Review for the Final

Stephen A. Edwards

Columbia University

Fall 2014
Table of Contents I

The Final

Structure of a Compiler

Scanning
  Languages and Regular Expressions
  NFAs
  Translating REs into NFAs
  Building a DFA from an NFA: Subset Construction

Parsing
  Resolving Ambiguity
  Rightmost and Reverse-Rightmost Derivations
  Building the LR(0) Automaton
  Building an SLR Parsing Table
  Shift/Reduce Parsing
Table of Contents II

Types and Static Semantics
- Static Semantic Analysis
- Scope
- Applicative vs. Normal-Order Evaluation

Runtime Environments
- Storage Classes and Memory Layout
- The Stack and Activation Records
- In-Memory Layout Issues
- The Heap
- Automatic Garbage Collection
- Objects and Inheritance
- Exceptions
Table of Contents III

Code Generation
  Intermediate Representations
  Optimization and Basic Blocks
  Separate Compilation and Linking
Table of Contents IV

The Lambda Calculus
  Lambda Expressions
  Beta-reduction
  Alpha-conversion
  Reduction Order
  Normal Form
  The Y Combinator

Logic Programming in Prolog
  Prolog Execution
  The Prolog Environment
  Unification
  The Searching Algorithm
  Prolog as an Imperative Language
The Final

75 minutes

Closed book

One double-sided sheet of notes of your own devising

Comprehensive: Anything discussed in class is fair game, including things from before the midterm

Little, if any, programming

Details of O’Caml/C/C++/Java syntax not required

Broad knowledge of languages discussed
Compiling a Simple Program

```c
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;
}
void 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;
}
Semantic Analysis Resolves Symbols and Checks Types

```
int a
int b

func gcd(args arg a, arg b) {
    while a != b {
        if a > b {
            a -= b
        } else {
            b -= a
        }
    }
    return a
}
```

Symbol Table

- int a
- int b
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;
}
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, \ldots, Z \}, ASCII, Unicode

**String**: A finite sequence of symbols from an alphabet
Examples: \( \epsilon \) (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
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

\( M^* = \{ \epsilon \} \cup M \cup MM \cup MMM \cdots = \{ \epsilon, \text{man}, \text{men}, \text{manman}, \text{manmen}, \text{menman}, \text{menmen}, \text{manmanman}, \text{manmanmen}, \text{manmenman}, \ldots \} \)
A standard way to express languages for tokens.

1. $\epsilon$ 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 $L(s)$,
   - $(r) | (s)$ denotes $L(r) \cup L(s)$
   - $(r)(s)$ denotes $\{tu : t \in L(r), u \in L(s)\}$
   - $(r)^*$ denotes $\bigcup_{i=0}^{\infty} L^i$ ($L^0 = \{\epsilon\}$ 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 $S:\{A, B, C, D\}$
2. Set of input symbols $\Sigma: \{0, 1\}$
3. Transition function $\sigma: S \times \Sigma \epsilon \rightarrow 2^S$

<table>
<thead>
<tr>
<th>state</th>
<th>$\epsilon$</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$\emptyset$</td>
<td>${B}$</td>
<td>${C}$</td>
</tr>
<tr>
<td>B</td>
<td>$\emptyset$</td>
<td>${A}$</td>
<td>${D}$</td>
</tr>
<tr>
<td>C</td>
<td>$\emptyset$</td>
<td>${D}$</td>
<td>${A}$</td>
</tr>
<tr>
<td>D</td>
<td>$\emptyset$</td>
<td>${C}$</td>
<td>${B}$</td>
</tr>
</tbody>
</table>

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

Symbol

Sequence

Choice

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 $\epsilon$-closure of the start state
2. For each character $c$,
   - New states: follow all transitions labeled $c$
   - Form the $\epsilon$-closure of the current states
3. Accept if any final state is accepting
Simulating an NFA: ·aabb, Start
Simulating an NFA: $a \cdot abb$
Simulating an NFA: $aa \cdot bb$
Simulating an NFA: $aab \cdot b$
Simulating an NFA: $aabb$, Done
Deterministic Finite Automata

Restricted form of NFAs:

- No state has a transition on $\epsilon$
- 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

```plaintext
{  
    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"
  | ['a'-'z'] ['a'-'z' '0'-'9']* as lit { ID(lit) }
  | ['0'-'9']* as num { NUM(num) }

\begin{center}
\includegraphics[width=\textwidth]{automata_diagram.png}
\end{center}
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 \((a \mid b)^* abb\)
Subset construction for \((a \mid b)^* abb\)
Subset construction for \((a | b)^* abb\)
Subset construction for \((a \mid b)^* abb\)
Result of subset construction for \((a \mid b)^* abb\)

Is this minimal?
Ambiguous Arithmetic

Ambiguity can be a problem in expressions. Consider parsing

\[ 3 - 4 \times 2 + 5 \]

with the grammar

\[ e \rightarrow e + e \mid e - e \mid e \times e \mid e / e \mid N \]
Operator Precedence

Defines how “sticky” an operator is.

\[ 1 \times 2 + 3 \times 4 \]

* at higher precedence than +:

\[(1 \times 2) + (3 \times 4)\]

+ at higher precedence than *:

\[1 \times (2 + 3) \times 4\]
Associativity

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

Most operators are left-associative

\[ 1 \ - \ 2 \ - \ 3 \ - \ 4 \]

\[ (((1 \ - \ 2) \ - \ 3) \ - \ 4) \]

left associative

\[ 1 \ - \ (2 \ - \ (3 \ - \ 4)) \]

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

\[
\begin{align*}
\text{expr} & : \text{expr} \ \text{PLUS} \ \text{expr} \\
& \quad | \ \text{expr} \ \text{MINUS} \ \text{expr} \\
& \quad | \ \text{term} \\
\text{term} & : \text{term} \ \text{TIMES} \ \text{term} \\
& \quad | \ \text{term} \ \text{DIVIDE} \ \text{term} \\
& \quad | \ \text{atom} \\
\text{atom} & : \text{NUMBER}
\end{align*}
\]

Still ambiguous: associativity not defined

Ocamlyacc’s complaint: “8 shift/reduce conflicts.”
Assigning Associativity

Make one side the next level of precedence

\[
\begin{align*}
expr & : \ expr \ PLUS \ term \\
   & \mid \ expr \ MINUS \ term \\
   & \mid \ term \\

term & : \ term \ TIMES \ atom \\
   & \mid \ term \ DIVIDE \ atom \\
   & \mid \ atom \\
atom & : \ NUMBER
\end{align*}
\]

This is left-associative.

