Review for the Final Exam

Stephen A. Edwards

Columbia University

Fall 2021

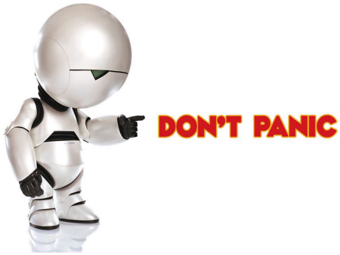


Table of Contents I

Table of Contents II

The Final

150 minutes

1:10 - 3:40, Friday, December 10th

Closed book, notes, Internet

One double-sided $8.5'' \times 11''$ sheet of notes of your own devising.

Comprehensive: Anything discussed in class is fair game

Little, if any, programming. This is not a test on OCaml

Details of OCaml/C/C++/Java syntax not required

Compiling a Simple Program

```
int gcd(int a, int b)
{
  while (a != b) {
    if (a > b) a -= b;
    else b -= a;
  }
  return a;
}
```

What the Compiler Sees

```
int gcd(int a, int b)
{
  while (a != b) {
    if (a > b) a -= b;
    else b -= a;
  }
  return a;
}
```

i	n	t	sp	g	c	d	(	i	n	t	sp	a	,	sp	i
n	t	sp	b	)	nl	{	nl	sp	sp	w	h	i	l	e	sp
(	a	sp	!	=	sp	b	)	sp	{	nl	sp	sp	sp	sp	i
f	sp	(	a	sp	>	sp	b	)	sp	a	sp	-	=	sp	b
;	nl	sp	sp	sp	sp	e	l	s	e	sp	b	sp	-	=	sp
a	;	nl	sp	sp	}	nl	sp	sp	r	e	t	u	r	n	sp
a	;	nl	}	nl											

Text file is a sequence of characters

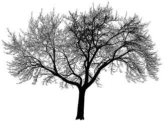
Lexical Analysis Gives Tokens

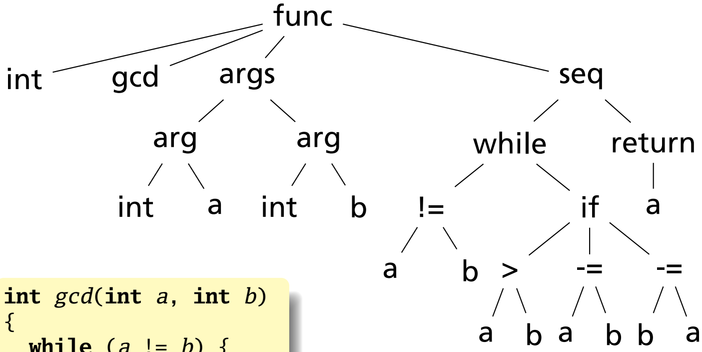
```
int gcd(int a, int b)
{
  while (a != b) {
    if (a > b) a -= b;
    else b -= a;
  }
  return a;
}
```

int	gcd	(	int	a	,	int	b	)	{	while	(	a
!=	b	)	{	if	(	a	>	b	)	a	-=	b
;	else	b	-=	a	;	}	return	a	;	}		

A stream of tokens. Whitespace, comments removed.

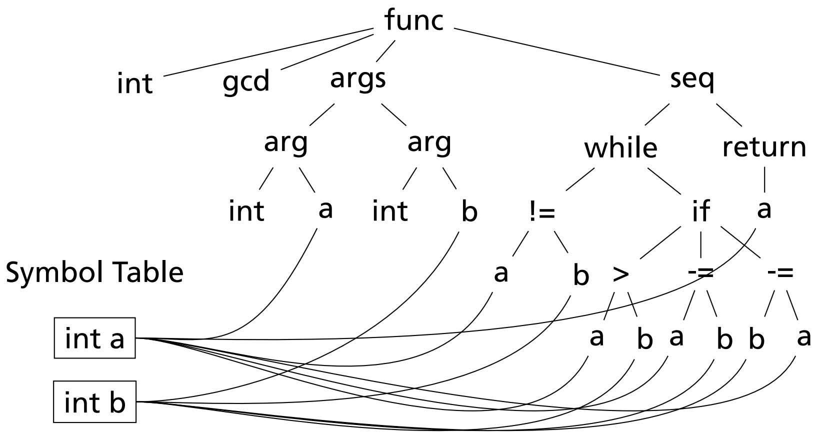
Parsing Gives an Abstract Syntax Tree






```
int gcd(int a, int b)
{
  while (a != b) {
    if (a > b) a -= b;
    else b -= a;
  }
  return a;
}
```

Semantic Analysis Resolves Symbols and Checks Types



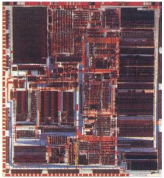
Translation into 3-Address Code

```
L0: sne   $1,  a, b
    seq   $0, $1, 0
    btrue $0, L1      # while (a != b)
    sl    $3,  b, a
    seq   $2, $3, 0
    btrue $2, L4      # if (a < b)
    sub   a,   a, b # a -= b
    jmp   L5
L4: sub   b,   b, a # b -= a
L5: jmp   L0
L1: ret    a
```

```
int gcd(int a, int b)
{
  while (a != b) {
    if (a > b) a -= b;
    else b -= a;
  }
  return a;
}
```

Idealized assembly language w/ infinite registers

Generation of 80386 Assembly



```
gcd:  pushl %ebp            # Save BP
      movl  %esp,%ebp
      movl  8(%ebp),%eax    # Load a from stack
      movl  12(%ebp),%edx   # Load b from stack
.L8:  cmpl  %edx,%eax
      je    .L3             # while (a != b)
      jle   .L5             # if (a < b)
      subl  %edx,%eax       # a -= b
      jmp   .L8
.L5:  subl  %eax,%edx       # b -= a
      jmp   .L8
.L3:  leave                 # Restore SP, BP
      ret
```

Describing Tokens

Alphabet: A finite set of symbols

Examples: $\{$ 0, 1 $\}$, $\{$ A, B, C, . . . , Z $\}$, ASCII, Unicode

String: A finite sequence of symbols from an alphabet

Examples: ϵ (the empty string), Stephen, $\alpha\beta\gamma$

Language: A set of strings over an alphabet

Examples: $\emptyset$ (the empty language), $\{$ 1, 11, 111, 1111 $\}$, all English words, strings that start with a letter followed by any sequence of letters and digits

Operations on Languages

Let $L = \{ \epsilon, \text{wo} \}$, $M = \{ \text{man}, \text{men} \}$

Concatenation: Strings from one followed by the other

$LM = \{ \text{man}, \text{men}, \text{woman}, \text{women} \}$

Union: All strings from each language

$L \cup M = \{\epsilon, \text{wo}, \text{man}, \text{men} \}$

Kleene Closure: Zero or more concatenations

$M^* = \{\epsilon\} \cup M \cup MM \cup MMM \cdots =$
$\{\epsilon$, man, men, manman, manmen, menman, menmen,
manmanman, manmanmen, manmenman, $\ldots\}$

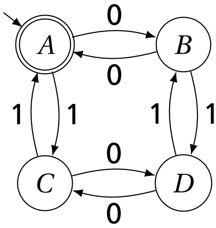
Regular Expressions over an Alphabet Σ

A standard way to express languages for tokens.

1. ϵ is a regular expression that denotes $\{\epsilon\}$
2. If $a \in \Sigma$, a is an RE that denotes $\{a\}$
3. If r and s denote languages $L(r)$ and $L(s)$,
   - $(r) \mid (s)$ denotes $L(r) \cup L(s)$
   - $(r)(s)$ denotes $\{tu : t \in L(r), u \in L(s)\}$
   - $(r)^*$ denotes $\cup_{i=0}^{\infty} L^i$ ($L^0 = \{\epsilon\}$ and $L^i = LL^{i-1}$)

Nondeterministic Finite Automata

"All strings containing an even number of 0's and 1's"



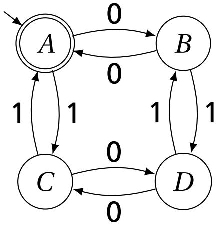
1. Set of states
   $S : \{ A \quad B \quad C \quad D \}$
2. Set of input symbols $\Sigma : \{0, 1\}$
3. Transition function $\sigma : S \times \Sigma_\epsilon \to 2^S$

state	ϵ	0	1
A	$\emptyset$	$\{B\}$	$\{C\}$
B	$\emptyset$	$\{A\}$	$\{D\}$
C	$\emptyset$	$\{D\}$	$\{A\}$
D	$\emptyset$	$\{C\}$	$\{B\}$

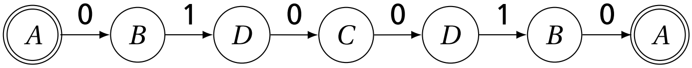
4. Start state $s_0 : A$
5. Set of accepting states
   $F : \{ A \}$

The Language induced by an NFA

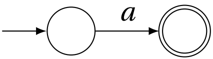
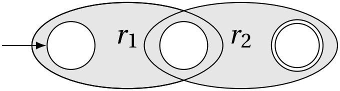
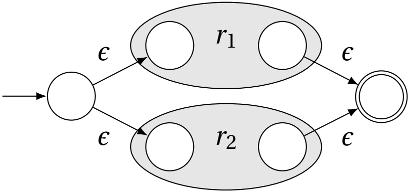
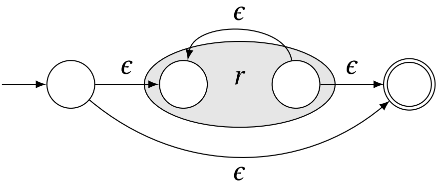
An NFA accepts an input string x iff there is a path from the start state to an accepting state that “spells out” x.



Show that the string “010010” is accepted.

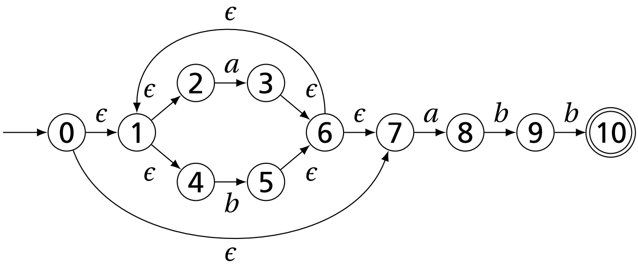


Translating REs into NFAs

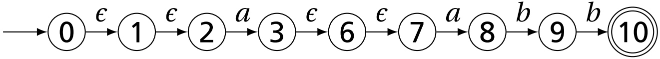
a		Symbol
$r_1 r_2$		Sequence
$r_1 \mid r_2$		Choice
$(r)^*$		Kleene Closure

Translating REs into NFAs

Example: Translate $(a \mid b)^* abb$ into an NFA. Answer:



Show that the string "$aabb$" is accepted. Answer:



Simulating NFAs

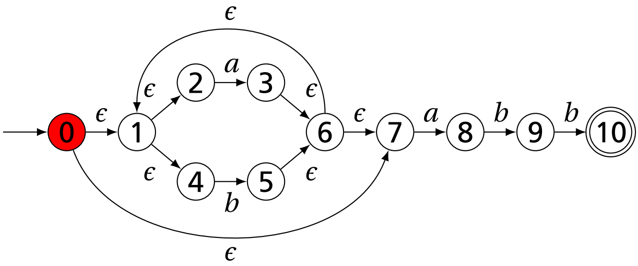
Problem: you must follow the “right” arcs to show that a string is accepted. How do you know which arc is right?

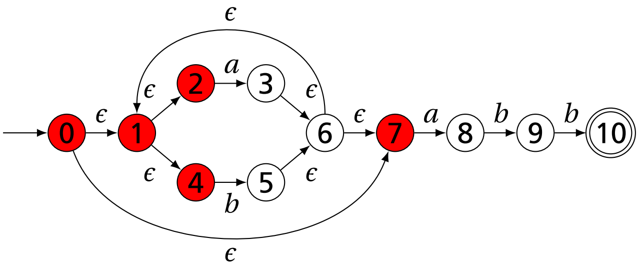
Solution: follow them all and sort it out later.

