

第八章 代码优化

# 流图



哈尔滨工业大学 陈鄞

#### 基本块(Basic Block)

- > 基本块是满足下列条件的最大的连续三地址指令序列
  - ▶ 控制流只能从基本块的第一个指令进入该块。也就是说,没有 跳转到基本块中间或末尾指令的转移指令
  - 除了基本块的最后一个指令,控制流在离开基本块之前不会跳转或者停机

如何划分基本块?

#### 基本块划分算法

- > 输入:
  - > 三地址指令序列
- >输出:
  - > 输入序列对应的基本块列表, 其中每个指令恰好被分配给一个基本块
- >方法:
  - ≥ 首先,确定指令序列中哪些指令是首指令(leaders),即某个基本块的第 一个指令
    - 1. 指令序列的第一个三地址指令是一个首指令
    - 2. 任意一个条件或无条件转移指令的目标指令是一个首指令
    - 3. 紧跟在一个条件或无条件转移指令之后的指令是一个首指令
  - 然后,每个首指令对应的基本块包括了从它自己开始,直到下一个首指令(不含)或者指令序列结尾之间的所有指令

## 例

```
i = m - 1; j = n; v = a[n];
while (1) {
   do i = i + 1; while (a[i] < v);
   do j = j - 1;while (a[j] > v);
   if (i \ge j) break;
   x=a[i]; a[i]=a[j]; a[j]=x;
x=a[i]; a[i]=a[n]; a[n]=x;
```

```
(1) i = m - 1
                                            (16) t_7 = 4 * i
      (2) j = n
                                            (17) t_8 = 4 * j
B_1 (3) t_1 = 4 * n
                                            (18) t_9 = a[t_8]
      (4)  v = a[t_1]
                                       B_{5} (19) a[t_{7}] = t_{9}
     \overline{(5)} \overline{i} = \overline{i} + \overline{1}
                                             (20) t_{10} = 4 * j
      (6) t_2 = 4 * i
                                             (21) \ a[t_{10}] = x
\frac{B_2}{(7)} t_3 = a[t_2]
                                            (22) goto (5)
      (8) if t_3 > v \ goto(5)
                                            (23) t_{11} = 4 * i
      (9) j = j - 1
                                            (24) x = a[t_{11}]
      (10) t_{\Delta} = 4 * j
                                             (25) t_{12} = 4 * i
B_3 (11) t_5 = a[t_4]
                                             (26) t_{13} = 4 * n
      (12) if t_5 > v \ goto(9)
                                             (27) t_{14} = a[t_{13}]
B_4 (13) if i > = j goto(23)
                                             (28) a[t_{12}] = t_{14}
      (14) t_6 = 4 * i
                                             (29) t_{15} = 4 * n
      (15) x = a[t_6]
                                             (30) a[t_{15}] = x
```

#### 流图(Flow Graphs)

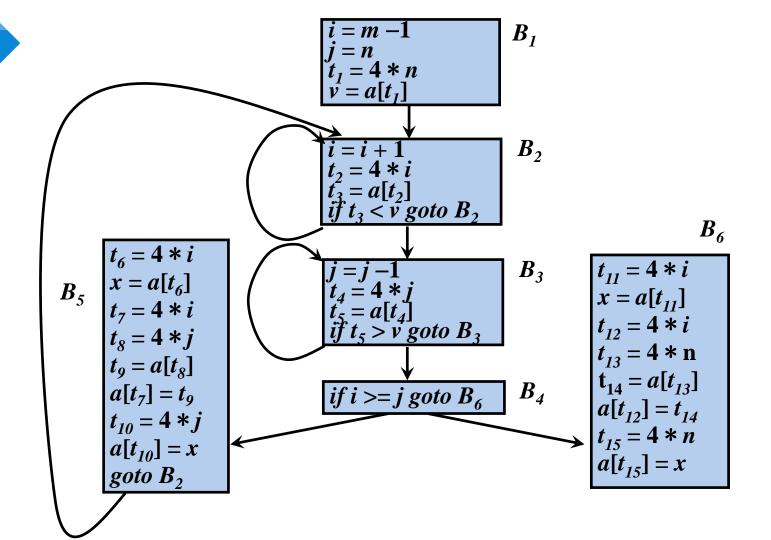
- ▶流图的结点是一些基本块
- ►从基本块B到基本块C之间有一条边当且仅当基本块C 的第一个指令可能紧跟在B的最后一条指令之后执行

此时称B是C的前驱(predecessor), C是B的后继(successor)

### 流图(Flow Graphs)

- ▶流图的结点是一些基本块
- ►从基本块B到基本块C之间有一条边当且仅当基本块C 的第一个指令可能紧跟在B的最后一条指令之后执行
  - >有两种方式可以确认这样的边:
    - ▶ 有一个从B的结尾跳转到C的开头的条件或无条件跳 转语句
    - $\triangleright$ 按照原来的三地址语句序列中的顺序,C紧跟在之B后,且B的结尾不存在无条件跳转语句

