

if-conversion 及其在 LLVM 上的实现

高翔

qasdfgyuiop@gmail.com

March 5, 2014

Abstract

人们追求程序运行的性能的脚步从来都没有停歇过，为了使得程序更快地运行，人们不断的改进处理器的性能。然而，处理器的性能只是程序性能的一部分，要实现高性能，编译器生成的代码的执行效率也至关重要。更先进的构架给编译器提出了更高的要求，人们也一直在探索如何让编译器生成更高效率的代码。由于早先的编译器无法很好地处理控制依赖，Allen 等人于 1983 年提出了 if 转换 (if-conversion) 的概念 [1]。if-conversion 能够消除程序中的除了后向分支以外的所有分支，这样就使得程序中的控制依赖被转化为数据依赖，编译器可以在此基础上进行进一步优化。1987 年，Ferrante 等人提出了程序依赖图 (PDG) 的概念 [4]，使得编译器可以轻易的处理控制依赖。尽管现代的编译器可以轻易处理控制依赖，分支语句的存在有时候仍然会严重影响性能。分支语句会中断指令流水，造成严重的性能损失。程序中大约 20% 的语句是分支语句，如果一个分支语句会造成 2 个周期的延迟，那么会有大约 40% 的机器时间浪费在分支上 [3]。此外，分支语句会将程序划分为若干基本块，从而阻断了进一步的优化。If 转换可以通过现代处理器提供的条件执行指令，实现分支消除，这样做不仅可以减少分支造成的延迟，而且可以将若干小的基本块合并成一个大的基本块，进而为后续优化提供便利。由于上述原因，虽然现代编译器能够很好地处理控制依赖，if 转换仍然显得尤为重要。

We have never stopped pursuing the performance of the program, to achieve this goal, the performance of the processor has been highly developed. However, the performance of the processor is only part of the performance of the program. The performance of the code generated by the compiler is also an important factor. The modern architecture has brought greater challenges to the development of compiler. How to design a compiler that generates high performance instructions has been a problem for a long time. Early compiler couldn't process control dependency. To solve this problem, Allen et al. came up with the concept of if-conversion [1], which can remove all branches in the program except for the backward branch. The deletion of branches converts control dependency to data dependency,

making it possible for the compiler to do further optimizations. In 1987, Ferrante et al. came up with the concept of program dependency graph (PDG) [4], making it easier for compilers to handle control dependency. Albeit this improvement, if-conversion is also important since the branch can significantly harm the performance. Branch interrupts the pipeline of instructions, causing great loss in performance. According to statistics, about 20% instructions are branches in a program. Assume that one branch can cause about 2 cycles delay, then about 40% of machine time will be wasted in branch [3]. What's more, branch divides the program to several basic blocks, making it harder for further optimization. If-conversion removes branches by taking use of predicted execution mechanism of modern processor. This can not only eliminate the delay caused by branch, but also merge several small basic blocks to a large basic block, making a lot of optimizations possible. Due to the reasons stated above, if-conversion is also very important.

目录

1	分支与性能的关系	1
1.1	分支对性能的阻碍	1
1.2	谓词执行对性能的改善	1
1.3	分支消除对性能的影响	2
2	if-conversion 的发展简史	2
3	if-conversion 的运行机制	4
3.1	控制依赖	4
3.2	RK 算法	5
3.3	Hyperblock	7
3.4	逆向 if-conversion	10
4	if-conversion 的执行策略	11
4.1	if-conversion 给编译器带来的挑战	11
4.2	August 的基于逆向 if-conversion 的平衡算法	11
5	LLVM 上 if-conversion 的实现	11
	附录 A Allen 的工作	11
	附录 B 后控制节点树的计算	21
	参考文献	21

1 分支与性能的关系

1.1 分支对性能的阻碍

在现代的高性能处理器中，分支指令是一个代价昂贵的指令 [11]。深度流水以及多发射极大的提高了处理器的性能，但是频繁出现的分支导致的控制流转移会中断指令流水线的连续性，从而降低性能。为了解决分支语句对性能的影响，现代处理器往往配置分支预测机制，即在条件分支执行前，处理器对分支方向进行预测，并按照预测的方向进行取指执行。不幸的是，分支预测经常出错，一旦预测失败就会造成非常严重的性能损失。研究显示，错误的分支预测会使得性能降低 2-10 倍 [17] [19] [10]。

分支预测失败对性能的影响主要有三个方面。首先，处理器会在预测的分支方向取出大量指令执行，一旦分支预测失败，所有执行的结果都必须被丢弃。这就意味着，预测失败时，处理器浪费了大量的指令槽来执行无用的指令。浪费的指令槽的数目与处理器的发射宽度有关，处理器发射宽度越高，浪费的指令槽的数目也就越多。其次，分支预测失败以后，处理器必须撤销预测执行对处理器的影响，处理器必须让流水线的排空，同时还要无效化处理器缓存，以避免其内容被更新到处理器状态。再次，当发现预测失败的时候，处理器必须恢复到正确的分支进行执行，这就需要计算正确的指令地址，以及初始化正确方向的取指过程。另外，由于分支语句大量存在于程序中，超标量处理器必须要能够在在一个周期内执行多个分支。假设指令流中包含 25% 的分支，那么一个 8 发射的超标量处理器必须具有每个周期执行至少 2 个分支的能力。一个周期执行多个分支增加了流水线的复杂度。在一个高发射率的处理器中，重复算术功能单元要比预测并执行多个分支要容易得多。

除了分支预测失败造成的处罚以外，分支语句会将程序分割为若干小的基本块，指令调度无法在基本块之间进行，从而影响性能。

1.2 谓词执行对性能的改善

为了解决造成的种种问题，现代处理器提供谓词执行的机制。谓词执行指的是指令的条件执行，它给指令增加一个布尔操作数，这个布尔操作数称为谓词。如果谓词的值为真，那么这个指令被正常地执行，而如果这个谓词的值为假，则这个指令就被当成空指令来执行。

有了处理器的支持，编译器就可以通过 if-conversion 来将分支转化为谓词执行。通过分支的移除，可以消除分支预测失败的惩罚，也可以避免让宽发射处理器在一个周期内处理多条分支从而提高了性能。

谓词执行的支持已经被非常广泛地应用在数值型程序以及非数值型程序的调度中 [15]。对于数值型的代码，使用软件流水调度对多个循环迭代进行重叠执行可以在超标量以及 VLIW 机器上获得高性能 [14] [8]。分支的移除使得软件流水线能够进行更紧凑的调度以及更简短的代码展开。对于非数值型的程序，谓词执行使得决定树调度在深流水以及多指令发射的处理器上获得更高的性

能 [6]。谓词执行允许多个控制路径上指令的并发执行，并且使指令在其依赖的分支被解析之前的执行成为可能。

1.3 分支消除对性能的影响

虽然适当地应用 if-conversion 可以对性能进行提升，但是过多地进行 if-conversion 也会对性能造成不利的影响，并且不多地进行 if-conversion 也不一定改善性能。[15] [18]。

