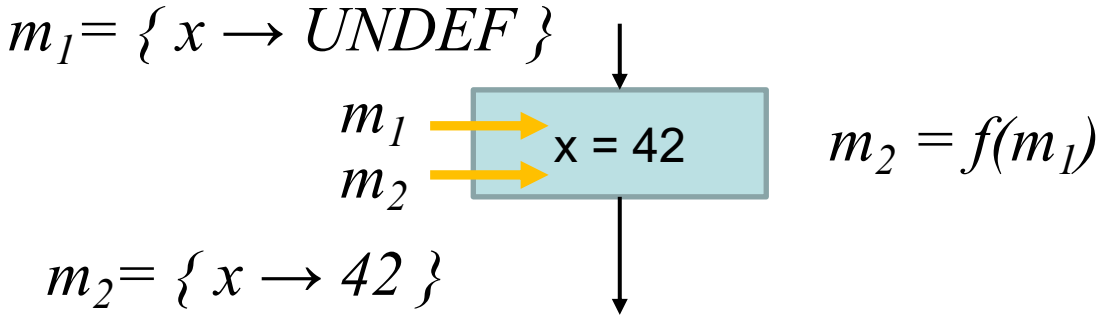
7. Распространение констант

7.1 Глобальное распространение констант

Основная идея

Пусть m_1 и m_2 – состояния программы до/после некоторой инструкции, f – передаточная функция



Состояние программы в каждой точке – отображение множества *SSA*-имен на полурешетку констант.

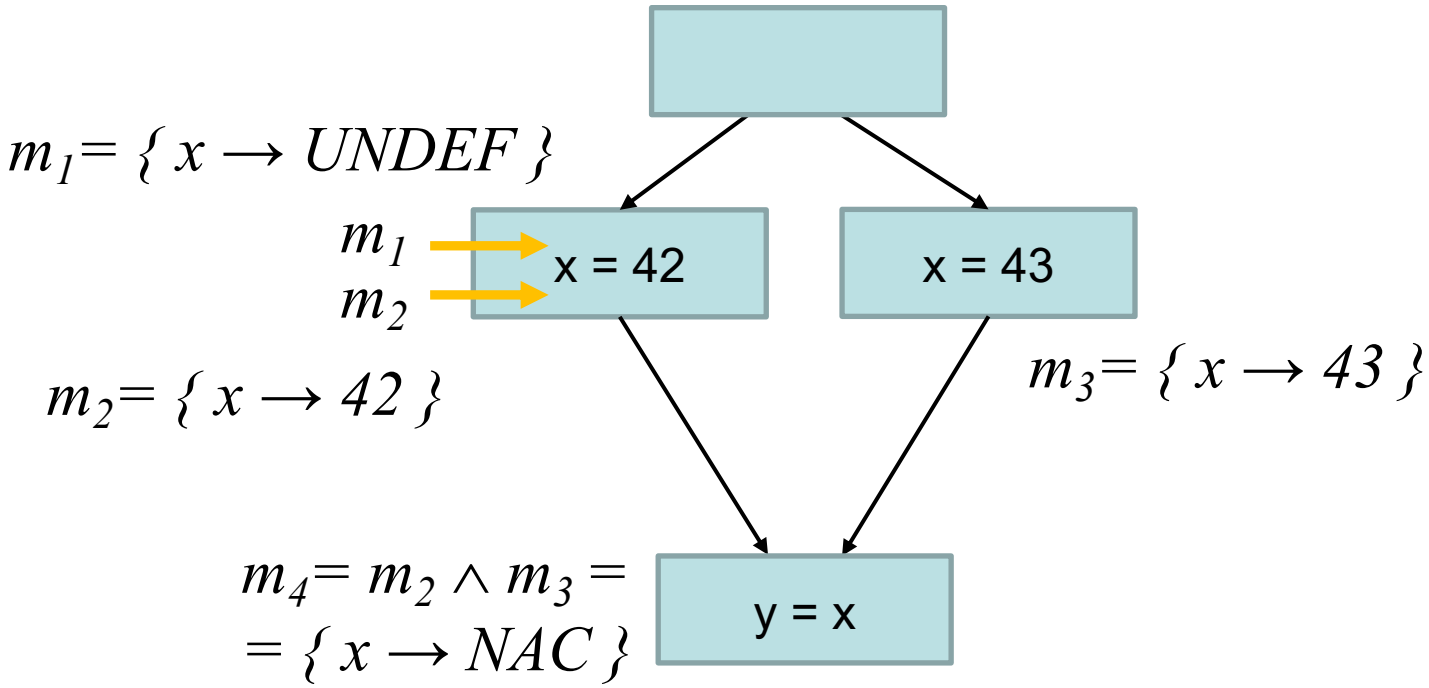
Альтернативная форма записи:

$m_2(x) = 42$

7.1 Глобальное распространение констант

Основная идея

Пусть m_1 и m_2 – состояния программы до/после некоторой инструкции, f – передаточная функция



7.1 Глобальное распространение констант

7.1.1 Постановка задачи

◊ **Распространение констант** – это оптимизирующее преобразование, заменяющее **выражения**, которые при выполнении всякий раз вычисляют одну и ту же константу, самой этой **константой**.

◊ Преобразование удобно выполнять над процедурой (программой) в SSA-форме.

7.1 Глобальное распространение констант

7.1.1 Постановка задачи

- ◊ **Распространение констант** – это оптимизирующее преобразование, заменяющее **выражения**, которые при выполнении всякий раз вычисляют одну и ту же константу, самой этой **константой**.
- ◊ Преобразование удобно выполнять над процедурой (программой) в SSA-форме.
- ◊ Распространение констант является прямой задачей потока данных.