“Two-stack” NFA simulation algorithm:

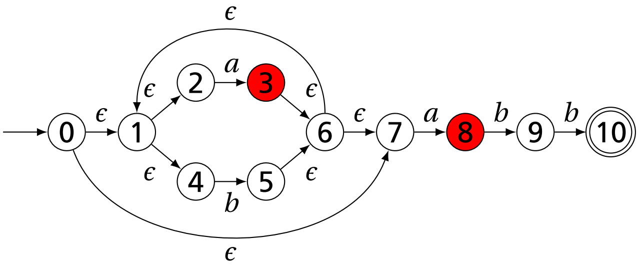
1. Initial states: the ϵ-closure of the start state
2. For each character c,
   - New states: follow all transitions labeled c
   - Form the ϵ-closure of the current states
3. Accept if any final state is accepting

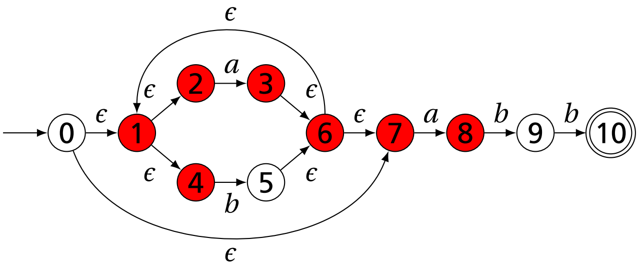
Simulating an NFA: $\cdot aabb$, Start



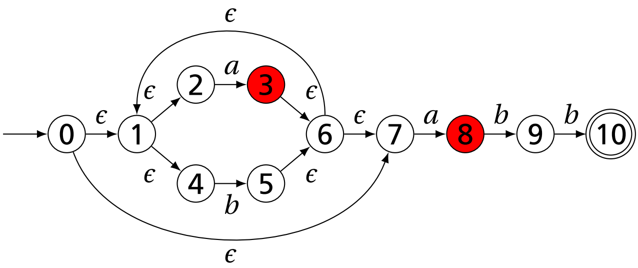


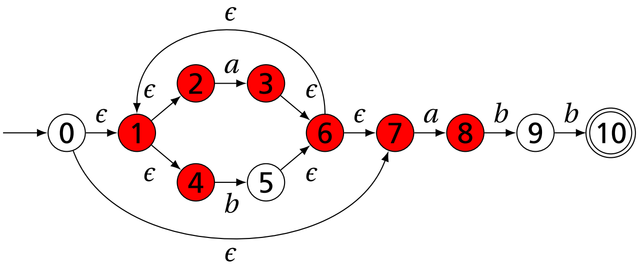
Simulating an NFA: $a{\cdot}abb$



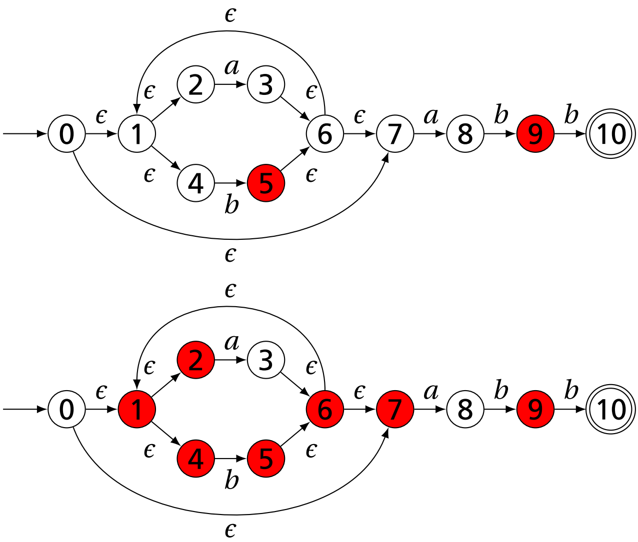


Simulating an NFA: $aa{\cdot}bb$

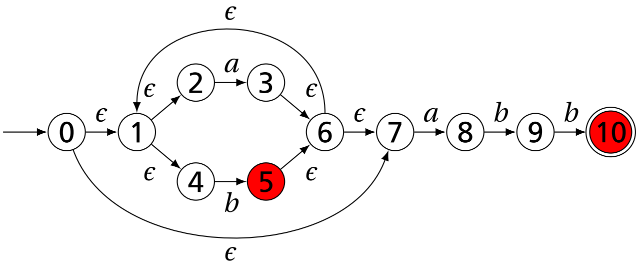


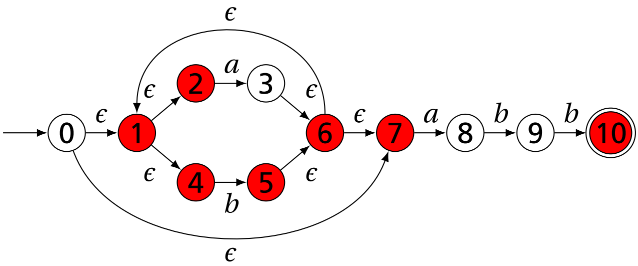


Simulating an NFA: $aab{\cdot}b$



Simulating an NFA: $aabb\cdot$, Done





Deterministic Finite Automata

Restricted form of NFAs:

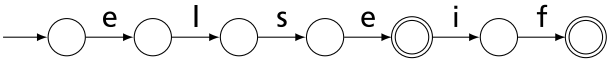
- No state has a transition on ϵ
- For each state s and symbol a, there is at most one edge labeled a leaving s.

Differs subtly from the definition used in COMS W3261 (Sipser, *Introduction to the Theory of Computation*)

Very easy to check acceptance: simulate by maintaining current state. Accept if you end up on an accepting state. Reject if you end on a non-accepting state or if there is no transition from the current state for the next symbol.

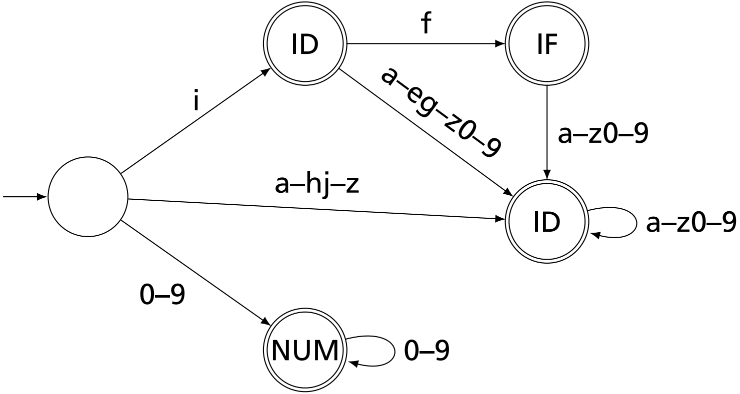
Deterministic Finite Automata

```
{
   type token = ELSE | ELSEIF
}

rule token =
  parse "else"   { ELSE }
      | "elseif" { ELSEIF }
```



Deterministic Finite Automata

```
{ type token = IF | ID of string | NUM of string }

rule token =
  parse "if"                                        { IF }
      | ['a'-'z'] ['a'-'z' '0'-'9']* as lit { ID(lit) }
      | ['0'-'9']+                         as num { NUM(num) }
```



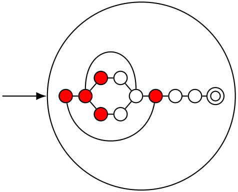
Building a DFA from an NFA

Subset construction algorithm

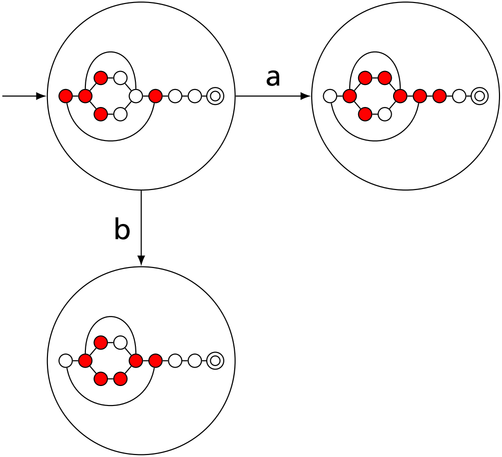
Simulate the NFA for all possible inputs and track the states that appear.

Each unique state during simulation becomes a state in the DFA.

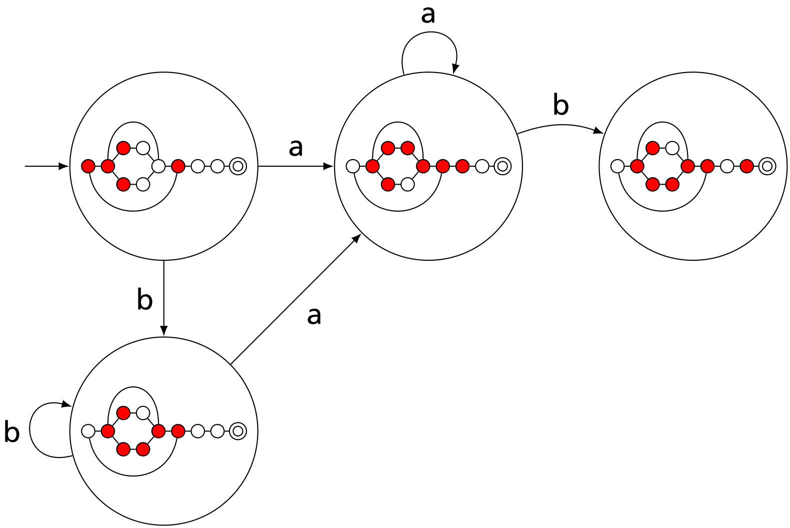
Subset construction for $(a \mid b)^* abb$



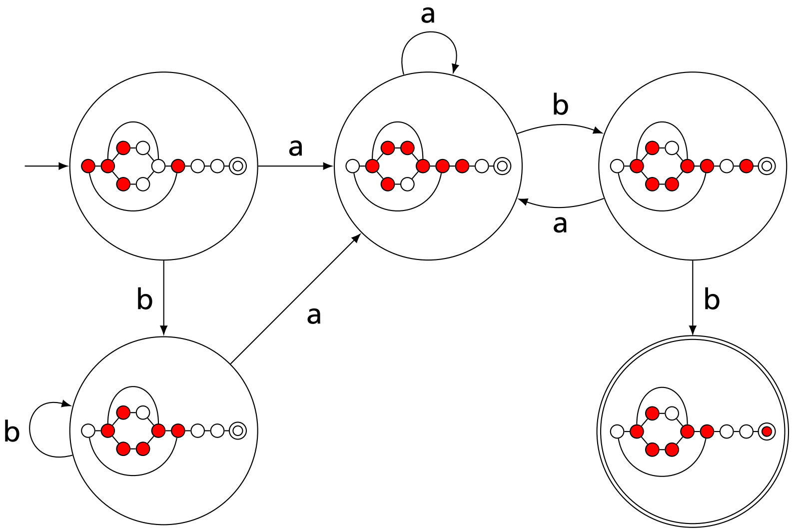
Subset construction for $(a\,|\,b)^*abb$



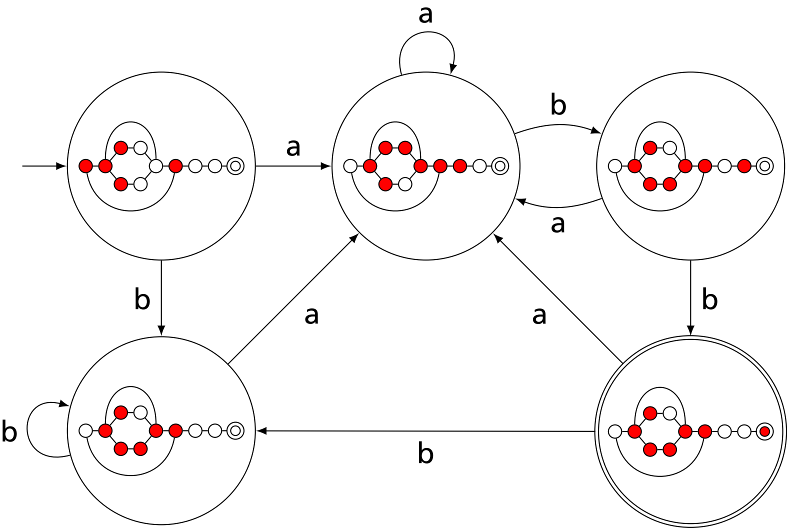
Subset construction for $(a \mid b)^* abb$



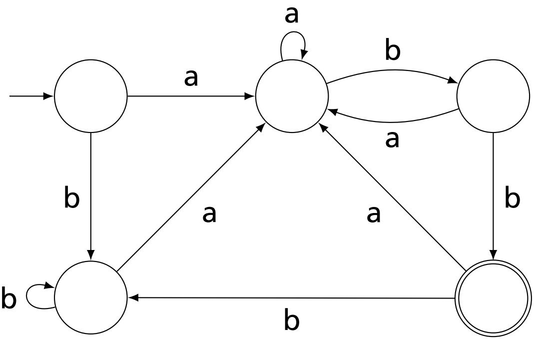
Subset construction for $(a \mid b)^* abb$



Subset construction for $(a \mid b)^* abb$



Result of subset construction for $(a \mid b)^* abb$



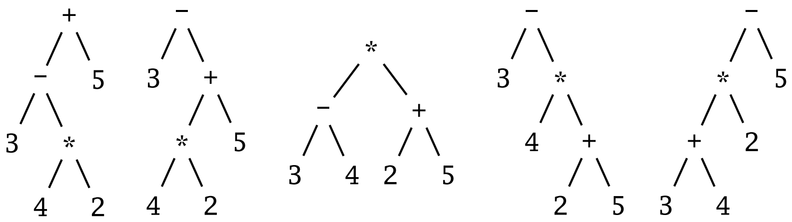
Ambiguous Arithmetic

Ambiguity can be a problem in expressions. Consider parsing

```
3 - 4 * 2 + 5
```

with the grammar

$$e \rightarrow e + e \mid e - e \mid e * e \mid e \mathbin{/} e \mid N$$



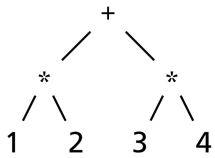
Operator Precedence

Defines how "sticky" an operator is.

