# CS 477: Operational Program Semantics <br> Sasa Misailovic 

Based on previous slides by Gul Agha, Elsa Gunter, Madhusudan Parthasarathy, Mahesh Viswanathan, and Vikram Adve

University of Illinois at Urbana-Champaign

## Previously, on CS 477

Propositional Logic:

- Syntax
- Semantics
- Proof
(Homework/Quiz \#1 is out: due next Thursday)

Simple Imperative Programming Language

- I $\in$ Identifiers
- $\mathrm{N} \in$ Numerals
- B ::= true | false
$|B \& B| B$ or $B|\operatorname{not} B| E<E \mid E=E$
- $\mathrm{E}::=\mathrm{N}|\mathrm{I}| \mathrm{E}+\mathrm{E}|\mathrm{E} * \mathrm{E}| \mathrm{E}-\mathrm{E} \mid-\mathrm{E}$
-S::= skip | S; S | I ::= E
| if B then S else S fi \| while B do S od


## Syntax -> Graphs

Reminder: Graph: (V, E)

- V is a set of vertices (nodes)
- $\mathrm{E} \subseteq V \times V$ is a relation denoting "connected" nodes. Elements $e \in E$ are edges: pairs of connected vertices $e=\left(v_{1}, v_{2}\right)$. Can be directed or undirected.
Common definitions:
- Post(v) - successor vertices of $\mathrm{v}, \operatorname{Pre}(\mathrm{v})$ - direct predecessor vertices of v
- Path: a sequence of vertices s.t. $v_{i} \in \operatorname{Pre}\left(v_{i+1}\right)$. Cycle when the same vertex multiple times in the path, else simple. Length: number of vertices in a path.
- Acyclic graphs: no cycles.
- Tree: exists $v_{\text {root }}$ (without predecessors) such that all other vertices reachable along unique paths
- Strongly connected component: all pairs of vertices mutually reachable
- Search: DFS, BFS; traversal: preorder, postorder, etc.


## Syntax -> Graphs

- Parse Tree (from CS 374)
- Abstract Syntax Tree
- Control-flow Graph


## Flow Graphs

- Flow Graph: A triple $\mathrm{G}=(\mathrm{N}, \mathrm{A}, \mathrm{s})$, where ( $\mathrm{N}, \mathrm{A}$ ) is a (finite) directed graph, $s \in N$ is a designated "initial" node, and there is a path from node $s$ to every node $n \in N$.
- An entry node in a flow graph has no predecessors.
- An exit node in a flow graph has no successors.
- There is exactly one entry node, s. We can modify a general DAG to ensure this. How?
- We can also transform the graph to have only one exit node. How?


## Control Flow Graph (CFG)

- Flow Graph: A triple $\mathrm{G}=(\mathrm{N}, \mathrm{A}, \mathrm{s})$, where $(\mathrm{N}, \mathrm{A})$ is a (finite) directed graph, $s \in N$ is a designated "initial" node, and there is a path from node $s$ to every node $n \in N$.
- Control Flow Graph (CFG) is a flow graph that represents all paths (sequences of statements) that might be traversed during program execution.
- Nodes in CFG are program statements, and edge ( $\mathrm{S}_{1}, \mathrm{~S}_{2}$ ) denotes that statement $S_{1}$ can be followed by $S_{2}$ in execution.
- In CFG, a node unreachable from s can be safely deleted. Why?
- Control flow graphs are usually sparse. I.e., $|\mathrm{A}|=\mathrm{O}(|\mathrm{N}|)$. In fact, if only binary branching is allowed $|\mathrm{A}| \leq 2|\mathrm{~N}|$.


## Control Flow Graph (CFG)

- Basic Block is a sequence of statements $S_{1} \ldots S_{n}$ such that execution control must reach $\mathrm{S}_{1}$ before $\mathrm{S}_{2}$, and, if $\mathrm{S}_{1}$ is executed, then $S_{2} \ldots S_{n}$ are all executed in that order
- Unless some statement $S_{i}$ causes the program to halt
- Leader is the first statement of a basic block
- Maximal Basic Block is a basic block with a maximum number of statements ( n )


## Control Flow Graph (CFG)

 Let us refine our previous definition- CFG is a directed graph in which:
- Each node is a single basic block
- There is an edge b1 $\rightarrow$ b2 if block b2 may be executed after block b1 in some execution
- We typically define it for a single procedure
- A CFG is a conservative approximation of the control flow! Why?