7.1 Глобальное распространение констант

7.1.1 Постановка задачи

- **Распространение констант** – это оптимизирующее преобразование, заменяющее **выражения**, которые при выполнении всякий раз вычисляют одну и ту же константу, самой этой **константой**.
- Преобразование удобно выполнять над процедурой (программой) в SSA-форме.
- Распространение констант является прямой задачей потока данных.
- Выполняя глобальный анализ потока данных (методом итераций) рассматриваем в процедуре все SSA-имена v и каждому из них *сопоставляем константу*; это может быть:
  - специальная константа $Undef$ – значение переменной не определено;
  - специальная константа NAC ($not\text{-}a\text{-}constant$) – переменная не может быть константой;
  - одна из множества рассматриваемых констант.

7.1 Глобальное распространение констант

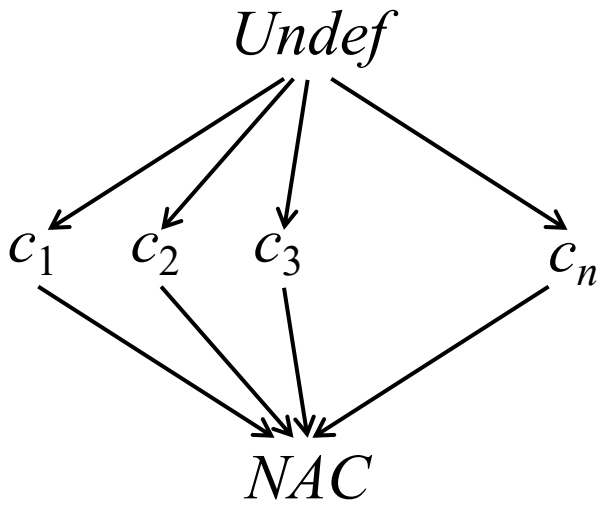
7.1.2 Константы *NAC* и *Undef*

- ***SSA*-имени *v* сопоставляется константа *NAC*** если установлено, что:
  - v может быть присвоено входное значение, т.е.
    - v – имя формального параметра процедуры или
    - v – имя целевой переменной функции ввода;
  - при вычислении v используется NAC
  - на различных путях вычисления v могут быть присвоены разные значения.
- ***SSA*-имени *v* сопоставляется константа *Undef***, если не найдено ни одного определения v, достигающего рассматриваемой точки. На самом деле всем SSA-именам сначала сопоставляется константа $Undef$.

7.1 Глобальное распространение констант

7.1.2 Полурешетка значений констант

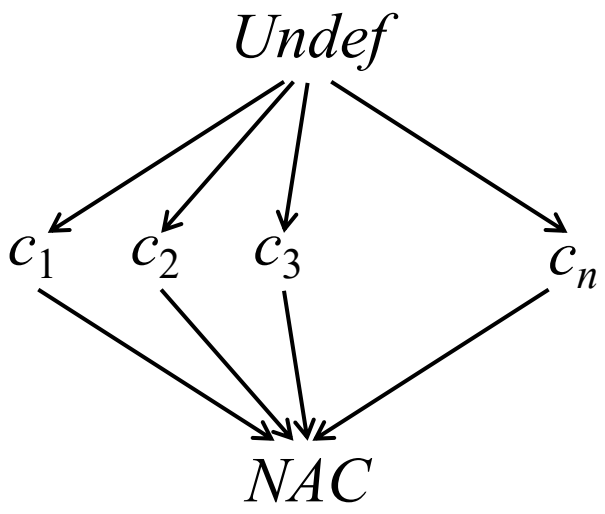
◊ Множество значений, сопоставляемых каждой переменной, (включая специальные значения $Undef$ и NAC) можно представить как полурешетку (см. рисунок).



7.1 Глобальное распространение констант

7.1.2 Полурешетка значений констант

◊ Множество значений, сопоставляемых каждой переменной, (включая специальные значения $Undef$ и NAC) можно представить как полурешетку (см. рисунок).

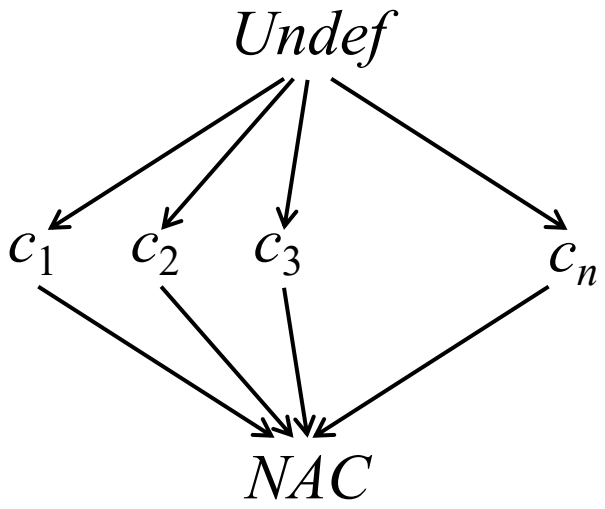


- верхний элемент $\top = Undef$
- нижний элемент $\bot = NAC$
- Значения констант (c_i) не сравнимы между собой, но все они «меньше» $Undef$ и «больше» NAC: $\forall i \;\; NAC \le c_i \le Undef$

7.1 Глобальное распространение констант

7.1.2 Полурешетка значений констант

◊ Множество значений, сопоставляемых каждой переменной, (включая специальные значения $Undef$ и NAC) можно представить как полурешетку (см. рисунок).



