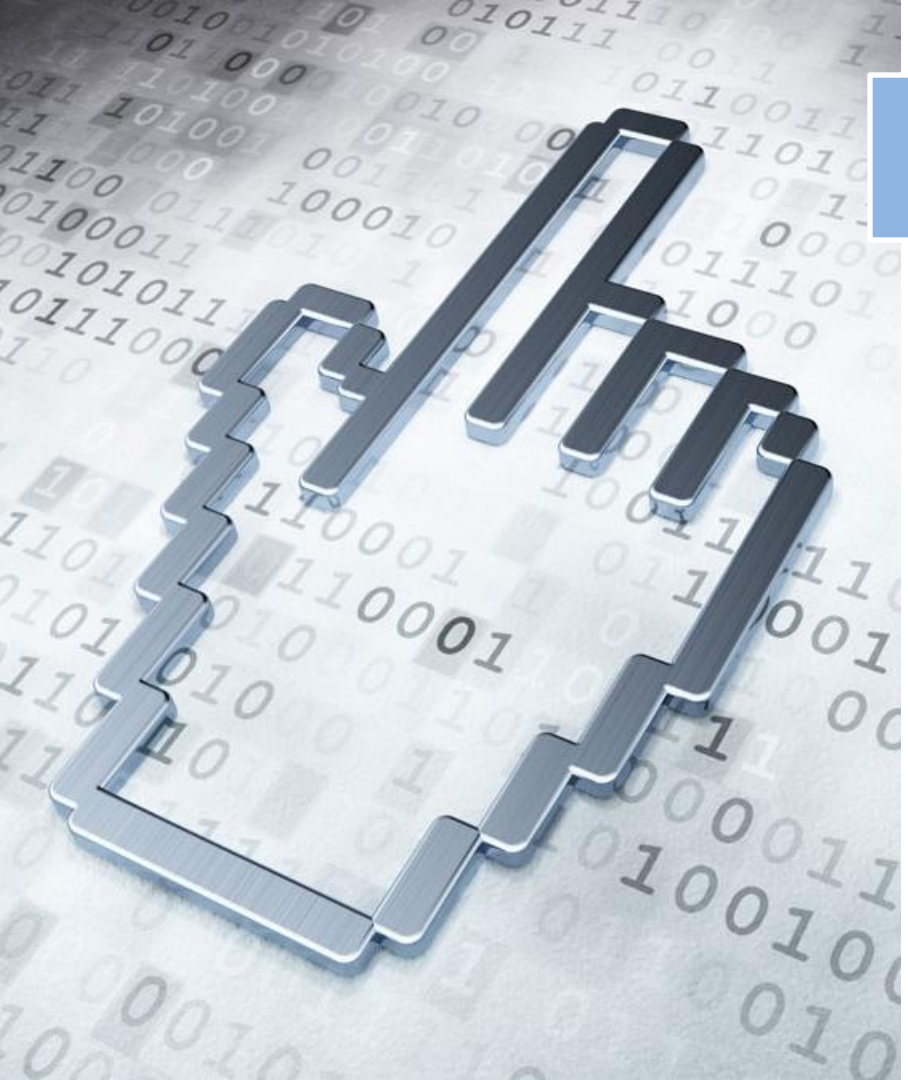


S →*if* { *B.true* = *fall*; *B.false* = *S.next*; } *B*
then { S_1.*next* = *S.next*; } S_1

B → { B_1.*true* = if *B.true* ≠*fall then B.true else newlabel*(); B_1.*false* = *fall*; }B_1 or {B_2.*true* = *B.true*; B_2.*false* = *B.false*; }B_2 { *label*(B_1.*true*); }

B → { B_1.*true* = *fall*; B_1.*false* = if *B.false* ≠ *fall then B.false else newlabel*();} B_1 and { B_2.*true* = *B.true*; B_2.*false* = *B.false*; }B_2 { *label*(B_1.*false*); }

ifFalse x>200 *goto* L_3
if x!=*y goto* L_2
L_3 : *ifFalse x*<100 *goto* L_1
L_2 : *x*=0
L_1 :



本章内容

6.5 回填 (*Backpatching*)

- 基本思想
  - 生成一个跳转指令时，暂时不指定该跳转指令的目标标号。具有相同的目标标号的跳转指令组成一个列表并以综合属性的形式进行传递。等到能够确定正确的目标标号时，才去填充这些指令的目标标号。

非终结符B的综合属性

- ***B.truelist***：指向一个包含跳转指令的列表，这些指令最终获得的目标标号就是当B为真时控制流应该转向的指令的标号
- ***B.falselist***：指向一个包含跳转指令的列表，这些指令最终获得的目标标号就是当B为假时控制流应该转向的指令的标号

- 将生成的指令按顺序编号，指令的序号作为标号

函数

- *makelist*(*i*)
  - 创建一个只包含*i*的列表，*i*是跳转指令的标号，函数返回指向新创建的列表的指针
- *merge*(p_1, p_2)
  - 将 p_1 和 p_2 指向的列表进行合并，返回指向合并后的列表的指针
- *backpatch*(*p*, *j*)
  - 将*j*作为目标标号插入到 *p*所指列表中的各跳转指令中

布尔表达式的回填

- $B \rightarrow E_1$ relop E_2

{

B.truelist = makelist(nextquad);

B.falselist = makelist(nextquad+1);

gen('if' E_1.addr relop E_2.addr 'goto _');

gen('goto _');

}

nextquad：即将生成的下一条指令的标号，也就是下一条跳转指令的标号

- 语义动作都在产生式末尾，可以在自底向上语法分析中实现的SDT

布尔表达式的回填

- $B \rightarrow E_1$ relop E_2
- $B \rightarrow$ true

{

B.truelist = makelist(nextquad);

gen('goto _');

}

布尔表达式的回填

- $B \rightarrow E_1$ relop E_2
- $B \rightarrow$ true
- $B \rightarrow$ false

{

B.falselist = makelist(nextquad);

gen('goto _');

}

布尔表达式的回填

- $B \rightarrow E_1$ relop E_2
- $B \rightarrow$ true
- $B \rightarrow$ false
- $B \rightarrow (B_1)$

{

$B.truelist = B_1.truelist$;

$B.falselist = B_1.falselist$;

}

布尔表达式的回填

- $B \rightarrow E_1$ relop E_2
- $B \rightarrow$ true
- $B \rightarrow$ false
- $B \rightarrow (B_1)$
- $B \rightarrow$ not B_1

{

$B.truelist = B_1.falselist$;

$B.falselist = B_1.truelist$;

}

$B \rightarrow B_1$ or B_2

$B \rightarrow B_1$ or M B_2

{

$B.truelist = merge(B_1.truelist, B_2.truelist)$;