例 (1) i = m - 1 $(16) t_7 = 4 * i$ (2) j = n $(17) t_8 = 4 * j$  $B_1$  (3)  $t_1 = 4 * n$  $(18) t_9 = a[t_8]$  $\boldsymbol{B_1}$  $B_{5}$  (19)  $a[t_{7}] = t_{9}$  $(4) \quad v = a[t_1]$ (5) i = i + 1 $(20) t_{10} = 4 * j$ (6)  $t_2 = 4 * i$  $(21) \ a[t_{10}] = x$  $\begin{array}{ccc} B_2 & (0) & t_2 \\ (7) & t_3 = a[t_2] \end{array}$ (22) goto (5) $B_3$ (8) if  $t_3 > v \ goto(5)$  $(23) t_{11} = 4 * i$ (9) j = j - 1 $(24) x = a[t_{11}]$  $(10) t_{4} = 4 * j$  $B_{4}$  $(25) t_{12} = 4 * i$  $B_3$  (11)  $t_5 = a[t_4]$  $B_6$  (26)  $t_{13} = 4 * n$  (27)  $t_{14} = a[t_{13}]$ (12) if  $t_5 > v \ goto(9)$  $\mathbf{B}_{4}$  (13) if i > = j goto(23) $(28) a[t_{12}] = t_{14}$  $(14) t_6 = 4 * i$ 无条件跳转  $(29) t_{15} = 4 * n$ 因此后面不能接B6  $(15) x = a[t_6]$  $(30) \ a[t_{15}] = x$ 





第八章 代码优化

# 流图



哈尔滨工业大学 陈鄞



### 优化的分类

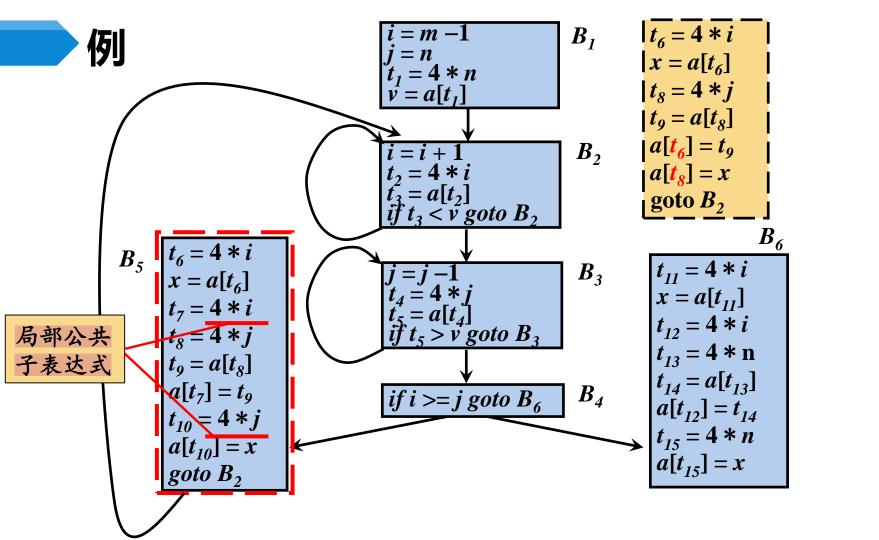
- ▶机器无关优化
  - ▶针对中间代码
- ▶机器相关优化
  - ▶针对目标代码
- ▶局部代码优化
  - ▶单个基本块范围内的优化
- ▶全局代码优化
  - ▶面向多个基本块的优化