No shift/reduce conflicts.
Rightmost Derivation of $\text{Id} \ast \text{Id} + \text{Id}$

1: $e \rightarrow t + e$
2: $e \rightarrow t$
3: $t \rightarrow \text{Id} \ast t$
4: $t \rightarrow \text{Id}$

At each step, expand the rightmost nonterminal.

Fun and interesting fact: there is exactly one rightmost expansion if the grammar is unambiguous.
Rightmost Derivation: What to Expand

1: \( e \rightarrow t + e \)
2: \( e \rightarrow t \)
3: \( t \rightarrow \text{Id} \ast t \)
4: \( t \rightarrow \text{Id} \)

Expand here

Terminals only
Reverse Rightmost Derivation

1: \( e \rightarrow t + e \)
2: \( e \rightarrow t \)
3: \( t \rightarrow \text{Id} \ast t \)
4: \( t \rightarrow \text{Id} \)

viable prefixes

terminals
Shift/Reduce Parsing Using an Oracle

1: $e \rightarrow t + e$
2: $e \rightarrow t$
3: $t \rightarrow \text{Id} * t$
4: $t \rightarrow \text{Id}$
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.
Building the Initial State of the LR(0) Automaton

1: $e \rightarrow t + e$
2: $e \rightarrow t$
3: $t \rightarrow \text{ld} \ast t$
4: $t \rightarrow \text{ld}$

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

<table>
<thead>
<tr>
<th>Rule</th>
<th>Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:</td>
<td>$e \rightarrow t + e$</td>
</tr>
<tr>
<td>2:</td>
<td>$e \rightarrow t$</td>
</tr>
<tr>
<td>3:</td>
<td>$t \rightarrow \text{Id}^* t$</td>
</tr>
<tr>
<td>4:</td>
<td>$t \rightarrow \text{Id}$</td>
</tr>
</tbody>
</table>

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

1: $e \rightarrow t + e$
2: $e \rightarrow t$
3: $t \rightarrow \text{Id} \ast t$
4: $t \rightarrow \text{Id}$

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 $\text{Id} \ast t$ or $\text{Id}$, so $t \rightarrow \cdot \text{Id} \ast t$ and $t \rightarrow \cdot \text{Id}$. 
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 $\text{ld}$.

$$e' \rightarrow \cdot e$$
$$e \rightarrow \cdot t + e$$

$S0: e \rightarrow \cdot t$
$$t \rightarrow \cdot \text{ld} \ast t$$
$$t \rightarrow \cdot \text{ld}$$
Building the LR(0) Automaton

"Just passed a string derived from e"

\[ S0 : e' \rightarrow e \cdot \]

\[ e' \rightarrow e \cdot \]
\[ e \rightarrow e \cdot t + e \]
\[ t \rightarrow t \cdot \text{Id} * t \]
\[ t \rightarrow t \cdot \text{Id} \]

"Just passed a prefix ending in a string derived from t"

\[ S2 : e \rightarrow t \cdot + e \]
\[ e \rightarrow t \cdot \]

The first state suggests a viable prefix can start as any string derived from \( e \), any string derived from \( t \), or \( \text{Id} \).

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 \( \text{Id} \cdot \cdot t \).
Building the LR(0) Automaton

S0: $e' \rightarrow e$
$e \rightarrow \cdot e$
$t \rightarrow \cdot t + e$
$t \rightarrow \cdot Id * t$
$t \rightarrow \cdot Id$

S1: $t \rightarrow Id \cdot * t$
$t \rightarrow Id \cdot$

S3: $t \rightarrow Id * \cdot t$
$t \rightarrow \cdot Id * t$
$t \rightarrow \cdot Id$

S7: $e' \rightarrow e$.

S2: $e \rightarrow t \cdot + e$
$e \rightarrow t \cdot$

S4: $e \rightarrow t + \cdot e$
$e \rightarrow \cdot t + e$
$t \rightarrow \cdot Id * t$
$t \rightarrow \cdot Id$

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

In S1, * may be next, giving $Id * \cdot t$ and two others.
Building the LR(0) Automaton

S0:  
- e' \rightarrow \cdot e
- e \rightarrow \cdot t + e
- t \rightarrow \cdot \text{Id} * t
- t \rightarrow \cdot \text{Id}

S1:  
- t \rightarrow \cdot \text{Id} * t
- \text{Id} \rightarrow \cdot * \text{Id}

S2:  
- e \rightarrow \cdot t + e
- e \rightarrow t \cdot e

S3:  
- t \rightarrow \cdot \text{Id} * t
- t \rightarrow \cdot \text{Id}

S4:  
- e \rightarrow \cdot t
- e \rightarrow t + \cdot e
- t \rightarrow \cdot \text{Id} * t
- t \rightarrow \cdot \text{Id}

S5:  
- t \rightarrow \cdot \text{Id} * t

S6:  
- e \rightarrow t + \cdot e
- e \rightarrow \cdot t + e

Just passed a prefix ending in a string derived from t

Just passed a prefix that ended in an Id

Just passed a string derived from e

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

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

Closure adds 4 more items.

In S1, \ast may be next, giving Id \cdot t and two others.
Converting the LR(0) Automaton to an SLR Parsing Table

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

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

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

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: $t \rightarrow Id \ast t$
4: $t \rightarrow Id$

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

<table>
<thead>
<tr>
<th>State</th>
<th>Action</th>
<th>Goto</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>s1</td>
<td>7 2</td>
</tr>
<tr>
<td>1</td>
<td>r4 s3 r4</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>s4 r2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>s1</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>s1</td>
<td>6 2</td>
</tr>
</tbody>
</table>
Converting the LR(0) Automaton to an SLR Parsing Table

<table>
<thead>
<tr>
<th>State</th>
<th>Action</th>
<th>Goto</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>s1</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>r4</td>
<td>s3</td>
</tr>
<tr>
<td>2</td>
<td>s4</td>
<td>r2</td>
</tr>
<tr>
<td>3</td>
<td>s1</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>s1</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>r3</td>
<td>r3</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>State</th>
<th>Action</th>
<th>Goto</th>
</tr>
</thead>
<tbody>
<tr>
<td>S0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1: \( e \rightarrow t + e \)
2: \( e \rightarrow t \)
3: \( t \rightarrow \text{Id} \ast t \)
4: \( t \rightarrow \text{Id} \)

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

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: \( t \rightarrow \text{Id} \ast t \)
4: \( t \rightarrow \text{Id} \)