首先，if-conversion 会把一个区域内的所有的执行路径合并成一条较长的路径，所以，每次基本块被执行的时候，所有指令都必须被执行。如果不同的路径具有相近的频率跟大小，这个时候代码的执行是高效的。但是不同路径的大小以及频率往往是不同的，这就导致程序每次都要浪费大量的时间用来执行不频繁的分支或者长的分支，这对性能是不利的影响。同时，一些分支中的函数调用或者未解析的内存访问也会限制转换后的大基本块的优化以及调度。

其次，预测执行跟谓词执行并不能很好地符合到一起。预测执行指的是在指令真正需要被执行之前将指令执行，对于谓词指令来说，预测执行表示的是在其谓词计算之前对指令的执行。预测执行是超标量以及 VLIW 处理器中非常重要的指令级并行的来源，因为它能使得长延迟指令在调度中非常早地被初始化。

再次，由于多个基本块被合并成了一个大的基本块，他们将共享处理器资源，这就造成处理器资源的争用，这会增大分配的寄存器的数量。如果寄存器的分配对谓词不敏感，这将会对寄存器的分配产生压力 [12] [13]。

2 if-conversion 的发展简史

if-conversion 最早由 Rice 大学的 Allen 等人在 1983 年提出 [1]。当时已有向量机的出现，并且新一代的 Fortran 也一定会提供向量操作来支持向量机。不幸的是，由于之前版本的 Fortran 中不支持向量操作，所以相当大的一部分旧有的 Fortran 程序无法充分利用处理器的新特性，这就意味着，自动向量化的研究势在必行。当时 Rice 大学的人正在开发一个翻译程序，称为并行 Fortran 转换器 (Parallel Fortran Converter, 简称 PFC)，它可将 Fortran 66 或者 Fortran 77 代码翻译成新一代 Fortran 的代码，并且通过自动向量化来充分利用向量操作。当时的编译器进行程序变换还是基于数据依赖，分支语句产生的控制依赖不能非常好的融入到当时的体系中去。与此同时，分支语句的存在也会使得自动向量化无法进行。由于这两个原因，Allen 就提出了将程序中的所有分支语句移除而转换成谓词执行的一些算法。Allen 的算法的目的比较看重向量化，这算法是直接源代码上进行操作的，它可以将所有的前向分支跟出口分支消除。但是 Allen 的算法没有考虑到如何最小化谓词的分配等问题。尽管 Allen 的算法可以将分支转换为谓词执行，但是编译器要想判断是否可以进行向量化，必须先对代码进行 if-conversion，这样就将所有的控制依赖转化

为数据依赖，然后才能使用基于数据依赖的算法判断是否能够进行向量化，一旦无法向量化，不单单是 if-conversion 执行的时间被浪费掉，而且分支的消除反而会对代码的执行效率产生影响，糟糕的是，if-conversion 已经将代码变得面目全非无法还原。

1987 年，Ferrante 等人提出了程序依赖图 (PDG) 的概念 [4]。这是一个非常重要的工作，因为 PDG 是一个新的框架，在这个框架下，数据依赖跟控制依赖被统一地进行处理，一切旧有的算法都可以高效地纳入新的框架下。由于不再存在之前的编译器无法处理控制依赖的问题，也就没有必要再利用 if-conversion 将控制依赖转化为数据依赖了。同时这个框架对 if-conversion 的发展也起到了承上启下的作用，后来对 if-conversion 的研究都是在这个框架下的。

虽然不再需要用 if-conversion 将控制依赖转化为数据依赖，但是 if-conversion 并没还有因此失去它的意义。正如上一节所说，程序中的分支语句已经称为高性能体系结构中性能提升的主要障碍，而 if-conversion 恰好可以用谓词执行来代替分支，进而提高效率。只不过问题的核心已经不是像 Allen 所说的那样，将控制依赖转化为数据依赖，也不是向量化，而是减少由于分支语句的存在对程序性能造成的影响，这也对经过 if-conversion 转换之后的代码的性能提出了要求。

1991 年，Park 跟 Schlansker 提出了 RK 算法 [7]，这个算法用于消除所有最内层循环的循环体中的分支语句，这个算法可以最小化谓词的数量，以及定义谓词的操作的数量。但问题是，谓词以及定义谓词的操作的数量最小并不意味着算法最优，之前也分析过（见第 1.3 节），分支消除也会对性能造成一定的负面影响，所以，到底是否进行分支消除，如果是，应该对哪部分分支进行消除也是需要研究的问题。

1992 年，Mahlke 等人提出了 Hyperblock 的概念 [16]。Hyperblock 是一个单入口多出口的谓词化的基本块的集合。这篇文章为选择性地 if-conversion 提供了一个框架，它告诉人们如何只对部分基本块进行 if-conversion 而保持一些其他的基本块不变：将基本块按照性质不同划分到不同的 Hyperblock 中去，然后对每个 Hyperblock 进行 if-conversion 以消除其中的分支。通过将基本块有选择地包括在 Hyperblock 之内或者排除在 Hyperblock 之外，就可以实现性能的最大化。这篇文章还提供了 Hyperblock 形成以后的一些优化算法。但是，这篇文章对 Hyperblock 选择的讨论比较少，并没有告诉人们如何划分 Hyperblock 才是最优的，自然，这也成为之后学者研究的内容。

至此，Hyperblock 概念已经提供了选择性地 if-conversion 的机制，并且 RK 算法已经告诉我们如何在选定 Hyperblock 以后产生最优代码。可以说，关于 if-conversion 的研究，剩下的只是策略的问题了。

1993 年，Warter 等人提出了逆向 if-conversion (reverse if-conversion) 的概念 [20]。在这算是一个小插曲，因为作者的工作重心并不是 if-conversion 本身，而只是提出可以利用 if-conversion 来进行全局指令调度：首先通过 if-conversion 将分支消除，然后进行局部调度，最后再用作者提出来的逆向 if-conversion 算

法将代码还原。单看工作本身，对 if-conversion 并没有大的贡献，而只是一个小小的应用，本来是不足以写进 if-conversion 的发展史的，但是由于逆向 if-conversion 的提出对后面的研究有着很大的启发作用，因此这篇文章还是对 if-conversion 有着深远影响的。

1997 年，August 等人对 if-conversion 策略进行了详细地讨论 [2]。August 详细讨论了处理器对谓词执行的支持给编译器带来的挑战，Hyperblock 选择等等问题，并且提出了一个平衡控制流跟谓词执行的框架：首先在编译过程的早期大量应用 if-conversion 来发掘谓词执行带来的全部好处，此时形成的 Hyperblock 比目标体系能处理的大得多，然后在比较靠后的编译阶段进行部分逆向 if-conversion，根据目标机器调整每个 Hyperblock 的谓词化代码的数量，以平衡控制流跟谓词数目。

