

# SSA form

Michel Schinz  
Advanced Compiler Construction – 2008-05-23

# Static single assignment (SSA) form

## Static single assignment

**Static single-assignment** (or **SSA**) form is an intermediate representation in which each variable has only one definition in the program.

That single definition can be executed many times when the program is run – if it is inside a loop – hence the qualifier *static*.

SSA form is interesting because it simplifies several optimisations and analysis, as we will see.

3

## Straight-line code

Transforming a piece of straight-line code – *i.e.* without branches – to SSA is trivial: each definition of a given name gives rise to a new version of that name, identified by a subscript:

x=12	→	x <sub>1</sub> =12
y=15		y <sub>1</sub> =15
x=x+y		x <sub>2</sub> =x <sub>1</sub> +y <sub>1</sub>
y=x+4		y <sub>2</sub> =x <sub>2</sub> +4
z=x+y		z <sub>1</sub> =x <sub>2</sub> +y <sub>2</sub>
y=y+1		y <sub>3</sub> =y <sub>2</sub> +1

4

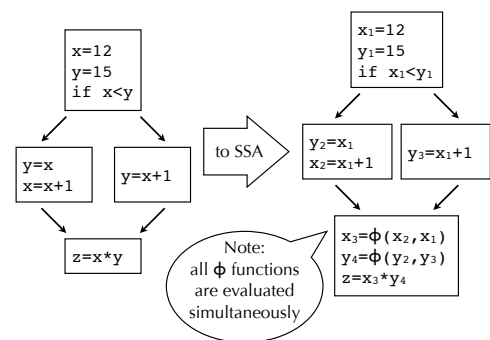
## $\phi$ -functions

Join-points in the CFG – nodes with more than one predecessors – are more problematic, as each predecessor can bring its own version of a given name.

To reconcile those different versions, a fictional  $\phi$ -function is introduced at the join point. That function takes as argument all the versions of the variable to reconcile, and automatically selects the right one depending on the flow of control.

5

## $\phi$ -functions example



6

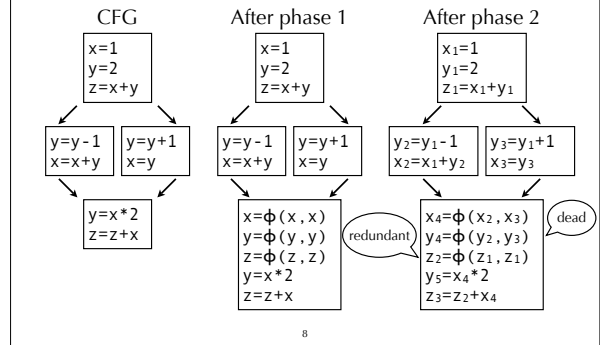
## (Naïve) building of SSA form

Naïve technique to build SSA form:

- for each variable  $x$  of the CFG, at each join point  $n$ , insert a  $\phi$ -function of the form  $x = \phi(x_1, \dots, x_n)$  with as many parameters as  $n$  has predecessors,
- compute reaching definitions, and use that information to rename any use of a variable according to the – now unique – definition reaching it.

7

## (Naïve) building of SSA form



8

## Smarter techniques

The naïve technique just presented works, in the sense that the resulting program is in SSA form and is equivalent to the original one.

However, it introduces too many  $\phi$ -functions – some dead, some redundant – to be useful in practice. It builds the **maximal** SSA form.

We will examine better techniques later, but to understand them we must first introduce the notion of dominance in a CFG.

9

## Dominance

## Dominance

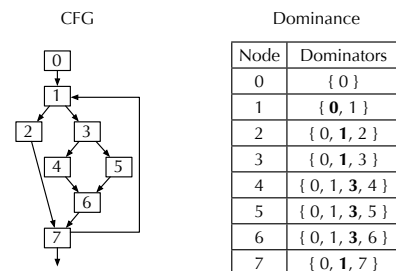
In a control-flow graph, a node  $n_1$  **dominates** a node  $n_2$  if all paths from the start node to  $n_2$  pass through  $n_1$ .