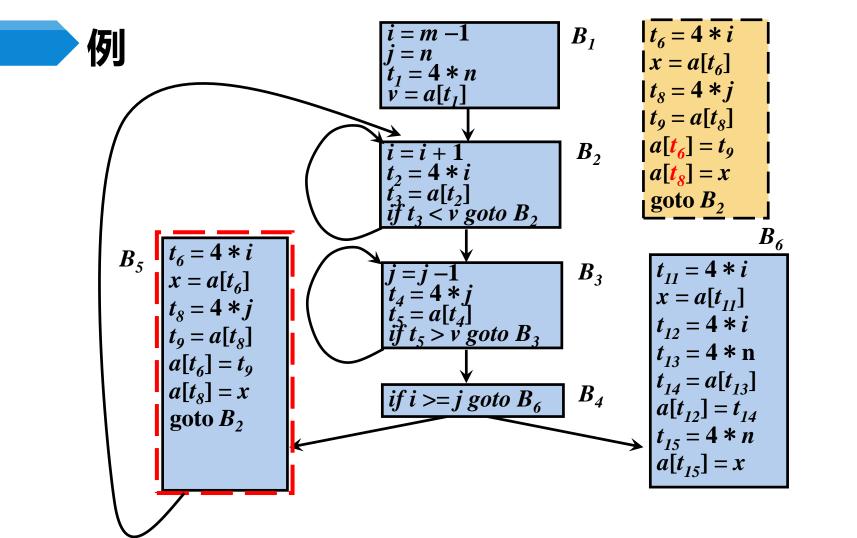
#### 常用的优化方法

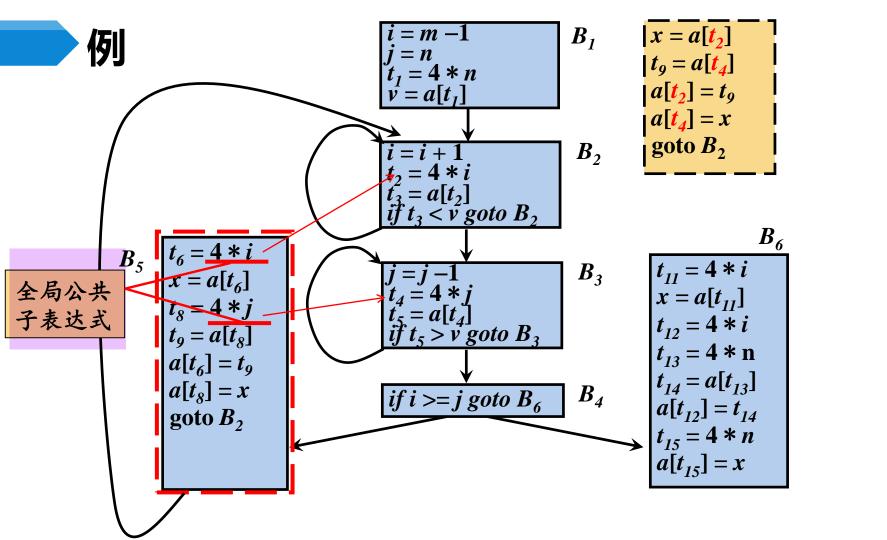
- > 删除公共子表达式
- ▶删除无用代码
- ▶常量合并
- 一代码移动
- ▶强度削弱
- ▶删除归纳变量

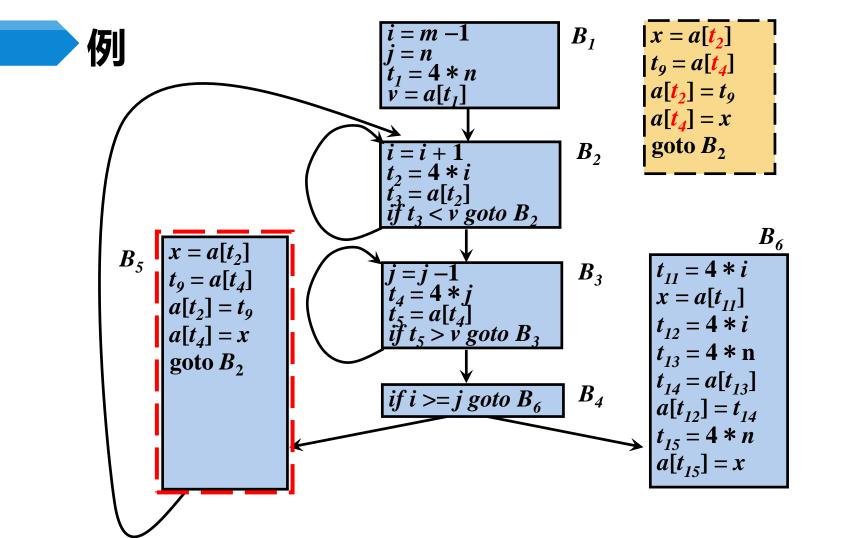
## ① 删除公共子表达式

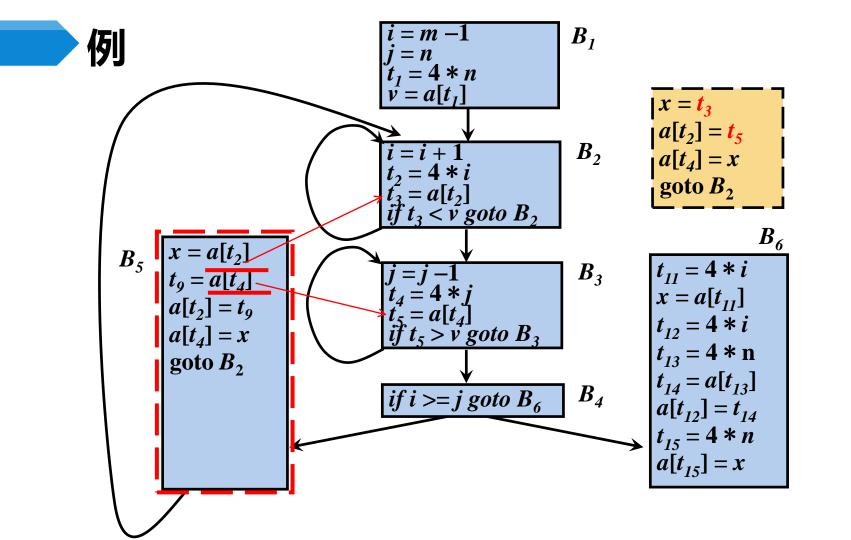
- **>公共子表达式** 
  - 一如果表达式x op y先前已被计算过,并且从先前的计算到现在,x op y中变量的值没有改变,那么x op y的这次出现就称为公共子表达式(common subexpression)