## Example

LLVM bitcode (ver 3.9.1)

## Source Code

```
unsigned fib(unsigned n) {
    int i;
    int f0 = 0, f1 = 1, f2;
    if (n <= 1) return n;
    for (i = 2; i <= n; i++) {
        f2 = f0 + f1;
        f0 = f1;
        f1 = f2;
    }
    return f2;
}
```

```
define i32 @fib(i32 %0) {
```

define i32 @fib(i32 %0) {
%2 = icmp ult i32 %0, 2
%2 = icmp ult i32 %0, 2
br i1 %2, label %12, label %3
br i1 %2, label %12, label %3
; <label>:3:
; <label>:3:
br label %4
br label %4
; <label>:4:
; <label>:4:
%5 = phi i32 [ %8, %4 ], [ 1, %3 ]
%5 = phi i32 [ %8, %4 ], [ 1, %3 ]
%5 = phi i32 [ %8, %4 ], [ 1, %3 ]
%5 = phi i32 [ %8, %4 ], [ 1, %3 ]
%5 = phi i32[ %8, %4 ], [ 1, %3 ]
%5 = phi i32[ %8, %4 ], [ 1, %3 ]
%8 = add i32 %5, %6
%8 = add i32 %5, %6
%9 = add i32 %7, 1
%9 = add i32 %7, 1
%10 = icmp ugt i32 %9, %0
%10 = icmp ugt i32 %9, %0
br i1 %10, label %11, label %4
br i1 %10, label %11, label %4
; <label>:11:
; <label>:11:
br label %12
br label %12
; <label>:12:
; <label>:12:
%13 = phi i32 [%0, %1], [%8, %11]
%13 = phi i32 [%0, %1], [%8, %11]
ret i32 %13
ret i32 %13
}

```
}
```


## Dominance in Flow Graphs

- Let $\mathrm{d}, \mathrm{d} 1, \mathrm{~d} 2, \mathrm{~d} 3, \mathrm{n}$ be nodes in G .
- d dominates n ("d dom n ") iff every path from s to n contains d
- $d$ properly dominates $n$ if $d$ dominates $n$ and $d \neq n$
- $d$ is the immediate dominator of $n$ ("d idom $n$ ") if $d$ is the last proper dominator on any path from initial node to $n$,
- DOM(x) denotes the set of dominators of $x$,
- Dominator tree: the children of each node d are the nodes n such that "d idom n" (immediately dominates)


## Dominator Properties

- Lemma 1: $\operatorname{DOM}(\mathrm{s})=\{\mathrm{s}\}$.
- Lemma 2: s dom d, for all nodes din G.
- Lemma 3: The dominance relation on nodes in a flow graph is a partial ordering
- Reflexive - $n$ dom $n$ is true for all $n$.
- Antisymmetric - If $d$ dom $n$, then cannot be $n$ dom $d$
- Transitive - d1 dom $d 2 \wedge d 2$ dom $d 3 \Rightarrow d 1$ dom d3
- Lemma 4: The dominators of a node form a list.
- Lemma 5: Every node except $s$ has a unique immediate dominator.


## Postdominance

Def. Postdomination: node $\rho$ postdominates a node $d$ iff all paths to the exit node of the graph starting at $d$ must go through $p$

Def. Reverse Control Flow Graph (RCFG) of a CFG has the same nodes as CFG and has edge $Y \rightarrow X$ if $X \rightarrow$ $Y$ is an edge in CFG.

- $p$ is a postdominator of $d$ iff $p$ dominates $d$ in the RCFG.