By definition, the domination relation is reflexive, that is a node  $n$  always dominates itself. We then say that node  $n_1$  **strictly dominates**  $n_2$  if  $n_1$  dominates  $n_2$  and  $n_1 \neq n_2$ .

The **immediate dominator** of a node  $n$  is the strict dominator of  $n$  closest to  $n$ .

11

## Dominance example



(immediate dominator in bold)

12

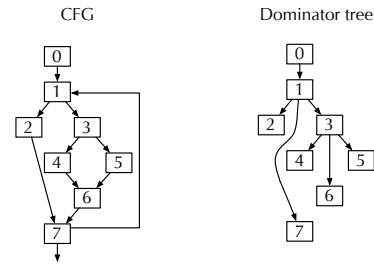
## Dominator tree

The **dominator tree** is a tree representing the dominance relation.

The nodes of the tree are the nodes of the CFG, and a node  $n_1$  is a parent of a node  $n_2$  if and only if  $n_1$  is the immediate dominator of  $n_2$ .

13

## Dominator tree example



14

## Computing dominance

Dominance can be computed using data-flow analysis.

To each node  $n$  of the CFG we attach a variable  $v_n$  giving the set of nodes that dominate  $n$ . The value of  $v_n$  is given by the following equation:

$$v_n = \{ n \} \cup (v_{p1} \cap v_{p2} \cap \dots \cap v_{pk})$$

where  $p_1, \dots, p_k$  are the predecessors of  $n$ .

15

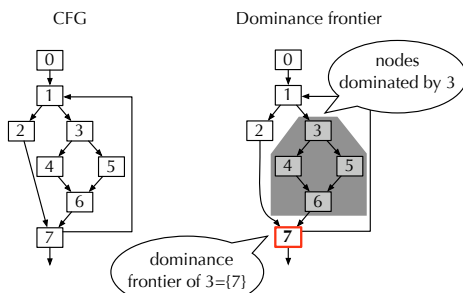
## Dominance frontier

The **dominance frontier** of a node  $n$  – written  $DF(n)$  – is the set of all nodes  $m$  such that  $n$  dominates a predecessor of  $m$ , but does not strictly dominate  $m$  itself.

Informally, the dominance frontier of  $n$  contains the first nodes which are reachable from  $n$  but which are not strictly dominated by  $n$ .

16

## Dominance frontier example



17

## Building SSA form

## Minimal SSA form

The naïve technique to build SSA form presented earlier inserts  $\phi$ -functions for every variable at the beginning of every join point.

Using dominance information, it is possible to do better, and compute **minimal** SSA form: for each definition of a variable  $x$  in a node  $n$ , insert a  $\phi$ -function for  $x$  in all nodes of  $DF(n)$ .

Notice that the inserted  $\phi$ -functions are definitions, and can therefore force the insertion of more  $\phi$ -functions.

19

## Improving on minimal SSA

Reminder: the naïve technique to build SSA form presented at the beginning computes maximal SSA form.

The better technique just presented computes minimal SSA form.

Unfortunately, minimal SSA form is not necessarily optimal, and can contain dead  $\phi$ -functions. To solve that problem, improved techniques have been developed to build semi-pruned – which is still not optimal – and pruned SSA form.

20

## Semi-pruned SSA form

Observation: a variable that is only live in a single node can never have a live  $\phi$ -function.

Therefore, the minimal technique can be further refined by first computing the set of **global names** – defined as the names that are live across more than one node – and producing  $\phi$ -functions for these names only.

This is called **semi-pruned SSA form**.

21

## Building semi-pruned SSA form

Like the naïve technique to build maximal SSA form, the algorithm to build semi-pruned SSA form is composed of two phases:

1.  $\phi$ -functions are inserted for global names, according to dominance information,
2. variables are renamed.