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: \( t \rightarrow \text{Id} \ast t \)  
4: \( t \rightarrow \text{Id} \)

<table>
<thead>
<tr>
<th>State</th>
<th>Action</th>
<th>Goto</th>
<th>Input</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>s1</td>
<td></td>
<td>\text{Id} \ast \text{Id} + \text{Id}$</td>
<td>Shift, goto 1</td>
</tr>
<tr>
<td>1</td>
<td>r4</td>
<td>s3</td>
<td>\text{Id} \ast \text{Id} + \text{Id}$</td>
<td>Shift, goto 3</td>
</tr>
<tr>
<td>2</td>
<td>s4</td>
<td></td>
<td>\text{Id} \ast \text{Id} + \text{Id}$</td>
<td>Reduce 2</td>
</tr>
<tr>
<td>3</td>
<td>s1</td>
<td></td>
<td>\text{Id} \ast \text{Id} + \text{Id}$</td>
<td>Reduce 3</td>
</tr>
<tr>
<td>4</td>
<td>s1</td>
<td></td>
<td>\text{Id} \ast \text{Id} + \text{Id}$</td>
<td>Reduce 1</td>
</tr>
<tr>
<td>5</td>
<td>r3</td>
<td></td>
<td>\text{Id} \ast \text{Id} + \text{Id}$</td>
<td>Accept</td>
</tr>
</tbody>
</table>

Here, the state is 1, the next symbol is \( \ast \), 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: $t \rightarrow \text{Id} \ast t$
4: $t \rightarrow \text{Id}$

<table>
<thead>
<tr>
<th>State</th>
<th>Action</th>
<th>Goto</th>
</tr>
</thead>
<tbody>
<tr>
<td>Id</td>
<td>+</td>
<td>e t</td>
</tr>
<tr>
<td>0</td>
<td>s1</td>
<td>7 2</td>
</tr>
<tr>
<td>1</td>
<td>r4 s3 r4</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>s4 r2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>s1</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>s1</td>
<td>6 2</td>
</tr>
<tr>
<td>5</td>
<td>r3 r3</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>r1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

Here, the state is 1, the next symbol is $\ast$. The table says reduce using rule 4.
# Shift/Reduce Parsing with an SLR Table

1: \( e \rightarrow t + e \)

2: \( e \rightarrow t \)

3: \( t \rightarrow \text{Id} \ast t \)

4: \( t \rightarrow \text{Id} \)

<table>
<thead>
<tr>
<th>State</th>
<th>Action</th>
<th>Goto</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>s1</td>
<td>7 2</td>
</tr>
<tr>
<td>1</td>
<td>r4</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>s4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>s1</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>s1</td>
<td>6 2</td>
</tr>
<tr>
<td>5</td>
<td>r3</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>r1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

Remove the RHS of the rule (here, just \( \text{Id} \)), 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:  \( t \rightarrow \text{id} \ast t \)
4:  \( t \rightarrow \text{id} \)

<table>
<thead>
<tr>
<th>State</th>
<th>Action</th>
<th>Goto</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>\text{id}</td>
<td>\text{+}</td>
</tr>
<tr>
<td>0</td>
<td>\text{s1}</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>\text{r4}</td>
<td>\text{s3}</td>
</tr>
<tr>
<td>2</td>
<td>\text{s4}</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>\text{s1}</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>\text{s1}</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>\text{r3}</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>\text{r1}</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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: \( e \to t + e \)
2: \( e \to t \)
3: \( t \to \text{Id} \ast t \)
4: \( t \to \text{Id} \)

<table>
<thead>
<tr>
<th>State</th>
<th>Action</th>
<th>Goto</th>
<th>Input</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>s1</td>
<td>e t</td>
<td>0 Id</td>
<td>Shift, goto 1</td>
</tr>
<tr>
<td>1</td>
<td>r4</td>
<td>s3</td>
<td>1 Id *</td>
<td>Shift, goto 3</td>
</tr>
<tr>
<td>2</td>
<td>s4</td>
<td>r2</td>
<td>1 Id +</td>
<td>Shift, goto 1</td>
</tr>
<tr>
<td>3</td>
<td>s1</td>
<td></td>
<td>1 Id *</td>
<td>Reduce 4</td>
</tr>
<tr>
<td>4</td>
<td>s1</td>
<td></td>
<td>1 Id +</td>
<td>Reduce 3</td>
</tr>
<tr>
<td>5</td>
<td>r3</td>
<td>r3</td>
<td>e 6</td>
<td>Reduce 1</td>
</tr>
<tr>
<td>6</td>
<td>r1</td>
<td></td>
<td>t 2</td>
<td>Accept</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

This time, we strip off the RHS for rule 3, \( \text{Id} \ast t \), 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: \( t \rightarrow \text{Id} \ast t \)
4: \( t \rightarrow \text{Id} \)

<table>
<thead>
<tr>
<th>State</th>
<th>Action</th>
<th>Goto</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>s1</td>
<td>7 2</td>
</tr>
<tr>
<td>1</td>
<td>r4 s3 r4</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>s4 r2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>s1</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>s1</td>
<td>6 2</td>
</tr>
<tr>
<td>5</td>
<td>r3 r3</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>r1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stack</th>
<th>Input</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 0 )</td>
<td>( \text{Id} \ast \text{Id} + \text{Id} )</td>
<td>Shift, goto 1</td>
</tr>
<tr>
<td>( 0 )</td>
<td>( \text{Id} )</td>
<td></td>
</tr>
<tr>
<td>( 0 )</td>
<td>( \ast \text{Id} + \text{Id} )</td>
<td>Shift, goto 3</td>
</tr>
<tr>
<td>( 0 )</td>
<td>( \text{Id} + \text{Id} )</td>
<td>Shift, goto 1</td>
</tr>
<tr>
<td>( 0 )</td>
<td>( \text{Id} \ast \text{Id} )</td>
<td>Reduce 4</td>
</tr>
<tr>
<td>( 0 )</td>
<td>( \text{Id} )</td>
<td>Reduce 3</td>
</tr>
<tr>
<td>( 0 )</td>
<td>( t + t )</td>
<td>Shift, goto 4</td>
</tr>
<tr>
<td>( 0 )</td>
<td>( \text{Id} )</td>
<td>Shift, goto 1</td>
</tr>
<tr>
<td>( 0 )</td>
<td>( t + t )</td>
<td>Reduce 4</td>
</tr>
<tr>
<td>( 0 )</td>
<td>( t + t )</td>
<td>Reduce 2</td>
</tr>
<tr>
<td>( 0 )</td>
<td>( t + e )</td>
<td>Reduce 1</td>
</tr>
<tr>
<td>( 0 )</td>
<td>( e )</td>
<td>Accept</td>
</tr>
</tbody>
</table>
Types

A restriction on the possible interpretations of a segment of memory or other program construct.

