第九章 代码生成

代码生成器的主要任务

哈尔滨工业大学　陈鄞

- 指令选择
  - 选择适当的目标机指令来实现中间表示 (*IR*) 语句
  - 例：
    - 三地址语句
      - *x = y + z*
    - 目标代码
      - *LD Ro , y /* 把 y 的值加载到寄存器 Ro 中 */*
      - *ADD Ro , Ro , z /* z 加到 Ro 上 */*
      - *ST x , Ro /* 把 Ro 的值保存到 x 中 */*

- **指令选择**
  - 选择适当的目标机指令来实现中间表示 (*IR*) 语句
  - 例：

三地址语句序列

- *a=b+c*
- *d=a+e*

目标代码

- *LD R0 , b // R0 = b*
- *ADD R0 , R0 , c // R0 = R0 + c*
- *ST a , R0 // a = R0*
- *LD R0 , a // R0 = a*
- *ADD R0 , R0 , e // R0 = R0 + e*

- **指令选择**
  - **选择适当的目标机指令来实现中间表示 (*IR*) 语句**
- 寄存器分配和指派
  - 把哪个值放在哪个寄存器中
- 指令排序
  - 按照什么顺序来安排指令的执行

第九章 代码生成

代码生成器的主要任务

哈尔滨工业大学　陈鄞

第九章 代码生成

一个简单的目标机模型

哈尔滨工业大学　陈鄞

- 三地址机器模型
  - 加载、保存、运算、跳转等操作
  - 内存按字节寻址
  - n 个通用寄存器 $R_0, R_1, \ldots, R_{n-1}$
  - 假设所有的运算分量都是整数
  - 指令之间可能有一个标号

- 加载指令　　*LD dst, addr*
  - *LD r, x*
  - *LD r1, r2*
- 保存　指令　　*ST x, r*
- 运算　指令　　*OP dst, src1, src2*
- 无条件跳转指令 *BR L*
- 条件跳转指令　*Bcond r, L*
  - 例 : *BLTZ r, L*

- 变量名 *a*
  - 例： *LD R1 , a*
    - *R1 = contents (a)*

- 变量名 *a*
- *a*(r)
  - *a* 是一个变量， r 是一个寄存器
  - 例： LD R1，*a*(R2)
    - R1 = contents (*a* + contents(R2))

- **变量名 *a***
- ***a*(*r*)**
- **c(r)**
  - **c 是一个整数**
  - **例： *LD R1 , 100 (R2)***
    - ***R1 = contents (contents(R2) + 100)***

- **变量名 *a***
- ***a*(*r*)**
- ***c*(*r*)**
- ****r***
  - 在寄存器 r 的内容所表示的位置上存放的内存位置
  - 例： *LD R1 , * R2*
    - *R1 = conents (contents (contents (R2)))*

- **变量名 *a***
- ***a*(*r*)**
- ***c*(*r*)**
- ***r**
- ***c*(*r*)**
  - 在寄存器 *r* 中内容加上 *c* 后所表示的位置上存放的内存位置
  - 例： *LD R1 , *100(R2)*
    - *R1 = conents (contents (contents(R2) + 100))*

- 变量名 *a*
- *a*(*r*)
- *c*(*r*)
- *r
- *c(*r*)
- #*c*
  - 例： LD R1，#100
    - R1 = 100

第九章 代码生成

一个简单的目标机模型

哈尔滨工业大学　陈鄞

第九章 代码生成

指令选择

哈尔滨工业大学　陈鄞

- 三地址语句
  - *x = y - z*
- 目标代码
  - *LD R1 , y // R1 = y*
  - *LD R2 , z // R2 = z*
  - *SUB R1 , R1 , R2 // R1 = R1 - R2*
  - *ST x , R1 // x = R1*