◈ для всех значений v:

$$Undef \wedge v = v$$

$$NAC \wedge v = NAC$$

◈ в частности

$$NAC \wedge Undef = NAC$$

◈ по определению

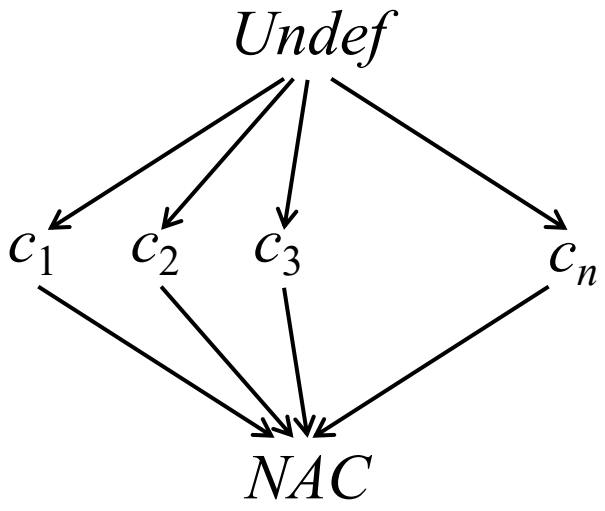
для любых двух констант c_1 и c_2

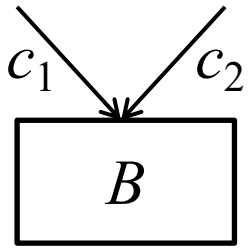
$$c_1 \wedge c_2 = \begin{cases} c_1 & \text{если } c_1 = c_2 \\ NAC & \text{если } c_1 \neq c_2 \end{cases}$$

7.1 Глобальное распространение констант

7.1.2 Полурешетка значений констант

◊ Множество значений, сопоставляемых каждой переменной, (включая специальные значения $Undef$ и NAC) можно представить как полурешетку (см. рисунок).





◈ для всех значений v:

$$Undef \wedge v = v$$

$$NAC \wedge v = NAC$$

◈ в частности

$$NAC \wedge Undef = NAC$$

◈ по определению

для любых двух констант c_1 и c_2

$$c_1 \wedge c_2 = \begin{cases} c_1 & \text{если } c_1 = c_2 \\ NAC & \text{если } c_1 \neq c_2 \end{cases}$$

7.1 Глобальное распространение констант

7.1.3 Операция сбора в структуре распространения констант

◊ Значение потока данных в структуре распространения констант представляет собой отображение

$$v: \mathcal{N} \rightarrow \mathcal{C},$$

где $\mathcal{N}$ – множество SSA-имен

$\mathcal{C}$ – полурешетка констант (для соответствующего типа переменной)

Значение, соответствующее SSA-имени n в отображении v, обозначается как $v(n)$

◊ Полурешетка значений потока данных представляет собой декартово произведение k полурешеток $\mathcal{C}$, где k – количество переменных в анализируемой процедуре:

$$v \wedge v' = v'' \quad \Leftrightarrow \quad \forall n: v''(n) = v(n) \wedge v'(n).$$

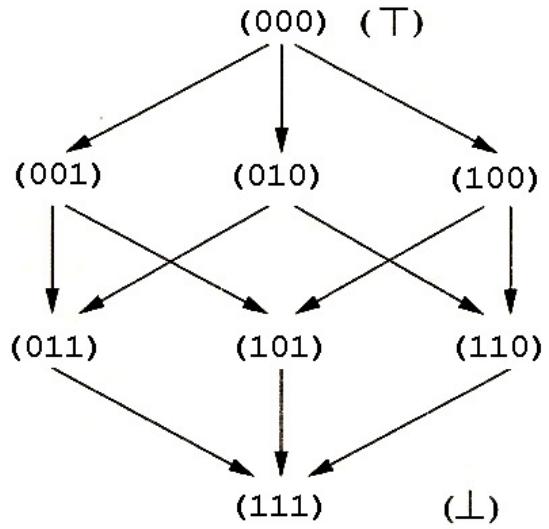
$$v \leq v' \quad \Leftrightarrow \quad \forall n: v(n) \leq v'(n)$$

(Повторение: полурешетки)

Диаграммы полурешеток

◊ *Диаграмма полурешетки* $\langle L, \wedge \rangle$ представляет собой граф, узлами которого являются элементы L, а ребра направлены от x к y, если $y \leq x$.

◊ **Пример**. На рисунке – диаграмма полурешетки $\langle U, \cup \rangle$, $|U| = 8$: элемент множества U представляется битовым 3-вектором.



(определение $\leq$ *:*
$x \leq y <=> x \wedge y = x$ *)*

7.1 Глобальное распространение констант

7.1.4 Передаточные функции структуры распространения констант

◊ Множество передаточных функций $\mathcal{F}_{Const}$ состоит из функций вида

$$f: \mathcal{V} \to \mathcal{V},$$

где $\mathcal{V}$ – множество отображений $v: \mathcal{N} \to C$.

◈ $\mathcal{F}_{Const}$ содержит тождественную функцию i:

$$\forall v \in \mathcal{V}\ i(v) = v$$

◈ $\mathcal{F}_{Const}$ содержит константную передаточную функцию c:

$$\forall v \in \mathcal{V}\ c(v) = v_0$$

где

$$\forall v \in \mathcal{V}\ v_0(v) = Undef$$

В частности, $c(Entry) = v_0$ означает, что перед началом выполнения программы не определена ни одна переменная.

7.1 Глобальное распространение констант

7.1.4 Передаточные функции структуры распространения констант

◊ Пусть $f_s \in \mathcal{F}_{Const}$ – передаточная функция для инструкции s и пусть v и $v' \in \mathcal{V}$ – значения потока данных в точках до и после s: $v' = f_s(v)$.