Two uses:

**Safety:** avoids data being treated as something it isn’t

**Optimization:** eliminates certain runtime decisions
# Types of Types

<table>
<thead>
<tr>
<th>Type</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic</td>
<td>Machine words, floating-point numbers,</td>
</tr>
<tr>
<td></td>
<td>addresses/pointers</td>
</tr>
<tr>
<td>Aggregate</td>
<td>Arrays, structs, classes</td>
</tr>
<tr>
<td>Function</td>
<td>Function pointers, lambdas</td>
</tr>
</tbody>
</table>
## Basic Types

Groups of data the processor is designed to operate on.

On an ARM processor,

<table>
<thead>
<tr>
<th>Type</th>
<th>Width (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unsigned/two’s-complement binary</strong></td>
<td></td>
</tr>
<tr>
<td>Byte</td>
<td>8</td>
</tr>
<tr>
<td>Halfword</td>
<td>16</td>
</tr>
<tr>
<td>Word</td>
<td>32</td>
</tr>
<tr>
<td><strong>IEEE 754 Floating Point</strong></td>
<td></td>
</tr>
<tr>
<td>Single-Precision scalars &amp; vectors</td>
<td>32, 64, .., 256</td>
</tr>
<tr>
<td>Double-Precision scalars &amp; vectors</td>
<td>64, 128, 192, 256</td>
</tr>
</tbody>
</table>
Derived types

**Array**: a list of objects of the same type, often fixed-length

**Record**: a collection of named fields, often of different types

**Pointer/References**: a reference to another object

**Function**: a reference to a block of code
C’s Declarations and Declarators

Declaration: list of specifiers followed by a comma-separated list of declarators.

```
static unsigned int(*f[10])(int, char*);
```

Specifiers
Declarator

Declarator’s notation matches that of an expression: use it to return the basic type.

Largely regarded as the worst syntactic aspect of C: both pre- (pointers) and post-fix operators (arrays, functions).
Structs are the precursors of objects:

Group and restrict what can be stored in an object, but not what operations they permit.

Can fake object-oriented programming:

```c
struct poly { ... };

struct poly *poly_create();
void poly_destroy(struct poly *p);
void poly_draw(struct poly *p);
void poly_move(struct poly *p, int x, int y);
int poly_area(struct poly *p);
```
A struct holds all of its fields at once. A union holds only one of its fields at any time (the last written).

```c
union token {
    int i;
    float f;
    char *string;
};

union token t;
t.i = 10;
t.f = 3.14159;    /* overwrite t.i */
char *s = t.string;  /* return gibberish */
```
Applications of Variant Records

A primitive form of polymorphism:

```c
struct poly {
    int x, y;
    int type;
    union {
        int radius;
        int size;
        float angle;
    } d;
};
```

If `poly.type == CIRCLE`, use `poly.d.radius`.
If `poly.type == SQUARE`, use `poly.d.size`.
If `poly.type == LINE`, use `poly.d.angle`. 
Name vs. Structural Equivalence

```c
struct f {
    int x, y;
} foo = { 0, 1 };

struct b {
    int x, y;
} bar;

bar = foo;
```

Is this legal in C? Should it be?
C’s declarators are unusual: they always specify a name along with its type.

Languages more often have type expressions: a grammar for expressing a type.

Type expressions appear in three places in C:

```
(int *) a /* Type casts */
sizeof(float [10]) /* Argument of sizeof() */
int f(int, char *, int (*)(int)) /* Function argument types */
```
Static Semantic Analysis

Lexical analysis: Make sure tokens are valid

```java
if i 3 "This" /* valid Java tokens */
#a1123 /* not a token */
```

Syntactic analysis: Makes sure tokens appear in correct order

```java
for ( i = 1 ; i < 5 ; i++ ) 3 + "foo"; /* valid Java syntax */
for break /* invalid syntax */
```

Semantic analysis: Makes sure program is consistent

```java
int v = 42 + 13; /* valid in Java (if v is new) */
return f + f(3); /* invalid */
```
What To Check

Examples from Java:

Verify names are defined and are of the right type.

```java
int i = 5;
int a = z;    /* Error: cannot find symbol */
int b = i[3]; /* Error: array required, but int found */
```

Verify the type of each expression is consistent.

```java
int j = i + 53;
int k = 3 + "hello";    /* Error: incompatible types */
int l = k(42);          /* Error: k is not a method */
if ("Hello") return 5;  /* Error: incompatible types */
String s = "Hello";
int m = s;              /* Error: incompatible types */
```
How To Check: Depth-first AST Walk

Checking function: environment → node → type

- $1 - 5$
- $1 + "Hello"

check(-)
  check(1) = int
  check(5) = int
  Success: int – int = int

check(+)
  check(1) = int
  check("Hello") = string
  FAIL: Can’t add int and string

Ask yourself: at each kind of node, what must be true about the nodes below it? What is the type of the node?
How To Check: Symbols

Checking function: environment $\rightarrow$ node $\rightarrow$ type

\[
\begin{array}{c}
1 + a \\
+ \\
1 \\
1 + a
\end{array}
\]

check(+)  
check(1) = int  
check(a) = int  
Success: int + int = int

The key operation: determining the type of a symbol when it is encountered.

The environment provides a “symbol table” that holds information about each in-scope symbol.
A name begins life where it is declared and ends at the end of its block.

From the CLRM, “The scope of an identifier declared at the head of a block begins at the end of its declarator, and persists to the end of the block.”

```c
void foo()
{
    int x;
}
```
Hiding a Definition

Nested scopes can hide earlier definitions, giving a hole.