尽可能避免使用上面的全部四个指令，如果
- ✓ **所需的运算分量已经在寄存器中了**
- ✓ **运算结果不需要存放回内存**

- **三地址语句**
  - *b = a[i]*
  - *a* 是一个实数数组，每个实数占 8 个字节
- 目标代码
  - *LD R1 , i // R1 = i*
  - *MUL R1 , R1, 8 // R1=R1 * 8*
  - *LD R2 , a(R1) // R2=contents (a + contents(R1))*
  - *ST b , R2 // b = R2*

- 三地址语句
  - *a* [*j*] = *c*
  - *a* 是一个实数数组，每个实数占 8 个字节
- 目标代码
  - *LD R1 , c // R1 = c*
  - *LD R2 , j // R2 = j*
  - *MUL R2 , R2 , 8 // R2 = R2 * 8*
  - *ST a(R2) , R1 // contents(a+contents(R2))=R1*

- 三地址语句
  - *x = *p*
- 目标代码
  - *LD R1, p // R1 = p*
  - *LD R2, 0 (R1) // R2 = contents (0 + contents (R1))*
  - *ST x , R2 // x = R2*

- 三地址语句
  - *p = y
- 目标代码
  - LD R1 , p // R1 = p
  - LD R2 , y // R2 = y
  - ST 0(R1), R2 //contents (0 + contents (R1)) = R2

- 三地址语句
  - *if x < y goto L*
- 目标代码
  - *LD R1 , x // R1 = x*
  - *LD R2 , y // R2 = y*
  - *SUB R1 , R1 , R2 // R1=R1 - R2*
  - *BLTZ R1 , M // if R1 < 0 jump to M*

M 是标号为 L 的三地址指令所产生的目标代码中的第一个指令的标号

静态存储分配

- 三地址语句
  - *call callee*
- 目标代码
  - *ST callee.staticArea, #here + 20*
  - *BR callee.codeArea*

- 三地址语句
  - *return*
- 目标代码
  - *BR *callee.staticArea*

callee 的活动记录在静态区中的起始位置

callee 的目标代码在代码区中的起始位置

栈式存储分配

- 三地址语句
  - *call callee*
- 目标代码
  - *ADD SP, SP, #caller.recordsize*
  - *ST 0(SP), #here + 16*
  - *BR callee.codeArea*

- 三地址语句
  - *return*
- 目标代码
  - 被调用过程
    - *BR *0(SP)*
  - 调用过程
    - *SUB SP , SP , #caller.recordsize*

第九章 代码生成

指令选择

哈尔滨工业大学　陈鄞

第九章 代码生成

寄存器的选择

哈尔滨工业大学　陈鄞

- 对每个形如 *x = y op z* 的三地址指令 *I*，执行如下动作
  - 调用函数 *getReg(I)* 来为 *x* 、 *y* 、 *z* 选择寄存器，把这些寄存器称为 *Rx* 、 *Ry* 、 *Rz*
  - 如果 *Ry* 中存放的不是 *y* ，则生成指令“ *LD Ry , y′* ”。 *y′* 是存放 *y* 的内存位置之一
  - 类似的，如果 *Rz* 中存放的不是 *z* ，生成指令“ *LD Rz, z′* ”
  - 生成目标指令“ *OP Rx, Ry, Rz*”

- 寄存器描述符 （*register descriptor*）
  - 记录每个寄存器当前存放的是哪些变量的值
- 地址描述符 （*address descriptor*）
  - 记录运行时每个名字的当前值存放在哪个或哪些位置
  - 该位置可能是寄存器、栈单元、内存地址或者是它们的某个集合
  - 这些信息可以存放在该变量名对应的符号表条目中

- 对于一个在基本块的出口处可能活跃的变量 x，如果它的地址描述符表明它的值没有存放在 x 的内存位置上，则生成指令“ ST x, R”（ R 是在基本块结尾处存放 x 值的寄存器 ）