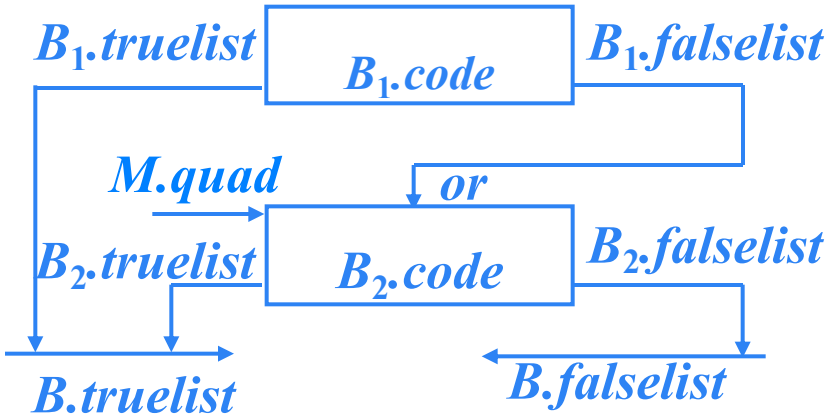
$B.falselist = B_2.falselist$;

$backpatch(B_1.falselist, M.quad)$;

}

$M \rightarrow \varepsilon$

{ $M.quad = nextquad$; }



$B \rightarrow B_1$ and B_2

$B \rightarrow B_1$ and M B_2

{

$B.truelist = B_2.truelist;$

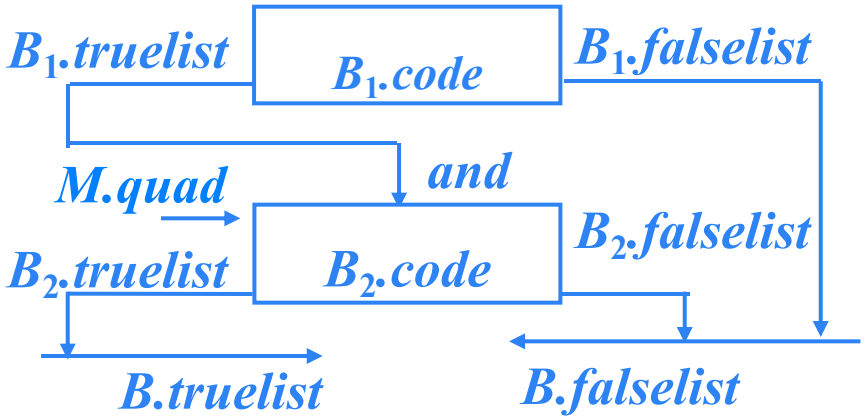
$B.falselist = merge(B_1.falselist, B_2.falselist);$

$backpatch(B_1.truelist, M.quad);$

}

$M \rightarrow \varepsilon$

{ $M.quad = nextquad$; }



例

$B \rightarrow E_1$ relop E_2

{ *B.truelist = makelist(nextquad);*

B.falselist = makelist(nextquad+1);

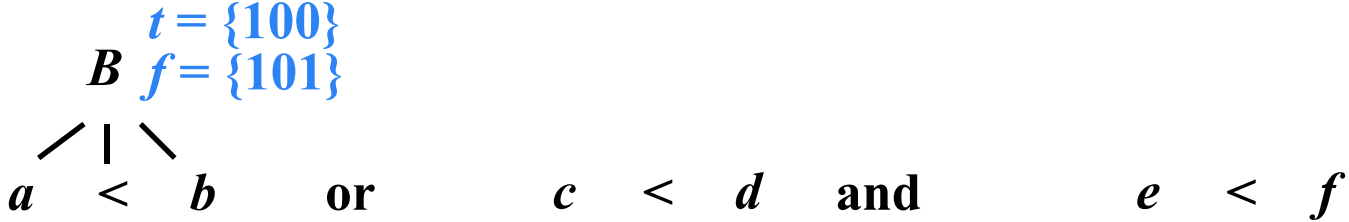
gen('if' E_1.addr relop E_2.addr 'goto _');

gen('goto _');

}

100: *if a<b goto _*

101: *goto _*



例

$B \rightarrow B_1$ or M B_2

{

backpatch(B_1.falselist, M.quad);

B.truelist = merge(B_1.truelist, B_2.truelist);

B.falselist = B_2.falselist ;

}

$M \rightarrow \varepsilon$

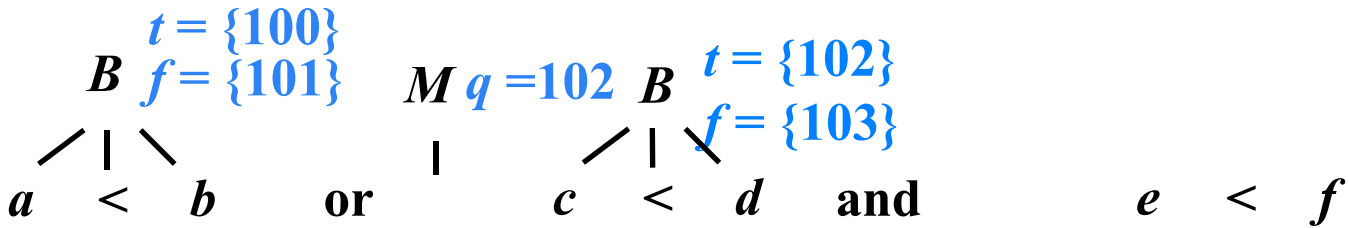
{ *M.quad = nextquad ;* }

100: *if a<b goto _*

101: *goto _*

102: *if c<d goto _*

103: *goto _*



例

$B \rightarrow B_1$ and M B_2

{

$B.truelist = B_2.truelist$;

$B.falselist = merge(B_1.falselist, B_2.falselist)$;

$backpatch(B_1.truelist, M.quad)$;

}

$M \rightarrow \varepsilon$

{ $M.quad = nextquad$; }

100: *if a<b goto* _

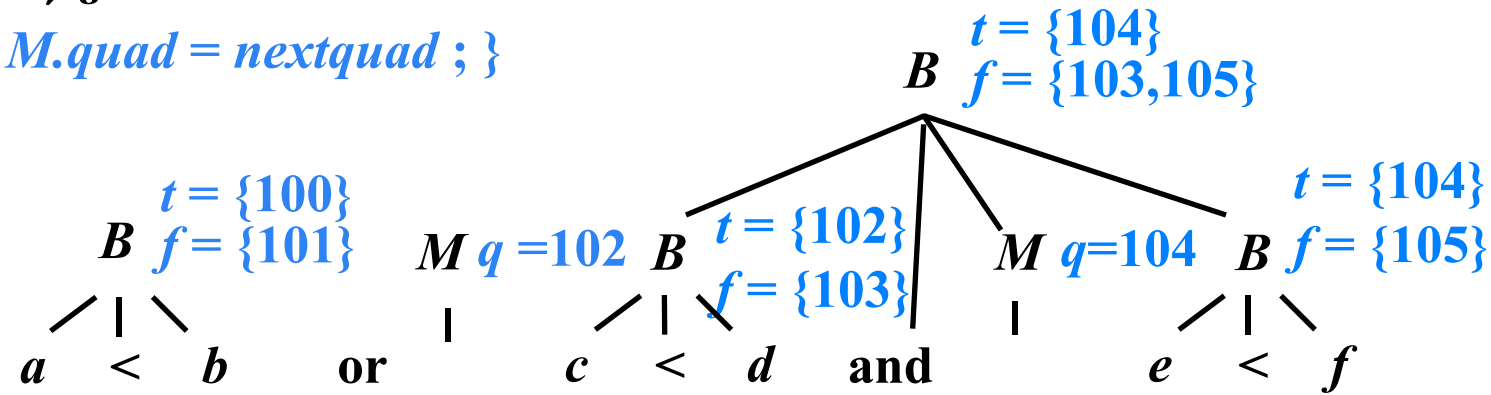
101: *goto* _

102: *if c<d goto* 104

103: *goto* _

104: *if e<f goto* _

105: *goto* _



例

$B \rightarrow B_1$ or M B_2

{

B.truelist = merge(B_1.truelist, B_2.truelist);