◊ Опишем f_s через v и v'.

(1) Если s – не содержит присваивания какому-либо SSA-имени, то

f_s – тождественная функция i, т.е. для всех v $v'(n) = v(n)$

(2) Если s – содержит присваивание SSA-имени x, то

для $n \neq x$ $v'(n) = v(n)$,

для $n = x$ $v'(n) = v'(x)$

7.1 Глобальное распространение констант

7.1.4 Передаточные функции структуры распространения констант

◊ Описание f_s через m и m'

(2) Пусть s – присваивание переменной x и пусть m и m' – значения потока данных, тогда $m' = m(x)$ определяется следующим образом:

(a) Если s – присваивание константы c: $x = c$, то $m'(x) = c$

(b) Если s – присваивание выражения: $x = y + z$, то

$$m'(x) = \begin{cases} m(y) + m(z), & \text{если } m(y) \text{ и } m(z) \text{ константы} \\ NAC, & \text{если } m(y) = NAC \text{ или } m(z) = NAC \\ Undef & \text{в остальных случаях} \end{cases}$$

(c) Если s – присваивание любого другого выражения (например, вызова функции или указателя), то $m'(x) = NAC$.

7.1 Глобальное распространение констант

7.1.5 Монотонность структуры распространения констант

◊ **Утверждение**. Структура распространения констант монотонна:
если для $m_1,\ m_2,\ m_1',\ m_2'$ выполняются соотношения:

$m_1 \le m_2, \quad m_1' = f_s(m_1) \quad m_2' = f_s(m_2)$
то $m_1' \le m_2'$

◊ **Доказательство**. Рассмотрим определение f_s через m и m'.

Случай 1): $\boldsymbol{f_s \equiv i} \Rightarrow\ m_1' = m_1, m_2' = m_2 \Rightarrow m_1' \le m_2'$

Случай 2a): $\forall m: f_s(m) = c \Rightarrow m_1' = m_2' = c \Rightarrow m_1' \le m_2'$

Случай 2b): на следующем слайде

Случай 2c): $\forall m : f_s(m) = NAC \Rightarrow m_1' = m_2' = NAC \Rightarrow m_1' \le m_2'$

7.1 Глобальное распространение констант

7.1.5 Монотонность структуры распространения констант

◊ **Доказательство утверждения (случай 2b)**

Для присваивания $x = y + z$ передаточная функция f_s описывается формулой

$$m'(x) = \begin{cases} m(y) + m(z), & \text{если } m(y) \text{ и } m(z) \text{ константы} \\ NAC, & \text{если } m(y) = NAC \text{ или } m(z) = NAC \\ Undef & \text{в остальных случаях} \end{cases}$$

которой соответствует таблица справа.

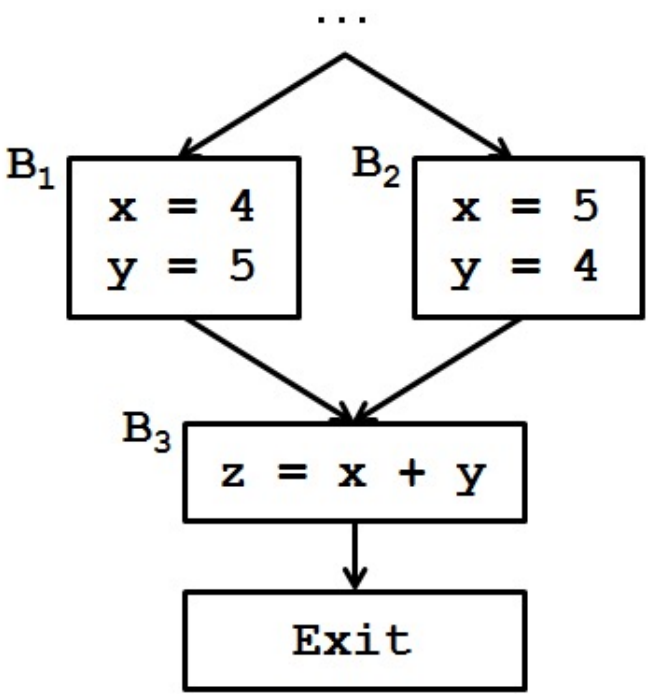
Из таблицы видно, что во всех случаях $m'(x) \le m(x)$

$m(y)$	$m(z)$	$m'(x)$
$Undef$	$Undef$	$Undef$
	c_2	$Undef$
	NAC	NAC
c_1	$Undef$	$Undef$
	c_2	$c_1 + c_2$
	NAC	NAC
NAC	$Undef$	NAC
	c_2	NAC
	NAC	NAC

7.1 Глобальное распространение констант

7.1.6 Недистрибутивность структуры распространения констант

◊ Для установления отсутствия дистрибутивности у передаточной функции структуры распространения констант рассмотрим пример



Пример.