## Semantics

- Expresses the meaning of syntax
- Static semantics
- Meaning based only on the form of the expression without executing it
- Usually restricted to type checking / type inference


## Dynamic semantics

- Method of describing meaning of executing a program
- Several different types:
- Operational Semantics
- Axiomatic Semantics
- Denotational Semantics
- Different languages better suited to different types of semantics
- Different types of semantics serve different purposes


## Operational Semantics

- Start with a simple notion of machine
- Describe how to execute (implement) programs of language on virtual machine, by describing how to execute each program statement (ie, following the structure of the program)
- Meaning of program is how its execution changes the state of the machine
- Useful as basis for implementations


## Denotational Semantics

- Construct a function $\mathscr{M}$ assigning a mathematical meaning to each program construct
- Lambda calculus often used as the range of the meaning function
- Meaning function is compositional: meaning of construct built from meaning of parts
- Useful for proving properties of programs


## Axiomatic Semantics

-Also called Floyd-Hoare Logic

- Based on formal logic (first order predicate calculus)
-Axiomatic Semantics is a logical system built from axioms and inference rules
- Mainly suited to simple imperative programming languages


## Axiomatic Semantics

- Used to formally prove a property (postcondition) of the state (the values of the program variables) after the execution of program, assuming another property (precondition) of the state before execution
-Written :

$$
\text { \{Precondition\} Program \{Postcondition\} }
$$

Much more about it later in the course!

## Modeling Program Environment



## Program Environment

Pair of code to execute + a valuation (aka state)
Code to execute: Next statement and program text that remains to be executed: Statement_1; Other_Statements
A valuation of program variables:

- Mapping m: Identifiers-> Value

Program statements (" $\mathrm{S}_{1} ; \mathrm{S}_{2} ; \ldots \mathrm{S}_{\mathrm{n}}$ ") transform the valuations. Execution is then:

- $m_{2}=\left[\left[S_{1}\right]\right]\left(m_{1}\right)$
- $m_{3}=\left[\left[S_{2}\right]\right]\left(m_{2}\right)$
- ...
- $m_{n+1}=\left[\left[S_{n}\right]\right]\left(m_{n}\right)$
- Also $\left(s_{1}, m_{1}\right) \rightarrow\left(s_{2}, m_{2}\right) \rightarrow\left(s_{3}, m_{3}\right) \rightarrow \ldots \rightarrow\left(s_{n}, m_{n}\right) \longrightarrow\left(\cdot, m_{n+1}\right)$. We can define the sequence $\left(s_{1}, m_{1}\right),\left(s_{2}, m_{2}\right),\left(s_{3}, m_{3}\right), \ldots,\left(s_{n}, m_{n}\right),\left(\cdot, m_{n+1}\right)$ or its projection $\left(m_{1}, \ldots m_{n}\right)$ as the trace of execution


## Natural Semantics ("Big-step Semantics")

- Aka Structural Operational Semantics, aka "Big Step Semantics"
- Provide value for a program by rules and derivations, similar to type derivations
- Rule conclusions look like

$$
(C, m) \Downarrow m^{\prime}
$$

"Evaluating a command C in the state m results in the new state $\mathrm{m}^{\prime}$ " or

$$
(E, m) \Downarrow v
$$

"Evaluating an expression E in the state $m$ results in the value $v$ ""

Simple Imperative Programming Language

- I $\in$ Identifiers
- $\mathrm{N} \in$ Numerals
- B ::= true | false
$|B \& B| B$ or $B|\operatorname{not} B| E<E \mid E=E$
- $\mathrm{E}::=\mathrm{N}|\mathrm{I}| \mathrm{E}+\mathrm{E}|\mathrm{E} * \mathrm{E}| \mathrm{E}-\mathrm{E} \mid-\mathrm{E}$
- C::= skip | C;C|I::=E
| if B then C else C fi | while B do C od


# Natural Semantics of Atomic Expressions 

- Identifiers: $(\mathrm{k}, \mathrm{m}) \Downarrow \mathrm{m}(\mathrm{k})$
- Numerals are values: $(N, m) \Downarrow N$
- Booleans: $\quad($ true, m$) \downarrow$ true
(false ,m) $\downarrow$ false


