More CFG
Static Single Assignment
Announcements

- HW 1 due next Monday (9/17)
- Lab 2 will be released tonight (due 9/24)
- Instructor OH cancelled this afternoon
Outline

- More on control flow analysis
  - Dominance frontier

- Dataflow analysis – static single assignment (SSA)
  - SSA Definition
  - PHI node (Φ-node) placement
  - Code optimizations with SSA
Revisiting Control Flow Analysis

- What becomes a leader statement of a basic block? (three cases)
- Does a node strictly (or properly) dominate itself?
- Does a predecessor of a node B always dominate B?
- Suppose A that dominates all of B’s predecessors. Does A always dominate B?
Review: Finding Loops

▸ Loop identification algorithm
  - Find an edge $B \rightarrow H$ where $H$ dominates $B$; This edge is called a back-edge

  - Find all nodes that (1) are dominated by $H$ and (2) can reach $B$ through nodes dominated by $H$; add them to the loop
    • $H$ and $B$ are naturally included
Finding Loops (1)

Find all back edges in this graph and the natural loop associated with each back edge

(9,1)
Finding Loops (1)

Find all back edges in this graph and the natural loop associated with each back edge

(9,1) Entire graph
Finding Loops (2)

Find all back edges in this graph and the natural loop associated with each back edge

(9,1) Entire graph
(10,7)
Intuition of Dominance Relation

Imagine a source of light at the entry node, and that the edges are optical fibers.

To find which nodes are dominated by a given node, place an opaque barrier at that node and observe which nodes become dark.
Finding Loops (2)

Find all back edges in this graph and the natural loop associated with each back edge

(9,1) Entire graph
(10,7)
Finding Loops (2)

Find all back edges in this graph and the natural loop associated with each back edge

(9,1) Entire graph
(10,7) \{7,8,10\}
Finding Loops (3)

Find all back edges in this graph and the natural loop associated with each back edge

- (9,1) Entire graph
- (10,7) \(\{7,8,10\}\)
- (7,4)
Finding Loops (3)

Find all back edges in this graph and the natural loop associated with each back edge

(9,1) Entire graph
(10,7) {7,8,10}
(7,4)
Finding Loops (3)

Find all back edges in this graph and the natural loop associated with each back edge

(9,1)  Entire graph
(10,7)  \{7,8,10\}
(7,4)  \{4,5,6,7,8,10\}
Static Single Assignment

- Static single assignment (SSA) form is a restricted IR where
  - Each variable \textit{definition} has a \textbf{unique} name
  - Each variable \textit{use} refers to a single definition

- SSA simplifies \textbf{data flow analysis} & many compiler optimizations
  - Eliminates artificial dependences (on scalars)
    - Write-after-write
    - Write-after-read
SSA within a Basic Block

- Assign each variable definition a unique name
- Update the uses accordingly

Original code
\[
\begin{align*}
x &= \text{read()} \\
x &= x \times 5 \\
x &= x + 1 \\
y &= x \times 9
\end{align*}
\]

SSA form
\[
\begin{align*}
x_0 &= \text{read()} \\
x_1 &= x_0 \times 5 \\
x_2 &= x_1 + 1 \\
y &= x_2 \times 9
\end{align*}
\]

Corresponding data flow graph
SSA with Control Flow

- Consider a situation where two control-flow paths merge
  - e.g., due to an if-then-else statement or a loop

\[ x = \text{read()} \]
\[ \text{if } (x > 0) \]
\[ y = 5 \]
\[ \text{else} \]
\[ y = 10 \]
\[ x = y \]

\[ x_{0} = \text{read()} \]
\[ \text{if } (x_{0} > 0) \]
\[ y_{0} = 5 \]
\[ y_{1} = 10 \]

\[ x_{1} = y \]

should this be \( y_{0} \) or \( y_{1} \)?
Introducing φ-Node

- Inserts special join functions (called φ-nodes or PHI nodes) at points where different control flow paths converge

\[
\begin{align*}
\text{if} \ (x_0 > 0) \\
\ y_0 &= 5 \\
\ y_1 &= 10 \\
\ y_2 &= \phi(y_0, y_1) \\
\ x_1 &= y_2
\end{align*}
\]

**Note:** φ is not an executable function!

To generate executable code from this form, appropriate copy statements need to be generated in the predecessors (in other words, reversing the SSA process for code generation)
SSA in a Loop