B.falselist = B_2.falselist ;

backpatch(B_1.falselist, M.quad);

}

100: *if a<b goto* _

101: *goto* 102

102: *if c<d goto* 104

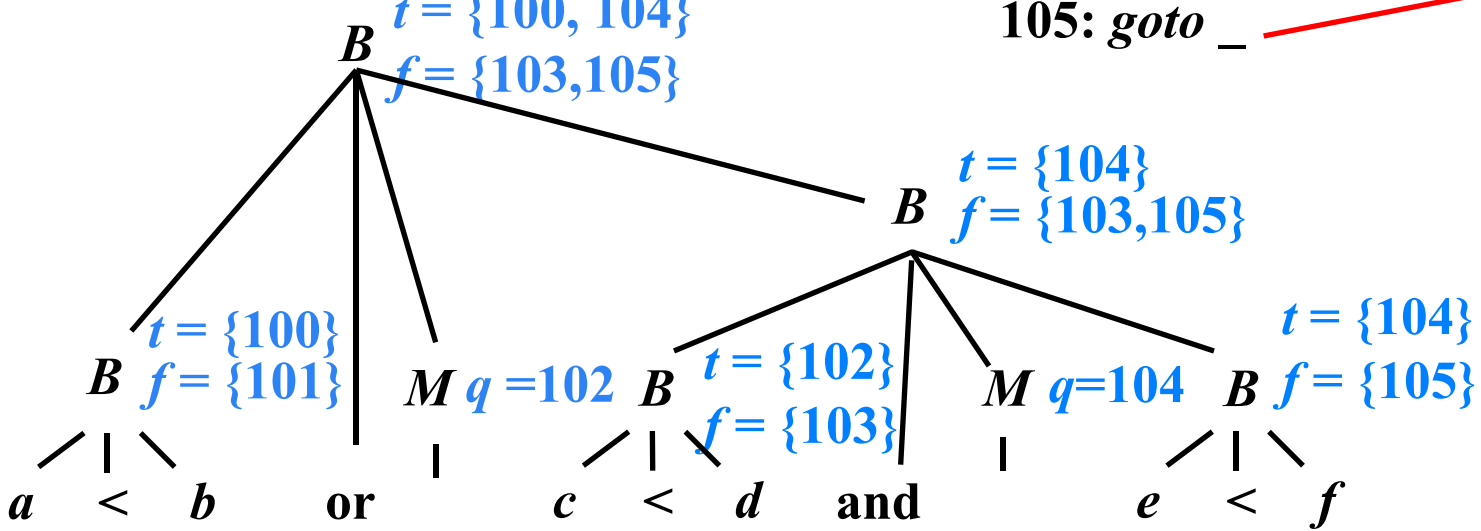
103: *goto* _

104: *if e<f goto* _

105: *goto* _

t

f



控制流语句的回填

- 文法
  - $S \rightarrow S_1 S_2$
  - $S \rightarrow \text{id} = E\ ; \mid L = E\ ;$
  - $S \rightarrow \text{if } B \text{ then } S_1$
    $\mid \text{if } B \text{ then } S_1 \text{ else } S_2$
    $\mid \text{while } B \text{ do } S_1$
- 综合属性
  - *S.next1ist*：指向一个包含跳转指令的列表，这些指令最终获得的目标标号就是按照运行顺序紧跟在*S*代码之后的指令的标号

$S \rightarrow$ if B then S_1

$S \rightarrow$ if B then M S_1

{

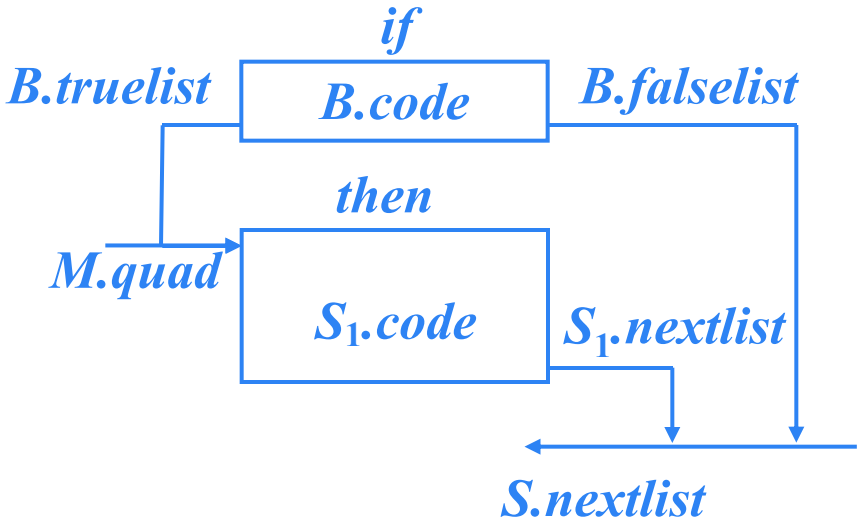
S.nextlist=merge(B.falselist, S_1.nextlist);

backpatch(B.truelist, M.quad);