`1 * 2 + 3 * 4`

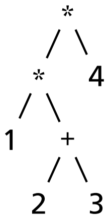
`*` at higher precedence than `+`:

`(1 * 2) + (3 * 4)`



`+` at higher precedence than `*`:

`1 * (2 + 3) * 4`



Associativity

Whether to evaluate left-to-right or right-to-left

Most operators are left-associative

```
1 - 2 - 3 - 4
```

```
          -                        -
         / \                      / \
        -   4                    1   -
       / \                          / \
      -   3                        2   -
     / \                              / \
    1   2                            3   4
```

$((1-2)-3)-4$ $\qquad$ $1-(2-(3-4))$

left associative $\qquad$ right associative

Fixing Ambiguous Grammars

A grammar specification:

```
expr :
    expr PLUS expr
  | expr MINUS expr
  | expr TIMES expr
  | expr DIVIDE expr
  | NUMBER
```

Ambiguous: no precedence or associativity.

Ocamlyacc's complaint: "16 shift/reduce conflicts."

Assigning Precedence Levels

Split into multiple rules, one per level

```
expr : expr PLUS expr
     | expr MINUS expr
     | term

term : term TIMES term
     | term DIVIDE term
     | atom

atom  : NUMBER
```

Still ambiguous: associativity not defined

Ocamlyacc's complaint: "8 shift/reduce conflicts."

Assigning Associativity

Make one side the next level of precedence

```
expr : expr PLUS term
     | expr MINUS term
     | term

term : term TIMES atom
     | term DIVIDE atom
     | atom

atom  : NUMBER
```

This is left-associative.

No shift/reduce conflicts.

Rightmost Derivation of $\mathbf{Id} * \mathbf{Id} + \mathbf{Id}$

$1: e \rightarrow t + e$

$2: e \rightarrow t$

$3: t \rightarrow \mathbf{Id} * t$

$4: t \rightarrow \mathbf{Id}$

$$e$$
$$\boxed{t + e}$$
$$t + \boxed{t}$$
$$t + \boxed{\mathbf{Id}}$$
$$\boxed{\mathbf{Id} * t} + \mathbf{Id}$$
$$\mathbf{Id} * \boxed{\mathbf{Id}} + \mathbf{Id}$$

At each step, expand the *rightmost* nonterminal.

nonterminal

"handle": The right side of a production

Fun and interesting fact: there is exactly one rightmost expansion if the grammar is unambigious.

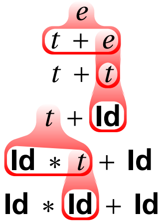
Rightmost Derivation: What to Expand

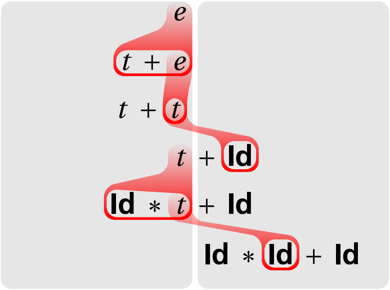
$1: e \rightarrow t + e$

$2: e \rightarrow t$

$3: t \rightarrow \textbf{Id} \; * \; t$

$4: t \rightarrow \textbf{Id}$





Expand here ↑ Terminals only

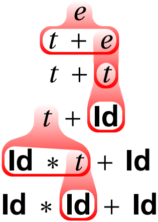
Reverse Rightmost Derivation

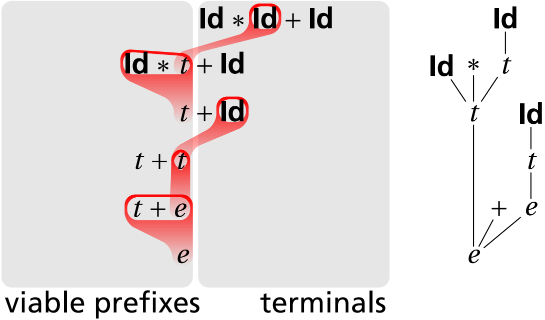
$1: e \rightarrow t + e$

$2: e \rightarrow t$

$3: t \rightarrow \textbf{Id} * t$

$4: t \rightarrow \textbf{Id}$





Shift/Reduce Parsing Using an Oracle

$1: e \rightarrow t + e$

$2: e \rightarrow t$

$3: t \rightarrow \textbf{Id} * t$

$4: t \rightarrow \textbf{Id}$

$$e$$
$$t + e$$
$$t + t$$
$$t + \textbf{Id}$$
$$\textbf{Id} * t + \textbf{Id}$$
$$\textbf{Id} * \textbf{Id} + \textbf{Id}$$

stack	input	
	$\textbf{Id} * \textbf{Id} + \textbf{Id}$	shift
$\textbf{Id}$	$* \textbf{Id} + \textbf{Id}$	shift
$\textbf{Id} *$	$\textbf{Id} + \textbf{Id}$	shift
$\textbf{Id} * \textbf{Id}$	$+ \textbf{Id}$	reduce 4
$\textbf{Id} * t$	$+ \textbf{Id}$	reduce 3
t	$+ \textbf{Id}$	shift
$t +$	$\textbf{Id}$	shift
$t + \textbf{Id}$		reduce 4
$t + t$		reduce 2
$t + e$		reduce 1
e		accept

Handle Hunting

Right Sentential Form: any step in a rightmost derivation

Handle: in a sentential form, a RHS of a rule that, when rewritten, yields the previous step in a rightmost derivation.

The big question in shift/reduce parsing:

When is there a handle on the top of the stack?

Enumerate all the right-sentential forms and pattern-match against them? *Usually infinite in number, but let's try anyway.*

The Handle-Identifying Automaton

Magical result, due to Knuth: *An automaton suffices to locate a handle in a right-sentential form.*

$\mathbf{Id} * \mathbf{Id} * \cdots * \boxed{\mathbf{Id} * t} \cdots$

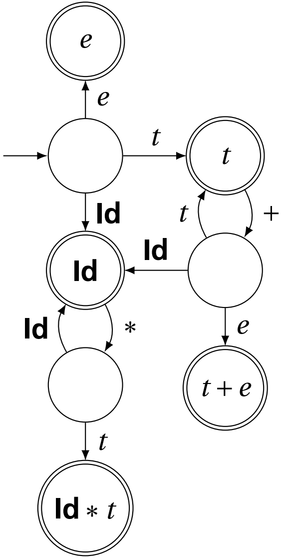
$\mathbf{Id} * \mathbf{Id} * \cdots * \boxed{\mathbf{Id}} \cdots$

$t + t + \cdots + \boxed{t + e}$

$t + t + \cdots + t + \boxed{\mathbf{Id}}$

$t + t + \cdots + t + \mathbf{Id} * \mathbf{Id} * \cdots * \boxed{\mathbf{Id} * t}$

$t + t + \cdots + \boxed{t}$



Building the Initial State of the LR(0) Automaton

$$\boxed{e' \rightarrow \cdot e}$$

$1 : e \rightarrow t + e$
$2 : e \rightarrow t$
$3 : t \rightarrow \textbf{Id} * t$
$4 : t \rightarrow \textbf{Id}$

Key idea: automata identify viable prefixes of right sentential forms. Each state is an equivalence class of possible places in productions.

At the beginning, any viable prefix must be at the beginning of a string expanded from e. We write this condition "$e' \rightarrow \cdot e$"

Building the Initial State of the LR(0) Automaton

$$1: e \rightarrow t + e$$
$$2: e \rightarrow t$$
$$3: t \rightarrow \textbf{Id} * t$$
$$4: t \rightarrow \textbf{Id}$$

$$\boxed{\begin{array}{l} e' \rightarrow \cdot e \\ e \rightarrow \cdot t + e \\ e \rightarrow \cdot t \end{array}}$$

Key idea: automata identify viable prefixes of right sentential forms. Each state is an equivalence class of possible places in productions.

At the beginning, any viable prefix must be at the beginning of a string expanded from e. We write this condition "$e' \rightarrow \cdot e$"

There are two choices for what an e may expand to: $t + e$ and t. So when $e' \rightarrow \cdot e$, $e \rightarrow \cdot t + e$ and $e \rightarrow \cdot t$ are also true, i.e., it must start with a string expanded from t.

Building the Initial State of the LR(0) Automaton

$$\boxed{\begin{array}{l} e' \rightarrow \cdot e \\ e \rightarrow \cdot t + e \\ e \rightarrow \cdot t \\ t \rightarrow \cdot \textbf{Id} * t \\ t \rightarrow \cdot \textbf{Id} \end{array}}$$

$1: e \rightarrow t + e$
$2: e \rightarrow t$
$3: t \rightarrow \textbf{Id} \; * t$
$4: t \rightarrow \textbf{Id}$

Key idea: automata identify viable prefixes of right sentential forms. Each state is an equivalence class of possible places in productions.

At the beginning, any viable prefix must be at the beginning of a string expanded from e. We write this condition "$e' \rightarrow \cdot e$"

There are two choices for what an e may expand to: $t + e$ and t. So when $e' \rightarrow \cdot e$, $e \rightarrow \cdot t + e$ and $e \rightarrow \cdot t$ are also true, i.e., it must start with a string expanded from t.

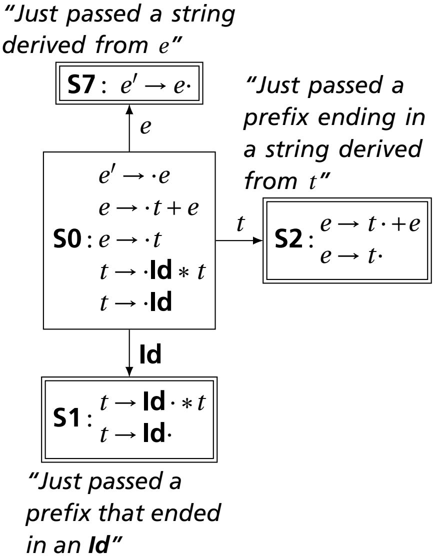
Similarly, t must be either $\textbf{Id} * t$ or $\textbf{Id}$, so $t \rightarrow \cdot \textbf{Id} * t$ and $t \rightarrow \cdot \textbf{Id}$.

Building the LR(0) Automaton

$$\mathbf{S0}: \begin{array}{l} e' \rightarrow \cdot e \\ e \rightarrow \cdot t + e \\ e \rightarrow \cdot t \\ t \rightarrow \cdot \mathbf{Id} * t \\ t \rightarrow \cdot \mathbf{Id} \end{array}$$

The first state suggests a viable prefix can start as any string derived from e, any string derived from t, or **Id**.

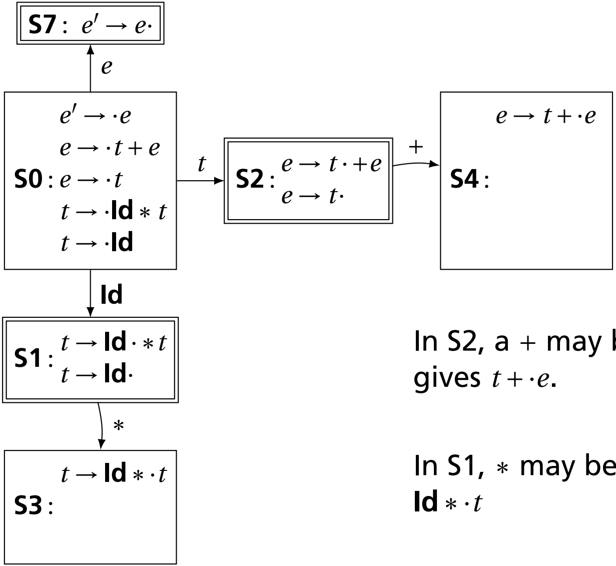
Building the LR(0) Automaton



The first state suggests a viable prefix can start as any string derived from e, any string derived from t, or **Id**.

The items for these three states come from advancing the $\cdot$ across each thing, then performing the closure operation (vacuous here).

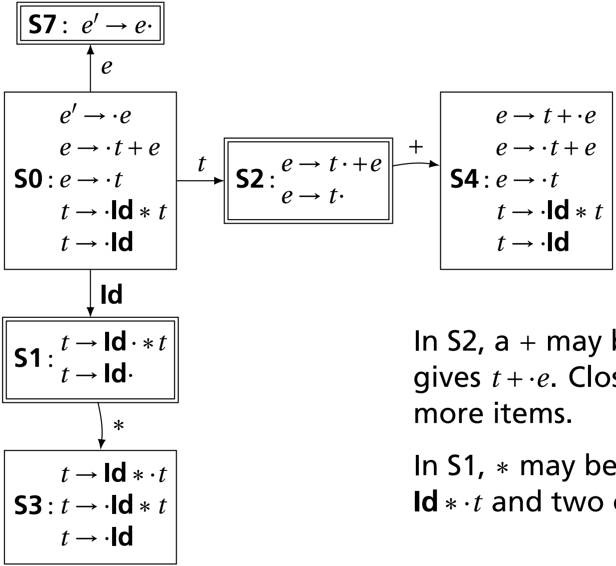
Building the LR(0) Automaton



In S2, a $+$ may be next. This gives $t + \cdot e$.

In S1, $*$ may be next, giving $\mathbf{Id} * \cdot t$

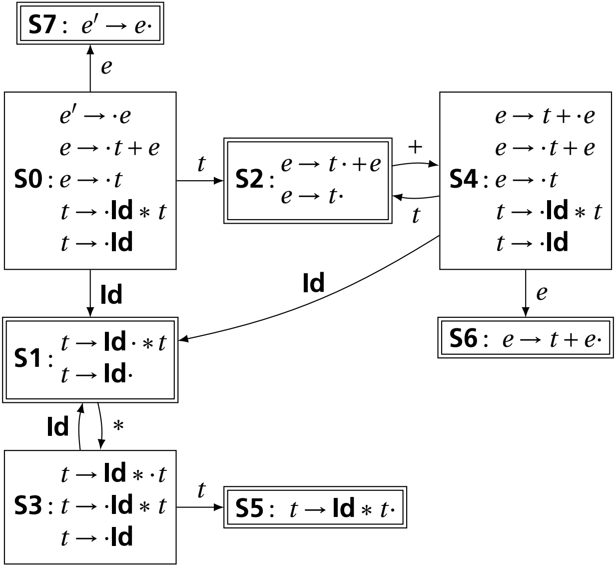
Building the LR(0) Automaton



In S2, a $+$ may be next. This gives $t + \cdot e$. Closure adds 4 more items.