From the CLRM, “If an identifier is explicitly declared at the head of a block, including the block constituting a function, any declaration of the identifier outside the block is suspended until the end of the block.”

```c
void foo()
{
    int x;
    while ( a < 10 ) {
        int x;
    }
}
```
public void example() {
  // x, y, z not visible

  int x;
  // x visible

  for ( int y = 1 ; y < 10 ; y++ ) {
    // x, y visible

    int z;
    // x, y, z visible
  }
  // x visible

}
Basic Static Scope in O’Caml

A name is bound after the “in” clause of a “let.” If the name is re-bound, the binding takes effect after the “in.”

```
let x = 8 in
let x = x + 1 in
```

Returns the pair (12, 8):

```
let x = 8 in
( let x = x + 2 in
  x + 2 ),
x
```
Let Rec in O’Caml

The “rec” keyword makes a name visible to its definition. This only makes sense for functions.

```ocaml
let rec fib i = if i < 1 then 1 else fib (i-1) + fib (i-2) in fib 5

(* Nonsensical *)
let rec x = x + 3 in
```
Let...and in O’Caml

Let...and lets you bind multiple names at once. Definitions are not mutually visible unless marked “rec.”

```ocaml
let x = 8
and y = 9 in

let rec fac n =
  if n < 2 then
   1
  else
   n * fac n
and fac1 n = fac (n - 1)
in
fac 5
```
Nesting Function Definitions

let articles words =

  let report w =

    let count = List.length (List.filter ((=) w) words)
    in w ^ ": " ^ string_of_int count

    in String.concat "", "
        (List.map report ["a"; "the"])

    in articles
        ["the"; "plt"; "class"; "is"; "a"; "pain"; "in";
         "the"; "butt"]

let count words w = List.length (List.filter ((=) w) words) in

let report words w = w ^ ": " ^ string_of_int (count words w) in

let articles words =
  String.concat "", "
    (List.map (report words)
        ["a"; "the"]) in

articles
    ["the"; "plt"; "class"; "is"; "a"; "pain"; "in";
     "the"; "butt"]

Produces “a: 1, the: 2”
int p(int i) {
    printf("%d ", i);
    return i;
}

void q(int a, int b, int c) {
    int total = a;
    printf("%d ", b);
    total += c;
    total += c;
}

q( p(1), 2, p(3) );

What does this print?
Applicative- vs. and Normal-Order

Most languages use applicative order.

Macro-like languages often use normal order.

```c
#define p(x) (printf("%d ",x), x)
#define q(a,b,c) total = (a), \n printf("%d ", (b)), \n total += (c)

q( p(1), 2, p(3) );
```

Prints 1 2 3.

Some functional languages also use normal order evaluation to avoid doing work. “Lazy Evaluation”
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.

Diagram:
- **Stack**
- **Heap**
- **Static**
- **Code**

Legend:
- **High memory**
- **Low memory**
- **Stack pointer**
- **Program break**
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`

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

(Real C)

```c
int fib(int n) {
    if (n<2)
        return 1;
    else
        return fib(n-1) + fib(n-2);
}
```

(Assembly-like C)

```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;
}
```

```
\[
\begin{align*}
\text{fib(3)} & \quad \text{fib(2)} \quad \text{fib(1)} \\
\text{fib(2)} & \quad \text{fib(1)} \\
\text{fib(1)} & \quad \text{fib(0)}
\end{align*}
\]
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;
}
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)

```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;
}
```

Executing fib(3)

```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;
}
```
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;
}
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)

```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;
}
```

```
n = 3
return address
last frame pointer
tmp1 = 2
tmp2 =
tmp3 =
n = 2
return address
last frame pointer
tmp1 = 2
tmp2 = 1
tmp3 = 1
```
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;
}
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;
}
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;
}
Local arrays with fixed size are easy to stack.

```
void foo()
{
    int a;
    int b[10];
    int c;
}
```
Variable-sized local arrays aren’t as easy.

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

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

Variables remain constant offset from frame pointer.
let articles words =

  let report w =

    let count = List.length (List.filter ((=) w) words)
    in w ^ "": " " ^ string_of_int count

    in String.concat ”, ”
      (List.map report ["a"; "the"])

  in String.concat ”, ”
    (List.map report ["a"; "the"])

  in articles
    ["the"; "plt"; "class"; "is"; "a"; "pain"; "in"; "the"; "butt"]

let count words w = List.length (List.filter ((=) w) words) in

let report words w = w ^ "": " " ^ string_of_int (count words w) in

let articles words =
    String.concat ”, ”
      (List.map (report words) ["a"; "the"])

  in articles
    ["the"; "plt"; "class"; "is"; "a"; "pain"; "in"; "the"; "butt"]

Produces “a: 1, the: 2”
What does “a 5 42” give?
Implementing Nested Functions with Access Links

```
let a x s =
  let b y =
    let c z = z + s in
    let d w = c (w+1) in
    d (y+1) in (* b *)
  in
  let e q = b (q+1) in
  e (x+1) (* a *)
```

What does “a 5 42” give?

<table>
<thead>
<tr>
<th>a:</th>
<th>x = 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>s = 42</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>e:</th>
<th>q = 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(access link)</td>
</tr>
</tbody>
</table>
Implementing Nested Functions with Access Links

```
let a x s =
    let b y =
        let c z = z + s in
        let d w = c (w+1) in
        d (y+1) in (* b *)
    let e q = b (q+1) in
    e (x+1) (* a *)
```

What does “a 5 42” give?
Implementing Nested Functions with Access Links

```plaintext
let a x s =
    let b y =
        let c z = z + s in
        let d w = c (w+1) in
        d (y+1) (* b *)
    let e q = b (q+1) in
    e (x+1) (* a *)
```

What does “a 5 42” give?

### Variables:
- `x = 5`
- `s = 42`
- `q = 6`
- `y = 7`
- `w = 8`
```
let a x s =
  let b y =
    let c z = z + s in
    let d w = c (w+1) in
    d (y+1) in (* b *)
  let e q = b (q+1) in
  e (x+1) in (* a *)
```

What does “a 5 42” give?

- **a:**
  - $x = 5$
  - $s = 42$

- **b:**
  - $y = 7$

- **c:**
  - $z = 9$

- **d:**
  - $w = 8$

- **e:**
  - $q = 6$
Layout of Records and Unions

Modern processors have byte-addressable memory.

The IBM 360 (c. 1964) helped to popularize byte-addressable memory.

Many data types (integers, addresses, floating-point numbers) are wider than a byte.

16-bit integer: 1 0
32-bit integer: 3 2 1 0
Modern memory systems read data in 32-, 64-, or 128-bit chunks:

Reading an aligned 32-bit value is fast: a single operation.

It is harder to read an unaligned value: two reads plus shifting.

SPARC and ARM prohibit unaligned accesses.

MIPS has special unaligned load/store instructions.

x86, 68k run more slowly with unaligned accesses.
Padding

To avoid unaligned accesses, the C compiler pads the layout of unions and records.

Rules:

- Each \( n \)-byte object must start on a multiple of \( n \) bytes (no unaligned accesses).
- Any object containing an \( n \)-byte object must be of size \( mn \) for some integer \( m \) (aligned even when arrayed).