}

$M \rightarrow \varepsilon$

{ *M.quad = nextquad ;* }



$S \rightarrow$ if B then S_1 else S_2

$S \rightarrow$ if B then M_1 S_1 N else M_2 S_2

{

$S.nextlist = merge(\ merge(S_1.nextlist,$
$N.nextlist), S_2.nextlist\ );$

$backpatch(B.truelist, M_1.quad);$

$backpatch(B.falselist, M_2.quad);$

}

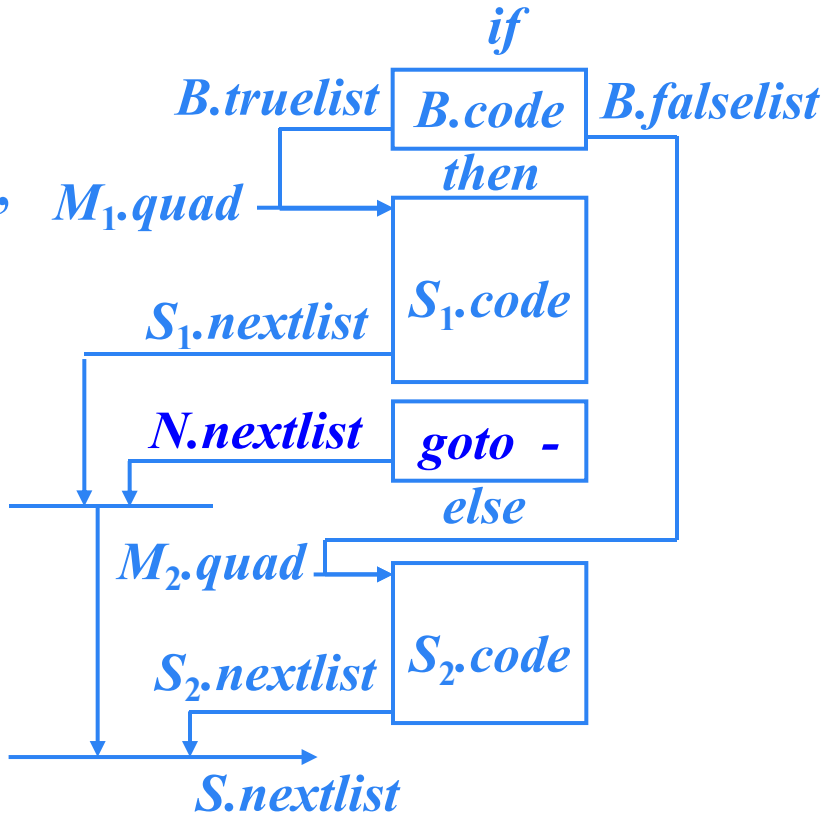
$N \rightarrow \varepsilon$

{

$N.nextlist = makelist(nextquad);$

$gen(\text{`}goto\ \_\text{'});$

}



$S \rightarrow$ while B do S_1

$S \rightarrow$ while M_1 B do M_2 S_1

{

$S.nextlist = B.falselist;$

$backpatch(\ S_1.nextlist, M_1.quad\ );$

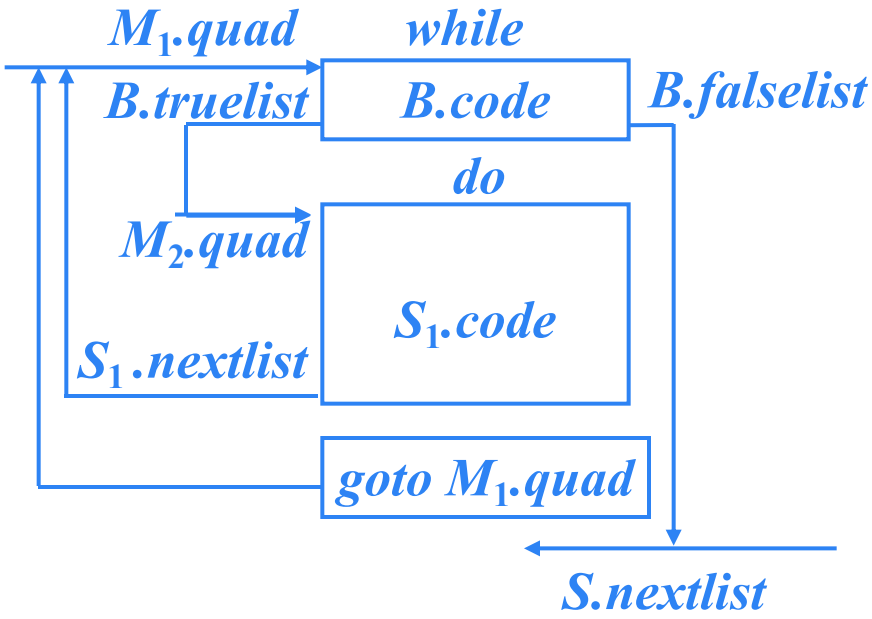
$backpatch(\ B.truelist, M_2.quad\ );$

$gen(\text{'goto'}\ M_1.quad);$

}

$M \rightarrow \varepsilon$

{ $M.quad = nextquad\ ;$ }



$S \rightarrow S_1\ S_2$

$S \rightarrow S_1\ M\ S_2$

{

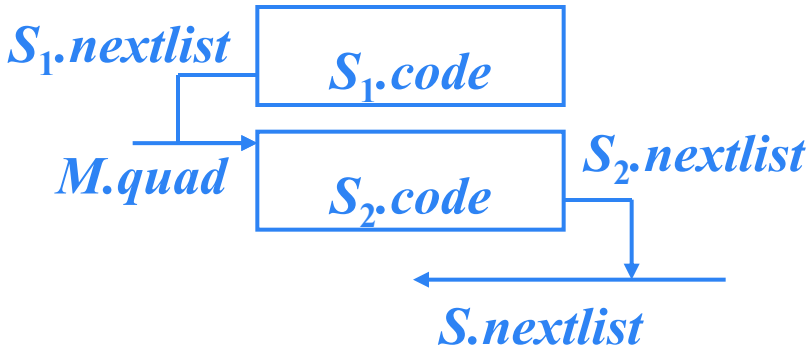
$S.nextlist = S_2.nextlist$;

$backpatch(\ S_1.nextlist,\ M.quad\ )$;

}

$M \rightarrow \varepsilon$

{ $M.quad = nextquad$; }



$S \rightarrow \mathbf{id} = E\ ; \mid L = E\ ;$

$S \rightarrow \mathbf{id} = E\ ; \mid L = E\ ;$ { *S.nextlist = null;* }

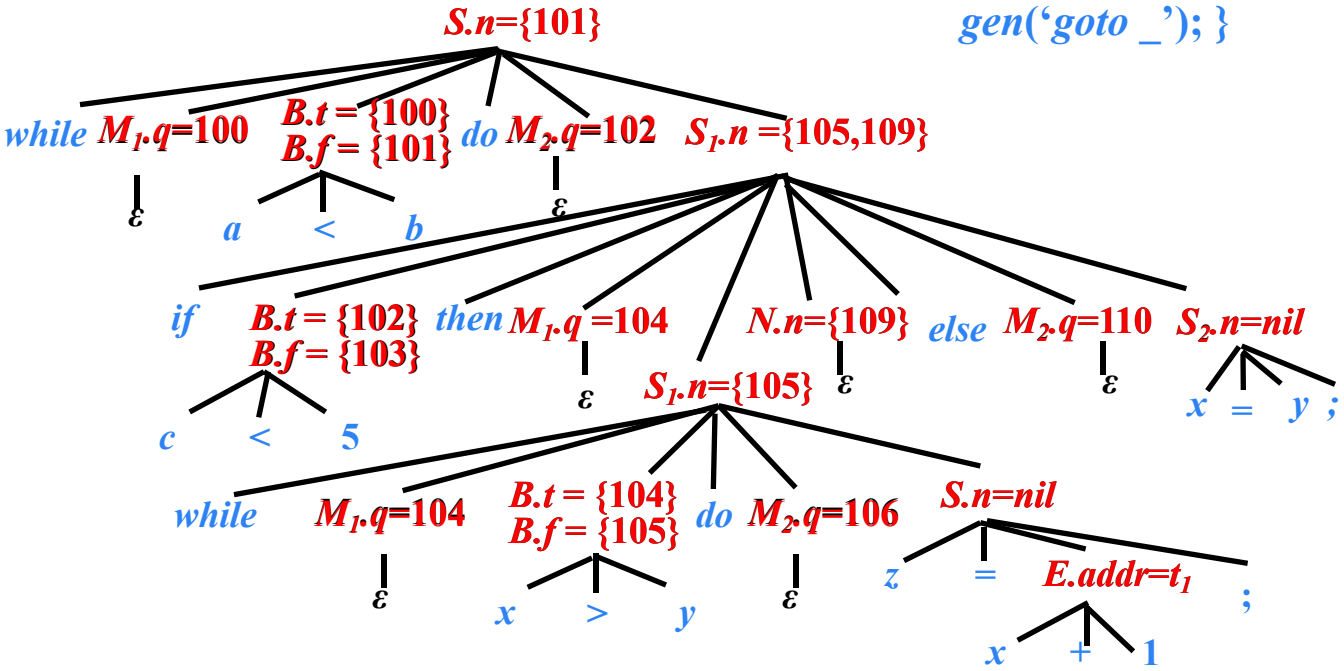
例

```
while  a < b  do

    if  c < 5  then

            while x > y do z = x + 1;
    else

            x = y;
```