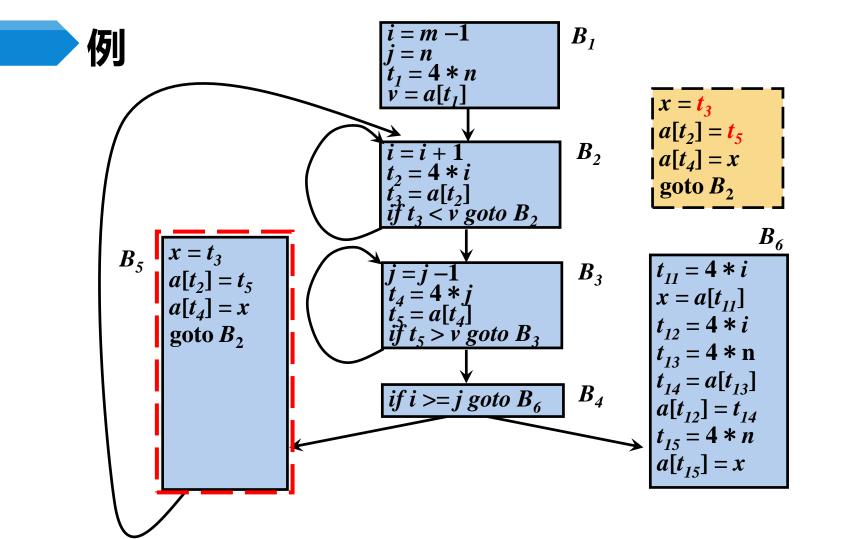


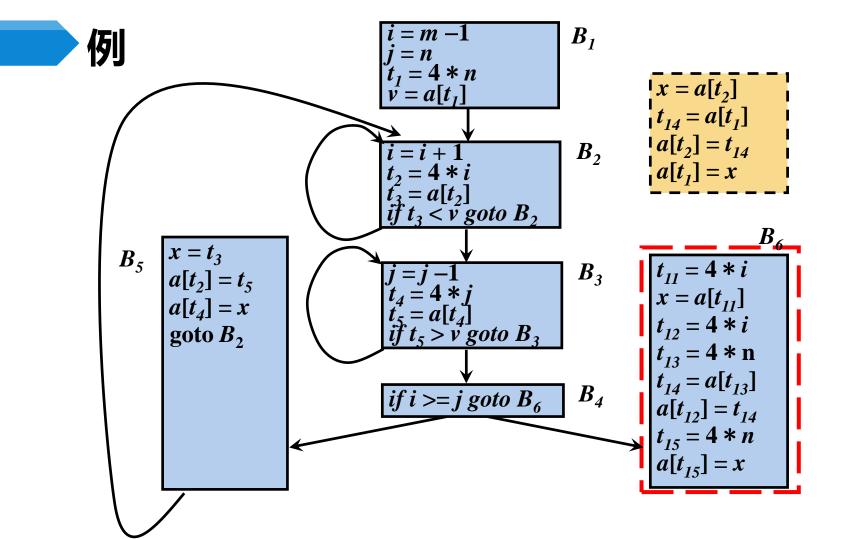


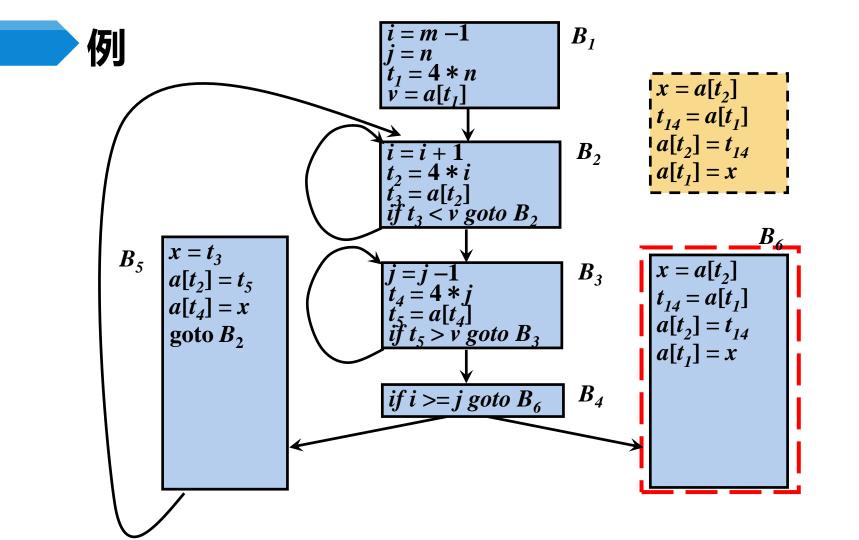


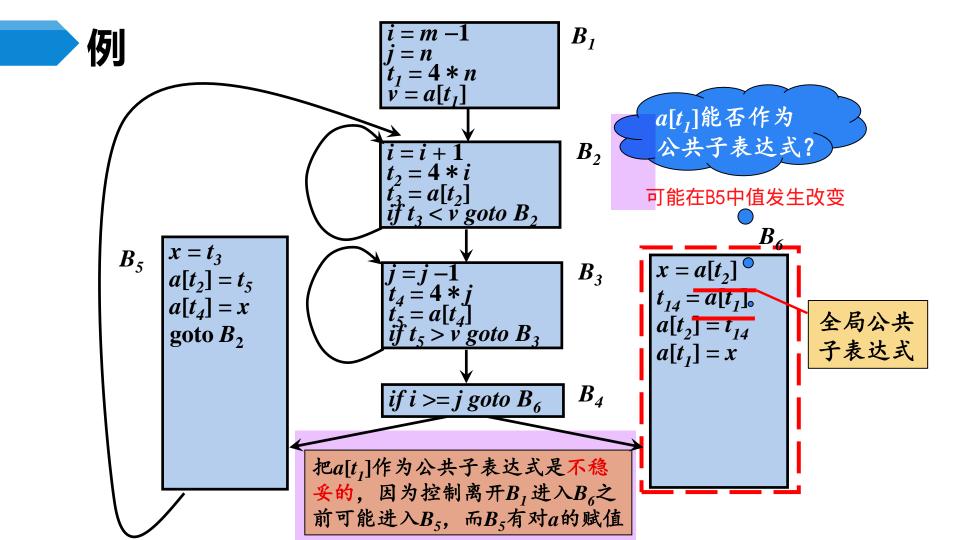


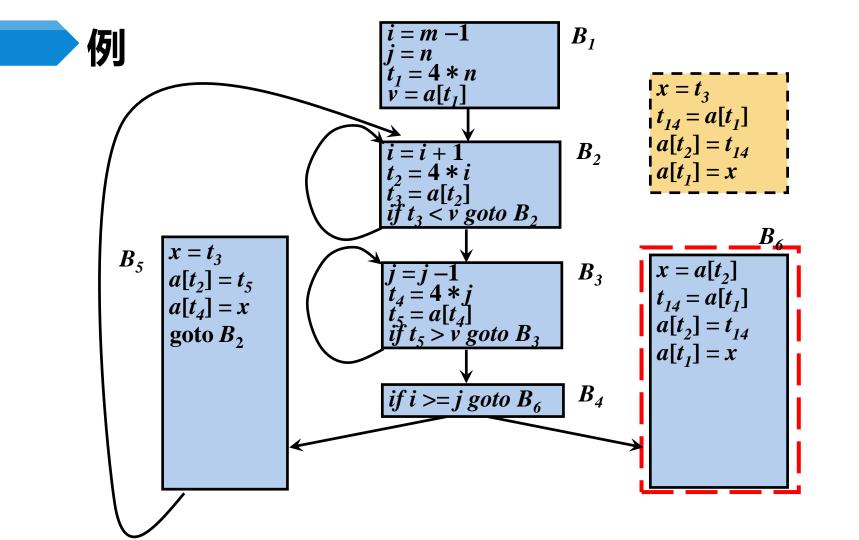


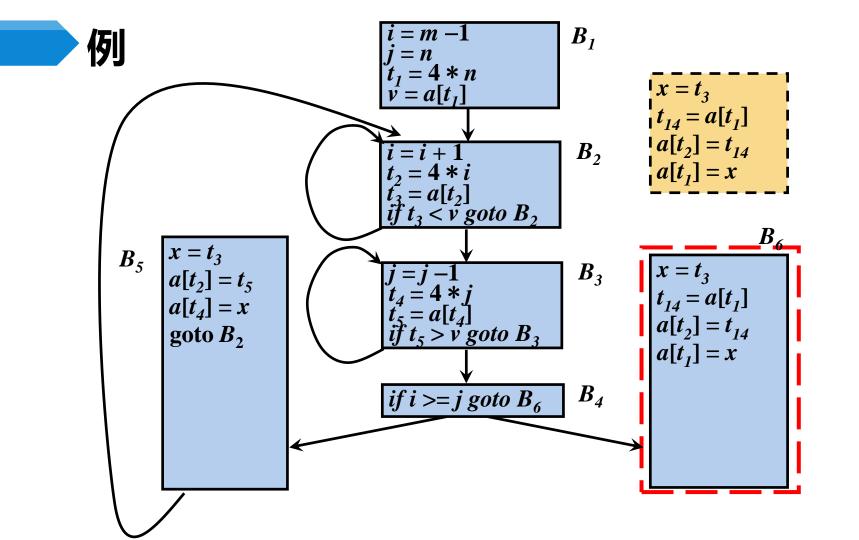