- 当代码生成算法生成加载、保存和其他指令时，它必须同时更新寄存器和地址描述符
  - 对于指令“ *LD R , x*”
    - 修改 *R* 的寄存器描述符，使之只包含 *x*
    - 修改 *x* 的地址描述符，把 *R* 作为新增位置加入到 *x* 的位置集合中
    - 从任何不同于 *x* 的地址描述符中删除 *R*

- 当代码生成算法生成加载、保存和其他指令时，它必须同时更新寄存器和地址描述符
  - 对于指令“ *LD R , x*”
  - 对于指令“ OP Rx，Ry，Rz”
    - 修改 Rx 的寄存器描述符，使之只包含 x
    - 从任何不同于 Rx 的寄存器描述符中删除 x
    - 修改 x 的地址描述符，使之只包含位置 Rx
    - 从任何不同于 x 的地址描述符中删除 Rx

- 当代码生成算法生成加载、保存和其他指令时，它必须同时更新寄存器和地址描述符
  - 对于指令“ *LD R , x*”
  - 对于指令“ *OP Rx , Ry , Rz*”
  - 对于指令“ *ST x , R*”
    - 修改 x 的地址描述符，使之包含自己的内存位置

- 当代码生成算法生成加载、保存和其他指令时，它必须同时更新寄存器和地址描述符
  - 对于指令“ *LD R , x*”
  - 对于指令“ *OP Rx , Ry , Rz*”
  - 对于指令“ *ST x , R*”
  - 对于复制语句 *x=y* ，如果需要生成加载指令“ *LD Ry , y′* ”则
    - 修改 *Ry* 的寄存器描述符，使之只包含 *y*
    - 修改 *y* 的地址描述符，把 *Ry* 作为新增位置加入到 *y* 的位置集合中
    - 从任何不同于 *y* 的变量的地址描述符中删除 *Ry*
    - 修改 *Ry* 的寄存器描述符，使之也包含 *x*