$S \rightarrow$ while M_1 B do M_2 S_1
{ $S.nextlist = B.falselist$;
$backpatch(S_1.nextlist, M_1.quad)$;
$backpatch(B.truelist, M_2.quad)$;
$gen(\text{'goto'}\ M_1.quad)$; }

$S \rightarrow$ if B then M_1 S_1 N else M_2 S_2
{ $S.nextlist = merge(merge(S_1.nextlist, N.nextlist), S_2.nextlist)$;
$backpatch(B.truelist, M_1.quad)$;
$backpatch(B.falselist, M_2.quad)$; }

$N \rightarrow \varepsilon$ { $N.nextlist = makelist(nextquad)$;
$gen(\text{'goto \_'})$; }



100: *if* $a < b$ *goto* 102
101: *goto* _
102: *if* $c < 5$ *goto* 104
103: *goto* 110
104: *if* $x > y$ *goto* 106
105: *goto* 100
106: $t_1 = x + 1$
107: $z = t_1$
108: *goto* 104
109: *goto* 100
110: $x = y$
111: *goto* 100
112:

语句“*while a<b do if c<5 then while x>y do z=x+1; else x=y;*”的注释分析树

***while a<b do if c<5 then while x>y* do *z*=*x*+1; *else x*=*y*;**

100: *if* $a < b$ *goto* **102**
101: *goto* _
102: *if* $c < 5$ *goto* **104**
103: *goto* **110**
104: *if* $x > y$ *goto* **106**
105: *goto* **100**
106: $t_1 = x + 1$
107: $z = t_1$
108: *goto* **104**
109: *goto* **100**
110: $x = y$
111: *goto* **100**
112:

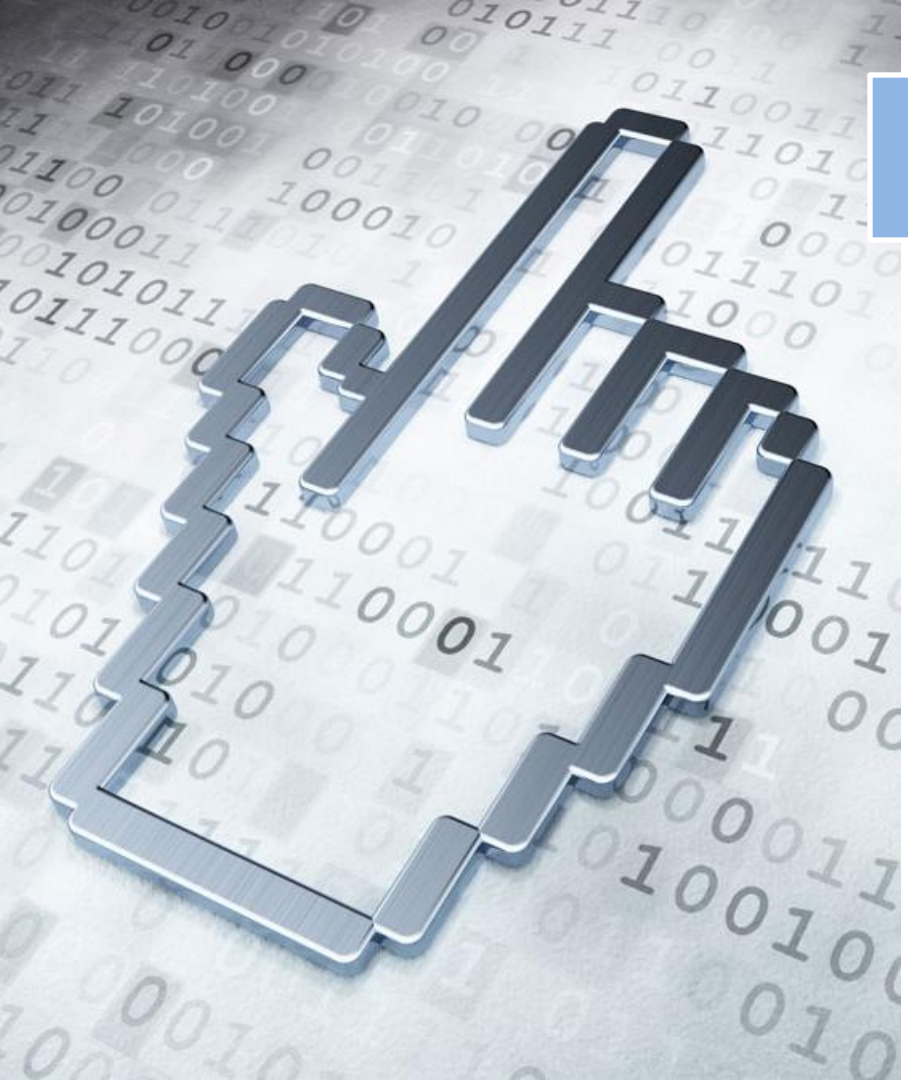
100: (*j*<, *a* ,*b* , **102**)
101: (*j* , - , - , -)
102: (*j*<, *c* , **5** , **104**)
103: (*j* , - , - , **110**)
104: (*j*>, *x* , *y* , **106**)
105: (*j* , - , - , **100**)
106: (+ , *x* , **1** , t_1)
107: (= , t_1 , - , *z*)
108: (*j* , - , - , **104**)
109: (*j* , - , - , **100**)
110: (= , *y* , - , *x*)
111: (*j* , - , - , **100**)
112:

练习3

- 试将下面的语句翻译成三地址码和对应的四元式序列（指令标号从1开始，跳转指令目标标号是数字代表的序号）。不必考虑避免生成冗余的goto语句。

（1）while a<c and b<d do

```
        if a=1 then c=c+1
        else while a<=d do
                a=a+2
```