In S1, $*$ may be next, giving $\mathbf{Id} * \cdot t$ and two others.

Building the LR(0) Automaton



The first function

If you can derive a string that starts with terminal t from a sequence of terminals and nonterminals α, then $t \in \text{first}(\alpha)$.

1. If X is a terminal, $\text{first}(X) = \{X\}$.
2. If $X \to \epsilon$, then add ϵ to $\text{first}(X)$.
3. If $X \to Y_1 \cdots Y_k$ and $\epsilon \in \text{first}(Y_1)$, $\epsilon \in \text{first}(Y_2)$, . . . , and $\epsilon \in \text{first}(Y_{i-1})$ for $i = 1, \ldots, k$ for some k,
   add $\text{first}(Y_i) - \{\epsilon\}$ to $\text{first}(X)$
   X starts with anything that appears after skipping empty strings. Usually just $\text{first}(Y_1) \in \text{first}(X)$
4. If $X \to Y_1 \cdots Y_K$ and $\epsilon \in \text{first}(Y_1)$, $\epsilon \in \text{first}(Y_2)$, . . . , and $\epsilon \in \text{first}(Y_k)$, add ϵ to $\text{first}(X)$
   If all of X can be empty, X can be empty

$1: e \to t + e$
$2: e \to t$
$3: t \to \textbf{Id} * t$
$4: t \to \textbf{Id}$

$\text{first}(\textbf{Id}) = \{\textbf{Id}\}$

$\text{first}(t) = \{\textbf{Id}\}$ because $t \to \textbf{Id} * t$ and $t \to \textbf{Id}$

$\text{first}(e) = \{\textbf{Id}\}$ because $e \to t + e$, $e \to t$, and $\text{first}(t) = \{\textbf{Id}\}$.

First and ϵ

$\epsilon \in \text{first}(\alpha)$ means α can derive the empty string.

1. If X is a terminal, $\text{first}(X) = \{X\}$.
2. If $X \to \epsilon$, then add ϵ to $\text{first}(X)$.
3. If $X \to Y_1 \cdots Y_k$ and
   $\epsilon \in \text{first}(Y_1)$, $\epsilon \in \text{first}(Y_2)$, ..., and $\epsilon \in \text{first}(Y_{i-1})$
   for $i = 1, \ldots, k$ for some k,
   add $\text{first}(Y_i) - \{\epsilon\}$ to $\text{first}(X)$
4. If $X \to Y_1 \cdots Y_K$ and
   $\epsilon \in \text{first}(Y_1)$, $\epsilon \in \text{first}(Y_2)$, ..., and $\epsilon \in \text{first}(Y_k)$,
   add ϵ to $\text{first}(X)$

$$
\begin{aligned}
X &\to Y Z a \\
Y &\to \\
Y &\to b \\
Z &\to c \\
Z &\to W \\
W &\to \\
W &\to d
\end{aligned}
$$

$$
\begin{aligned}
&\text{first}(b) = \{b\} \quad \text{first}(c) = \{c\} \quad \text{first}(d) = \{d\} && (1) \\
&\text{first}(W) = \{\epsilon\} \cup \text{first}(d) = \{\epsilon, d\} && (2,3) \\
&\text{first}(Z) = \text{first}(c) \cup (\text{first}(W) - \{\epsilon\}) \cup \{\epsilon\} = \{\epsilon, c, d\} && (3,3,4) \\
&\text{first}(Y) = \{\epsilon\} \cup \{b\} = \{\epsilon, b\} && (2,3) \\
&\text{first}(X) = (\text{first}(Y) - \{\epsilon\}) \cup (\text{first}(Z) - \{\epsilon\}) \cup \\
&\qquad\qquad \text{first}(a) = \{a, b, c, d\} && (3,3,3)
\end{aligned}
$$

The follow function

If t is a terminal, A is a nonterminal, and $\cdots At \cdots$ can be derived, then $t \in \text{follow}(A)$.

1. Add \$ ("end-of-input") to $\text{follow}(S)$ (start symbol).
   End-of-input comes after the start symbol
2. For each prod. $\rightarrow \cdots A\alpha$, add $\text{first}(\alpha) - \{\epsilon\}$ to $\text{follow}(A)$.
   A is followed by the first thing after it
3. For each prod. $A \rightarrow \cdots B$ or $A \rightarrow \cdots B\alpha$ where $\epsilon \in \text{first}(\alpha)$, then add everything in $\text{follow}(A)$ to $\text{follow}(B)$.
   If B appears at the end of a production, it can be followed by whatever follows that production

---

$1: e \rightarrow t + e$

$2: e \rightarrow t$

$3: t \rightarrow \textbf{Id} * t$

$4: t \rightarrow \textbf{Id}$

$\text{first}(t) = \{\textbf{Id}\}$

$\text{first}(e) = \{\textbf{Id}\}$

$\text{follow}(e) = \{\$\}$

$\text{follow}(t) = \{ \quad \}$

1. Because e is the start symbol

The follow function

If t is a terminal, A is a nonterminal, and $\cdots At\cdots$ can be derived, then $t \in \text{follow}(A)$.

1. Add \$ ("end-of-input") to follow(S) (start symbol).
   End-of-input comes after the start symbol
2. For each prod. $\rightarrow \cdots A\alpha$, add first$(\alpha) - \{\epsilon\}$ to follow(A).
   A is followed by the first thing after it
3. For each prod. $A \rightarrow \cdots B$ or $A \rightarrow \cdots B\alpha$ where $\epsilon \in \text{first}(\alpha)$, then add everything in follow(A) to follow(B).
   If B appears at the end of a production, it can be followed by whatever follows that production

---

$1: e \rightarrow t + e$

$2: e \rightarrow t$

$3: t \rightarrow \textbf{Id} * t$

$4: t \rightarrow \textbf{Id}$

first$(t) = \{\textbf{Id}\}$

first$(e) = \{\textbf{Id}\}$

follow$(e) = \{\$\}$

follow$(t) = \{+ \quad\}$

2. Because $e \rightarrow \underline{t} + e$ *and* first$(+) = \{+\}$

The follow function

If t is a terminal, A is a nonterminal, and $\cdots At \cdots$ can be derived, then $t \in \text{follow}(A)$.

1. Add \$ ("end-of-input") to $\text{follow}(S)$ (start symbol).
   End-of-input comes after the start symbol
2. For each prod. $\rightarrow \cdots A\alpha$, add $\text{first}(\alpha) - \{\epsilon\}$ to $\text{follow}(A)$.
   A is followed by the first thing after it
3. For each prod. $A \rightarrow \cdots B$ or $A \rightarrow \cdots B\alpha$ where $\epsilon \in \text{first}(\alpha)$, then add everything in $\text{follow}(A)$ to $\text{follow}(B)$.
   If B appears at the end of a production, it can be followed by whatever follows that production

---

$1: e \rightarrow t + e$

$2: e \rightarrow t$

$3: t \rightarrow \textbf{Id} * t$

$4: t \rightarrow \textbf{Id}$

$\text{first}(t) = \{\textbf{Id}\}$

$\text{first}(e) = \{\textbf{Id}\}$

$\text{follow}(e) = \{\$\}$

$\text{follow}(t) = \{+, \$\}$

3. Because $e \rightarrow \underline{t}$ *and* $\$ \in \text{follow}(e)$

The follow function

If t is a terminal, A is a nonterminal, and $\cdots At \cdots$ can be derived, then $t \in \text{follow}(A)$.

1. Add $ ("end-of-input") to follow(S) (start symbol).
   End-of-input comes after the start symbol
2. For each prod. $\rightarrow \cdots A\alpha$, add first$(\alpha) - \{\epsilon\}$ to follow(A).
   A is followed by the first thing after it
3. For each prod. $A \rightarrow \cdots B$ or $A \rightarrow \cdots B\alpha$ where $\epsilon \in \text{first}(\alpha)$, then add everything in follow(A) to follow(B).
   If B appears at the end of a production, it can be followed by whatever follows that production

---

$1: e \rightarrow t + e$

$2: e \rightarrow t$

$3: t \rightarrow \textbf{Id} * t$

$4: t \rightarrow \textbf{Id}$

first$(t) = \{\textbf{Id}\}$

first$(e) = \{\textbf{Id}\}$

follow$(e) = \{\$\}$

follow$(t) = \{+, \$\}$

Fixed-point reached: applying any rule does not change any set

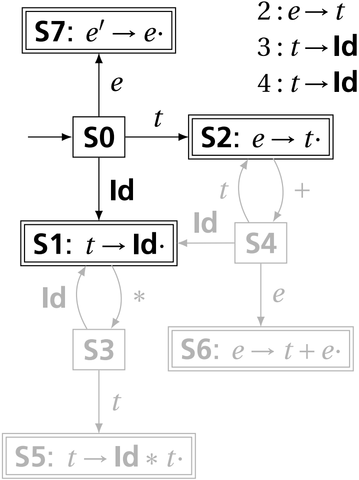
Converting the LR(0) Automaton to an SLR Parsing Table

$1: e \rightarrow t + e$

$2: e \rightarrow t$

$3: t \rightarrow \mathbf{Id} * t$

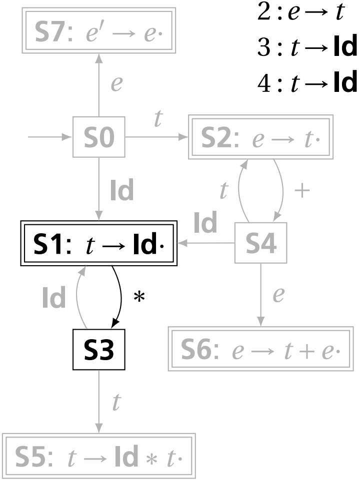
$4: t \rightarrow \mathbf{Id}$



State	Action				Goto	
	Id	$+$	$*$	\$	e	t
0	s1				7	2

From S0, shift an **Id** and go to S1; or cross a t and go to S2; or cross an e and go to S7.

Converting the LR(0) Automaton to an SLR Parsing Table



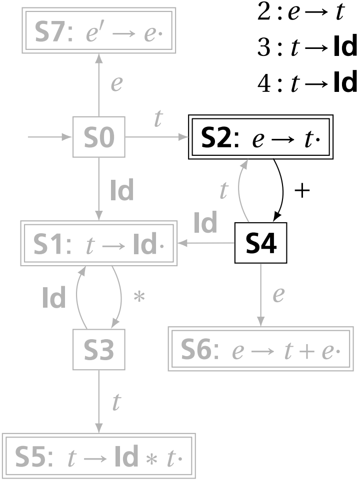
$1: e \to t + e$
$2: e \to t$
$3: t \to \textbf{Id} * t$
$4: t \to \textbf{Id}$

State	Action				Goto	
	Id	$+$	$*$	$	e	t
0	s1				7	2
1		r4	s3	r4		

From S1, shift a $*$ and go to S3; or, if the next input could follow a t, reduce by rule 4. According to rule 1, $+$ could follow t; from rule 2, $ could.

Converting the LR(0) Automaton to an SLR Parsing Table

$$1 : e \rightarrow t + e$$
$$2 : e \rightarrow t$$
$$3 : t \rightarrow \textbf{Id} * t$$
$$4 : t \rightarrow \textbf{Id}$$



State	Action				Goto	
	Id	+	*	$	e	t
0	s1				7	2
1		r4	s3	r4		
2		s4		r2		

From S2, shift a + and go to S4; or, if the next input could follow an e (only the end-of-input $), reduce by rule 2.

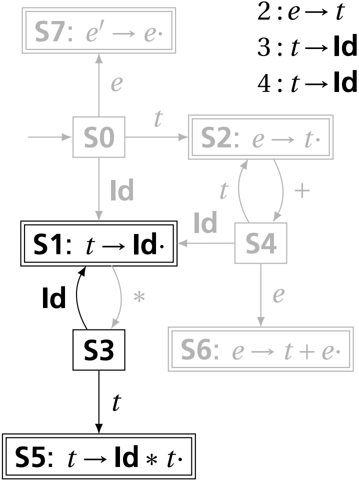
Converting the LR(0) Automaton to an SLR Parsing Table

$1: e \to t + e$

$2: e \to t$

$3: t \to \textbf{Id} * t$

$4: t \to \textbf{Id}$

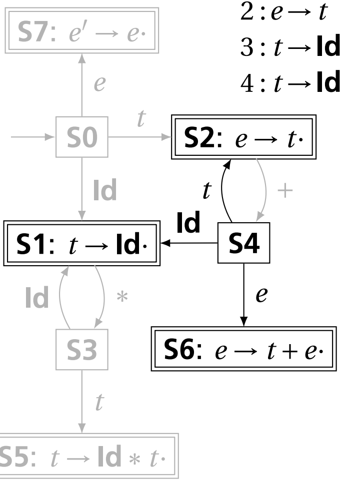


State	Action				Goto	
	Id	$+$	$*$	\$	e	t
0	s1				7	2
1		r4	s3	r4		
2		s4		r2		
3	s1					5

From S3, shift an **Id** and go to S1; or cross a t and go to S5.