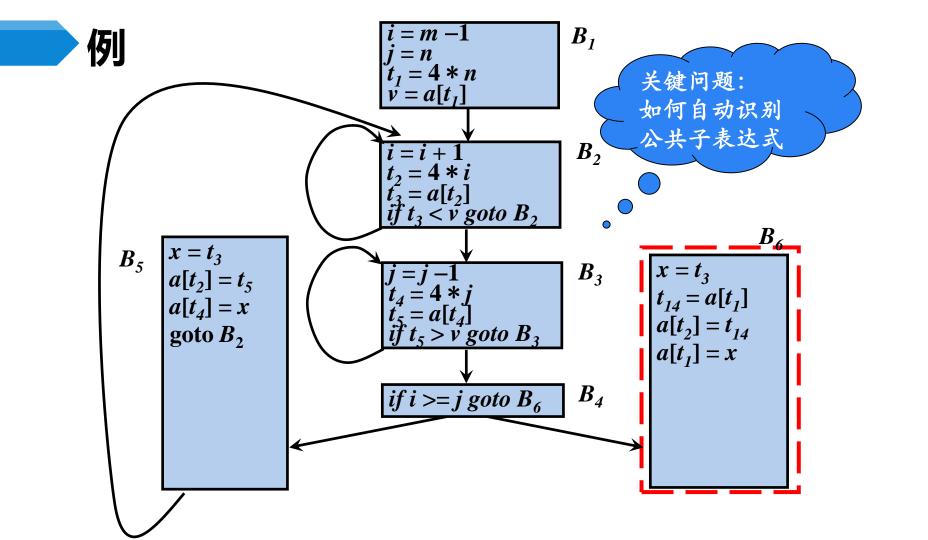












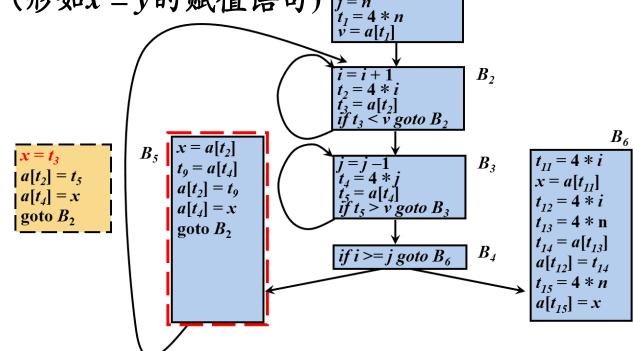




### ② 删除无用代码

- ▶复制传播
  - 户常用的公共子表达式消除算法和其它一些优化算法会引入

一些复制语句(形如x = y的赋值语句)  $\begin{bmatrix} i = m - 1 \\ j = n \\ t_1 = 4 * n \end{bmatrix}$   $B_1$ 