有了机制，又有了策略，可以说，此时 if-conversion 的研究最重要的工作已经完成，if-conversion 像是一个完成的大厦，剩下的就是修修补补了。从 1983 年 Allen 的开山之作，到 1997 年 August 为大厦添上最后一块砖，整个过程历时 14 年，5 篇意义重大的文章。

3 if-conversion 的运行机制

控制依赖的概念最早由 Ferrante 等人提出 [4]，RK 算法由 Park 跟 Schlansker 提出 [7]，Hyperblock 的概念由 Mahlke 等人给出 [16]，逆向 if-conversion 则是 Warter 等人的研究成果 [20]。

3.1 控制依赖

定义 1 (控制流图 (control flow graph)) 控制流图是一个有向图，它有唯一的入口节点 $START$ 以及唯一的出口节点 $STOP$ ，图中每一个节点最多有两个后继。对于那些有两个后继的节点，它的两条出边分别被赋予属性 T (真) 跟 F (假)。对于图中的每个节点 N ，都存在从 $START$ 到 N ，以及从 N 到 $STOP$ 的路径。

定义 2 (后控制 (post-dominate)) 设 V 与 W 是控制流图 G 中的节点，如果从 V 到 $STOP$ 的每条路径 (起始位置的 V 并不计算在路径中) 都包含 W ，则称 V 被 W 后控制，或者 W 后控制 V ，记为 $W \text{ pdom } V$ 。

定义 3 (直接后控制 (immediate post-dominate)) 设 X, Y 以及 Z 是控制流图 G 中的节点，如果 $Y \text{ pdom } X$ ，并且对于任意的满足 $Z \text{ pdom } X$ 以及 $Z \neq Y$ 的节点 Z ，都有 $Z \text{ pdom } Y$ ，则称 Y 直接后控制 X ，或者 X 被 Y 直接后控制，记为 $Y \text{ ipdom } X$ 。

定义 4 (pdom 函数) 设 N 是节点的集合， pdom 是一个映射，它将 N 中的节点 x 映射为所有后控制 x 的节点的集合，即 $\text{pdom} : N \rightarrow 2^N$ 。 pdom 的定义的数学表述为： $\text{pdom}(x) := \{y \in N : y \text{ pdom } x\}$

定义 5 (ipdom 函数) 设 N 是节点的集合, $ipdom$ 是一个映射, 它将 N 中的节点 x 映射为 x 直接后控制的节点的节点, 即 $pdom : N \rightarrow N$ 。 $ipdom$ 的定义的数学表述为: $ipdom(x) := y \in N$, 其中 $y ipdom x$

定义 6 (后控制节点树) 直接后控制关系构成一个树, 称为后控制节点树。树的节点为控制流图 G 中的节点, 若 $y ipdom x$, 则 y 是 x 的双亲节点。显然, 求 $ipdom(x)$ 即为求 x 在树中的双亲节点, 求 $pdom(x)$ 即为求 x 在树中的所有祖先节点的集合。

计算图 G 的后控制节点树的算法见附录 B

定义 7 (控制依赖 (control dependent)) 设 G 为控制流图, X 以及 Y 是图中节点, 称 Y 控制依赖于 X 当且仅当:

1. 存在从 X 到 Y 的路径 P , 使得 Y 后控制 P 上除了 X 与 Y 的所有节点。
2. X 不被 Y 后控制

从定义可以看出, 如果 Y 控制依赖于 X , 则那么 X 必然有两个出口, 其中一条出口导致 Y 必然被执行, 另一条出口则导致 Y 可能不被执行。

定义 8 (CD 函数) 设 N 是节点的集合, C 是控制依赖的集合。 CD 函数是一个映射, 它将 N 中的节点 x 映射为 x 的所有控制依赖的集合, 即 $CD : N \rightarrow 2^C$ 。若将 C 中的元素 $c \in C$ 表示为 $\pm y$, 其中 $+y$ 表示 y 的 *true* 边, $-y$ 表示 y 的 *false* 边, 则 CD 函数的定义可以为: $CD(x) := \{\pm y : x \text{ 控制依赖于 } \pm y\}$

CD 函数的计算算法如算法 1

算法 1: ComputeCD

Input: G 是控制流图, N 是 G 中节点的集合, E 是 G 中边的集合

Output: CD 函数

```

1  计算函数  $pdom(x)$  以及  $ipdom(x)$ ;
2  for 每个  $[x, y, label] \in E$  并且  $y \notin pdom(x)$  do
3       $Lub = ipdom(x)$ ;
4      if  $\neg label$  then
5           $x = -x$ ;
6      end
7       $t = y$ ;
8      while  $t \neq Lub$  do
9           $CD(t) = CD(t) \cup \{x\}$ ;
10          $t = ipdom(t)$ ;
11     end
12 end

```

3.2 RK 算法

CD 函数并非是一个单值函数, 不同的 x 可能对应相同的 $CD(x)$ 。由于具有相同控制依赖的不同节点的执行条件相同, 所以将节点转化为谓词执行以后,

他们的谓词也应该相同，这样就可以为 CD 函数的值域中的每个元素分配一个与之一一对应的谓词 p ，谓词跟值域中元素的集合的对应关系记为 K 函数。显然， K 函数是个单值函数，它的逆 K^{-1} 必然存在，记函数 $R = CD \cdot K^{-1}$ 。可以看出，这么做将 CD 函数分解为 R 函数跟 K 函数： $CD = R \cdot K$ 。其中 R 函数负责将每个节点 x 与谓词 p 对应起来，而 K 则将每个谓词跟其对应的控制依赖对应起来。通过 R 函数，具有相同控制依赖（因而具有相同执行条件）的不同节点被映射到同一个谓词，也就是说， R 函数为每个节点实现的谓词分配，节点 x 对应的谓词 $p = R(x)$ 。至于 K 函数，则将每个谓词跟其控制依赖对应起来，这也为谓词的分配提供了参考。一种不错的谓词分配策略是，对于每一个谓词，在其对应的依赖，的相应节点末尾插入相应的赋值语句。例如 $K(p) = \{y, -z\}$ ，则在 y 与 z 节点插入 p 的定义语句。插入的时候，如果对该点依赖值为正，则谓词的值就是分支条件，如果为负，作为谓词的值就是该点分支条件的非。例如上面例子中，在 y 点插入 $p = t_y$ ，而在 z 点插入 $p = \neg t_z$ 。 R 函数以及 K 函数的计算算法如算法 2 所示。

算法 2: ComputeRK

Input: CD 是控制依赖函数， N 是 G 中节点的集合

Output: R 函数以及 K 函数

```

1  $p = 1$ ;                                     /* 谓词命名从 1 开始顺序命名 */
2 for  $x \in N$  do
3    $t = CD(x)$ ;
4   if  $t \in K$  then
5      $R(x) = q$ , 其中  $q$  为使得  $K(q) = t$  的谓词;
6   else
7      $K(p) = t$ ;
8      $R(x) = p++$ ;
9   end
10 end
```