$t = a - b$

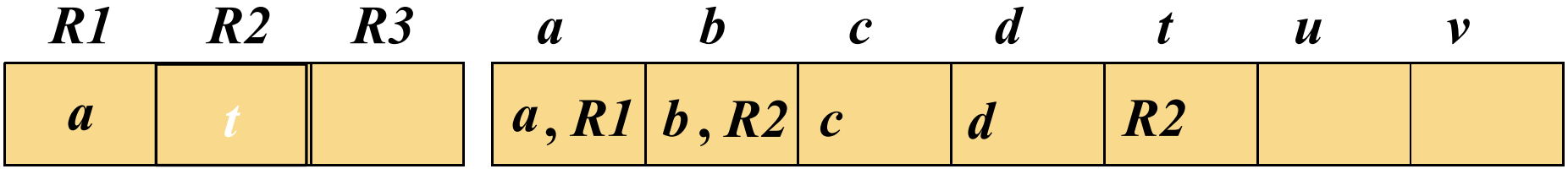
$u = a - c$

$v = t + u$

$a = d$

$d = v + u$

```
LD   R1 , a
LD   R2 , b
SUB R2 , R1 , R2
```



$t = a - b$

$u = a - c$

$v = t + u$

$a = d$

$d = v + u$

```
LD   R3 , c
SUB R1 , R1 , R3
```

R1	R2	R3
u	t	c

a	b	c	d	t	u	v
a , R1	b	c ,R3	d	R2	R1	

$$t = a - b$$

$$u = a - c$$

$$v = t + u$$

$$a = d$$

$$d = v + u$$

R1	R2	R3
u	t	v

a	b	c	d	t	u	v
a	b	c ,R3	d	R2	R1	R3