Итеративный алгоритм не может обнаружить, что в программе на рисунке слева значение **x + y** на входе в блок B_3 всегда равно 9

И **x**, и **y** на входе в B_3 имеют значение NAC

Этот результат безопасен (консервативен), но неточен: не удается отследить зависимость:

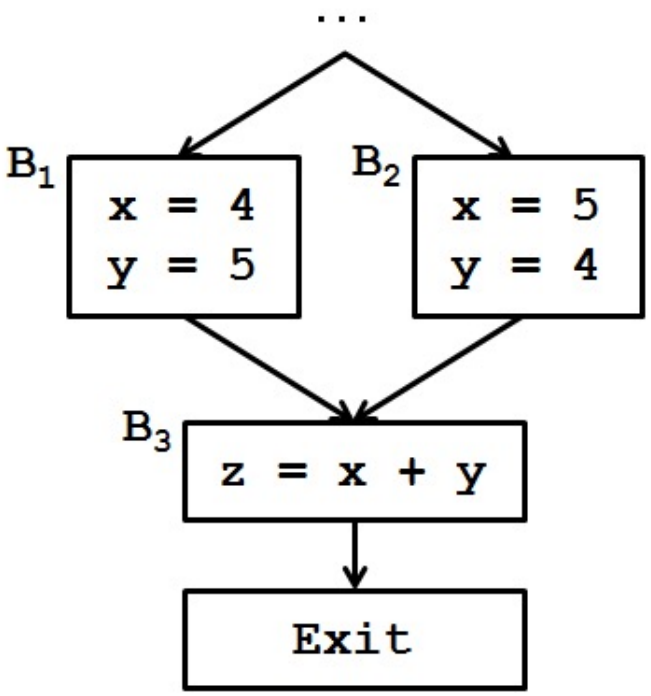
когда **x** равно 4, **y** равно 5,

а когда **x** равно 5, **y** равно 4.

7.1 Глобальное распространение констант

7.1.6 Недистрибутивность структуры распространения констант

◊ Таким образом, структура распространения констант не дистрибутивна, т.е. MFP-решение безопасно, но необязательно равно MOP-решению и может оказаться «меньше» (грубее) него.



Пример.

Итеративный алгоритм не сможет обнаружить, что в программе на рисунке слева значение **x + y** на входе в блок B_3 всегда равно **9**,

так как и **x**, и **y** на входе в B_3 имеют значение **NAC**

Этот результат безопасен (консервативен), но неточен: из-за недистрибутивности структуры распространения констант итеративный алгоритм не может отследить зависимость:

когда **x** равно 4, **y** равно 5,

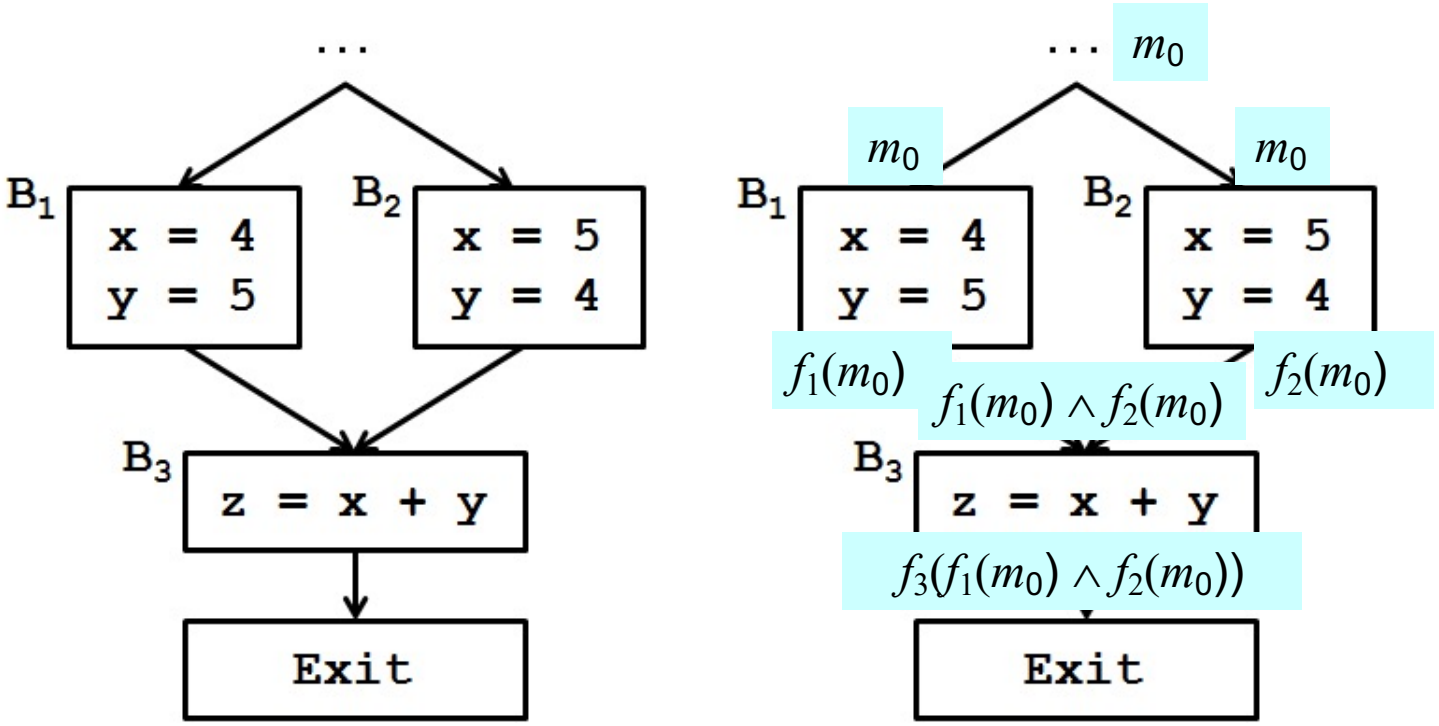
а когда **x** равно 5, **y** равно 4.