计算好 R 函数以及 K 函数，并按照 R 函数跟 K 函数进行谓词分配与定义以后，只要对控制流图进行拓扑排序即可得到转换后的代码。if-conversion 的

算法如算法 3。

算法 3: If-Conversion

Input: G 是控制流图, N 是 G 中节点的集合, E 是 G 中边的集合

Output: 转换后的代码

```
1 计算 CD 函数;
2 计算 RK 函数;
3 for  $x \in N$  do
4    $p = R(x)$ ;
5   if  $K(p) \neq \emptyset$  then
6     给  $x$  分配谓词  $p$ ;
7   end
8 end
9 for 每个谓词  $p$  do
10  for  $\pm y \in K(p)$  do
11    若是  $+y$  则在  $y$  尾部插入  $p = t_y$ ;
12    若是  $-y$  则在  $y$  尾部插入  $p = \neg t_y$ ;
13  end
14 end
15 对  $G$  进行拓扑排序, 删掉分支语句, 并对已分配谓词的基本块进行谓词化;
```

3.3 Hyperblock

定义 9 (Hyperblock) *Hyperblock* 是一个谓词化的基本块的集合, 这个集合只有一个入口, 但是可以有多个出口。

哪些基本块被包含在 Hyperblock 中, 哪些不被包含在 Hyperblock 中, 这个的选择比较自由, 可以根据需要选择, 比如说把不经常被执行的基本块, 或者太过长的基本块排除在外, 而把经常执行又短小精悍的基本块包含在内。自由选择的基本块不一定满足 Hyperblock 的定义, 比如图 1 以及图 2 的 (a), 这时候就需要对控制流图进行一定的处理, 这样才能使选择的基本块能够构成一个 Hyperblock。

对不满足定义的基本块的集合处理的一种变换是尾复制 (Tail Duplication), 另一种是循环剥离 (loop peeling)。尾复制过程如图 1 所示, 图中 (a) 是分支选择的结果, (b) 是尾复制以后的结果, (c) 是 if-conversion 之后的结果。