本章内容

6.6 *switch*语句的翻译

- switch语句语法

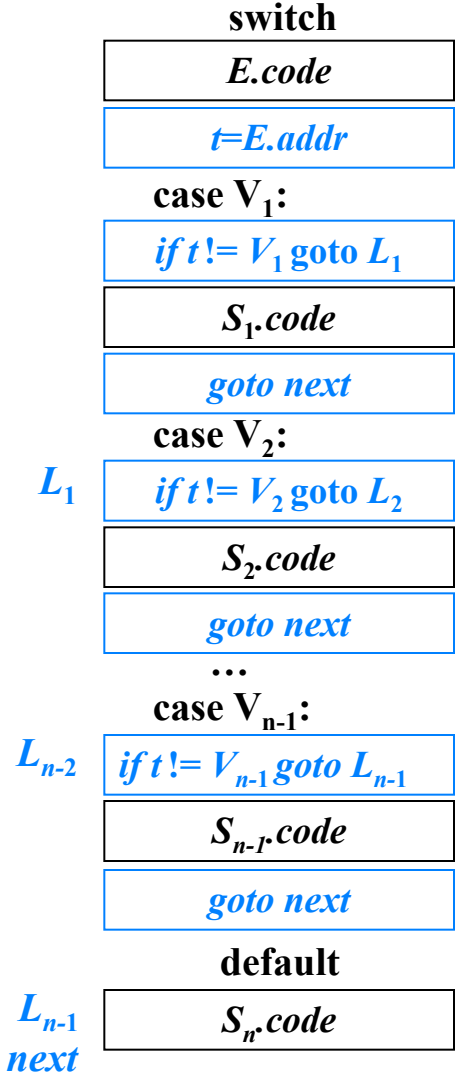
```
switch E
    begin
            case V1: S1
            case V2: S2
            . . .
            case Vn – 1: Sn – 1
            default: Sn
    end
```

- switch语句文法

$S \rightarrow$ switch E begin $Clist$ end

$Clist \rightarrow$ case V: S $Clist$

$Clist \rightarrow$ default: S

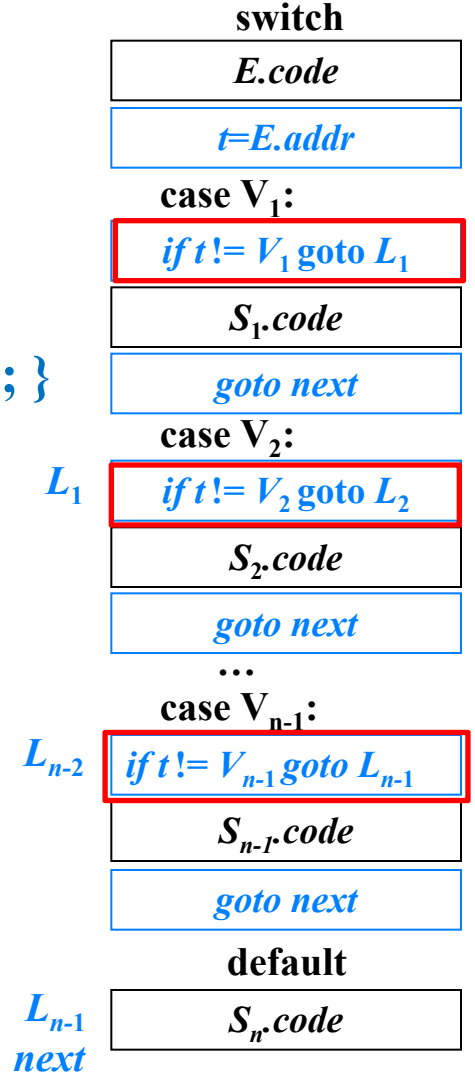


6.6 *switch*语句的翻译

$S \rightarrow$ switch E { t = *newtemp*();
gen(t '=' *E.addr*);
next = *newlabel*(); *l* = *newlabel*(); }
begin *Clist* end

Clist $\rightarrow$ case V: { *label*(*l*); *l* = *newlabel*();
gen('*if*' t '!=' V '*goto*' *l*); }
S { *gen*('*goto*' *next*); }
Clist

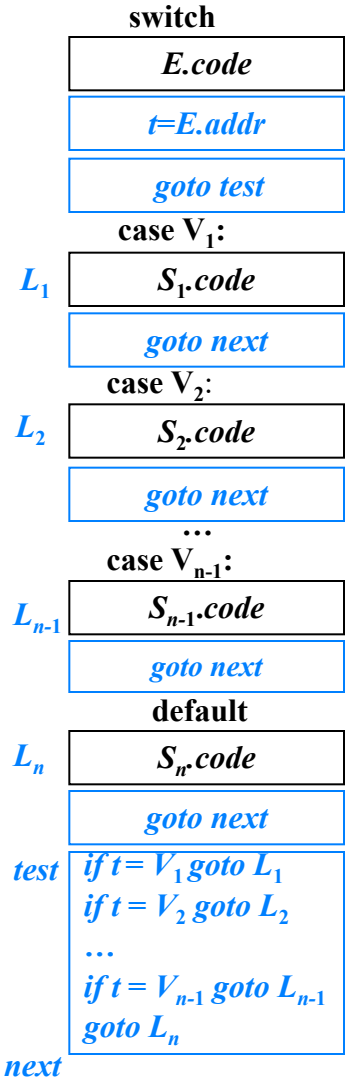
Clist $\rightarrow$ default: { *label*(*l*); }
S { *label* (*next*); }



*switch*语句的另一种翻译

switch *E*
  begin
    case V_1: S_1
    case V_2: S_2
    . . .
    case V_{n-1}: S_{n-1}
    default: S_n
  end

优势：根据分支的个数以及这些值是否在一个较小的范围内，这种条件跳转指令序列可以被翻译成最高效的*n*路分支(散列表)



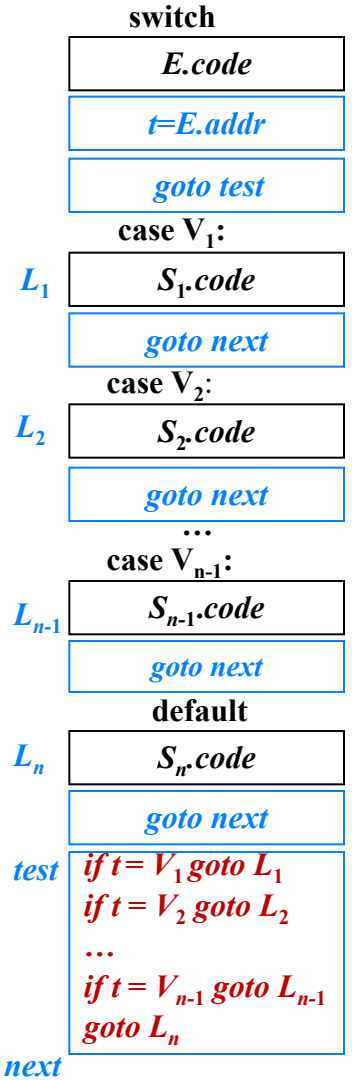
*switch*语句的另一种翻译

$S \rightarrow$ switch E { *t = newtemp(); gen(t '=' E.addr);*
test = newlabel(); next = newlabel();
gen('goto' test); }
begin *Clist* end