7.1 Глобальное распространение констант

7.1.6 Недистрибутивность структуры распространения констант

◊ **Окончание примера.**

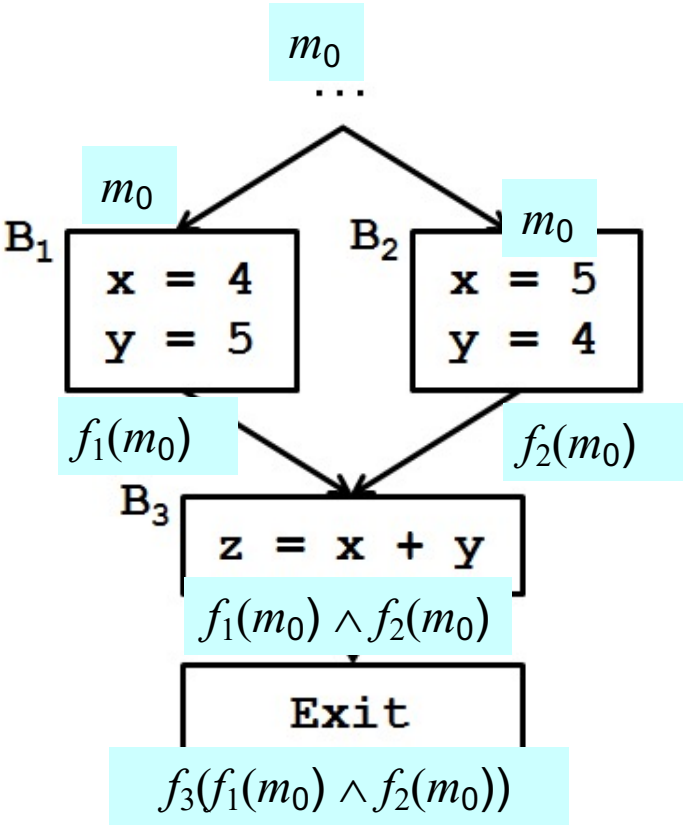
Пусть f_1, f_2 и f_3 – передаточные функции блоков B_1, B_2 и B_3 соответственно, а m_0 – состояние на входе в блоки B_1 и B_2. Тогда



7.1 Глобальное распространение констант

7.1.6 Недистрибутивность структуры распространения констант

◊ **Пример.** Таким образом, как видно из таблицы



m	$m(x)$	$m(y)$	$m(z)$
m_0	*Undef*	*Undef*	*Undef*
$f_1(m_0)$	4	5	*Undef*
$f_2(m_0)$	5	4	*Undef*
$f_1(m_0) \wedge f_2(m_0)$	*NAC*	*NAC*	*Undef*
$f_3(f_1(m_0) \wedge f_2(m_0))$	*NAC*	*NAC*	*NAC*
$f_3(f_1(m_0))$	4	5	9
$f_3(f_2(m_0))$	5	4	9
$f_3(f_1(m_0)) \wedge f_3(f_2(m_0))$	*NAC*	*NAC*	9

$f_3(f_1(m_0) \wedge f_2(m_0)) = NAC$, а $f_3(f_1(m_0)) \wedge f_3(f_2(m_0)) = 9$,

откуда следует **недистрибутивность** структуры:

$$f_3(f_1(m_0) \wedge f_2(m_0)) \neq f_3(f_1(m_0)) \wedge f_3(f_2(m_0))$$

7.1 Глобальное распространение констант

7.1.7 Интерпретация значения *Undef*

◊ Значение *Undef* используется в итеративном алгоритме:

(1) для инициализации входного узла (граничное условие).

(2) для инициализации внутренних точек программы на нулевой итерации

◊ Смысл случаев (1) и (2):

⟐ **Случай (1)**: в начале выполнения программы значения ее переменных не определены.

В конце итеративного процесса переменные на выходе из *Entry* сохраняют значение *Undef*, поскольку $Out[Entry]$ не изменяется.

7.1 Глобальное распространение констант

7.1.7 Интерпретация значения *Undef*

◊ Смысл случаев (1) и (2):

- ◈ **Случай (2):** из-за недостатка информации в начале итерационного процесса решение апроксимируется верхним элементом *Undef*.

  В конце итерационного процесса эти *Undef* сохранятся только в тех точках, для которых не найдется определений соответствующих переменных ни на одном пути, ведущем к этим точкам.

  Если же найдется хотя бы один путь, содержащий определение переменной, достигающее рассматриваемой точки программы, то эта переменная не будет иметь значение *Undef*.

7.1 Глобальное распространение констант

7.1.7 Интерпретация значения *Undef*

◊ Если все определения какой-либо переменной, достигающие некоторой точки программы, имеют одно и то же константное значение (но не $Undef$ и не NAC), то эта переменная рассматривается как константа, **даже если может оказаться, что она не определена на одном из путей**.

◊ Если предположить, что оптимизируемая программа корректна, то итеративный алгоритм может найти большее количество констант, т.к. он может считать константами и те переменные, которые на части путей имеют значения $Undef$.

Такой подход правомерен для языков программирования, которые исходят из того, что неопределенная ($Undef$) переменная может иметь любое значение.