## Boolean:

$\frac{(B, m) \Downarrow \text { false }}{\left(B \& B^{\prime}, m\right) \Downarrow \text { false }} \frac{(B, m) \Downarrow \text { true }\left(B^{\prime}, m\right) \Downarrow b}{\left(B \& B^{\prime}, m\right) \Downarrow b}$
$\frac{(B, m) \Downarrow \text { true }}{\left(B \text { or } B^{\prime}, m\right) \Downarrow \text { true }} \frac{(B, m) \Downarrow \text { false }\left(B^{\prime}, m\right) \Downarrow b}{\left(B \text { or } B^{\prime}, m\right) \Downarrow_{b}}$


## Binary Relations



- By U rop V = b, we mean does (the meaning of) the relation rop hold on the meaning of $U$ and $V$
- May be specified by a mathematical expression/equation or rules matching U and V


## Arithmetic Expressions


where N is the specified value for (mathematical) U op V

## Commands

Skip: $\quad($ skip,$m) \Downarrow m$

Assignment:
$\frac{(E, m) \Downarrow V}{(k:=E, m) \Downarrow m[k<--V]}$

Sequencing: $\frac{(C, m) \Downarrow m^{\prime}\left(C^{\prime}, m^{\prime}\right) \Downarrow m^{\prime \prime}}{\left(C ; C^{\prime}, m\right) \Downarrow m^{\prime \prime}}$

## If Then Else Command


$\frac{(B, m) \Downarrow \text { false } \quad\left(C^{\prime}, m\right) \Downarrow m^{\prime}}{\text { (if } B \text { then } C \text { else } C^{\prime} \text { fi, } m \text { ) } \text { m }^{\prime}}$

## Example: If Then Else Rule

(if $x>5$ then $y:=2+3$ else $y:=3+4$ fi, $\{x->7\}$ )
$\Downarrow$ ?

## Example: If Then Else Rule

$(x>5,\{x->7\}) \downarrow ?$
(if $x>5$ then $y:=2+3$ else $y:=3+4$ fi, $\{x->7\}$ )
$\Downarrow$ ?

## Example: Arith Relation

? > ? $=$ ?
$(x,\{x->7\}) \downarrow ? \quad(5,\{x->7\}) \downarrow ?$
$(x>5,\{x->7\}) \downarrow$ ?
(if $x>5$ then $y:=2+3$ else $y:=3+4$ fi, $\{x->7\}$ )

$$
\Downarrow ?
$$

## Example: Identifier(s)

$7>5=$ true
$(x,\{x->7\}) \Downarrow 7 \quad(5,\{x->7\}) \downarrow 5$
$(x>5,\{x->7\}) \downarrow$ ?
(if $x>5$ then $y:=2+3$ else $y:=3+4$ fi, $\{x->7\}$ )
$\Downarrow$ ?

## Example: Arith Relation

$$
7>5=\text { true }
$$

$(x,\{x->7\}) \Downarrow_{7} \quad(5,\{x->7\}) \Downarrow_{5}$
$(x>5,\{x->7\}) \$ true
(if $x>5$ then $y:=2+3$ else $y:=3+4$ fi, $\{x->7\}$ )
$\Downarrow$ ?

## Example: If Then Else Rule

$7>5=$ true
$(x,\{x->7\}) \Downarrow 7 \quad(5,\{x->7\}) \Downarrow 5 \quad(y:=2+3,\{x->7\}$
$(x>5,\{x->7\}) \backslash$ true
$\Downarrow$ ?
(if $x>5$ then $y:=2+3$ else $y:=3+4$ fig, $\{x->7\}$ )
$\Downarrow$ ?

## Example: Assignment

$$
\left.\begin{array}{cc}
7>5=\text { true } & (2+3,\{x->7\}) \downarrow ? \\
\frac{(x,\{x->7\}) \Downarrow 7}{(x>5,\{x->7\}) \backslash \text { true }} \quad(5,\{x->7\}) \Downarrow 5
\end{array}\right) \quad \begin{gathered}
(y:=2+3,\{x->7\} \\
\Downarrow ? \\
\hline
\end{gathered}
$$

(if $x>5$ then $y:=2+3$ else $y:=3+4$ fi, $\{x->7\}$ )

$$
\Downarrow ?
$$

## Example: Arith Op

? + ? = ?

$$
(2,\{x->7\}) \downarrow ? \quad(3,\{x->7\}) \downarrow ?
$$

$$
7>5=\text { true }
$$

$$
(2+3,\{x->7\}) \downarrow ?
$$

$(x,\{x->7\}) \Downarrow 7 \quad(5,\{x->7\}) \| 5$
$(y:=2+3,\{x->7\}$
$(x>5,\{x->7\}) \backslash$ true
$\downarrow$ ?
(if $x>5$ then $y:=2+3$ else $y:=3+4$ fig, $\{x->7\}$ )
$\Downarrow$ ?