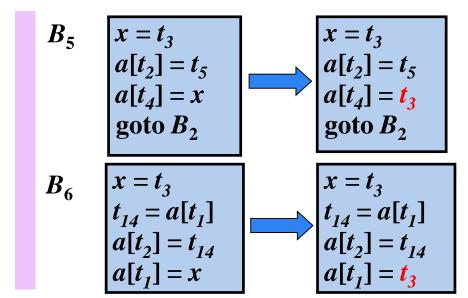


### ② 删除无用代码

▶复制传播

〉例

- ▶常用的公共子表达式消除算法和其它一些优化算法会引入 一些复制语句(形如x=y的赋值语句)
- ▶ 复制传播: 在复制语句x = y之后尽可能地用y代替x



## ② 删除无用代码

- ▶复制传播
  - ▶常用的公共子表达式消除算法和其它一些优化算法会引入 一些复制语句(形如x=y的赋值语句)
  - ≥ 复制传播:在复制语句x=y之后尽可能地用y代替x
    - > 复制传播给删除无用代码带来机会
- ► 无用代码(死代码Dead-Code): 其计算结果永远不会被使用的语句

 $\boldsymbol{B}_1$ = m - 1例 如何自动识别  $t_1 = 4 * n$ 无用代码?  $\vec{v} = a[t_1]$  $\begin{vmatrix} t_{14} = a[t_1] \\ a[t_2] = t_{14} \end{vmatrix}$  $a[t_2] = t_5$  $B_2$  $a[t_4] = t_3$  $t_2 = 4 * i$  $a[t_1] = t_3$  $f_3 = a[t_2]$  $goto B_2$  $t_3 < \tilde{v} goto B_2$  $x = t_3$  $B_5$  $\boldsymbol{B_3}$  $a[t_2] = t_5$  $t_{14} = a[t_1]$  $a[t_4] = t_3$  $a[t_2] = t_{14}$  $t_5 > v goto B_3$ goto  $B_2$  $a[t_1] = t_3$  $B_4$  $|if i\rangle = j goto B_6$ 程序员不大可能有意引入无用代码, 无用代码 通常是因为前面执行过的某些转换而造成的

 $\boldsymbol{B_1}$ 例  $t_1 = 4 * n$  $t_{14} = a[t_1]$  $a[t_2] = t_{14}$  $a[t_2] = t_5$  $a[t_4] = t_3$  $\boldsymbol{B}_2$  $a[t_1] = t_3$  $\begin{aligned}
 t_3^2 &= a[t_2] \\
 if t_3 &< v \ goto \ B_2
 \end{aligned}$  $goto B_2$  $\boldsymbol{B_3}$  $t_5 > v goto B_3$  $t_{14} = a[t_{13}]$  $if i >= j goto B_6$  $a[t_{10}] = x$ goto B<sub>2</sub>

## ③ 常量合并(Constant Folding)

如果在编译时刻推导出一个表达式的值是常量,就可以使用该常量来替代这个表达式。该技术被称为常量合并

 $\triangleright$ 例: l = 2\*3.14\*r

$$t_1 = 2 * 3.14$$
 $t_2 = t_1 * r$ 
 $l = t_2$ 
 $t_1 = 2 * 3.14$ 
 $t_2 = 6.28 * r$ 
 $l = t_2$ 

### 4 代码移动(Code Motion)

- 一代码移动
  - ➤ 这个转换处理的是那些不管循环执行多少次都得到相同 结果的表达式(即循环不变计算, loop-invariant computation), 在进入循环之前就对它们求值

## 例

 $(5) t_2 = t_1 * pi$ 

(6)  $t_3 = t_2 * r$ 

 $(7) t_{\Delta} = t_3 * r$ 

```
▶ 优化后程序
> 原始程序
                                    C = 1/360*pi*r*r;
  for( n=10; n<360; n++)
                                    for(n=10; n<360; n++)
  \{ S=1/360*pi*r*r*n;
                                    \{ S=C*n;
    printf("Area is \%f", S);
                                     printf("Area is \%f", S);
                   循环不变计算
  (1) n = 1
                        (8) t_5 = t_4 * n
                                             如何自动识别
  (2) if n > 360 got 6(21)
                        (9) S = t_5
                                             循环不变计算?
  (3) goto (4)
  (4) t_1 = 1 / 360
                        (18) t_0 = n + 1
```

 $(19) n = t_0$ 

(21)

(20) goto (4)

#### 循环不变计算的相对性

▶对于多重嵌套的循环,循环不变计算是相对于某个循环而言的。可能对于更加外层的循环,它就不是循环不变计算

>例:

```
for(i = 1; i < 10; i++)
for(n=1; n < 360/(5*i); n++)
{ S=(5*i)/360*pi*r*r*n; ...}
```