值-标号对加入队列q

Clist → case V: { *l = newlabel(); label(l); q(V, l);* }
S { *gen('goto' next);* } *C*

Clist → default: {*l = newlabel(); label(l);* }
S { *gen('goto' next) ; label(test);*
for 队列 *q* 中的每对 *(v, l) do*
{
gen('if' t '=' V 'goto' ln) ;
}
gen('goto' l) ; // 跳转到 default S
label (next); }

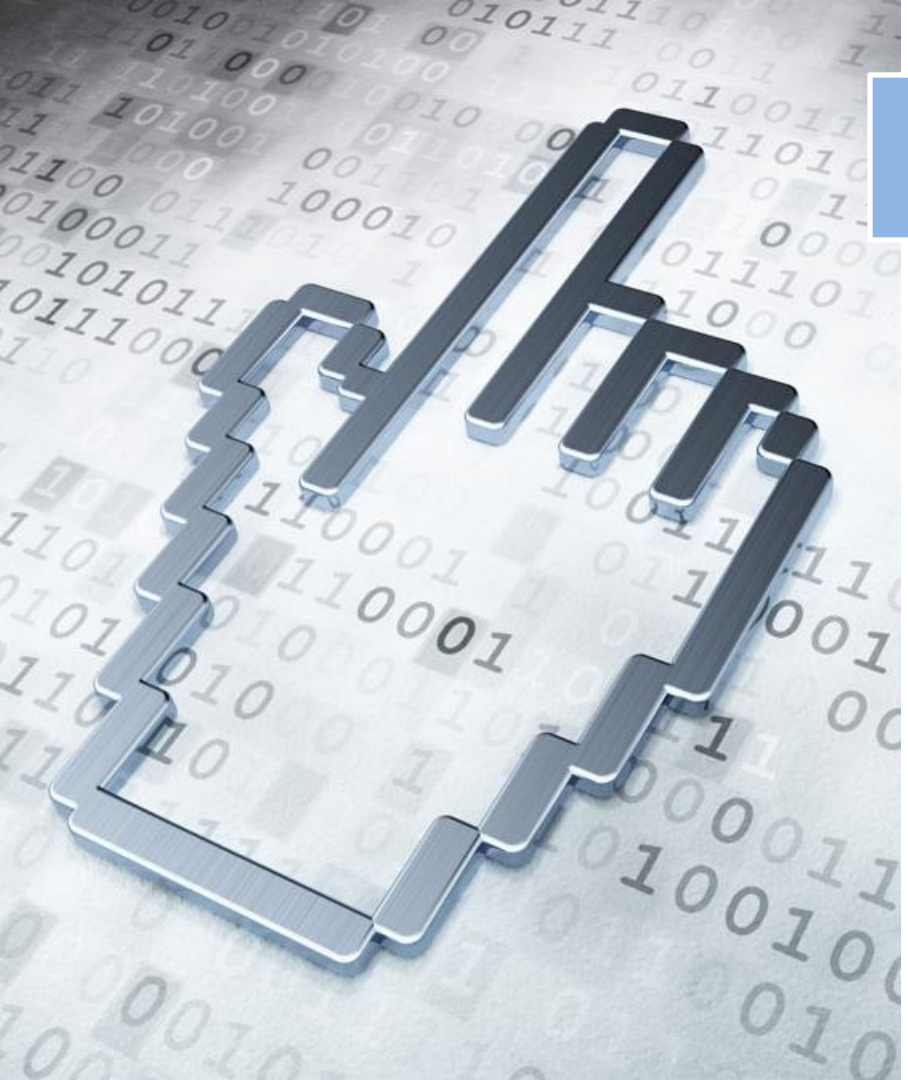


增加一种*case*指令

```
test :  if t = V1 goto L1
        if t = V2 goto L2
        . . .
        if t = Vn-1 goto Ln-1
        goto Ln
next :
```

```
test :  case t  V1    L1
        case t  V2    L2
           . . .
        case t  Vn-1  Ln-1
        case t  t     Ln
next :
```

指令 *case t V_i L_i* 和 *if t = V_i goto L_i* 的含义相同，
但是*case*指令更加容易被最终的代码生成器探测到，
从而对这些指令进行特殊处理



本章内容

6.7 过程调用的翻译

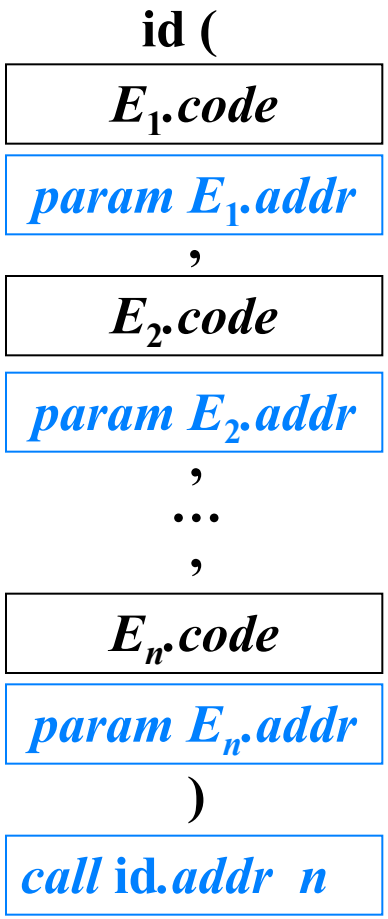
- 文法
  - $S \rightarrow$ call id ($Elist$)