22

## Phase 1: inserting $\phi$ -functions

Before inserting  $\phi$ -functions, the set  $G$  of global names must be computed. Once this is done, insertion of  $\phi$ -functions is done as follows:

```
for each name  $x$  in  $G$ 
  work list = all nodes in which  $x$  is defined
  for each node  $n$  in work list
    for each node  $m$  in  $DF(n)$ 
      insert a  $\phi$ -function for  $x$  in  $m$ 
      work list = work list  $\cup$  {  $m$  }
```

23

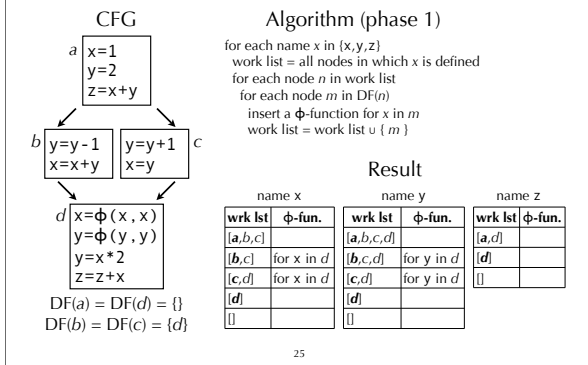
## Phase 2: renaming variables

Renaming is done by a pre-order traversal of the dominator tree, as follows:

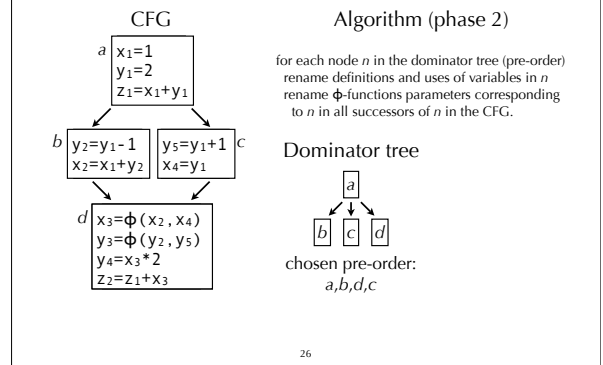
```
for each node  $n$  in the dominator tree
  rename definitions and uses of variables in  $n$ 
  rename  $\phi$ -functions parameters corresponding to  $n$  in all
  successors of  $n$  in the CFG.
```

24

## Example: phase 1



## Example: phase 2



## Generating code from SSA form

## Generating code from SSA

After the program has been turned into SSA form and the various optimisations performed on that representation, it must be transformed into executable form.

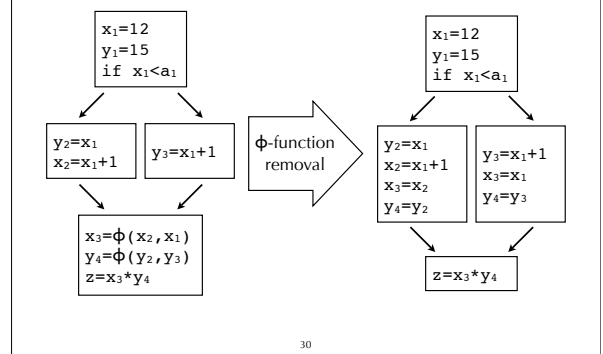
This implies in particular that  $\phi$ -functions must be removed, as they cannot be implemented on standard machines.

## Removing $\phi$ -functions

A  $\phi$ -function of the form  $x_i = \phi(x_{i_1}, \dots, x_{i_n})$  can be removed by inserting appropriate assignments to  $x_i$  in all predecessors of the node containing that function.

This will introduce many assignments of the form  $x_i = x_j$  (that is, move instructions), but most of them will be removed later during register allocation, thanks to coalescing.

## Removing $\phi$ -functions



## Critical edges

CFG edges that go from a node with multiple successors to a node with multiple predecessors are called **critical edges**.