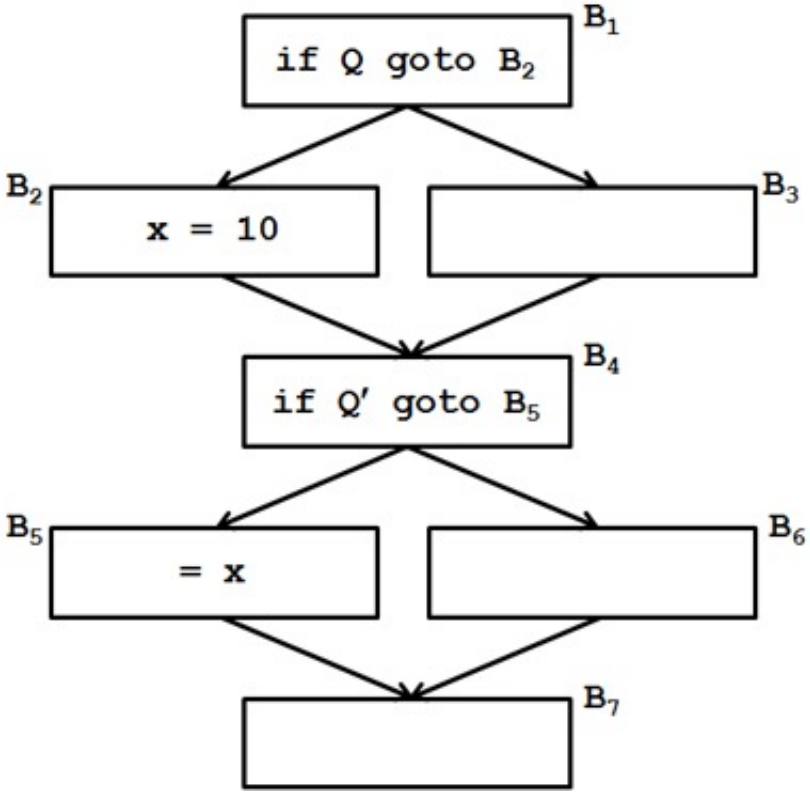
Если семантика языка программирования требует, чтобы все неопределенные переменные принимали некоторое конкретное значение, необходимо это ограничение включить в формулировку задачи анализа потока данных.

7.1 Глобальное распространение констант

7.1.7 Интерпретация значения *Undef*

◊ На рисунке справа значения x на выходе из блоков B_2 и B_3 – соответственно 10 и $Undef$. Из $Undef \wedge 10 = 10$ следует, что значение x на входе в B_4 равно 10, и в B_5 можно заменить x на 10.

◊ Если выполнение пойдет по пути $B_1 \to B_3 \to B_4 \to B_5$, то блока B_5 будет достигать значение $x = Undef$, и замена x на 10 будет корректной не для всех исходных языков.

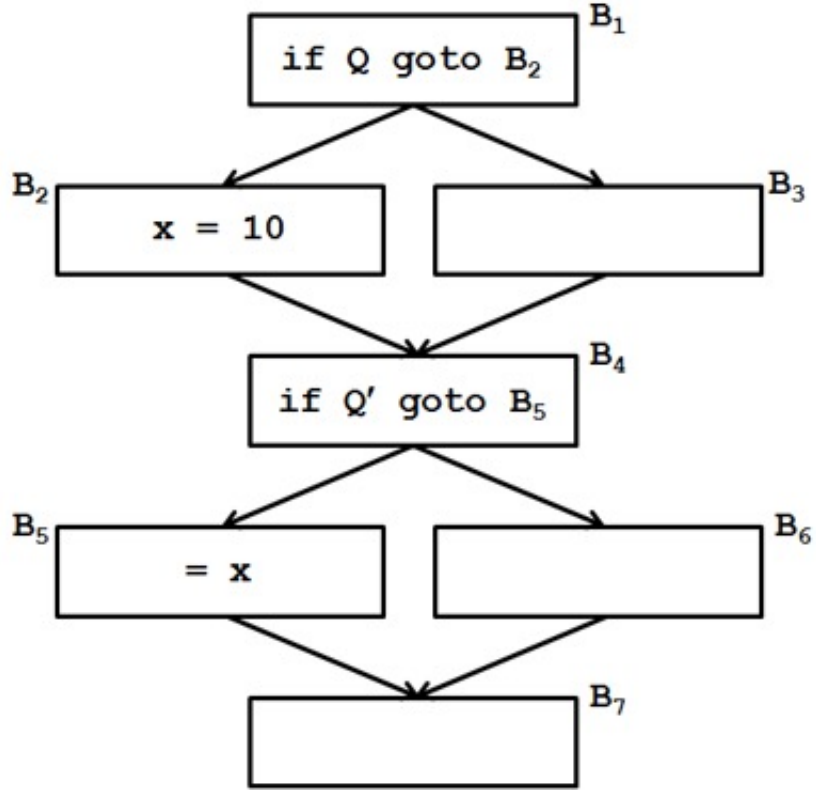


7.1 Глобальное распространение констант

7.1.7 Интерпретация значения *Undef*

Пример.

◊ Даже если можно доказать, что предикат Q не может иметь значение $false$, когда предикат Q' имеет значение $true$, и поэтому путь $B_1 \to B_3 \to B_4 \to B_5$ никогда не будет пройден, следует понимать, что **такой вывод находится за пределами возможностей анализа потоков данных.**



◊ Таким образом, если считать, что исходная программа корректна и что все ее переменные определяются до их использования, то корректно считать, что x при входе в B_5 определена, и ее значение равно 10.

7.1 Глобальное распространение констант

7.1.7 Интерпретация значения *Undef*

В операции сбора:

$\forall c$: $Undef \wedge c = c$,

При этом при вычислении констант в выражениях:

$c + Undef = Undef$

В чем смысл таких определений?

```
void foo(bool p) {
  int z, x, y = 20; // x = UNDEF

  if (p) {
    x = 10;
    z = x + y; // z = 30
  } else {
    z = x + y;
    // UNDEF + 20 = UNDEF
  }
}

// 30 ∧ UNDEF = 30
printf("%d\n", z); // <- z = 30
```

В операции сбора компилятор исходит из того, что конкретное значение «30» не хуже любого другого неинициализированного значения, которое представляет «UNDEF», а в сложении – что при выполнении арифметической операции над неопределенным значением оно остается неопределенным.

В данном случае удается выполнить распространение констант вдоль одной ветви.