循环剥离的过程如图 2 所示, 图中 (a) 是分支选择的结果, (b) 是循环剥离以后的结果。

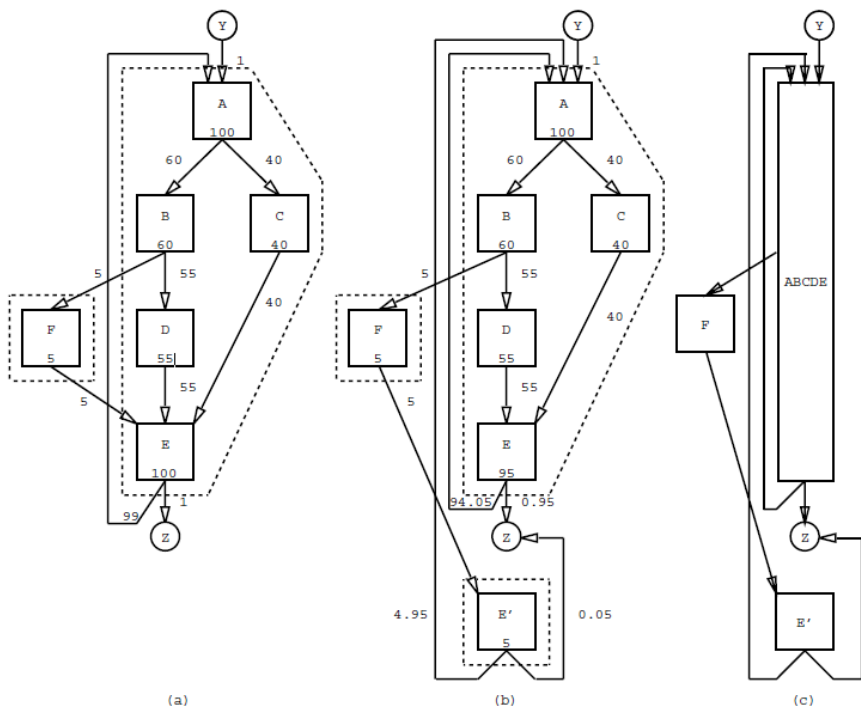


图 1: 尾复制

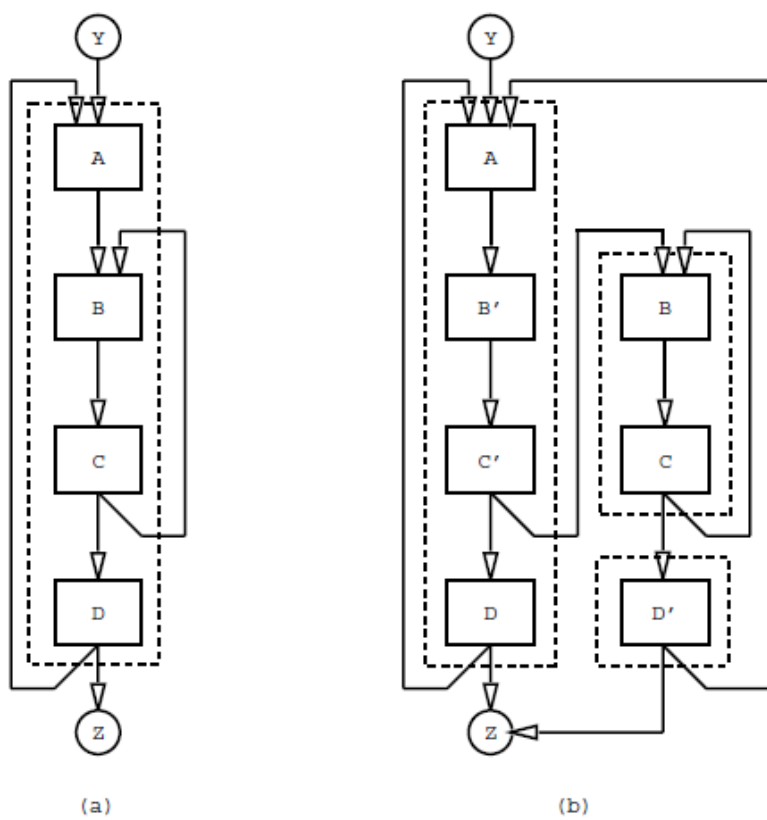


图 2: 循环剥离

3.4 逆向 if-conversion

逆向 if-conversion 能够将谓词化的语句重新转换成分支语句，其算法如下：

Algorithm RIC: Given hyperblock H , regenerate a correct control flow graph. For VLIW processors, insert_no-op fills each empty operation slot in an instruction with a no-op operation.

```

Let  $\forall X \in H$ 
 $P(X)$  = predicate that  $X$  is control dependent upon

create root node
 $L = \{\text{root}\}$ 
 $\rho(\text{root}) = \{\text{p0}\}$ 
for  $\text{op} \in H$  in scheduled order
  for  $X \in L$ 
    if  $\text{op}$  is predicate define operation
      if  $P(\text{op}) \in \rho(X)$ 
        insert conditional branch operation in  $X$ 
        create successor nodes  $\text{Succ}_t$  and  $\text{Succ}_f$ 
         $\rho(\text{Succ}_t) = \rho(X) \cup \text{true}(\text{op})$ 
         $\rho(\text{Succ}_f) = \rho(X) \cup \text{false}(\text{op})$ 
         $L = (L - X) \cup \text{Succ}_t \cup \text{Succ}_f$ 
    if  $\text{op}$  is predicate merge
      for  $p \in (\text{jump}(\text{op}) \cup \text{no\_jump}(\text{op}))$ 
         $\rho_{\text{new}}(X) = \rho(X) - \text{no\_jump}(\text{op}) - \text{jump}(\text{op})$ 
        if  $p \in \rho(X)$ 
          if  $p \in \text{jump}(\text{op})$ 
            insert jump operation in  $X$ 
          for  $Y \in L$ 
            if  $\rho_{\text{new}}(X) \equiv \rho(Y)$ 
              if  $p \in \text{jump}(\text{op})$ 
                 $\text{Succ}_t = Y$ 
              else
                 $\text{Succ}_f = Y$ 
             $L = L - X$ 
          if successor not found
            if  $p \in \text{jump}(\text{op})$ 
              create successor node  $\text{Succ}_t$ 
               $\rho(\text{Succ}_t) = \rho_{\text{new}}(X)$ 
               $L = (L - X) \cup \text{Succ}_t$ 
            else
              create successor node  $\text{Succ}_f$ 
               $\rho(\text{Succ}_f) = \rho_{\text{new}}(X)$ 
               $L = (L - X) \cup \text{Succ}_f$ 
    if  $\text{op}$  is predicated operation
      if  $P(\text{op}) \in \rho(X)$ 
        insert  $\text{op}$  in  $X$ 
  if target processor is VLIW and last op of instruction
    insert_no-op

```

4 if-conversion 的执行策略

4.1 if-conversion 给编译器带来的挑战

1997 年, August 在文章中详细讨论了处理器对谓词执行的支持给编译器带来的挑战, 以及 if-conversion 的研究需要解决的问题 [2]。现将 August 的分析总结如下:

4.2 August 的基于逆向 if-conversion 的平衡算法

August 建议, 在编译过程的早期大量应用 if-conversion 来发掘谓词执行带来的全部好处, 此时形成的 Hyperblock 比目标体系能处理的大得多, 然后在比较靠后的编译阶段进行部分逆向 if-conversion, 根据目标机器调整每个 Hyperblock 的谓词化代码的数量, 以平衡控制流跟谓词数目。整个过程的工作流程如图 3 所示。

5 LLVM 上 if-conversion 的实现

附录

附录 A Allen 的工作

if-conversion 最早由 Rice 大学的 Allen 等人于 1983 年提出 [1], 但是由于 Allen 的工作与后来的研究关联性不大, 所以正文中对 Allen 的工作并没有叙述, 现将 Allen 的工作在此附录中介绍一下。Allen 的算法直接工作在 Fortran 源代码上, 它将代码中的所有分支移除, 进而为接下来的向量化做铺垫。

Allen 将 Fortran 中的语句分为如下几种类型:

定义 10 (动作语句 (action statements)) 动作语句指的是能导致计算状态改变或者产生重要的副作用的语句, 比如: 赋值、读、写、函数调用。

定义 11 (分支语句 (branch statements)) 分支语句指的是能够显式的引起控制转移的语句, 比如: *goto* 语句。注意函数调用不是分支语句而是动作语句, 因为在给定的模块中, 函数调用可以看成是一个宏操作。

定义 12 (迭代语句 (iterative statements)) 迭代语句指的是能造成其他语句循环执行的语句, 例如: *do* 语句。

定义 13 (占位符语句 (placeholder statements)) 占位符语句指的是没有任何动作, 只是用来占位的语句, 例如: Fortran 中的 *continue* 语句。

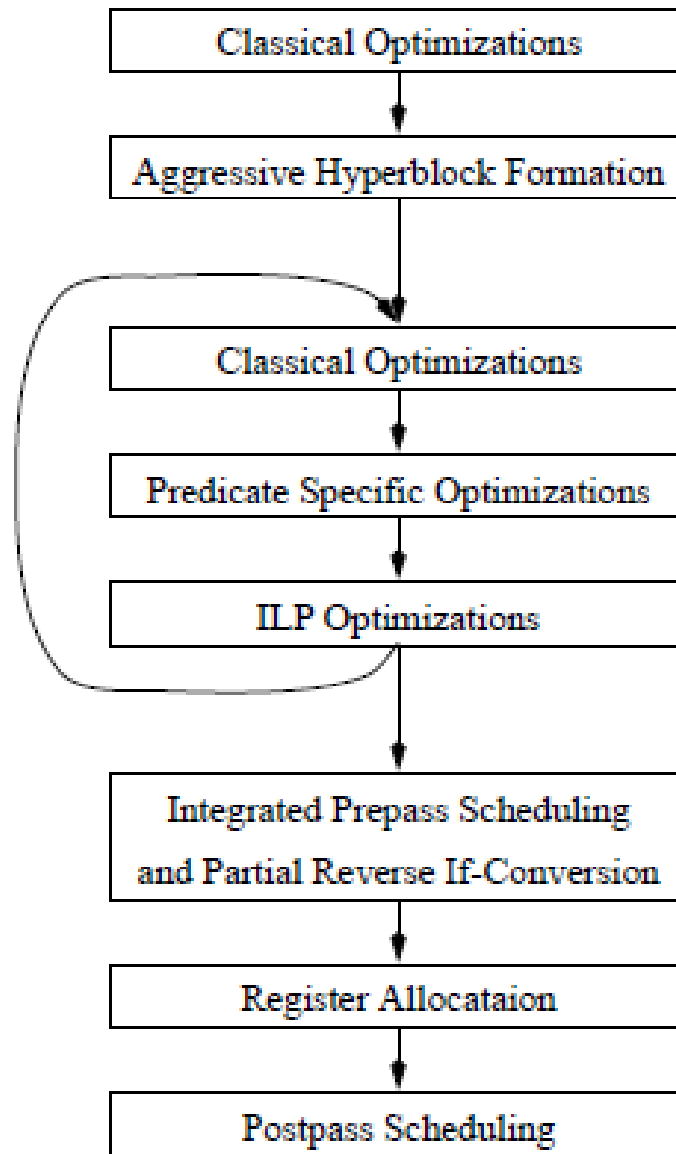


图 3: if-conversion 的执行策略

if-conversion 的目标是消除所有的分支语句，分支的消除会导致一些语句被替代为条件执行语句，这些谓词执行语句的条件被称为谓词 (guard):