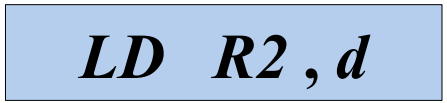
$$t = a - b$$

$$u = a - c$$

$$v = t + u$$

$$a = d$$

$$d = v + u$$



R1	R2	R3
u	d , a	v

a	b	c	d	t	u	v
R2	b	c	d ,R2	R2	R1	R3

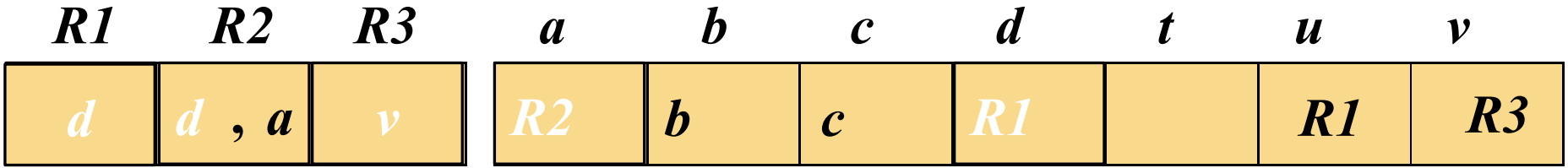
$$t = a - b$$

$$u = a - c$$

$$v = t + u$$

$$a = d$$

$$d = v + u$$



$t = a - b$

$u = a - c$

$v = t + u$

$a = d$

$d = v + u$

exit

```
ST  a , R2
ST  d , R1
```

R1	R2	R3	a	b	c	d	t	u	v
d	a	v	R2 ,a	b	c	R1 ,d			R3

第九章 代码生成