## Example: Numerals

$$
\begin{aligned}
& 2+3=5 \\
& (2,\{x->7\}) \downarrow_{2} \quad(3,\{x->7\}) \downarrow 3 \\
& 7>5=\text { true } \\
& (2+3,\{x->7\}) \downarrow \text { ? } \\
& (x,\{x->7\}) \Downarrow 7 \quad(5,\{x->7\}) \|_{5} \\
& (y:=2+3,\{x->7\} \\
& (x>5,\{x->7\}) \backslash \text { true } \\
& \Downarrow \text { ? } \\
& \text { (if } x>5 \text { then } y:=2+3 \text { else } y:=3+4 \text { ai, }\{x->7\} \text { ) } \\
& \Downarrow \text { ? }
\end{aligned}
$$

## Example: Arith Op

$$
\begin{gathered}
2+3=5 \\
7>5=\text { true } \\
\frac{(2,\{x->7\}) \Downarrow_{2} \quad(3,\{x->7\}) \Downarrow_{3}}{(x,\{x->7\}) \Downarrow 7 \quad(5,\{x->7\}) \Downarrow_{5}} \frac{(2+3,\{x->7\}) \Downarrow_{5}}{(x:=2+3,\{x->7\}} \\
\frac{(x>5,\{x->7\}) \downarrow \text { true }}{\Downarrow} \downarrow ? \\
\hline \text { (if } x>5 \text { then } y:=2+3 \text { else } y:=3+4 \text { fig, }\{x->7\}) \\
\Downarrow ?
\end{gathered}
$$

## Example: Assignment

$$
\begin{aligned}
& 2+3=5 \\
& (2,\{x->7\}) \downarrow 2 \quad(3,\{x->7\}) \downarrow 3 \\
& 7>5=\text { true } \\
& (2+3,\{x->7\}) \downarrow 5 \\
& (x,\{x->7\}) \downarrow 7 \quad(5,\{x->7\}) \downarrow ل_{5} \\
& (y:=2+3,\{x->7\} \\
& (x>5,\{x->7\}) \backslash \text { true } \\
& \Downarrow\{x->7, y->5\} \\
& \text { (if } x>5 \text { then } y:=2+3 \text { else } y:=3+4 \text { ai, }\{x->7\} \text { ) } \\
& \downarrow \text { ? }
\end{aligned}
$$

## Example: If Then Else Rule

$$
\begin{aligned}
& 2+3=5 \\
& (2,\{x->7\}) \downarrow 2 \quad(3,\{x->7\}) \downarrow 3 \\
& 7>5=\text { true } \\
& (2+3,\{x->7\}) ل_{5} \\
& (x,\{x->7\}) \Downarrow 7 \quad(5,\{x->7\}) \downarrow ل_{5} \\
& (y:=2+3,\{x->7\} \\
& (x>5,\{x->7\}) \downarrow \text { true } \\
& \Downarrow\{x->7, y->5\} \\
& \text { (if } x>5 \text { then } y:=2+3 \text { else } y:=3+4 \text { fin, }\{x->7\} \text { ) } \\
& \Downarrow\{x->7, y->5\}
\end{aligned}
$$

## While Command


(while B do Cod, m) $\downarrow \mathrm{m}$
(B,m) $\Downarrow_{\text {true }}^{2}(C, m) \Downarrow m^{\prime}{ }^{3}$ (while $B$ do $C$ od, $\left.m^{\prime}\right) \Downarrow m^{\prime \prime}$ (while B do Cod, m) $\downarrow \mathrm{m}^{\prime \prime}$