Converting the LR(0) Automaton to an SLR Parsing Table

$1: e \rightarrow t + e$
$2: e \rightarrow t$
$3: t \rightarrow \textbf{Id} * t$
$4: t \rightarrow \textbf{Id}$



State	Action				Goto	
	Id	$+$	$*$	**\$**	e	t
0	s1				7	2
1		r4	s3	r4		
2		s4		r2		
3	s1					5
4	s1				6	2

From S4, shift an **Id** and go to S1; or cross an e or a t.

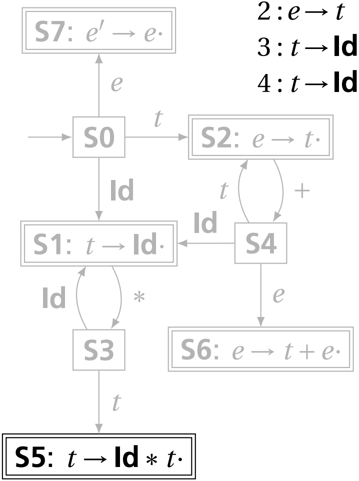
Converting the LR(0) Automaton to an SLR Parsing Table

$1: e \to t + e$

$2: e \to t$

$3: t \to \textbf{Id} * t$

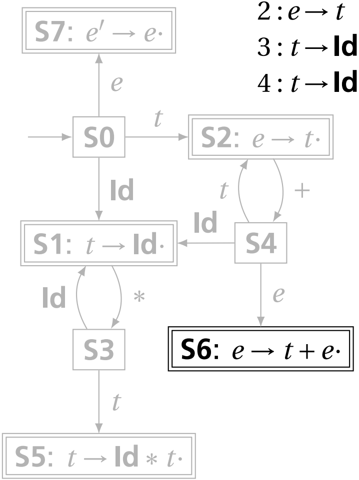
$4: t \to \textbf{Id}$



State	Action				Goto	
	Id	$+$	$*$	\$	e	t
0	s1				7	2
1		r4	s3	r4		
2		s4		r2		
3	s1					5
4	s1				6	2
5		r3		r3		

From S5, reduce using rule 3 if the next symbol could follow a t (again, $+$ and \$).

Converting the LR(0) Automaton to an SLR Parsing Table



$1: e \rightarrow t + e$
$2: e \rightarrow t$
$3: t \rightarrow \mathbf{Id} * t$
$4: t \rightarrow \mathbf{Id}$

State	Action				Goto	
	Id	$+$	$*$	$	e	t
0	s1				7	2
1		r4	s3	r4		
2		s4		r2		
3	s1					5
4	s1				6	2
5		r3		r3		
6				r1		

From S6, reduce using rule 1 if the next symbol could follow an e ($ only).

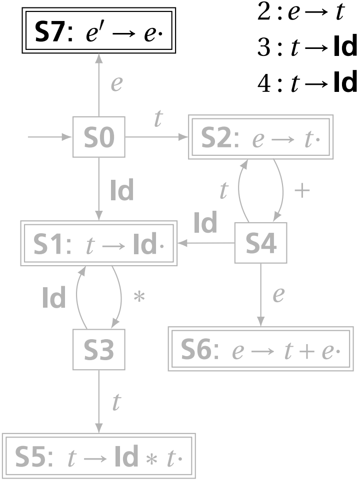
Converting the LR(0) Automaton to an SLR Parsing Table

$1: e \rightarrow t + e$

$2: e \rightarrow t$

$3: t \rightarrow \textbf{Id} * t$

$4: t \rightarrow \textbf{Id}$



State	Action				Goto	
	Id	$+$	$*$	$	e	t
0	s1				7	2
1		r4	s3	r4		
2		s4		r2		
3	s1					5
4	s1				6	2
5		r3		r3		
6				r1		
7				✓		

If, in S7, we just crossed an e, accept if we are at the end of the input.

Shift/Reduce Parsing with an SLR Table

$1: e \rightarrow t + e$
$2: e \rightarrow t$
$3: t \rightarrow \mathbf{Id} * t$
$4: t \rightarrow \mathbf{Id}$

State	Action				Goto	
	Id	$+$	$*$	$	e	t
0	s1				7	2
1		r4	s3	r4		
2		s4		r2		
3	s1					5
4	s1				6	2
5		r3		r3		
6				r1		
7				✓		

Stack	Input	Action
0	$\mathbf{Id} * \mathbf{Id} + \mathbf{Id}\,\$$	Shift, goto 1

Look at the state on top of the stack and the next input token.

Find the action (shift, reduce, or error) in the table.

In this case, shift the token onto the stack and mark it with state 1.

Shift/Reduce Parsing with an SLR Table

$1: e \rightarrow t + e$
$2: e \rightarrow t$
$3: t \rightarrow \textbf{Id} * t$
$4: t \rightarrow \textbf{Id}$

State	Action				Goto	
	Id	$+$	$*$	\$	e	t
0	s1				7	2
1		r4	s3	r4		
2		s4		r2		
3	s1					5
4	s1				6	2
5		r3		r3		
6				r1		
7				✓		

Stack	Input	Action
0	**Id** $*$ **Id** $+$ **Id** \$	Shift, goto 1
0 **Id** 1	$*$ **Id** $+$ **Id** \$	Shift, goto 3

Here, the state is 1, the next symbol is $*$, so shift and mark it with state 3.

Shift/Reduce Parsing with an SLR Table

$1: e \rightarrow t + e$

$2: e \rightarrow t$

$3: t \rightarrow \textbf{Id} * t$

$4: t \rightarrow \textbf{Id}$

State	Action				Goto	
	Id	$+$	$*$	**$**	e	t
0	s1				7	2
1		r4	s3	r4		
2		s4		r2		
3	s1					5
4	s1				6	2
5		r3		r3		
6				r1		
7				✓		

Stack	Input	Action
0	**Id** $*$ **Id** $+$ **Id** $\$$	Shift, goto 1
0 **Id** 1	$*$ **Id** $+$ **Id** $\$$	Shift, goto 3
0 **Id** 1 $*$ 3	**Id** $+$ **Id** $\$$	Shift, goto 1
0 **Id** 1 $*$ 3 **Id** 1	$+$ **Id** $\$$	Reduce 4

Here, the state is 1, the next symbol is $+$. The table says reduce using rule 4.

Shift/Reduce Parsing with an SLR Table

$1: e \rightarrow t + e$
$2: e \rightarrow t$
$3: t \rightarrow \textbf{Id} * t$
$4: t \rightarrow \textbf{Id}$

State	Action				Goto	
	Id	+	*	$	e	t
0	s1				7	2
1		r4	s3	r4		
2		s4		r2		
3	s1					5
4	s1				6	2
5		r3		r3		
6				r1		
7				✓		

Stack	Input	Action
0	**Id** * **Id** + **Id** $	Shift, goto 1
0 **Id**/1	* **Id** + **Id** $	Shift, goto 3
0 **Id**/1 */3	**Id** + **Id** $	Shift, goto 1
0 **Id**/1 */3 **Id**/1	+ **Id** $	Reduce 4
0 **Id**/1 */3	+ **Id** $	

Remove the RHS of the rule (here, just **Id**), observe the state on the top of the stack, and consult the "goto" portion of the table.

Shift/Reduce Parsing with an SLR Table

$1 : e \rightarrow t + e$
$2 : e \rightarrow t$
$3 : t \rightarrow \textbf{Id} \; * \; t$
$4 : t \rightarrow \textbf{Id}$

State	Action				Goto	
	Id	$+$	$*$	**$**	e	t
0	s1				7	2
1		r4	s3	r4		
2		s4		r2		
3	s1					5
4	s1				6	2
5		r3		r3		
6				r1		
7				✓		

Stack	Input	Action
0	**Id** $*$ **Id** $+$ **Id** $\$$	Shift, goto 1
0 **Id** 1	$*$ **Id** $+$ **Id** $\$$	Shift, goto 3
0 **Id** 1 $*$ 3	**Id** $+$ **Id** $\$$	Shift, goto 1
0 **Id** 1 $*$ 3 **Id** 1	$+$ **Id** $\$$	Reduce 4
0 **Id** 1 $*$ 3 t 5	$+$ **Id** $\$$	Reduce 3

Here, we push a t with state 5. This effectively "backs up" the LR(0) automaton and runs it over the newly added nonterminal.

In state 5 with an upcoming $+$, the action is "reduce 3."

Shift/Reduce Parsing with an SLR Table

$1: e \rightarrow t + e$

$2: e \rightarrow t$

$3: t \rightarrow \textbf{Id} * t$

$4: t \rightarrow \textbf{Id}$

State	Action				Goto	
	Id	$+$	$*$	**$**	e	t
0	s1				7	2
1		r4	s3	r4		
2		s4		r2		
3	s1					5
4	s1				6	2
5		r3		r3		
6				r1		
7				✓		

Stack	Input	Action
0	$\textbf{Id} * \textbf{Id} + \textbf{Id}\,\$$	Shift, goto 1
0, Id 1	$* \textbf{Id} + \textbf{Id}\,\$$	Shift, goto 3
0, Id 1, $*$ 3	$\textbf{Id} + \textbf{Id}\,\$$	Shift, goto 1
0, Id 1, $*$ 3, Id 1	$+ \textbf{Id}\,\$$	Reduce 4
0, Id 1, $*$ 3, t 5	$+ \textbf{Id}\,\$$	Reduce 3
0, t 2	$+ \textbf{Id}\,\$$	Shift, goto 4

This time, we strip off the RHS for rule 3, $\textbf{Id} * t$, exposing state 0, so we push a t with state 2.

Shift/Reduce Parsing with an SLR Table

$1: e \rightarrow t + e$

$2: e \rightarrow t$

$3: t \rightarrow \textbf{Id} * t$

$4: t \rightarrow \textbf{Id}$

State	Action				Goto	
	Id	$+$	$*$	**$**	e	t
0	s1				7	2
1		r4	s3	r4		
2		s4		r2		
3	s1					5
4	s1				6	2
5		r3		r3		
6				r1		
7				✓		

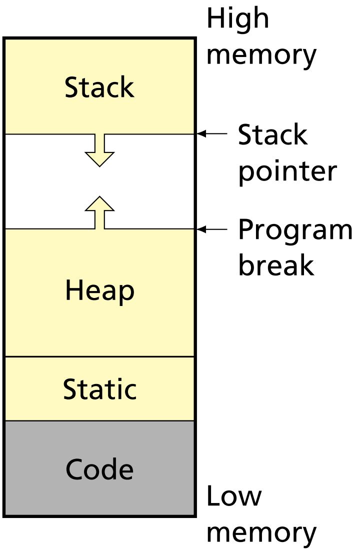
Stack	Input	Action
0	**Id** $*$ **Id** $+$ **Id** $	Shift, goto 1
0 **Id** 1	$*$ **Id** $+$ **Id** $	Shift, goto 3
0 **Id** 1 $*$ 3	**Id** $+$ **Id** $	Shift, goto 1
0 **Id** 1 $*$ 3 **Id** 1	$+$ **Id** $	Reduce 4
0 **Id** 1 $*$ 3 t 5	$+$ **Id** $	Reduce 3
0 t 2	$+$ **Id** $	Shift, goto 4
0 t 2 $+$ 4	**Id** $	Shift, goto 1
0 t 2 $+$ 4 **Id** 1	$	Reduce 4
0 t 2 $+$ 4 t 2	$	Reduce 2
0 t 2 $+$ 4 e 6	$	Reduce 1
0 e 7	$	Accept

Storage Classes and Memory Layout

Stack: objects created/destroyed in last-in, first-out order

Heap: objects created/destroyed in any order; automatic garbage collection optional

Static: objects allocated at compile time; persist throughout run



Static Objects

```
class Example {
  public static final int a = 3;

  public void hello() {
    System.out.println("Hello");
  }
}
```

Examples

Static class variable

Code for hello method

String constant “Hello”

Information about the Example class

Advantages

Zero-cost memory management

Often faster access (address a constant)

No out-of-memory danger

Disadvantages

Size and number must be known beforehand

Wasteful if sharing is possible

Stack-Allocated Objects

Natural for supporting recursion.

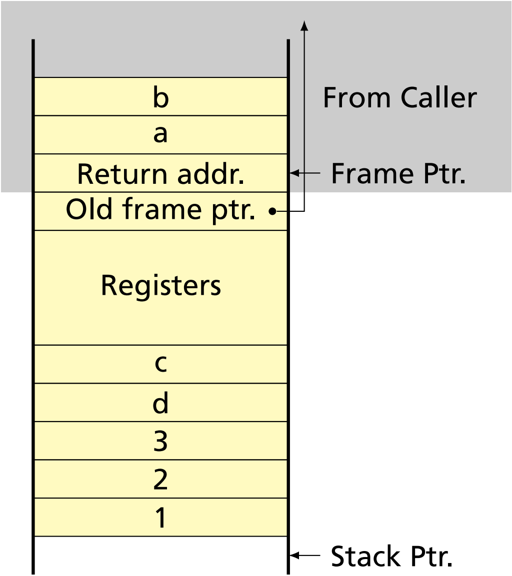
Idea: some objects persist from when a procedure is called to when it returns.