```c
struct padded {
    int x;    /* 4 bytes */
    char z;   /* 1 byte */
    short y;  /* 2 bytes */
    char w;   /* 1 byte */
};
```

```c
struct padded {
    char a;   /* 1 byte */
    short b;  /* 2 bytes */
    short c;  /* 2 bytes */
};
```
Unions

A C struct has a separate space for each field; a C union shares one space among all fields

```c
union intchar {
    int i;  /* 4 bytes */
    char c; /* 1 byte */
};

union twostructs {
    struct {
        char c;    /* 1 byte */
        int i;     /* 4 bytes */
    } a;
    struct {
        short s1; /* 2 bytes */
        short s2; /* 2 bytes */
    } b;
};
```

or

```c

or

```c
short s2; /* 2 bytes */
short s1; /* 2 bytes */
```
Arrays

Basic policy in C: an array is just one object after another in memory.

```c
int a[10];
```

This is why you need padding at the end of `structs`.

```c
struct {
    int a;
    char c;
} b[2];
```
Arrays and Aggregate types

The largest primitive type dictates the alignment

```c
struct {
    short a;
    short b;
    char c;
} d[4];
```
Arrays of Arrays

```c
char a[4];
char a[3][4];
```
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)
```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

↓ free( )
Dynamic Storage Allocation

```
↓ free()

↓ malloc()
```
Dynamic Storage Allocation

\[
\text{free(} \quad \text{malloc(}
\]

\[
\downarrow \text{free(} \quad \text{malloc(}
\]

\[
\downarrow \text{malloc(}
\]
Dynamic Storage Allocation

\[
\downarrow \text{free( )}
\]

\[
\downarrow \text{malloc( )}
\]
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

```c
malloc()
free()
```
Simple Dynamic Storage Allocation

malloc( )

free( )
Simple Dynamic Storage Allocation

malloc

malloc( )
Simple Dynamic Storage Allocation

malloc( )

free( ; )
Simple Dynamic Storage Allocation

malloc( )

free( )
Fragmentation

malloc( ) seven times give

free() four times gives

malloc( ) ?

Need more memory; can’t use fragmented memory.

Hockey smile
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

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

```
|   | 0 | 42, 17 |
```
Reference Counting

What and when to free?

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

```ml
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

```ocaml
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

```ml
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

```plaintext
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

```ml
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

```plaintext
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:

![Diagram](display)  

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

```plaintext
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

```ocaml
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++

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

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

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

Equivalent C

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

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

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

```cpp
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;
```
C++’s Exceptions

```c++
struct Except {} ex; // This struct functions as an exception

void top(void) {
    try {
        child();
    } catch (Except e) { // throw sends control here
        printf("oops\n");
    }
}

void child() {
    child2();
}

void child2() {
    throw ex; // Pass control up to the catch block
}
```
C’s setjmp/longjmp: Idiosyncratic Exceptions

```c
#include <setjmp.h>

jmp_buf closure;  /* return address, stack & frame ptrs. */

void top(void) {
    switch ( setjmp(closure) ) { /* normal: store closure, return 0 */
        /* longjmp jumps here, returns 1 */
    case 0:
        child();  /* unexceptional case */
        break;
    case 1:
        break;  /* longjmp( ,1) called */
    }
}

void child() {
    child2();
}

void child2() {
    longjmp(closure, 1);
}
```
Implementing Exceptions

One way: maintain a stack of exception handlers

```c
try {
    child();
} catch (Ex e) {
    foo();
}

void child() {
    child2();
}

void child2() {
    throw ex;
}
```

```c
push(Ex, Handler); // Push handler on stack

child();
pop();            // Normal termination
goto Exit;        // Jump over "catch"

Handler:
    foo();          // Body of "catch"
Exit:

void child() {
    child2();
}

void child2() {
    throw(ex);       // Unroll stack; find handler
}
```

Incurs overhead, even when no exceptions thrown
Q: When an exception is *thrown*, where was the last *try*?

A: Consult a table: relevant handler or “pop” for every PC

```
void foo() {
    try {
        bar();
    } catch (Ex1 e) {
        a();
    }
}

void bar() {
    baz();
}

void baz() {
    try {
        throw ex1;
    } catch (Ex2 e) {
        b();
    }
}
```

<table>
<thead>
<tr>
<th>Lines</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–2</td>
<td>Pop stack</td>
</tr>
<tr>
<td>3–5</td>
<td>Handler @ 5 for Ex1</td>
</tr>
<tr>
<td>6–15</td>
<td>Pop stack</td>
</tr>
<tr>
<td>16–18</td>
<td>Handler @ 14 for Ex2</td>
</tr>
<tr>
<td>19–21</td>
<td>Pop stack</td>
</tr>
</tbody>
</table>
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
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
```c
int gcd(int a, int b) {
    while (a != b) {
        if (a < b) b -= a;
        else a -= b;
    }
    return a;
}
```

### GCC on SPARC
```assembly
    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
    .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
```
Typical Optimizations

- Folding constant expressions
  \[1 + 3 \rightarrow 4\]

- Removing dead code
  \[
  \text{if (0)} \{ \ldots \} \rightarrow \text{nothing}
  \]

- Moving variables from memory to registers
  \[
  \text{ld} \quad [%fp+68], \%i1 \\
  \text{sub} \quad \%i0, \%i1, \%i0 \rightarrow \text{sub} \quad \%o1, \%o0, \%o1 \\
  \text{st} \quad \%i0, [%fp+72]
  \]

- 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.

\[
\begin{align*}
a &= c + 5 + 3; \\
\text{if } (0 + 3) \{ & \quad \rightarrow \quad b = a = c + 8; \\
& \quad b = c + 8;
\}
\end{align*}
\]

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
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.
A CFG illustrates the flow of control among basic blocks.

A:
  sne t, a, b
  bz E, t
  slt t, a, b
  bnez B, t
  sub b, b, a
  jmp C

B:
  sub a, a, b

C:
  jmp A

E:
  ret a

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

slt t, a, b
bnez B, t

sub b, b, a
jmp C

B:
  sub a, a, b

C:
  jmp A

E:
  ret a
Separate Compilation and Linking

- **Compiler**: foo.c ➔ foo.s ➔ foo.o
- **Assembler**: bar.c ➔ bar.s ➔ bar.o
- **Archiver**: printf.o ➔ fopen.o ➔ malloc.o ➔ libc.a
- **Linker**: foo.o ➔ bar.o ➔ printf.o ➔ fopen.o ➔ malloc.o ➔ libc.a ➔ foo
- **Id**: foo ➔ bar

**Tools**:
- cc
- as
- ar
Linking

Goal of the linker is to combine the disparate pieces of the program into a coherent whole.

file1.c:
```c
#include <stdio.h>
char a[] = "Hello";
extern void bar();