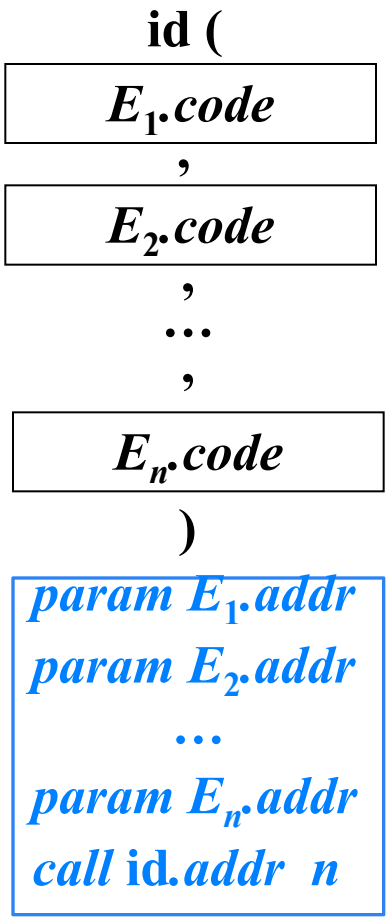
    $Elist \rightarrow Elist, E$

    $Elist \rightarrow E$

过程调用语句的代码结构

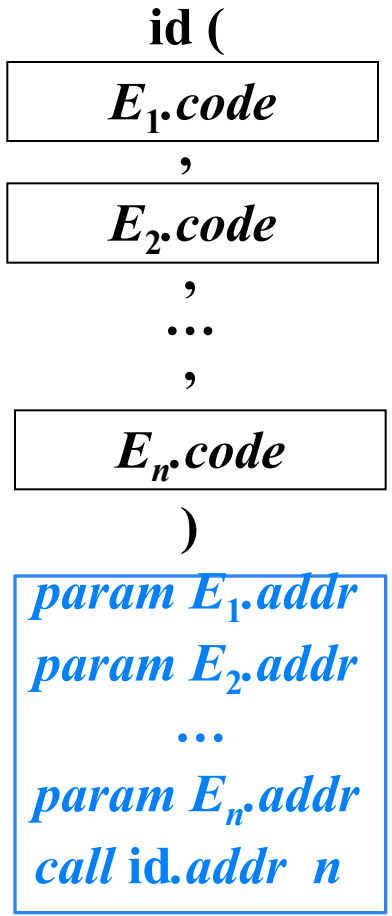
id(E_1, E_2, … , E_n)





过程调用语句的代码结构

id(E_1, E_2, … , E_n)



需要一个队列q存放E_1.addr 、E_2.addr、…、E_n.addr，以生成

过程调用语句的*SDD*

- $S \rightarrow$ call id (*Elist*)

  { n=0;
  *for q*中的每个*t do*
  { *gen*(‘*param*’ *t*);
  $n = n$+1;
  }
  gen(‘*call*’ id.*addr* ‘,’ *n*);
  }

- *Elist* → *E*

  { 将*q*初始化为只包含*E.addr*; }

- *Elist* →*Elist*$_1$, *E*

  { 将*E.addr*添加到*q*的队尾; }

id (

E_1.*code*

,

E_2.*code*

,

…

,

E_n.*code*

)

param E_1.*addr*
param E_2.*addr*
…
param E_n.*addr*
call id.*addr* *n*

例：翻译以下语句$f(b*c-1, x+y, x, y)$

$t_1 = b*c$

$t_2 = t_1 - 1$

$t_3 = x+y$

param t_2

param t_3

param x

param y

call f, 4

语义分析中的错误检测

- 变量或过程未经声明就使用 （赋值/表达式/过程调用语句翻译）

$S \rightarrow$ id = E;
{ p = *lookup*(id.*lexeme*); *if* p == *nil then error* ;
gen(p '=' *E.addr*); }
$E \rightarrow$ id
{ *E.addr* = *lookup*(id.*lexeme*); *if E.addr* == *nil then error* ; }

$S \rightarrow$ call id (*Elist*)
{ n=0;
*for q*中的每个*t do*
{ *gen*('*param*' *t*); $n = n+1$; }
gen('*call*' id.*addr* ',' n);
}

语义分析中的错误检测

- 变量或过程未经声明就使用 （赋值/表达式/过程调用语句翻译）
- 变量或过程名重复声明 （声明语句翻译）

$D \rightarrow T\ \mathrm{id};\{\ enter(\ \mathrm{id}.lexeme,\ T.type,\ offset\ );\ \ offset = offset + T.width;\ \}D$

语义分析中的错误检测

- 变量或过程未经声明就使用 （赋值/表达式/过程调用语句翻译）
- 变量或过程名重复声明 （声明语句翻译）
- 运算分量类型不匹配 （表达式翻译）

$E \rightarrow E_1 + E_2$
{ $E.addr = newtemp()$
if $E_1.type == integer$ and $E_2.type == integer$ then
{ $gen(E.addr$ '=' $E_1.addr$ 'int+' $E_2.addr)$; $E.type = integer$; }
else if $E_1.type == integer$ and $E_2.type == real$ then
{ $u = newtemp()$;
$gen(u$ '=' 'inttoreal' $E_1.addr)$;
$gen(E.addr$ '=' u 'real+' $E_2.addr)$;
$E.type = real$; }
...
}

语义分析中的错误检测

- 变量或过程未经声明就使用（赋值/表达式/过程调用语句翻译）
- 变量或过程名重复声明（声明语句翻译）
- 运算分量类型不匹配（表达式翻译）
- 操作符与操作数之间的类型不匹配
  - 数组下标不是整数（数组引用翻译）

$$L \rightarrow \mathrm{id}\,[E] \mid L_1\,[E]$$

语义分析中的错误检测

- 变量或过程未经声明就使用（赋值/表达式/过程调用语句翻译）
- 变量或过程名重复声明（声明语句翻译）
- 运算分量类型不匹配（表达式翻译）
- 操作符与操作数之间的类型不匹配
  - 数组下标不是整数（数组引用翻译）
  - 对非数组变量使用数组访问操作符（数组引用翻译）

$$L \rightarrow \mathrm{id}\,[E] \mid L_1\,[E]$$

语义分析中的错误检测

- 变量或过程未经声明就使用（赋值/表达式/过程调用语句翻译）
- 变量或过程名重复声明（声明语句翻译）
- 运算分量类型不匹配（表达式翻译）
- 操作符与操作数之间的类型不匹配
  - 数组下标不是整数（数组引用翻译）
  - 对非数组变量使用数组访问操作符（数组引用翻译）
  - 对非过程名使用过程调用操作符（过程调用翻译）

$S \rightarrow \textbf{call id} (\textit{Elist})$

语义分析中的错误检测

- 变量或过程未经声明就使用（赋值/表达式/过程调用语句翻译）
- 变量或过程名重复声明（声明语句翻译）
- 运算分量类型不匹配（表达式翻译）
- 操作符与操作数之间的类型不匹配
  - 数组下标不是整数（数组引用翻译）
  - 对非数组变量使用数组访问操作符（数组引用翻译）
  - 对非过程名使用过程调用操作符（过程调用翻译）
  - 过程调用的参数类型或数目不匹配

$S \rightarrow$ call id (*Elist*)
{ *n*=0;
*for q*中的每个*t do*
{ 检查 t 与 id 的形参类型？ }
检查数目是否匹配？ }

语义分析中的错误检测

- 变量或过程未经声明就使用 （赋值/表达式/过程调用语句翻译）
- 变量或过程名重复声明 （声明语句翻译）
- 运算分量类型不匹配 （表达式翻译）
- 操作符与操作数之间的类型不匹配
  - 数组下标不是整数 （数组引用翻译）
  - 对非数组变量使用数组访问操作符（数组引用翻译）
  - 对非过程名使用过程调用操作符 （过程调用翻译）
  - 过程调用的参数类型或数目不匹配
  - 函数返回类型有误

$D \rightarrow$ define T id (F) { S }
{ 检查id的返回值类型与S.type }
.....
S $\rightarrow$ return E; { S.type = E.type}

本章小结

- 声明语句的翻译
  - 变量声明的翻译
  - 数组声明的翻译
- 赋值语句的翻译
  - 简单赋值语句的翻译
  - 数组引用的翻译
- 控制语句的翻译
- 回填
- switch语句的翻译
- 过程调用语句的翻译

结束