Naturally implemented with a stack: linear array of memory that grows and shrinks at only one boundary.

Each invocation of a procedure gets its own *frame* (*activation record*) where it stores its own local variables and bookkeeping information.

An Activation Record: The State Before Calling *bar*

```
int foo(int a, int b) {
  int c, d;
  bar(1, 2, 3);
}
```



Recursive Fibonacci

(Real C)

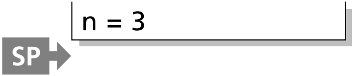
```
int fib(int n) {

  if (n<2)

    return 1;
  else
    return
        fib(n-1)
        +
        fib(n-2);

}
```

(Assembly-like C)

```
int fib(int n) {
     int tmp1, tmp2, tmp3;
     tmp1 = n < 2;
     if (!tmp1) goto L1;
     return 1;
L1:  tmp1 = n - 1;
     tmp2 = fib(tmp1);
L2:  tmp1 = n - 2;
     tmp3 = fib(tmp1);
L3:  tmp1 = tmp2 + tmp3;
     return tmp1;
}
```

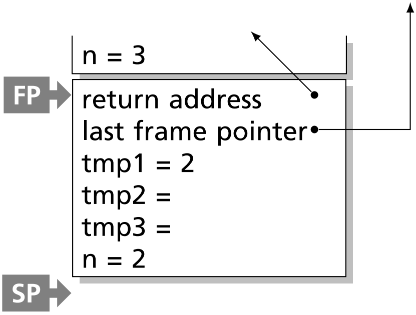
```
          fib(3)
         /      \
     fib(2)    fib(1)
    /      \
fib(1)    fib(0)
```

Executing fib(3)




```
int fib(int n) {
    int tmp1, tmp2, tmp3;
    tmp1 = n < 2;
    if (!tmp1) goto L1;
    return 1;
L1: tmp1 = n - 1;
    tmp2 = fib(tmp1);
L2: tmp1 = n - 2;
    tmp3 = fib(tmp1);
L3: tmp1 = tmp2 + tmp3;
    return tmp1;
}
```

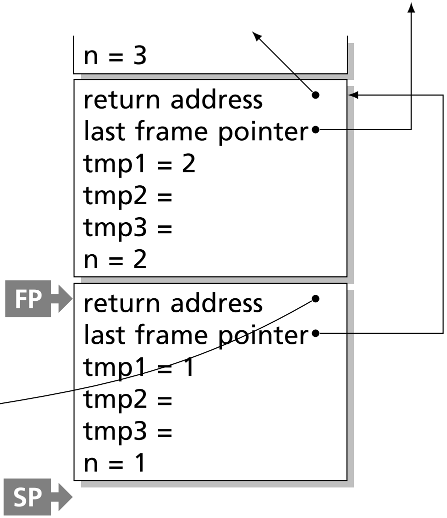
Executing fib(3)




```
int fib(int n) {
    int tmp1, tmp2, tmp3;
    tmp1 = n < 2;
    if (!tmp1) goto L1;
    return 1;
L1: tmp1 = n - 1;
    tmp2 = fib(tmp1);
L2: tmp1 = n - 2;
    tmp3 = fib(tmp1);
L3: tmp1 = tmp2 + tmp3;
    return tmp1;
}
```

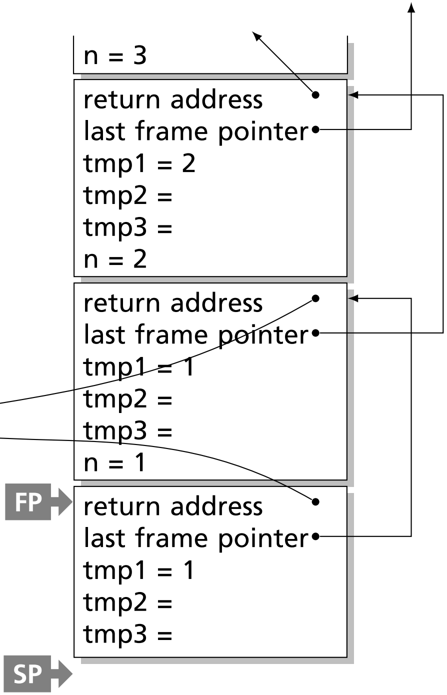
Executing fib(3)

```
int fib(int n) {
    int tmp1, tmp2, tmp3;
    tmp1 = n < 2;
    if (!tmp1) goto L1;
    return 1;
L1: tmp1 = n - 1;
    tmp2 = fib(tmp1);
L2: tmp1 = n - 2;
    tmp3 = fib(tmp1);
L3: tmp1 = tmp2 + tmp3;
    return tmp1;
}
```



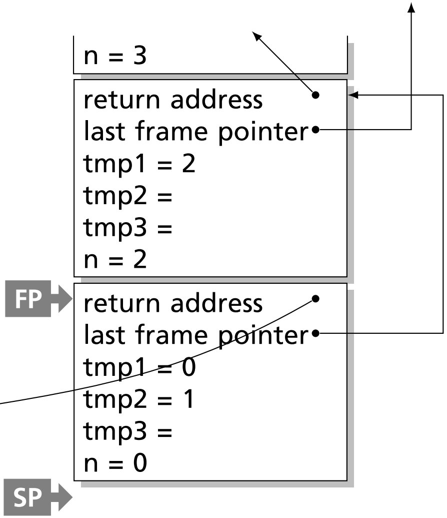
Executing fib(3)

```
int fib(int n) {
    int tmp1, tmp2, tmp3;
    tmp1 = n < 2;
    if (!tmp1) goto L1;
    return 1;
L1: tmp1 = n - 1;
    tmp2 = fib(tmp1);
L2: tmp1 = n - 2;
    tmp3 = fib(tmp1);
L3: tmp1 = tmp2 + tmp3;
    return tmp1;
}
```



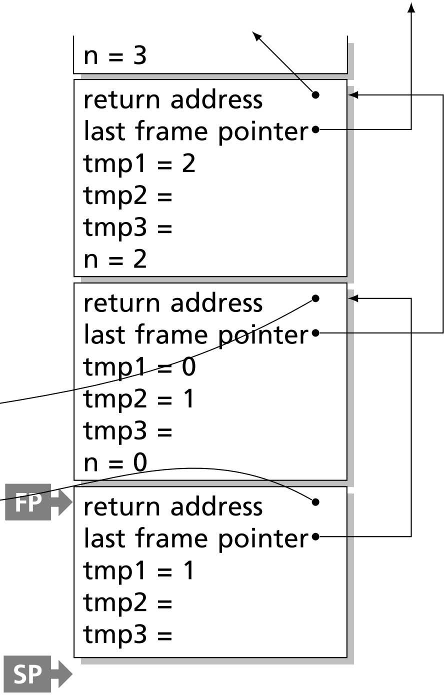
Executing fib(3)

```
int fib(int n) {
    int tmp1, tmp2, tmp3;
    tmp1 = n < 2;
    if (!tmp1) goto L1;
    return 1;
L1: tmp1 = n - 1;
    tmp2 = fib(tmp1);
L2: tmp1 = n - 2;
    tmp3 = fib(tmp1);
L3: tmp1 = tmp2 + tmp3;
    return tmp1;
}
```



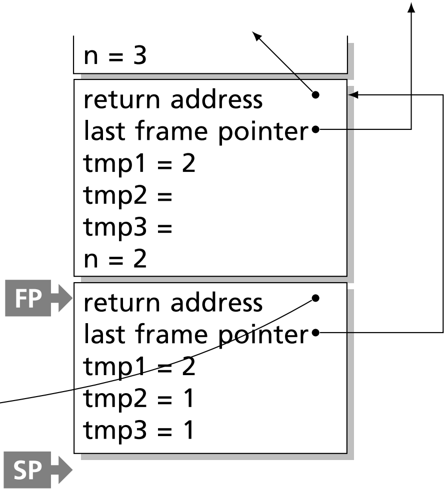
Executing fib(3)

```
int fib(int n) {
    int tmp1, tmp2, tmp3;
    tmp1 = n < 2;
    if (!tmp1) goto L1;
    return 1;
L1: tmp1 = n - 1;
    tmp2 = fib(tmp1);
L2: tmp1 = n - 2;
    tmp3 = fib(tmp1);
L3: tmp1 = tmp2 + tmp3;
    return tmp1;
}
```

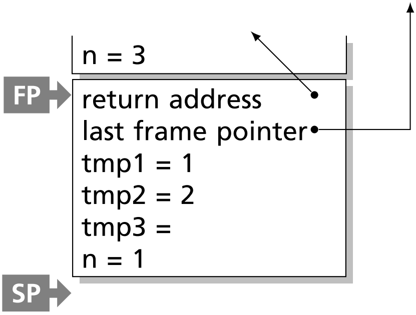


Executing fib(3)

```
int fib(int n) {
    int tmp1, tmp2, tmp3;
    tmp1 = n < 2;
    if (!tmp1) goto L1;
    return 1;
L1: tmp1 = n - 1;
    tmp2 = fib(tmp1);
L2: tmp1 = n - 2;
    tmp3 = fib(tmp1);
L3: tmp1 = tmp2 + tmp3;
    return tmp1;
}
```



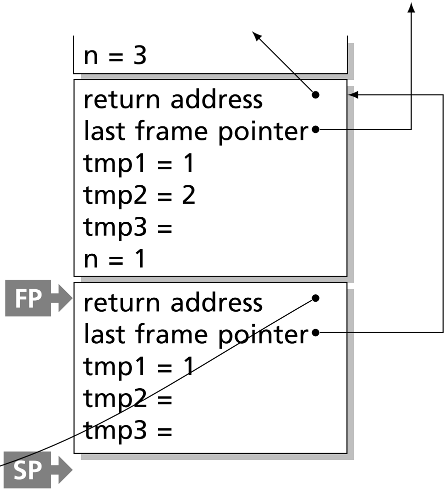
Executing fib(3)



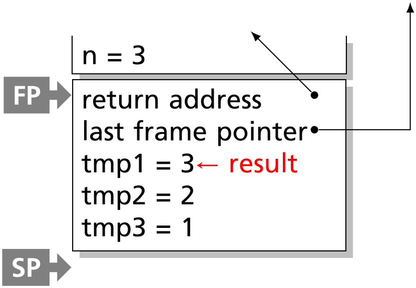

```
int fib(int n) {
    int tmp1, tmp2, tmp3;
    tmp1 = n < 2;
    if (!tmp1) goto L1;
    return 1;
L1: tmp1 = n - 1;
    tmp2 = fib(tmp1);
L2: tmp1 = n - 2;
    tmp3 = fib(tmp1);
L3: tmp1 = tmp2 + tmp3;
    return tmp1;
}
```

Executing fib(3)

```
int fib(int n) {
    int tmp1, tmp2, tmp3;
    tmp1 = n < 2;
    if (!tmp1) goto L1;
    return 1;
L1: tmp1 = n - 1;
    tmp2 = fib(tmp1);
L2: tmp1 = n - 2;
    tmp3 = fib(tmp1);
L3: tmp1 = tmp2 + tmp3;
    return tmp1;
}
```



Executing fib(3)

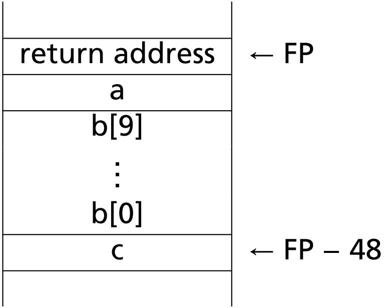



```
int fib(int n) {
    int tmp1, tmp2, tmp3;
    tmp1 = n < 2;
    if (!tmp1) goto L1;
    return 1;
L1: tmp1 = n - 1;
    tmp2 = fib(tmp1);
L2: tmp1 = n - 2;
    tmp3 = fib(tmp1);
L3: tmp1 = tmp2 + tmp3;
    return tmp1;
}
```

Allocating Fixed-Size Arrays

Local arrays with fixed size are easy to stack.

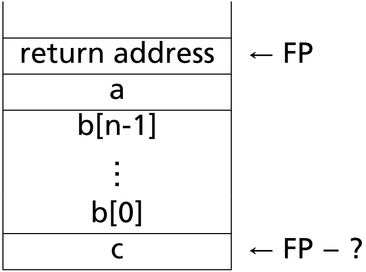
```
void foo()
{
  int a;
  int b[10];
  int c;
}
```



Allocating Variable-Sized Arrays

Variable-sized local arrays aren't as easy.

```
void foo(int n)
{
  int a;
  int b[n];
  int c;
}
```

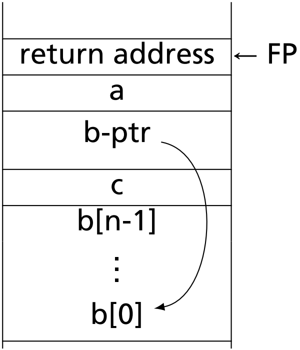


Doesn't work: generated code expects a fixed offset for c.
Even worse for multi-dimensional arrays.

Allocating Variable-Sized Arrays

As always:
add a level of indirection

```
void foo(int n)
{
  int a;
  int b[n];
  int c;
}
```



Variables remain constant offset from frame pointer.

Heap-Allocated Storage

Static works when you know everything beforehand and always need it.

Stack enables, but also requires, recursive behavior.