寄存器的选择

哈尔滨工业大学　陈鄞

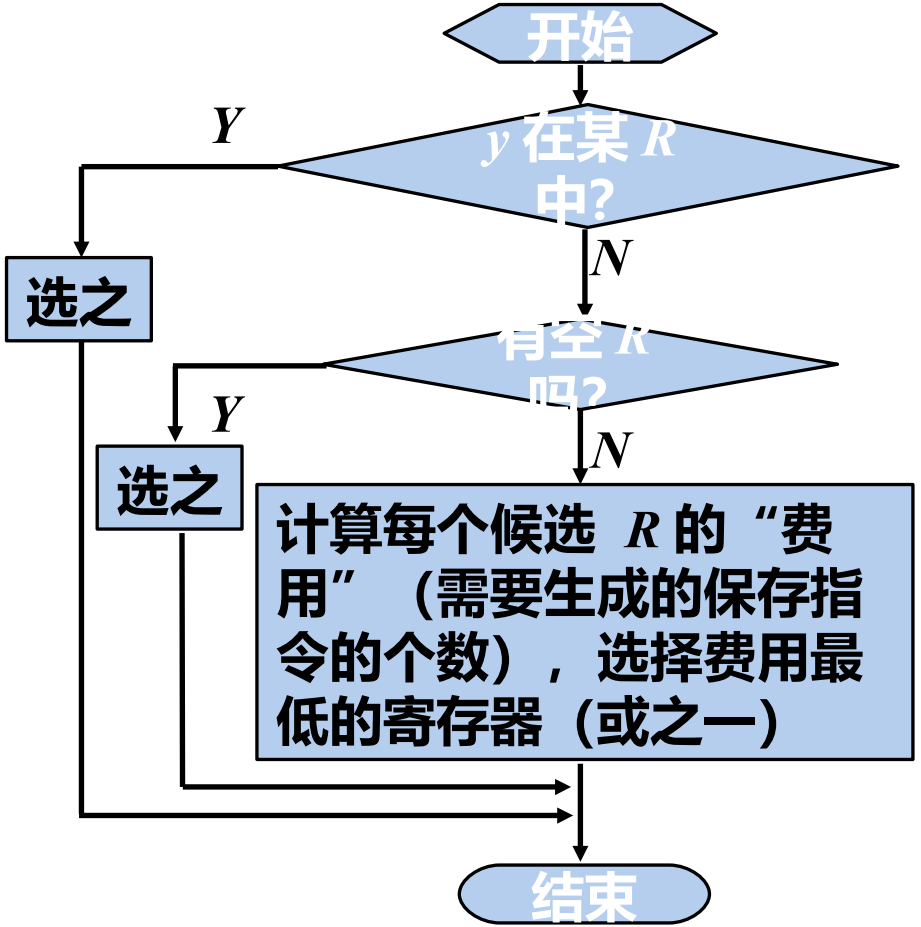
第九章 代码生成

寄存器选择函数 getReg 的设计

哈尔滨工业大学　陈鄞

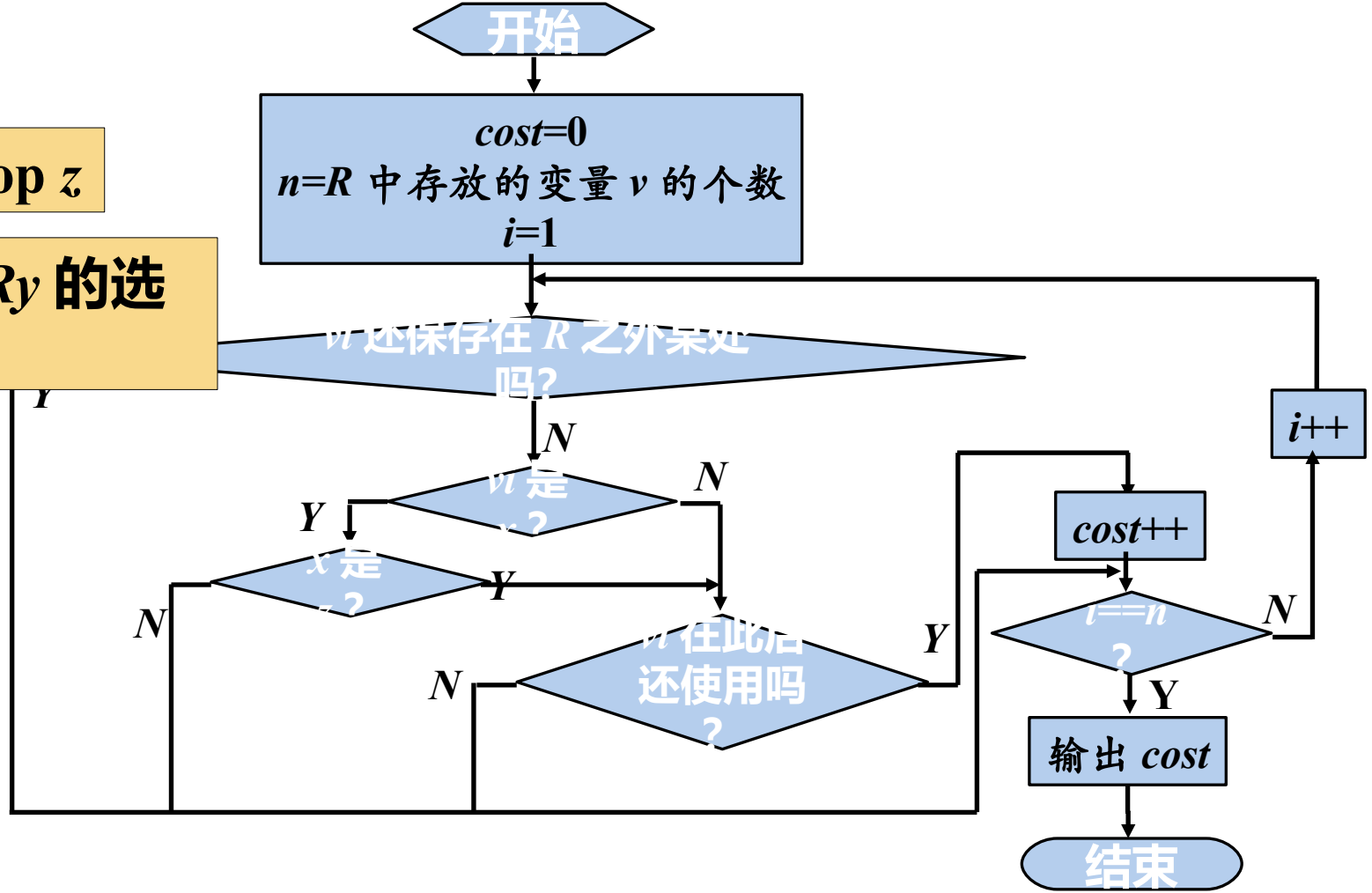
$x = y \text{ op } z$

寄存器 *Ry* 的选择



$x = y \text{ op } z$

寄存器 *Ry* 的选择



$x = y$ op z

- 选择方法与 Ry 类似，区别之处在于
  - 因为 x 的一个新值正在被计算，因此只存放了 x 的值的寄存器对 Rx 来说总是可接受的，即使 x 就是 y 或 z 之一 (因为我们的机器指令允许一个指令中的两个寄存器相同)
  - 如果 y 在指令 I 之后不再使用，且 (在必要时加载 y 之后) Ry 仅仅保存了 y 的值，那么， Ry 同时也可以用作 Rx 。对 z 和 Rz 也有类似选择