## ⑤ 强度削弱(Strength Reduction)

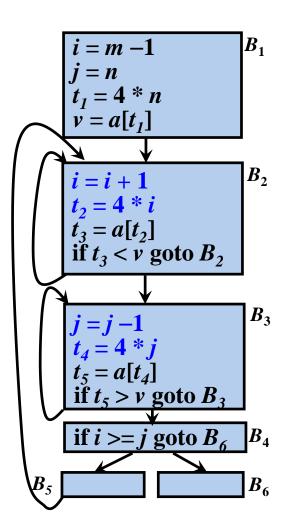
#### > 强度削弱

>用较快的操作代替较慢的操作,如用加代替乘

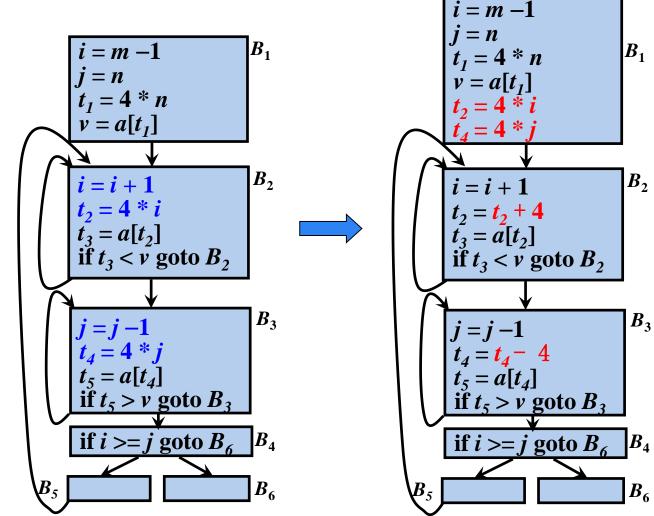
〉例

#### 循环中的强度削弱

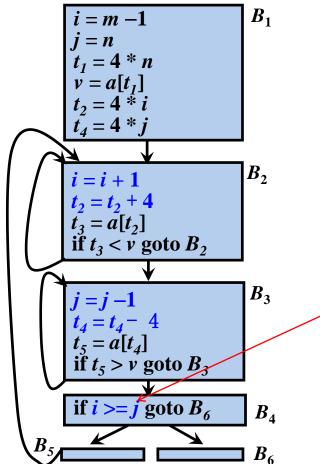
- ▶归纳变量
  - 一对于一个变量x,如果存在一个正的或负的常数c使得每次x被赋值时它的值总增加c,那么x就称为归纳变量(Induction Variable)



归纳变量可以通过在每次循环迭代中进行一次简单的增量运算(加法或减法)来计算



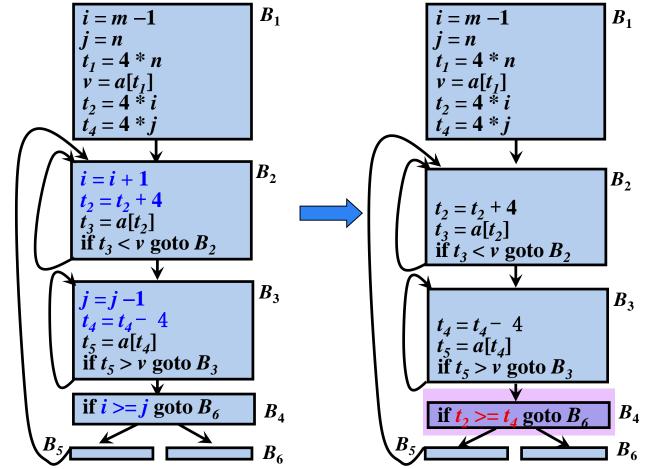
## ⑥ 删除归纳变量



在沿着循环运行时,如果有一组归纳变量的值的变化保持步调一致,常常可以将这组变量删除为只剩一个

i\_j控制t2\_t4之后就没有用了,因此删除

## ⑥ 删除归纳变量







第八章 代码优化

# 基本块的优化



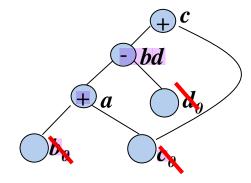
哈尔滨工业大学 陈鄞

## 基本块的优化

- ▶很多重要的局部优化技术首先把一个基本块转换成为
  - 一个无环有向图(directed acyclic graph, DAG)

#### 基本块的 DAG 表示

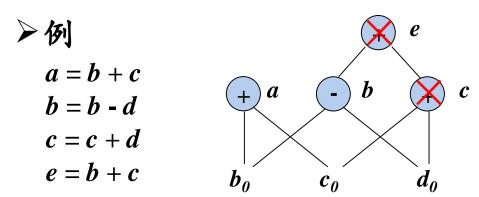
对于形如x=y+z的三地址指令,如果已经有一个结点表示y+z,就不往DAG中增加新的结点,而是给已经存在的结点附加定值变量x