- Insert $\phi$-nodes in the loop header block

$x = 0$
$i = 1$
while ($i < 10$) {
  $x = x + i$
  $i = i + 1$
}

$x_0 = 0$
$i_0 = 1$

$x_1 = \phi(x_0, x_2)$
$i_1 = \phi(i_0, i_2)$
if ($i_1 < 10$)

$x_2 = x_1 + 1$
$i_2 = i_1 + 1$
When and where to add $\phi$-nodes?
- If two control paths $A \to C$ and $B \to C$ converge at a node $C$, and both $A$ and $B$ contain assignments to variable $X$, then $\phi$-node for $X$ must be placed at $C$
  - We often call $C$ a join node or convergence point
  - Can be generalized to more than two converging control paths

Objective: Minimize the number of $\phi$-nodes
- Need to compute dominance frontier sets
Dominance Frontier

- A basic block $F$ is in the dominance frontier set (DF) of basic block $B$ if and only if
  - $B$ does NOT strictly dominate $F$
  - $B$ dominates some predecessor(s) of $F$
  If above two conditions hold, $F \in \text{DF}(B)$

- Intuitively, the basic blocks in the dominance frontier set of $B$ are *almost* strictly dominated by $B$

- Useful for efficiently computing the SSA form
Dominance Frontier and SSA

- **B7 is in the dominance frontier set of B3**
  - B3 does not dominate B7, but dominates one of its predecessors (i.e., B6)

- **For each variable definition in B3, a $\varphi$ node is needed in B7**
  - B7 is the destination of some edge(s) leaving an area dominated by B3
Example: Dominance Frontiers

$\text{ CFG }$

Dominance frontiers of $B$

- $B$ \(\text{DF}(B)\)
- 0  –
- 1  1
- 2  7
- 3  7
- 4  6
- 5  6
- 6  7
- 7  1

$\text{ BF(B)}$

- $F$ is in the dominance frontier set (DF) of $B$ iff
  - $B$ does NOT strictly dominate $F$
  - $B$ dominates some predecessor(s) of $F$

If above two conditions hold, $F \in \text{DF}(B)$
Iterative $\phi$-Node Insertion

- $a$ is defined in 0, 3
  $\Rightarrow$ need $\phi$ in 7,
  then $a$ defined in 7
  $\Rightarrow$ need $\phi$ in 1

- $b$ is defined in 0, 2, 6
  $\Rightarrow$ need $\phi$ in 7
  then $b$ defined in 7
  $\Rightarrow$ need $\phi$ in 1

- $c$ is defined in 0, 2, 5
  $\Rightarrow$ need $\phi$ in 6, 7
  then $c$ defined in 7
  $\Rightarrow$ need $\phi$ in 1
SSA Applications

- SSA form simplifies data flow analysis and many code transformations
  - Primarily due to explicit & simplified (sparse) def-use chains

- Here we show two simple examples
  - Dead code elimination
  - Loop induction variable detection
Dead Code in CDFG

- Dead code is either
  - Unreachable code
  - Definitions never used

- Dead statements?

```
x = a + b
y = c + d
x = a - b
z = z + 1
y = c - d
z = x + y
f(x, y)
```
Dead Code Elimination with SSA

Iteratively remove unused definitions:
remove y1, z2 (and B4) \(\rightarrow\) then remove z1
Loop Induction Variables

- An induction variable is a variable that
  - Gets increased or decreased by a fixed amount (loop invariant) on every iteration of a loop (basic induction variable)
    - \(i = i + c\)
  - or is an affine function of another induction variable (mutual induction variable)
    - \(j = a \times i + b\)
Identifying Basic Loop Induction Variable

- Find basic loop induction variable(s)
  1. Inspect back edges in the loop
  2. Each back edge points to a $\phi$ node in the loop header, which may indicate a basic induction variable
  3. $\phi$ is a function of an initialized variable and a definition in the form of “$i + c$” (i.e., increment operation)
Summary

- Importance of compilers
  - Essential component of SoC software development flow
  - Essential component of high-level synthesis

- A good intermediate representation (IR) enables efficient and effective analysis and optimization
  - Dominance relation helps effective CFG analysis
  - SSA form facilitates efficient IR-level optimization
Acknowledgements

- These slides contain/adapt materials developed by
  - Prof. Scott Mahlke (UMich)