当 *I* 是复制指令 $x=y$ 时，选择好 *Ry* 后，令 *Rx* =*Ry*

第九章 代码生成

寄存器选择函数 getReg 的设计

哈尔滨工业大学　陈鄞

第九章 代码生成

窥孔优化

哈尔滨工业大学　陈鄞

- 窥孔 (*peephole*) 是程序上的一个小的滑动窗口
- 窥孔优化是指在优化的时候，检查目标指令的一个滑动

  窗口 (即窥孔) ，并且只要有可能就在窥孔内用更快或更

  短的指令来替换窗口中的指令序列
- 也可以在中间代码生成之后直接应用窥孔优化来提高中

- 冗余指令删除
- 控制流优化
- 代数优化
- 机器特有指令的使用

- 消除冗余的加载和保存指令
  - 例

三地址指令序列

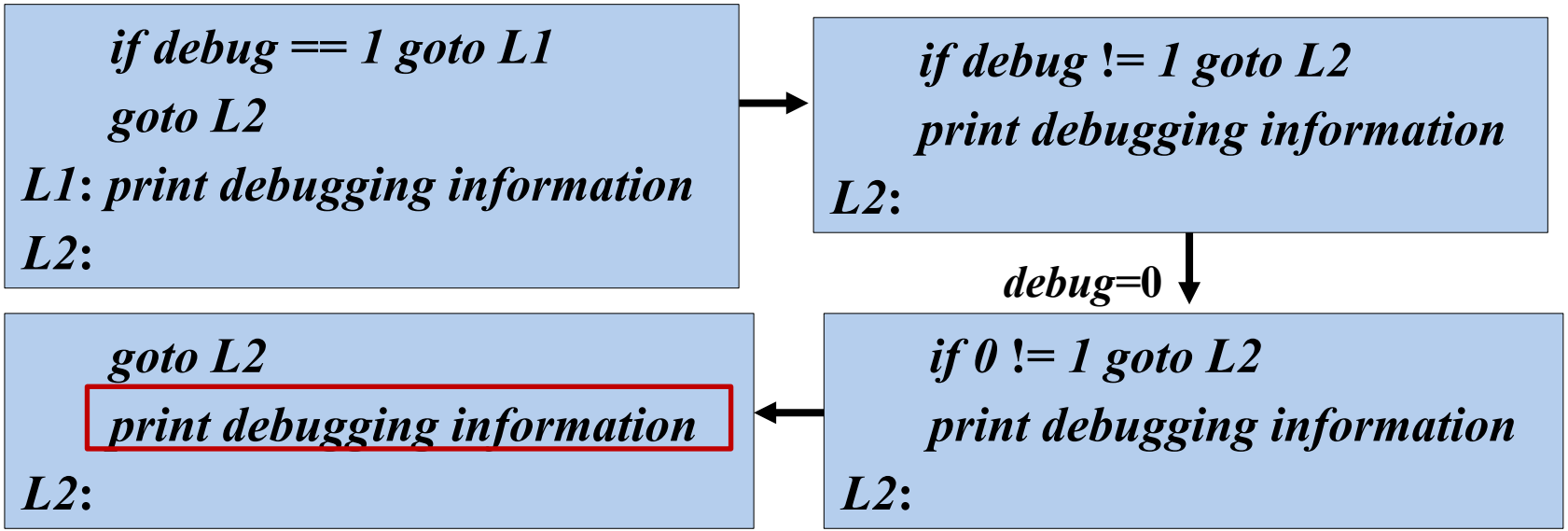
- $a=b+c$
- $d=a+e$

目标代码

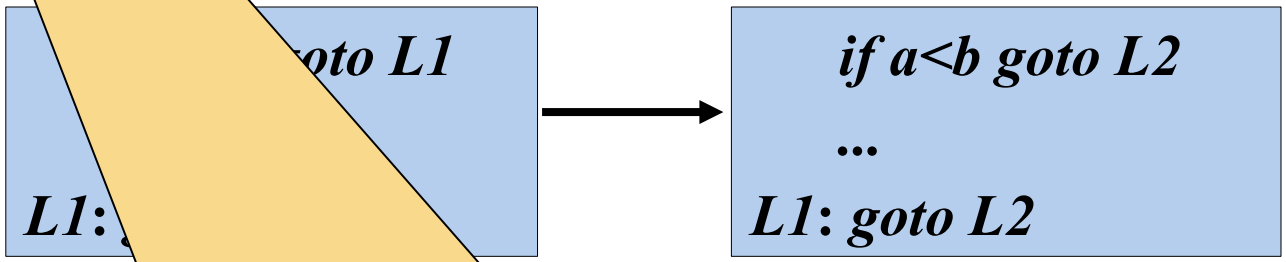
- $LD\ R0\ ,\ b$ // $R0 = b$
- $ADD\ R0\ ,\ R0\ ,\ c$ // $R0 = R0 + c$
- $ST\ a\ ,\ R0$ // $a = R0$
- $LD\ R0\ ,\ a$ // $R0 = a$
- $ADD\ R0\ ,\ R0\ ,\ e$ // $R0 = R0 + e$
- $ST\ d\ ,\ R0$ // $d = R0$

如果第四条指

- **消除冗余的加载和保存指令**
- 消除不可达代码
  - 一个紧跟在无条件跳转之后的不带标号的指令可以被删除
    - 例



- 在代[illegible]出现跳转到跳转指令的指令时，某些条件下可以使用一[illegible]转指令来代替
  - 例



如果不再有跳转到 *L1* 的指令，并且语句 *L1*: *goto L2* 之前是一个无条件跳转指令，则可以删除该语句

- 代数恒等式
  - 消除窥孔中类似于 x=x+0 或 x=x*1 的运算指令
- 强度削弱
  - 对于乘数（除数）是 2 的幂的定点数乘法（除法），用移位运算实现代价比较低
  - 除数为常量的浮点数除法可以通过乘数为该常量倒数的乘法来求近似值

- 充分利用目标系统的某些高效的特殊指令来提高代码效率
  - 例如： *INC* 指令可以用来替代加 1 的操作

第九章 代码生成

窥孔优化

哈尔滨工业大学　陈鄞

- 假设每个运算符 op 有且只有一个对应的机器指令。该指令对存放在寄存器中的所需的运算分量进行运算，并把结果存放在一个寄存器中。机器指令的形式如下
  - *LD reg, mem*
  - *OP reg, reg, reg*
  - *ST mem, reg*

- 对每个形如 *x=y* 的语句 *I*，假设 *getreg* 总是为 *x* 和 *y* 选择同一个寄存器
  - 如果 *Ry* 中存放的不是 *y*，则生成指令“ *LD Ry, y′* ”
  - 修改 *Ry* 的寄存器描述符，表明 *Ry* 中也存放了 *x* 的值