- ▶ 基本块中的每个语句s都对应一个内部结点N
  - ▶ 结点N的标号是s中的运算符;同时还有一个定值变量表被关联到N,表示s是 在此基本块内最晚对表中变量进行定值的语句
  - ▶ N的子结点是基本块中在s之前、最后一个对s所使用的运算分量进行定值的语句对应的结点。如果s的某个运算分量在基本块内没有在s之前被定值,则这个运算分量对应的子结点就是代表该运算分量初始值的叶结点(为区别起见,叶节点的定值变量表中的变量加上下脚标0)
  - $\triangleright$  在为语句x=y+z构造结点N的时候,如果x已经在某结点M的定值变量表中,则从M的定值变量表中删除变量x

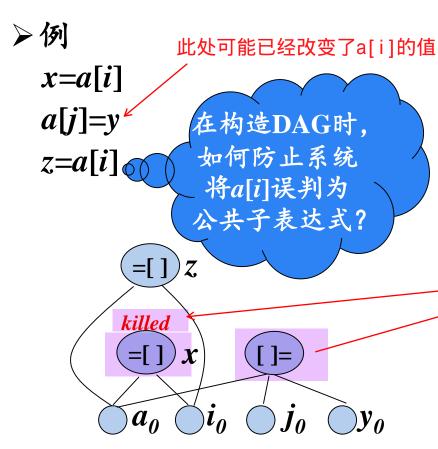
#### 基于基本块的 DAG 删除无用代码

► 从一个DAG上删除所有没有附加活跃变量(活跃变量是指其值可能会在以后被使用的变量)的根结点(即没有父结点的结点)。重复应用这样的处理过程就可以从DAG中消除所有对应于无用代码的结点



假设a和b是活跃变量,但c和e不是

### 数组元素赋值指令的表示



- 》对于形如a[j]=y的三地址指令,创建一个运算符为"[]="的结点,这个结点有3个子结点,分别表示a、j和y
- > 该结点没有定值变量表
- ≥该结点的创建将杀死所有已经 建立的、其值依赖于a的结点
- 一个被杀死的结点不能再获得 任何定值变量,也就是说,它 不可能成为一个公共子表达式

### 根据基本块的DAG可以获得一些非常有用的信息

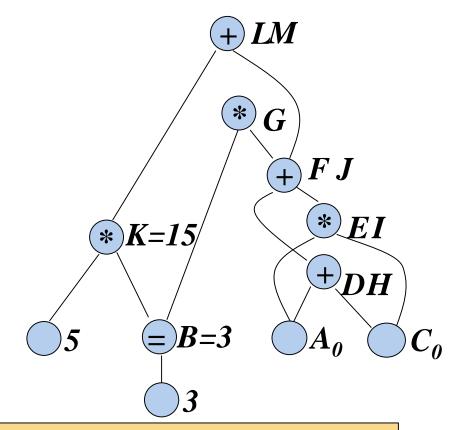
- >确定哪些变量的值在该基本块中赋值前被引用过
  - ▶在DAG中创建了叶结点的那些变量
- 确定哪些语句计算的值可以在基本块外被引用
  - 上在DAG构造过程中为语句S(该语句为变量x定值)创建的节点N,在DAG构造结束时x仍然是N的定值变量

#### 从 DAG 到基本块的重组

- ▶对每个具有若干定值变量的节点,构造一个三地址语句 来计算其中某个变量的值
  - ➤ 倾向于把计算得到的结果赋给一个在基本块出口处活跃的变量(如果没有全局活跃变量的信息作为依据,就要假设所有变量都在基本块出口处活跃,但是不包含编译器为处理表达式而生成的临时变量)
- 》如果结点有多个附加的活跃变量,就必须引入复制语句, 以便给每一个变量都赋予正确的值

- 〉给定一个基本块
  - ① B = 3
  - (2) D = A + C
  - ③ E = A \* C
  - $(4) \mathbf{F} = \mathbf{E} + \mathbf{D}$
  - $\bigcirc$  G = B \* F
  - $\bigcirc H = A + C$
  - (7) I = A \* C

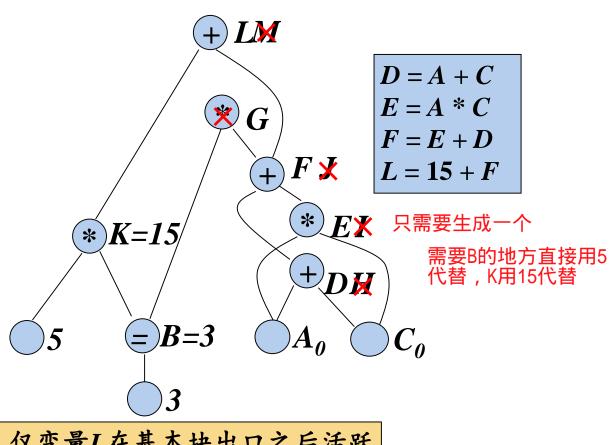
  - (9) K = B \* 5
  - 10 L = K + J
  - $\widehat{11} M = L$



假设: 仅变量L在基本块出口之后活跃

#### 〉给定一个基本块

- ① B = 3
- (2) D = A + C
- $(4) \mathbf{F} = \mathbf{E} + \mathbf{D}$
- $\bigcirc G = B * F$
- $\bigcirc H = A + C$
- (7) I = A \* C
- (8) J = H + I
- (9) K = B \* 5
- (10) L = K + J
- $\widehat{(11)} M = L$



假设: 仅变量L在基本块出口之后活跃



第八章 代码优化

# 基本块的优化



哈尔滨工业大学 陈鄞