定义 14 (谓词 (guard))：条件动作语句的谓词 (*guard*) 是一个布尔表达式，这个表达式表示这条语句执行的条件，当且仅当布尔表达式为真时，这条语句的动作才会执行。

Fortran 中条件动作语句形式如下：

```
1 IF(guard) statement
```

Allen 将分支分为如下 3 种类型：

定义 15 (出口分支 (exit branch)) 出口分支指的是能终止一个或者多个循环的分支，例如：

```
1 DO 100 I=1,100
2     IF (ABS(A(I)-B(I)).LE. DEL) GOTO 200
3 100 CONTINUE
4 ...
5 200 CONTINUE
```

定义 16 (前向分支 (forward branch)) 前向分支指的是跳转目标在当前语句之后但在相同的循环中的语句，例如：

```
1 DO 100 I = 1, 10
2     IF (A(I).EQ.0.0) GOTO 100
3     B(I) = B(I)/A(I)
4 100 CONTINUE
```

定义 17 (后向分支 (backward branch)) 后向分支指的是跳转目标在当前语句之前但在相同的循环中的语句，例如：

```
1 10 I = I + 1
2 A(I) = A(I) + B(I)
3 IF (I .LE. 100) GOTO 10
```

if-conversion 使用如下两种变换来消除程序中的分支：

1. 分支重定位 (**branch relocation**)：将分支语句一层层移出循环，直至分支语句与其目标在嵌套在相同的循环中。该过程把所有的出口分支转换为前向或者后向分支。
2. 分支移除 (**branch removal**)：将所有的前向分支移除，取而代之的是，计算每个语句的谓词，并将动作语句用条件动作语句来代替。

在分支重定位算法中，Allen 引入了出口标记 (**exit flag**) 的概念：

定义 18 (出口标记 (exit flag)) 出口标记是一个布尔变量，每一个出口分支都有与其对应的一个出口标记。在程序执行的过程中，该变量在其对应分支被执行之前的值始终为真，而在对应分支被执行以后其值变为假。在这里，出口分支 i 的出口标记记为 ex_i ，与其对应的 Fortran 变量名记为 EXi

在上面的定义中出现了两个容易混淆的概念，出口标记 ex_i 及其对应的 Fortran 变量 EXi 。这两个概念的区别是，小写字母 ex_i 是算法以及行文中出现的变量，为了更容易地写出算法，也为了行文方便，在对小写字母的变量进行操作的时候并不会考虑布尔表达式的求值效率。但是一旦要进行程序生成，布尔表达式的效率就变得尤为重要，因为复杂的布尔表达式会使生成的程序运行缓慢，于是对小写字母变量进行的操作，在真正被转化为代码的时候，必须进行布尔化简以后才能写进程序。为了区别，经过布尔化简以后的版本中的变量（也就是最终写进程序中的变量），用大写字母 EXi 表示。类似的表示方法在后文中还会经常出现。

在只有一个循环、一个出口分支的情况下，例如：

```

1      DO 100 I = 1,100
2          S1
3          IF (X(I)) GOTO 200
4          S2
5 100  CONTINUE
6      S3
7 200  S4

```

程序中唯一的一个出口分支对应的出口标记为 ex_1 。由出口标记的定义不难看出，在循环开始的时候，该变量的值应该为真，因为此时分支语句显然没有执行。此后运行过程中，该变量的值一直保持为真，直到分支被成功执行，即 $X(I)$ 为假时，该变量才为假。由于出口分支被消除，而循环中的每个动作语句将会被替换为条件动作语句。显然，语句的控制条件就是 ex_1 。

不难看出，重定位之后的代码应该为：

```

1      EX1 = .TRUE.
2      DO 100 I = 1,100
3          IF (EX1) S1
4          IF (EX1) EX1 = .NOT. X(I)
5          IF (EX1) S2
6 100  CONTINUE
7      IF (.NOT. EX1) GOTO 200
8      S3
9 200  S4

```

不难理解，在一个循环、 n 个出口分支的情况下，每个语句（包括所有新引入的对哥哥出口标记赋值的语句）的控制条件为： $ex_1 \wedge ex_2 \wedge \dots \wedge ex_n$

这个方法也可以非常容易的拓展到多个循环、多个出口分支的情况：

```

1      DO 200 I = 1,100
2 50    S1

```

```

3          DO 100 J=1,100
4              S2
5              IF X(I,J) GOTO 300
6              S3
7              IF Y(I,J) GOTO 50
8              S4
9 100      CONTINUE
10         S5
11 200 CONTINUE
12 300 S6

```

经过分支重定向以后变为：

```

1          EX1 = .TRUE.
2          DO 200 I =1,100
3 50        IF (EX1) S1
4          IF (EX1) EX2 = .TRUE.
5          DO 100 J=1,100
6              IF (EX1 .AND. EX2) S2
7              IF (EX1 .AND. EX2) EX1 = .NOT. X(I,J)
8              IF (EX1 .AND. EX2) S3
9              IF (EX1 .AND. EX2) EX2 = .NOT. Y(I,J)
10             IF (EX1 .AND. EX2) S4
11 100      CONTINUE
12          IF (EX1 .AND. .NOT. EX2) GOTO 50
13          IF (EX1) S5
14 200 CONTINUE
15          IF (.NOT. EX1) GOTO 300
16 300 S6

```

该算法的伪代码如算法 4 所示。

算法 4: relocate_branches(x)

Input: x 是一个 DO 循环

Output: 不含出口分支的版本

/* loop_guard 表示所有出口标记的合取 */

```

1 loop_guard = true;
2 for 跳出循环 x 的每个出口分支 IF(P) GOTO S1 do
3   | 创建一个新的出口标记  $ex_i$ ，其对应的 FORTRAN 变量为 EXi;
4   | 在 x 前插入语句 "EXi = .TRUE.";
5   | loop_guard = loop_guard  $\wedge$   $ex_i$ ;
6   | 在循环后插入分支语句 "IF (.NOT. EXi) GOTO S1" ;
7   | 将出口分支替换为赋值语句 "EXi = .NOT. P" ;
8 end
9 for 循环 x 中的每个 DO 循环 y do
10  | relocate_branches(y);
11 end
12 for 循环 x 中的每个非循环语句 y do
13  | guard(y) = guard(y)  $\wedge$  loop_guard;
14 end

```

前向分支是最简单的分支，它的移除称为前向分支移除 (forward branch removal)。为了进行前向分支移除，引入当前条件 (current condition) 概念：