A *heap* is a region of memory where blocks can be allocated and deallocated in any order.

(These heaps are different than those in, e.g., heapsort)

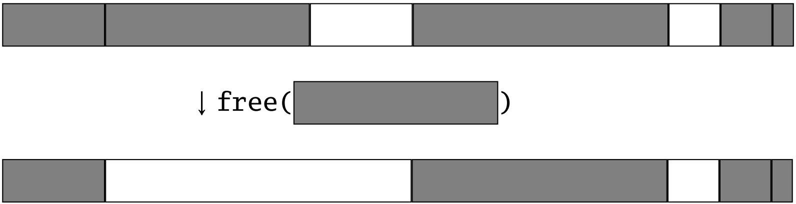
Dynamic Storage Allocation in C

```
struct point {
   int x, y;
};

int play_with_points(int n)
{
  int i;
  struct point *points;

  points = malloc(n * sizeof(struct point));

  for ( i = 0 ; i < n ; i++ ) {
    points[i].x = random();
    points[i].y = random();
  }

  /* do something with the array */

  free(points);
}
```

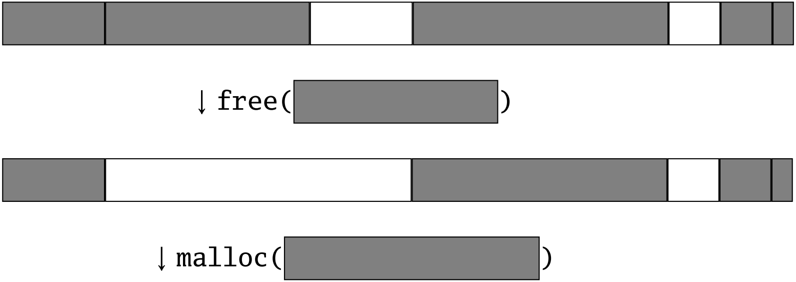
Dynamic Storage Allocation

Dynamic Storage Allocation

Dynamic Storage Allocation



Dynamic Storage Allocation



Dynamic Storage Allocation

Dynamic Storage Allocation

Rules:

- Each allocated block contiguous (no holes)
- Blocks stay fixed once allocated

`malloc()`

- Find an area large enough for requested block
- Mark memory as allocated

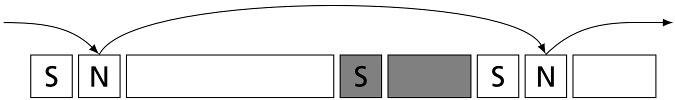
`free()`

- Mark the block as unallocated

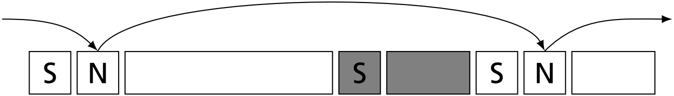
Simple Dynamic Storage Allocation

- Maintaining information about free memory
  - Simplest: Linked list
- The algorithm for locating a suitable block
  - Simplest: First-fit
- The algorithm for freeing an allocated block
  - Simplest: Coalesce adjacent free blocks

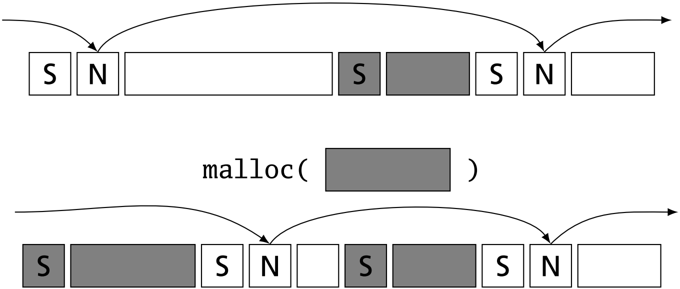
Simple Dynamic Storage Allocation



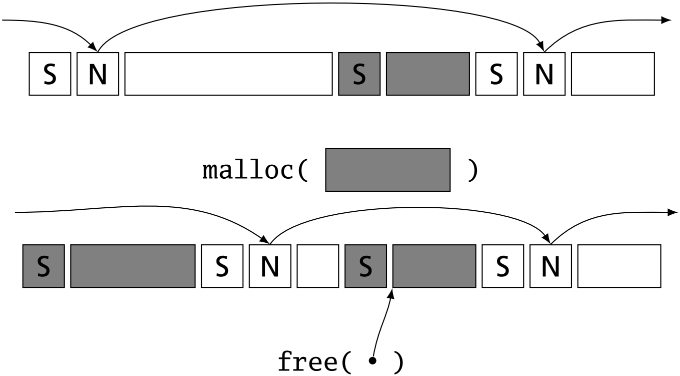
Simple Dynamic Storage Allocation



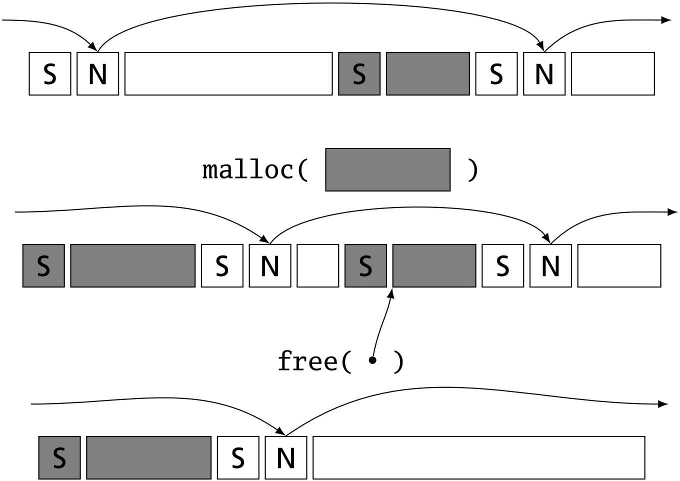
Simple Dynamic Storage Allocation



Simple Dynamic Storage Allocation



Simple Dynamic Storage Allocation



Fragmentation

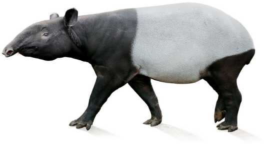
`malloc(`  `)` seven times give

`free()` four times gives

`malloc(`  `)` ?

Need more memory; can’t use fragmented memory.

Zebra

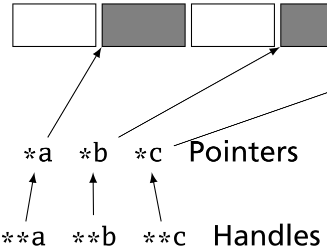


Tapir

Fragmentation and Handles

Standard CS solution: Add another layer of indirection.

Always reference memory through “handles.”

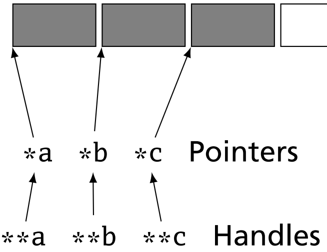


The original Macintosh did this to save memory.

Fragmentation and Handles

Standard CS solution: Add another layer of indirection.

Always reference memory through “handles.”



The original Macintosh did this to save memory.

Automatic Garbage Collection

Entrust the runtime system with freeing heap objects

Now common: Java, C#, Javascript, Python, Ruby, OCaml and most functional languages

Advantages

Much easier for the programmer

Greatly improves reliability: no memory leaks, double-freeing, or other memory management errors

Disadvantages

Slower, sometimes unpredictably so

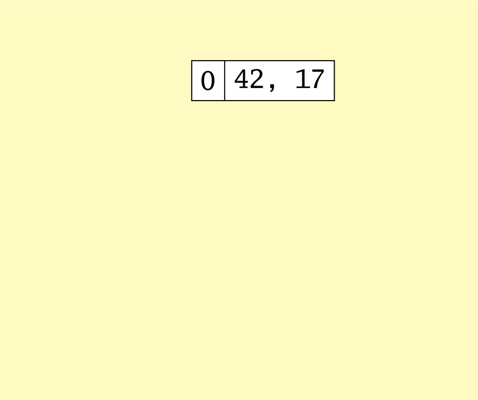
May consume more memory

Reference Counting

What and when to free?

- Maintain count of references to each object
- Free when count reaches zero

```
let a = (42, 17) in
let b = [a;a] in
let c = (1,2)::b in
b
```

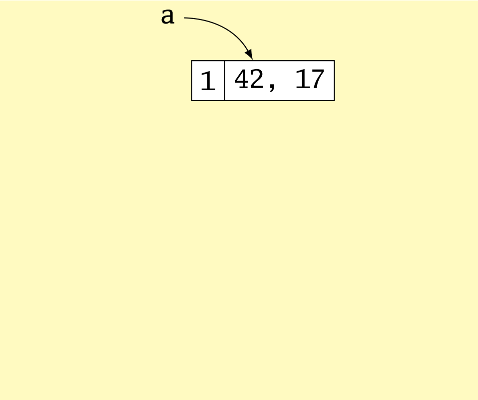


Reference Counting

What and when to free?

- Maintain count of references to each object
- Free when count reaches zero

```
let a = (42, 17) in
let b = [a;a] in
let c = (1,2)::b in
b
```

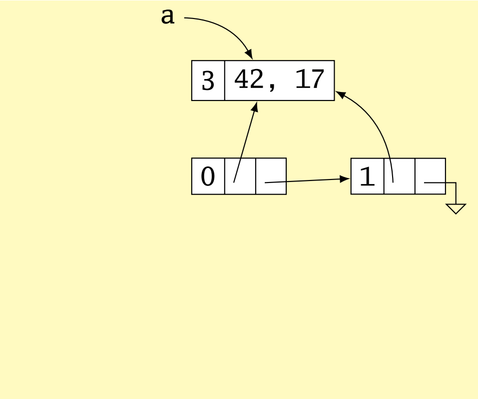


Reference Counting

What and when to free?

- Maintain count of references to each object
- Free when count reaches zero

```
let a = (42, 17) in
let b = [a;a] in
let c = (1,2)::b in
b
```

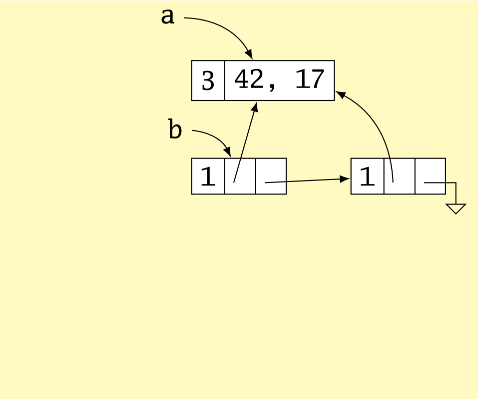


Reference Counting

What and when to free?

- Maintain count of references to each object
- Free when count reaches zero

```
let a = (42, 17) in
let b = [a;a] in
let c = (1,2)::b in
b
```

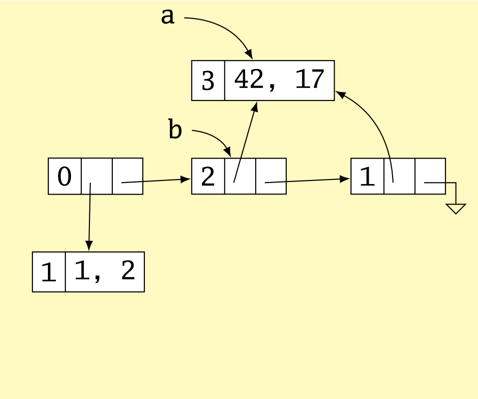


Reference Counting

What and when to free?

- Maintain count of references to each object
- Free when count reaches zero

```
let a = (42, 17) in
let b = [a;a] in
let c = (1,2)::b in
b
```

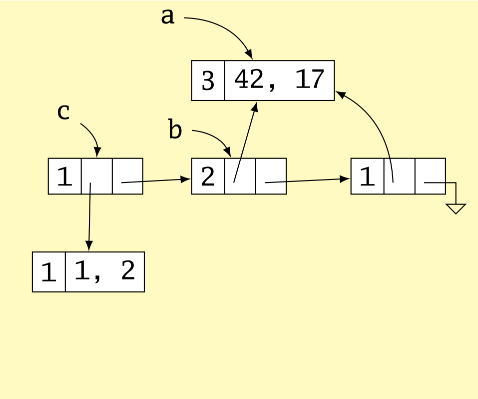


Reference Counting

What and when to free?

- Maintain count of references to each object
- Free when count reaches zero

```
let a = (42, 17) in
let b = [a;a] in
let c = (1,2)::b in
b
```

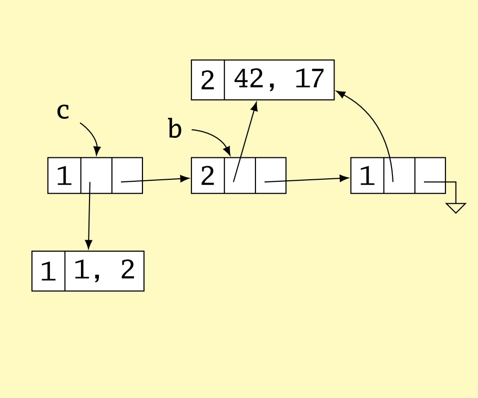


Reference Counting

What and when to free?