7.2 Алгоритм глобального распространения констант

7.2.1 Фаза инициализации

```
// Initialization Phase
WorkList = ∅;
for each SSA-name v {
     initialize Value(v) by rules specified in
the text;
     if Value(v) ≠ Undef then
          WorkList = WorkList ∪ {v};
}
```

На фазе инициализации устанавливаются начальные значения SSA-имен и формируется начальное состояние **`WorkList`**.

Для каждого SSA-имени **`v`** анализируется операция, определяющая **`v`** и устанавливающая значение **`Value(v)`** в соответствии со следующими простыми правилами (следующий слайд).

7.2 Алгоритм глобального распространения констант

7.2.1 Фаза инициализации

```
// Initialization Phase
WorkList = ∅;
for each SSA-name v
initialize Value(v) by rules specified in the
text;
if Value(v) ≠ Undef then
WorkList   WorkList ∪ {v};
```

The rules specified in the text

1. **v** определяется φ-функцией, **Value(v) = Undef**.
2. **v** равно одной из констант c_i, **Value(v) = c_i**.
3. **v** может быть присвоено входное значение, **Value(v) = NAC**.
4. Значение **v** неизвестно, **Value(v)** = **Undef**.

Если **Value(v) ≠ Undef**, алгоритм добавляет **v** в **WorkList**

7.2 Алгоритм глобального распространения констант

7.2.2 Фаза распространения

```
// Propagation Phase
// Выполнять итерации (до неподвижной точки)
while (WorkList ≠ ∅) {
  remove some v from WorkList
  //Выбрать произвольное имя
  for each operation op that uses v {
    let w be the SSA-name that op defines
    if Value(w) ≠ NAC then {
      // Recompute and test for change
      t ← Value(w)
      Value(w) ← result of interpreting op
                 over lattice values
      if Value(w) ≠ t then
        WorkList ← WorkList ∪ {w};
    }
  }}
```

7.2 Алгоритм глобального распространения констант

7.2.2 Фаза распространения

Фаза распространения проста:

Из **`WorkList`** извлекается (с удалением) произвольное SSA-имя **`v`**.

Алгоритм анализирует каждую операцию **`op`**,

которая использует **`v`**

и определяет SSA-имя **`w`**.

Если **`Value(w)`** == **`NAC`**, дальнейший анализ **`w`** не имеет смысла.

В противном случае, моделируется выполнение **`op`** с помощью интерпретации операции над полурешетками значений ее операндов.

Если результат отличен от **`Value(w)`** , он рассматривается как новое значение **`Value(w)`** , причем **`w`** добавляется в **`WorkList`**.

Алгоритм завершается, когда **`WorkList`** пуст.

7.2 Алгоритм глобального распространения констант

7.2.2 Фаза распространения

Интерпретация операции над значениями полурешетки требует некоторой осторожности.
Для φ-функции результаты операции получаются применением операции «сбор» ко всем аргументам φ-функции; правила вычисления результатов операции «сбор» на рисунке справа (в порядке предпочтения)

Правила для операции $\wedge$

$\top \wedge x = x \quad \forall x$

$\bot \wedge x = \bot \quad \forall x$

$c_i \wedge c_j = c_i \quad c_i = c_j$

$c_i \wedge c_j = \bot \quad c_i = c_j$

Что касается других операций компилятор должен «знать» особенности их выполнения. Если любой операнд имеет значение T, необходимо вернуть результат T. Если ни один из операндов не имеет значения T, необходимо вернуть соответствующее значение.
Для каждой операции, вычисляющей значение, алгоритму необходимо иметь набор правил, моделирующих поведение операндов.
Рассмотрим, например операцию $a \times b$. Если a = 4, а b = 17, необходимо выдать в качестве результата $a \times b$ значение 68. Однако, если $a = \bot$, необходимо вернуть результат $\bot$ для всех значений b, кроме b = 0, потому что $a \times 0 = 0$ независимо от значения a.

7.2 Алгоритм глобального распространения констант

7.2.3 Заключительные замечания

◊ **Сложность.**

Фаза распространения – это классическая схема нахождения фиксированной точки. Доказательство сходимости и завершимости, а также оценка сложности алгоритма опирается на факт конечности нисходящих цепочек.

Переменная, связанная с каждым SSA-именем может иметь одно из трех начальных значений: какая-нибудь константа c_i, отличная от $Undef$, или $Undef$, или NAC.

Фаза распространения может только «уменьшить» ее значение. Для данного SSA-имени это может произойти не более, чем дважды: от $Undef$ к c_i и от c_i к NAC. Следовательно, каждое SSA-имя может встретиться в $Worklist$ не более, чем два раза: алгоритм интерпретирует операцию, когда один из операндов удаляется из $Worklist$. Так что общее число интерпретаций не больше, чем удвоенное число операций.

7.2 Алгоритм глобального распространения констант

7.2.3 Заключительные замечания

◊ **Ценность SSA-формы**

SSA-форма позволяет разработать простой, эффективный и понятный алгоритм глобального распространения констант.

Рассмотрим для сравнения классический подход к проблеме распространения констант методом анализа потока данных. Для его разработки потребуется рассмотреть множество Constants в каждом базовом блоке B_i, составить систему уравнений относительно неизвестных $In[B_i]$, определить передаточную функцию для вычисления $Out[B_i]$. Несомненно представленный выше алгоритм гораздо проще. Наибольшую сложность вызывает разработка традиционного механизма интерпретации операций, но в остальном это простой итерационный алгоритм нахождения неподвижной точки над чрезвычайно разреженной решеткой.

Алгоритм работает гораздо быстрее традиционного анализа потока данных.