定义 19 (当前条件 (current condition)) 当前条件是一个逻辑表达式，它表示现在正在处理的表达式被执行的条件。

前向分支在两个地方影响控制流：

1. 分支处：在没有其他控制流改变的情况下，前向分支之后紧跟的语句只有在分支没有发生的条件下才会被执行。所以，如果分支处的当前条件为 cc_1 ，分支的条件 p ，那么紧跟分支之后的语句的控制条件为 $cc_1 \wedge \neg p$
2. 目标处：控制流到达分支目标处有两种情况，一种是分支语句没有执行，控制流直接顺序执行到达目标处，另一种则是通过分支语句直接到达目标处。如果目标前面的那条语句的控制条件是 cc_2 ，则目标处的控制条件应该为 $cc_2 \vee (cc_1 \wedge p)$

引入分支标记 (branch flags) 的概念：

定义 20 (分支标记 (branch flags)) 每个前向分支 (编号为 i) 对应于一个被称为分支标记 (branch flags) 的变量 br_i ， br_i 为真当且仅当对应分支语句的分支条件为真。

例如代码：

1	DO 100 I = 1, 100
---	-------------------

```

2      IF (A(I).GT.10) GOTO 60
3      A(I) = A(I) + 10
4      IF (B(I).GT.10) GOTO 80
5      B(I) = B(I) + 10
6 60    A(I) = B(I) + A(I)
7 80    B(I) = A(I) - 5
8 100 CONTINUE

```

为两个分支分别引入一个布尔变量 br_1 跟 br_2 ，他们对应的 Fortran 变量分别为 BR1 和 BR2。为了计算 BR1 和 BR2，我们分别在两个分支处插入代码：

```

1  BR1 = A(I).GT.10

```

以及

```

1  BR2 = B(I).GT.10

```

由之前的描述，不难发现示例中的各个动作语句的控制条件条件为：

行号	控制条件
3	$\neg br_1$
5	$\neg br_1 \wedge \neg br_2$
6	$br_1 \vee (\neg br_1 \wedge \neg br_2)$
7	$br_1 \vee (\neg br_1 \wedge br_2) \vee (\neg br_1 \wedge \neg br_2)$

为了防止控制条件的不断增长，编译器必须具备布尔化简的能力，例如第七行对应的布尔表达式恒真。我们用 $\mu(p)$ 表示逻辑表达式 p 经过布尔化简以后转换成逻辑表达式的 Fortran 语句表示。经过前向分支消除以及布尔优化后的结果为：

```

1      DO 100 I = 1, 100
2          BR1 = A(I).GT.10
3          IF (.NOT. BR1) A(I) = A(I) + 10
4          IF (.NOT. BR1) BR2 = B(I).GT.10
5          IF (.NOT. BR1 .AND. .NOT. BR2) B(I) = B(I) + 10
6          IF (BR1 .OR. .NOT. BR2) A(I) = B(I) + A(I)
7          B(I) = A(I) + 5
8 100 CONTINUE

```

前向分支消除的伪代码如5所示。

算法 5: *forward_convert*(x, cc_0)

Input: x 是要考虑的分支, cc_0 是 x 前的控制条件

Output: 返回 x 的控制条件, 输出不含前向分支的版本

/ predicate_list*(x) 是一个队列, 队列包含所有的由于分支到 x 而必须在 x 处分离的分支条件 **/*

```

1  $cc_1 \leftarrow cc_0$ ;
2 while not_empty(predicate_list( $x$ )) do
3    $p \leftarrow \text{get\_from\_queue}(\text{predicate\_list}(x))$ ;
4    $cc_1 \leftarrow cc_1 \vee p$ ;
5 end
6 switch statement_type( $x$ ) do
7   case 前向分支 IF( $P$ ) GOTO  $y$ 
8     创建一个新的分支标记  $br_i$ , 其对应的 FORTRAN 变量为  $BR_i$ ;
9     把  $x$  替换为 “IF( $\mu(cc_1)$ )  $BR_i = P$ ”;
10    add_to_queue(predicate_list( $y$ ),  $cc_1 \wedge br_i$ );
11     $cc_1 \leftarrow cc_1 \wedge \neg br_i$ ;
12  end
13  case 前向分支 GOTO  $y$ 
14    add_to_queue(predicate_list( $y$ ),  $cc_1$ );
15     $cc_1 \leftarrow \text{false}$ ;
16    删除分支  $x$ ;
17  end
18  otherwise
19    设置  $x$  的控制条件为  $x$ ;
20  end
21 endsw
22 return  $cc_1$ 

```

后向分支会产生隐式循环, 因此无法被 if-conversion 消除。然而更严重的是, 后向分支导致从循环外跳转到循环内成为可能, 这就导致5产生不正确的代码。一种简单的做法是, 保持隐式循环区内的所有代码不变, 不对隐式循环区内的代码进行优化, 然而这是一个非常严重的限制, 因此必须考虑后向分支存在的情况下代码的优化。本节介绍完全分支移除算法6, 这个算法可以在后向分支存在的前提下删除所有的前向分支。

考虑下面这个例子:

1	IF (X) GOTO 200
2	...
3	100 S1
4	...
5	200 S2
6	...
7	IF (Y) GOTO 100

S1 的控制条件必须反映下面事实:

1. 仅当 X 为假的时候, S1 才会在在第一轮得到执行
2. 如果是从后向分支跳转过来的, 那么 S1 将永远得到执行

这暗示我们引入一个新的标记来表示 S1 的执行是否是由于后向分支跳转到此而执行的, 这个标记称之为向后分支标记 (**branch back flag**), 记为 bb, 其对应的 FORTRAN 变量为 BB。

这样, 例子中的程序就可以化为: 考虑下面这个例子:

```
1      BR1 = X
2      IF(.NOT. BR1) ...
3      BB1 = .FALSE.
4 100  IF(.NOT. BR1 .OR. BR1 .AND. BB1) S1
5      IF(.NOT. BR1 .OR. BR1 .AND. BB1) ...
6 200  S2
7      ...
8      IF(Y) THEN
9          BB1 = .TRUE.
10         GOTO 100
11     ENDIF
```

可以看出, 后向分支 j 的目标语句 y 被执行, 可以分为两种情况:

1. 顺序执行: y 前面的语句执行完成以后, 顺序执行到 y, 此时 y 的控制条件就是 y 前面语句的控制条件 cc_y 。
2. 后向分支: 控制流通过分支 i 进入隐式循环区, 然后通过后向分支 j 跳转至此。这种情况发生的条件是 $br_i \wedge bb_j$

于是, y 的控制条件应为 $cc_y \vee (br_i \wedge bb_j)$ 。如果有不止一个分支能够进入隐式循环区, 则将式中的 br_i 换为所有这些进入隐式循环区的分支的分支标记的析取。

考虑另外一个例子:

```
1      IF(X) GOTO 200
2 100  S1
3      GOTO 300
4 200  S2
5      IF(Y) GOTO 100
6 300  S3
```