## Example: While Rule


(3) $(x>5,\{x->2\}) \Downarrow$ false
$(x>5,\{x->7\}) \downarrow$ true $\quad$ while $x>5$ do $x:=x-5$ od;
2. $(x:=x-5,\{x->7\}) \Downarrow\{x->2\} \quad\{x->2\}) \Downarrow\{x->2\}$
(while $x>5$ do $x:=x-5$ od, $\{x->7\}) \Downarrow\{x->2\}$

## While Command


$(B, m) \Downarrow$ true $\quad(C, m) \Downarrow m^{\prime} \quad\left(\right.$ while $B$ do $C$ od, $\left.m^{\prime}\right) \Downarrow m^{\prime \prime}$
(while B do C od, m) $\downarrow \mathrm{m}^{\prime \prime}$

The rule assumes the loop terminates!

## While Command

$$
\begin{gathered}
\frac{(B, m) \Downarrow \text { false }}{(\text { while } B \text { do C od, m) } \Downarrow \mathrm{m}} \\
\frac{(B, m) \Downarrow \text { true }(C, m) \Downarrow m^{\prime} \quad\left(\text { while B do C od, } \mathrm{m}^{\prime}\right) \Downarrow \mathrm{m}^{\prime \prime}}{(\text { while } B \text { do C od, } \mathrm{m}) \Downarrow \mathrm{m}^{\prime \prime}}
\end{gathered}
$$

The rule assumes the loop terminates!
? ? ?
while $(x>0)$ do $x:=x+10 d,\{x->1\} \downarrow$ ? ? ?

## Interpretation Versus Compilation

- A compiler from language L 1 to language L 2 is a program that takes an L1 program and for each piece of code in L1 generates a piece of code in L2 of same meaning
- An interpreter of L1 in L2 is an L2 program that executes the meaning of a given L1 program
- Compiler would examine the body of a loop once; an interpreter would examine it every time the loop was executed


## Interpreter

- An Interpreter represents the operational semantics of a language L1 (source language) in the language of implementation L2 (target language)
- Built incrementally
- Start with literals
- Variables
- Primitive operations
- Evaluation of expressions
- Evaluation of commands/declarations


## Interpreter

- Takes abstract syntax trees as input
- In simple cases could be just strings
- One procedure for each syntactic category (nonterminal)
- eg one for expressions, another for commands
- If Natural semantics used, tells how to compute final value from code
- If Transition semantics used, tells how to compute next "state"
- To get final value, put in a loop


## Natural Semantics Interpreter Implementation

- Identifiers: $(k, m) \Downarrow m(k)$
- Numerals are values: $(\mathrm{N}, \mathrm{m}) \Downarrow \mathrm{N}$
- Conditionals: $\quad \frac{(B, m) \Downarrow \text { true }(C, m) \Downarrow m^{\prime}}{\left(\text { (if } B \text { then } C \text { else } C^{\prime} \text { fi, } m\right) \Downarrow m^{\prime}} \quad \frac{(B, m) \Downarrow \text { false } \quad\left(C^{\prime}, m\right) \Downarrow m^{\prime}}{\left.\text { (if } B \text { then } C \text { else } C^{\prime} \text { fi, } m\right) \Downarrow m^{\prime}}$

```
compute_exp (Var(v), m) = look_up v m
compute_exp (Int(n), _) = Num (n)
```

```
compute_com (IfExp(b,c1,c2), m) =
    if compute_exp (b,m) = Bool(true)
    then compute_com (c1,m)
    else compute_com (c2,m)
```


## Natural Semantics Interpreter Implementation

- Loop: $\frac{(B, m) \Downarrow \text { false }}{\text { (while } B \text { do } C \text { od, } m) \Downarrow m}$
$\frac{(B, m) \Downarrow \text { true }(C, m) \Downarrow m^{\prime}\left(\text { while } B \text { do } C \text { od, } m^{\prime}\right) \Downarrow m^{\prime \prime}}{(\text { while } B \text { do } C \text { od, } m) \Downarrow m^{\prime}}$
compute_com (While(b, c), m) =
if compute_exp (b,m) = Bool(false)
then m
else compute_com
(While(b, c), compute_com(c,m))
- May fail to terminate - exceed stack limits
- Returns no useful information then