- Maintain count of references to each object
- Free when count reaches zero

```
let a = (42, 17) in
let b = [a;a] in
let c = (1,2)::b in
b
```

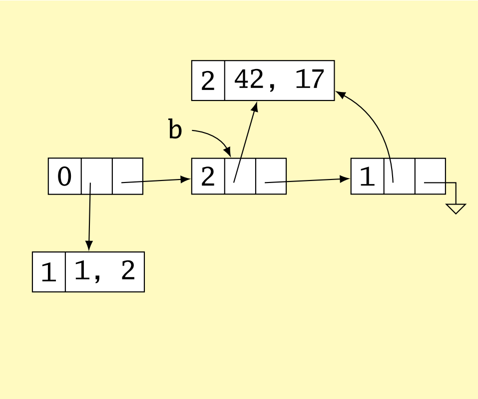


Reference Counting

What and when to free?

- Maintain count of references to each object
- Free when count reaches zero

```
let a = (42, 17) in
let b = [a;a] in
let c = (1,2)::b in
b
```

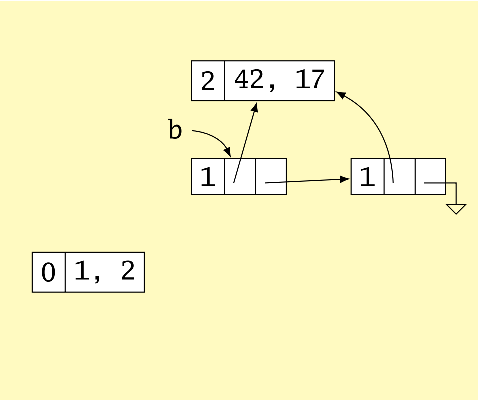


Reference Counting

What and when to free?

- Maintain count of references to each object
- Free when count reaches zero

```
let a = (42, 17) in
let b = [a;a] in
let c = (1,2)::b in
b
```

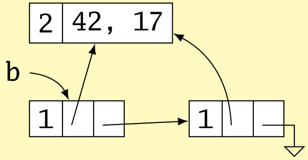


Reference Counting

What and when to free?

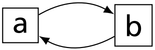
- Maintain count of references to each object
- Free when count reaches zero

```
let a = (42, 17) in
let b = [a;a] in
let c = (1,2)::b in
b
```



Issues with Reference Counting

Circular structures defy reference counting:



Neither is reachable, yet both have non-zero reference counts.

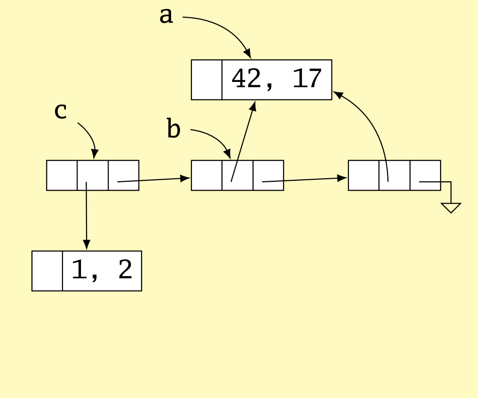
High overhead (must update counts constantly), although incremental

Mark-and-Sweep

What and when to free?

- Stop-the-world algorithm invoked when memory full
- Breadth-first-search marks all reachable memory
- All unmarked items freed

```
let a = (42, 17) in
let b = [a;a] in
let c = (1,2)::b in
b
```

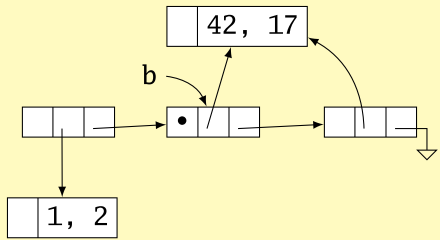


Mark-and-Sweep

What and when to free?

- Stop-the-world algorithm invoked when memory full
- Breadth-first-search marks all reachable memory
- All unmarked items freed

```
let a = (42, 17) in
let b = [a;a] in
let c = (1,2)::b in
b
```

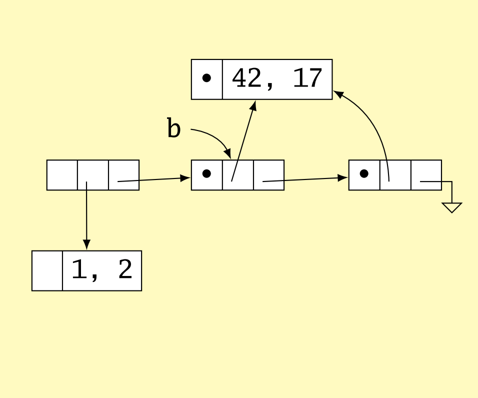


Mark-and-Sweep

What and when to free?

- Stop-the-world algorithm invoked when memory full
- Breadth-first-search marks all reachable memory
- All unmarked items freed

```
let a = (42, 17) in
let b = [a;a] in
let c = (1,2)::b in
b
```

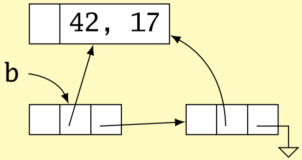


Mark-and-Sweep

What and when to free?

- Stop-the-world algorithm invoked when memory full
- Breadth-first-search marks all reachable memory
- All unmarked items freed

```
let a = (42, 17) in
let b = [a;a] in
let c = (1,2)::b in
b
```



Mark-and-Sweep

Mark-and-sweep is faster overall; may induce big pauses

Mark-and-compact variant also moves or copies reachable objects to eliminate fragmentation

Incremental garbage collectors try to avoid doing everything at once

Most objects die young; generational garbage collectors segregate heap objects by age

Parallel garbage collection tricky

Real-time garbage collection tricky

Single Inheritance

Simple: Add new fields to end of the object

Fields in base class always at same offset in derived class (compiler never reorders)

Consequence: Derived classes can never remove fields

C++

```
class Shape {
  double x, y;
};

class Box : Shape {
  double h, w;
};


class Circle : Shape {
  double r;
};
```

Equivalent C

```
struct Shape {
  double x, y;
};

struct Box {
  double x, y;
  double h, w;
};

struct Circle {
  double x, y;
  double r;
};
```

Virtual Functions

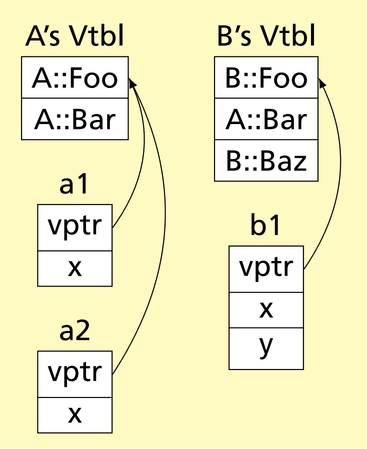
```
class Shape {
  virtual void draw(); // Invoked by object's run-time class
};                     // not its compile-time type.

class Line : public Shape {
  void draw();
}

class Arc : public Shape {
  void draw();
};

Shape *s[10];
s[0] = new Line;
s[1] = new Arc;
s[0]->draw();   // Invoke Line::draw()
s[1]->draw();   // Invoke Arc::draw()
```

Virtual Functions

Trick: add to each object a pointer to the virtual table for its type, filled with pointers to the virtual functions.

Like the objects themselves, the virtual table for each derived type begins identically.

```
struct A {
  int x;
  virtual void Foo();
  virtual void Bar();
};

struct B : A {
  int y;
  virtual void Foo();
  virtual void Baz();
};

A a1;
A a2;
B b1;
```



Stack-Based IR: Java Bytecode

```
int gcd(int a, int b) {
  while (a != b) {
    if (a > b)
      a -= b;
    else
      b -= a;
  }
  return a;
}
```

```
# javap -c Gcd

Method int gcd(int, int)
   0 goto 19

   3 iload_1         // Push a
   4 iload_2         // Push b
   5 if_icmple 15    // if a <= b goto 15

   8 iload_1         // Push a
   9 iload_2         // Push b
  10 isub            // a - b
  11 istore_1        // Store new a
  12 goto 19

  15 iload_2         // Push b
  16 iload_1         // Push a
  17 isub            // b - a
  18 istore_2        // Store new b

  19 iload_1         // Push a
  20 iload_2         // Push b
  21 if_icmpne 3     // if a != b goto 3

  24 iload_1         // Push a
  25 ireturn         // Return a
```

Stack-Based IRs

Advantages:

- Trivial translation of expressions
- Trivial interpreters
- No problems with exhausting registers
- Often compact

Disadvantages:

- Semantic gap between stack operations and modern register machines
- Hard to see what communicates with what
- Difficult representation for optimization

Register-Based IR: Mach SUIF

```
int gcd(int a, int b) {
  while (a != b) {
    if (a > b)
      a -= b;
    else
      b -= a;
  }
  return a;
}
```

```
gcd:
gcd._gcdTmp0:
  sne   $vr1.s32 <- gcd.a,gcd.b
  seq   $vr0.s32 <- $vr1.s32,0
  btrue $vr0.s32,gcd._gcdTmp1   // if!(a != b) goto Tmp1

  sl    $vr3.s32 <- gcd.b,gcd.a
  seq   $vr2.s32 <- $vr3.s32,0
  btrue $vr2.s32,gcd._gcdTmp4   // if!(a < b) goto Tmp4

  mrk   2, 4    // Line number 4
  sub   $vr4.s32 <- gcd.a,gcd.b
  mov   gcd._gcdTmp2 <- $vr4.s32
  mov   gcd.a <- gcd._gcdTmp2   // a = a - b
  jmp   gcd._gcdTmp5
gcd._gcdTmp4:
  mrk   2, 6
  sub   $vr5.s32 <- gcd.b,gcd.a
  mov   gcd._gcdTmp3 <- $vr5.s32
  mov   gcd.b <- gcd._gcdTmp3   // b = b - a
gcd._gcdTmp5:
  jmp   gcd._gcdTmp0

gcd._gcdTmp1:
  mrk   2, 8
  ret   gcd.a   // Return a
```

Register-Based IRs

Most common type of IR

Advantages:

- Better representation for register machines
- Dataflow is usually clear

Disadvantages:

- Slightly harder to synthesize from code
- Less compact
- More complicated to interpret

Optimization In Action

```
int gcd(int a, int b) {
  while (a != b) {
    if (a < b) b -= a;
    else a -= b;
  }
  return a;
}
```

GCC on SPARC

```
gcd:  save %sp, -112, %sp
      st   %i0, [%fp+68]
      st   %i1, [%fp+72]
.LL2: ld   [%fp+68], %i1
      ld   [%fp+72], %i0
      cmp  %i1, %i0
      bne  .LL4
      nop
      b    .LL3
      nop
.LL4: ld   [%fp+68], %i1
      ld   [%fp+72], %i0
      cmp  %i1, %i0
      bge  .LL5
      nop
      ld   [%fp+72], %i0
      ld   [%fp+68], %i1
      sub  %i0, %i1, %i0
      st   %i0, [%fp+72]
      b    .LL2
      nop
.LL5: ld   [%fp+68], %i0
      ld   [%fp+72], %i1
      sub  %i0, %i1, %i0
      st   %i0, [%fp+68]
      b    .LL2
      nop
.LL3: ld   [%fp+68], %i0
      ret
      restore
```

GCC -O7 on SPARC

```
gcd:  cmp    %o0, %o1
      be     .LL8
      nop
.LL9: bge,a  .LL2
      sub    %o0, %o1, %o0
      sub    %o1, %o0, %o1
.LL2: cmp    %o0, %o1
      bne    .LL9
      nop
.LL8: retl
      nop
```

Typical Optimizations

- Folding constant expressions
  1+3 → 4
- Removing dead code
  if (0) { . . . } → nothing
- Moving variables from memory to registers

  ```
  ld   [%fp+68], %i1
  sub  %i0, %i1, %i0  → sub   %o1, %o0, %o1
  st   %i0, [%fp+72]
  ```

- Removing unnecessary data movement
- Filling branch delay slots (Pipelined RISC processors)
- Common subexpression elimination

Machine-Dependent vs. -Independent Optimization

No matter what the machine is, folding constants and eliminating dead code is always a good idea.

```
a = c + 5 + 3;
if (0 + 3) {
  b = c + 8;
}
```

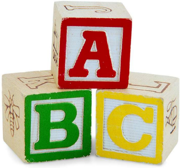
→

```
b = a = c + 8;
```

However, many optimizations are processor-specific:

- Register allocation depends on how many registers the machine has
- Not all processors have branch delay slots to fill
- Each processor’s pipeline is a little different

Basic Blocks



```
int gcd(int a, int b) {
  while (a != b) {
    if (a < b) b -= a;
    else a -= b;
  }
  return a;
}
```

lower →

```
A: sne t, a, b
   bz  E, t
   slt t, a, b
   bnz B, t
   sub b, b, a
   jmp C
B: sub a, a, b
C: jmp A
E: ret a
```

split →

```
A: sne t, a, b
   bz  E, t

   slt t, a, b
   bnz B, t

   sub b, b, a
   jmp C

B: sub a, a, b

C: jmp A

E: ret a
```

The statements in a basic block all run if the first one does.

Starts with a statement following a conditional branch or is a branch target.

Usually ends with a control-transfer statement.

Control-Flow Graphs

A CFG illustrates the flow of control among basic blocks.

```
A:
sne t, a, b
bz  E, t

slt t, a, b
bnz B, t

sub b, b, a
jmp C

B:
sub a, a, b

C:
jmp A

E:
ret a
```