显然, 正确的 S2 的控制条件应该是 $br_1 \wedge \neg bb_1$, 要移除 S2 前面的分支语句, $\neg bb_1$ 必须被添加到相应语句的控制条件里面去。考虑到这一点, 转换后的代码应该为:

```
1      BR1 = X
2      BB1 = .FALSE.
3 100  IF(.NOT. BR1 .OR. BR1 .AND. BB1) S1
4      /*GOTO 300 已经被移除*/
5 200  IF(.NOT. BB1 .AND. BR1) S2
6      IF(.NOT. BB1 .AND. BR1 .AND. Y) THEN
7          BB1 = .TRUE.
8          GOTO 100
```

完全分支移除的算法如算法 6 所示。

算法 6: $remove_branches(x, cc_0)$

Input: x 是要考虑的分支, cc_0 是 x 前的控制条件

Output: 返回 x 后面语句的控制条件 cc_1 , 输出不含前向分支的版本

```

1  $cc_1 \leftarrow cc_0$ ;
2 while  $not\_empty(predicate\_list(x))$  do
3    $p \leftarrow get\_from\_queue(predicate\_list(x))$ ;
4    $cc_1 \leftarrow cc_1 \vee p$ ;
5 end
6 switch  $statement\_type(x)$  do
7   case 前向分支  $IF(P) GOTO y$ 
8     创建一个新的分支标记  $br_i$ , 其对应的 FORTRAN 变量为  $BR_i$ ;
9     把  $x$  替换为 “ $IF(\mu(cc_1)) BR_i = P$ ” ;
10     $process\_branch(x, y, cc_1 \wedge br_i)$ ;
11     $cc_1 \leftarrow cc_1 \wedge \neg br_1$ ;
12  end
13  case 前向分支  $GOTO y$ 
14     $process\_branch(x, y, cc_1)$ ;
15     $cc_1 \leftarrow false$ ;
16    删除分支  $x$ ;
17  end
18  case 后向分支  $IF(P) GOTO y$ 
19    创建一个新的分支标记  $bb_j$ , 其对应的 FORTRAN 变量为  $BR_j$ ;
20    在  $y$  之前插入 “ $BB_j = .FALSE.$ ” ;
21    将  $x$  替换为代码段:
22    “ $TP_k = \mu(cc_1)$ ”
23    “ $IF(TP_k) TP_k = P$ ”
24    “ $IF(TP_k) BB_j = .TRUE.$ ”
25    “ $IF(TP_k) GOTO y$ ” ;
26  end
27  otherwise
28     $guard(x) \leftarrow guard(x) \wedge cc_1$ ;
29  end
30 endsw
31 return  $cc_1$ 

```

其中的 *process_branch* 如7所示。

算法 7: *process_branch* (x, y, br)

Input: x 是要考虑的分支, y 是分支的目标, br 是分支的条件

```

1 stmt_guard  $\leftarrow$  true;
2 for  $x$  跳入的每个隐式循环区 do
    /* 设  $bb_j$  是控制这个隐式循环区的向后分支标记,  $x_j$  是相应后向分支的目标 */
3     add_to_queue (predicate_list ( $x_j$ ),  $br \wedge bb_j$ );
4     stmt_guard  $\leftarrow$  stmt_guard  $\wedge$   $\neg bb_j$ ;
5 end
6 add_to_queue (predicate_list ( $y$ ),  $br \wedge$  stmt_guard);
```

附录 B 后控制节点树的计算

简单的后控制节点树的计算算法可以参见 Allen 的《现代体系结构优化编译器》P102。被广泛应用的是 Lengauer 和 Tarjan 研究出来的一个算法 [9], 它在最坏情况下的时间复杂度为 $O(E\alpha(N, E))$, 其中 α 是一个增长非常缓慢的函数。后来, Harel 发表了一个算法达到了线性时间界限 [5]。

参考

- [1] John R Allen, Ken Kennedy, Carrie Porterfield, and Joe Warren. Conversion of control dependence to data dependence. In *Proceedings of the 10th ACM SIGACT-SIGPLAN symposium on Principles of programming languages*, pages 177–189. ACM, 1983.
- [2] David I. August, Wen mei W. Hwu, and Scott A. Mahlke. A framework for balancing control flow and predication, 1997.
- [3] Roger A. Bringmann, Scott A. Mahlke, Richard E. Hank, John C. Gyllenhaal, and Wen mei W. Hwu. Speculative execution exception recovery using write-back suppression. In *MICRO*, pages 214–223. ACM/IEEE, 1993.
- [4] J. Ferrante, K. Ottenstein, and J. Warren. The program dependence graph and its use in optimization. *ACM Transactions on Programming Languages and Systems*, 9(3):319–349, July 1987.
- [5] D Harel. A linear algorithm for finding dominators in flow graphs and related problems. *Proceedings of the seventeenth annual ACM symposium on Theory of computing - STOC 1985*, 1985.
- [6] Peter Y.-T. Hsu and Edward S. Davidson. Highly concurrent scalar processing. In *ISCA*, pages 386–395, 1986.
- [7] Mike Schlansker Joseph C. H. Park. On predicated execution, 1991.
- [8] Ardent Computer Ttt Joseph P. Brutt. Overlapped loop support in the cydra 5.

- [9] Thomas Lengauer, Robert Endre, and Tar Jan. A fast algorithm for finding dominators in a flowgraph, 1979.
- [10] Y. Patt M. Alsup H. Scales M. Shebanow M. Butler, T. Yeh. Single instruction stream parallelism is greater than two, 1991.
- [11] Scott A. Mahlke, Richard E. Hank, Roger A. Bringmann, John C. Gyllenhaal, David M. Gallagher, and Wen mei W. Hwu. Characterizing the impact of predicated execution on branch prediction, 1994.
- [12] Eduardo Quiñones. Selective predicate prediction for out-of-order processors, 2006.
- [13] Eduardo Quiñones, Departament D' arquitectura De Computadors, Antonio González, and Joan manuel Parcerisa. Improving branch prediction and predicate execution in out-of-order processors, 2007.
- [14] B. R. Rau, D. W. L. Yen, W. Yen, and J. P. Bratt. The cydra 5 departmental supercomputer: design philosophies, decisions, and trade-offs. *Computer*, 22(1):12–35, 1989.
- [15] et al. Scott A. Mahlke. Effective compiler support for predicated execution using the hyperblock, 1992.
- [16] et al. Scott A. Mahlke. Effective compiler support for predicated execution using the hyperblock, 1992.
- [17] Michael D. Smith, Mike Johnson, and Mark A. Horowitz. Limits on multiple instruction issue, 1989.
- [18] Zuwei Tian and Guang Sun. An if-conversion algorithm based on predication execution. *Information Technology Journal*, 9(5):984–988, May 2010.
- [19] David W. Wall. Limits of instruction-level parallelism, 1991.
- [20] Nancy J. Warter, Scott A. Mahlke, Wen mei W. Hwu, and B. Ramakrishna Rau. Reverse if-conversion, 1993.