While removing  $\phi$ -functions, the presence of a critical edge from  $n_1$  to  $n_2$  leads to the insertion of redundant move instructions in  $n_1$ , corresponding to the  $\phi$ -functions of  $n_2$ . Ideally, they should be executed only if control reaches  $n_2$  later, but this is not certain when  $n_1$  executes.

31

## Edge splitting

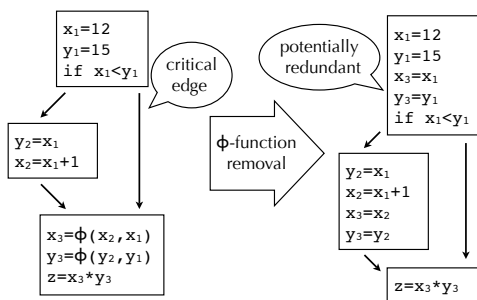
Critical edges can easily be avoided completely using **edge splitting**.

Edge splitting consists in replacing all critical edges leading from a node  $n_1$  to a node  $n_2$  by two edges: one from  $n_1$  to a new empty node  $n_3$ , and one from  $n_3$  to  $n_2$ .

Since the new empty block  $n_3$  has only one predecessor and one successor, this effectively removes the critical edge.

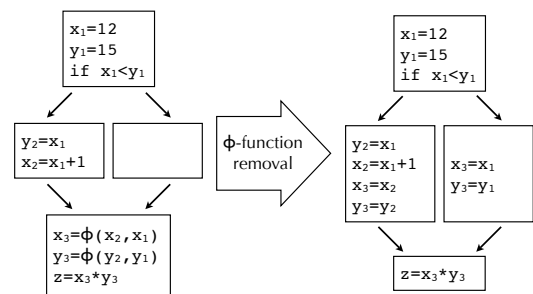
32

## Without edge splitting



33

## With edge splitting



34

## Using SSA form

## Dead-code elimination

Basic dead-code elimination is trivial in SSA form: if a variable  $x_i$  is not used in some expression, then its definition – of the form  $x_i=y_j$  or  $x_i=\phi(x_j, \dots, x_k)$  – can be deleted. Of course, this is only true if that definition does not have side-effects.

The deletion of a definition can remove the last use of some other variable, in which case its definition can be deleted too, and so on...

36

## Simple constant propagation

SSA form also simplifies constant propagation: whenever a definition of the form  $x_i=c$  – where  $c$  is a constant – is encountered, then all uses of  $x_i$  can be replaced by  $c$ . Moreover, the definition itself can be deleted from the program, as it is now dead.

Also, a  $\phi$ -function of the form  $x_i=\phi(c_1, \dots, c_n)$  where  $c_1=\dots=c_n$  can be replaced by  $x_i=c_1$ , which is then simplified as above.

37

## Copy propagation

Copy propagation can be handled in a similar fashion as constant propagation: definitions of the form  $x_i=y_j$ , and single-argument  $\phi$ -functions of the form  $x_i=\phi(y_j)$  can be deleted, and all uses of  $x_i$  replaced by uses of  $y_j$ .

The same is true of constant folding: a definition of the form  $x_i=c_1 \text{ op } c_2$  – where  $c_1$  and  $c_2$  are constants – can be deleted and all uses of  $x_i$  replaced by the value of  $c_1 \text{ op } c_2$ .

38

## Liveness analysis

SSA form also simplifies liveness analysis, and hence the construction of the interference graph needed by register allocation.

To compute the region where a variable  $x_i$  is live in SSA form, it is sufficient to start from all uses of  $x_i$  and walk backwards in the CFG until the definition of  $x_i$  is encountered. The statements encountered during that walk are those during which  $x_i$  is live.

39

## Summary

Static single-assignment (SSA) form is an intermediate representation where all names are defined exactly once. To enable this,  $\phi$ -functions have to be inserted at join points in the CFG.

Transforming a program to SSA form is not completely trivial since unnecessary  $\phi$ -functions should be avoided.

SSA encodes the data-flow of the program in its names, making several optimisations easier.

40