int main() {
    bar();
}

void baz(char *s) {
    printf("%s", s);
}
```

file2.c:
```c
#include <stdio.h>
extern char a[];
static char b[6];

void bar() {
    strcpy(b, a);
    baz(b);
}
```

libc.a:
```c
int printf(char *s, ...)
{
    /* ... */
}
char *
strcpy(char *d, char *s)
{
    /* ... */
}
```
Linking

Goal of the linker is to combine the disparate pieces of the program into a coherent whole.

file1.c:
#include <stdio.h>
char a[] = "Hello";
extern void bar();
int main() {
    bar();
}
void baz(char *s) {
    printf("%s", s);
}

file2.c:
#include <stdio.h>
extern char a[];
static char b[6];
void bar() {
    strcpy(b, a);
    baz(b);
}

libc.a:
int printf(char *s, ...)
{
    /* ... */
}
char *
strcpy(char *d, char *s)
{
    /* ... */
}
Linking

file1.o
- a="Hello"
- main()
- baz()

file2.o
- char b[6]
- bar()
Linking

- `file1.o`:
  - `.text` segment:
    - `main()`
    - `baz()`
    - `a="Hello"`
  - `.data` segment:
    - `bar()`
  - `.bss` segment:
    - `char b[6]`

- `file2.o`:
  - `.text` segment:
    - `bar()`
  - `.data` segment:
    - `bar()`
  - `.bss` segment:
    - `char b[6]`

- `a.out`:
  - `.text` segment:
    - `main()`
    - `baz()`
    - `bar()`
  - `.data` segment:
    - `a="Hello"`
  - `.bss` segment:
    - `char b[6]`

Code of program

- `.text`
- `.data`
- `.bss`

Initialized data

Uninitialized data

"Block Started by Symbol"
Relocatable: Many need to be pasted together. Final in-memory address of code not known when program is compiled

Object files contain

- imported symbols (unresolved “external” symbols)
- relocation information (what needs to change)
- exported symbols (what other files may refer to)
file1.c:

```c
#include <stdio.h>
char a[] = "Hello";
extern void bar();

int main() {
    bar();
}

void baz(char *s) {
    printf("%s", s);
}
```

exported symbols

imported symbols
# Object Files

## file1.c:

```c
#include <stdio.h>
char a[] = "Hello";
extern void bar();

int main() {
    bar();
}

void baz(char *s) {
    printf("%s", s);
}
```

```bash
# objdump -x file1.o
Sections:
Idx Name     Size  VMA LMA Offset  Algn
  0 .text  038 0  0  034 2**2
  1 .data  008 0  0  070 2**3
  2 .bss   000 0  0  078 2**0
  3 .rodata 008 0  0  078 2**3

SYMBOL TABLE:
0000 g O .data  006  a
0000 g F .text  014  main
0000  *UND*  000  bar
0014 g F .text  024  baz
0000  *UND*  000  printf

RELOCATION RECORDS FOR [.text]:
OFFSET TYPE      VALUE
0004  R_SPARC_WDISP30  bar
001c  R_SPARC_HI22     .rodata
0020  R_SPARC_LO10     .rodata
0028  R_SPARC_WDISP30  printf
```
# include <stdio.h>
char a[] = "Hello";
extern void bar();

int main() {
    bar();
}

void baz(char *s) {
    printf("%s", s);
}

# objdump -d file1.o
0000 <main>:
  0: 9d e3 bf 90 save %sp, -112, %sp
  4: 40 00 00 00 call 4 <main+0x4>
     4: R_SPARC_WDISP30 bar
  8: 01 00 00 00 nop
c: 81 c7 e0 08 ret
10: 81 e8 00 00 restore

0014 <baz>:
 14: 9d e3 bf 90 save %sp, -112, %sp
 18: f0 27 a0 44 st %i0, [ %fp + 0x44 ]
1c: 11 00 00 00 sethi %hi(0), %o0
     1c: R_SPARC_HI22 .rodata
 20: 90 12 20 00 mov %o0, %o0
     20: R_SPARC_LO10 .rodata
 24: d2 07 a0 44 ld [ %fp + 0x44 ], %o1
 28: 40 00 00 00 call 28 <baz+0x14>
     28: R_SPARC_WDISP30 printf
 2c: 01 00 00 00 nop
30: 81 c7 e0 08 ret
34: 81 e8 00 00 restore
Before and After Linking

- Combine object files
- Relocate each function’s code
- Resolve previously unresolved symbols

```
int main() {
    bar();
}

void baz(char *s) {
    printf("%s", s);
}
```

105f8 <main>:
0: 9d e3 bf 90 save %sp, -112, %sp
4: 40 00 00 00 call 4 <main+0x4>
   4: R_SPARC_WDISP30 bar
8: 01 00 00 00 nop
c: 81 c7 e0 08 ret
10: 81 e8 00 00 restore

1060c <baz>:
14: 9d e3 bf 90 save %sp, -112, %sp
18: f0 27 a0 44 st %i0, [%fp + 0x44]
1c: 11 00 00 00 sethi %hi(0), %o0
   1c: R_SPARC_HI22 .rodata
20: 90 12 23 00 mov %o0, %o0
   20: R_SPARC_LO10 .rodata
24: d2 07 a0 44 ld [%fp + 0x44], %o1
28: 40 00 00 00 call 28 <baz+0x14>
   28: R_SPARC_WDISP30 printf
2c: 01 00 00 00 nop
30: 81 c7 e0 08 ret
34: 81 e8 00 00 restore
Linking Resolves Symbols

file1.c:

```c
#include <stdio.h>
char a[] = "Hello";
extern void bar();

int main() {
    bar();
}

void baz(char *s) {
    printf("%s", s);
}
```

file2.c:

```c
#include <stdio.h>
extern char a[];
static char b[6];

void bar() {
    strcpy(b, a);
    baz(b);
}
```
Lambda Expressions

Function application written in prefix form. “Add four and five” is

\[ (+ \ 4 \ 5) \]

Evaluation: select a redex and evaluate it:

\[ (+ \ (* \ 5 \ 6) \ (* \ 8 \ 3)) \rightarrow (+ \ 30 \ (* \ 8 \ 3)) \]
\[ \rightarrow (+ \ 30 \ 24) \]
\[ \rightarrow 54 \]

Often more than one way to proceed:

\[ (+ \ (* \ 5 \ 6) \ (* \ 8 \ 3)) \rightarrow (+ \ (* \ 5 \ 6) \ 24) \]
\[ \rightarrow (+ \ 30 \ 24) \]
\[ \rightarrow 54 \]

Function Application and Currying

Function application is written as juxtaposition:

\[ f \ x \]

Every function has exactly one argument. Multiple-argument functions, e.g., \(+\), are represented by currying, named after Haskell Brooks Curry (1900–1982). So,

\[(+ \ x)\]

is the function that adds \(x\) to its argument.

Function application associates left-to-right:

\[ (+ 3 4) = ((+ 3) 4) \rightarrow 7 \]
The only other thing in the lambda calculus is *lambda abstraction*: a notation for defining unnamed functions.

\[(\lambda x . + x 1)\]

That function of \(x\) that adds \(x\) to 1
The Syntax of the Lambda Calculus

\[
\text{expr} ::= \text{expr expr} \\
| \lambda \text{variable . expr} \\
| \text{constant} \\
| \text{variable} \\
| (\text{expr})
\]

Constants are numbers and built-in functions; variables are identifiers.
Beta-Reduction

Evaluation of a lambda abstraction—*beta-reduction*—is just substitution:

\[(\lambda x . \, + \, x \, 1) \, 4 \rightarrow (+ \, 4 \, 1) \rightarrow 5\]

The argument may appear more than once

\[(\lambda x . \, + \, x \, x) \, 4 \rightarrow (+ \, 4 \, 4) \rightarrow 8\]

or not at all

\[(\lambda x . \, 3) \, 5 \rightarrow 3\]
Free and Bound Variables

Here, $x$ is like a function argument but $y$ is like a global variable.

Technically, $x$ occurs bound and $y$ occurs free in

$$(\lambda x . + x y)$$

However, both $x$ and $y$ occur free in

$$(+ x y)$$
Beta-Reduction More Formally

\[(\lambda x . E) \ F \rightarrow_\beta \ E'\]

where \( E' \) is obtained from \( E \) by replacing every instance of \( x \) that appears free in \( E \) with \( F \).

The definition of free and bound mean variables have scopes. Only the rightmost \( x \) appears free in

\[(\lambda x . \ + \ (\ - \ x \ 1)) \ x \ 3\]

so

\[(\lambda x . \ (\lambda x . \ + \ (\ - \ x \ 1)) \ x \ 3) \ 9 \rightarrow (\lambda x . \ + \ (\ - \ x \ 1)) \ 9 \ 3 \rightarrow \ + \ (-9 \ 1) \ 3 \rightarrow \ + \ 8 \ 3 \rightarrow \ 11\]
One way to confuse yourself less is to do $\alpha$-conversion: renaming a $\lambda$ argument and its bound variables.

Formal parameters are only names: they are correct if they are consistent.

\[
(\lambda x . (\lambda x . + (− x 1)) x 3) 9 \leftrightarrow (\lambda x . (\lambda y . + (− y 1)) x 3) 9 \\
\rightarrow (((\lambda y . + (− y 1)) 9 3) \\
\rightarrow (+ (− 9 1) 3) \\
\rightarrow (+ 8 3) \\
\rightarrow 11
\]
Beta-Abstraction and Eta-Conversion

Running $\beta$-reduction in reverse, leaving the “meaning” of a lambda expression unchanged, is called *beta abstraction*:

\[ + 4 1 \leftarrow (\lambda x \cdot + x 1) 4 \]

Eta-conversion is another type of conversion that leaves “meaning” unchanged:

\[ (\lambda x \cdot + 1 x) \leftrightarrow_\eta (+ 1) \]

Formally, if $F$ is a function in which $x$ does not occur free,

\[ (\lambda x \cdot F x) \leftrightarrow_\eta F \]
Reduction Order

The order in which you reduce things can matter.

\[(\lambda x . \lambda y . y) \ ( (\lambda z . z \ z) \ (\lambda z . z \ z) )\]

Two things can be reduced:

\[(\lambda z . z \ z) \ (\lambda z . z \ z)\]

\[(\lambda x . \lambda y . y) \ (\cdots)\]

However,

\[(\lambda z . z \ z) \ (\lambda z . z \ z) \rightarrow (\lambda z . z \ z) \ (\lambda z . z \ z)\]

\[(\lambda x . \lambda y . y) \ (\cdots) \rightarrow (\lambda y . y)\]
Normal Form

A lambda expression that cannot be $\beta$-reduced is in *normal form*. Thus,

\[
\lambda y \cdot y
\]

is the normal form of

\[
(\lambda x . \lambda y . y) \ ((\lambda z . z \ z) \ (\lambda z . z \ z))
\]

Not everything has a normal form. E.g.,

\[
(\lambda z . z \ z) \ (\lambda z . z \ z)
\]

can only be reduced to itself, so it never produces a non-reducible expression.
Normal Form

Can a lambda expression have more than one normal form?

**Church-Rosser Theorem I:** If $E_1 \leftrightarrow E_2$, then there exists an expression $E$ such that $E_1 \rightarrow E$ and $E_2 \rightarrow E$.

**Corollary.** No expression may have two distinct normal forms.

**Proof.** Assume $E_1$ and $E_2$ are distinct normal forms for $E$: $E \leftrightarrow E_1$ and $E \leftrightarrow E_2$. So $E_1 \leftrightarrow E_2$ and by the Church-Rosser Theorem I, there must exist an $F$ such that $E_1 \rightarrow F$ and $E_2 \rightarrow F$. However, since $E_1$ and $E_2$ are in normal form, $E_1 = F = E_2$, a contradiction.
Normal-Order Reduction

Not all expressions have normal forms, but is there a reliable way to find the normal form if it exists?

**Church-Rosser Theorem II:** If $E_1 \rightarrow E_2$ and $E_2$ is in normal form, then there exists a *normal order* reduction sequence from $E_1$ to $E_2$.

*Normal order reduction:* reduce the leftmost outermost redex.
Normal-Order Reduction

\[
\left(\left(\lambda x \bdv (\lambda w \cdot \lambda z \cdot w \cdot z) \cdot 1\right) \cdot (\lambda x \cdot x \cdot x) \cdot (\lambda y \cdot y) \cdot 1\right) + 2
\]
Recursion

Where is recursion in the lambda calculus?

\[
FAC = \left( \lambda n . \ IF \ (= \ n \ 0) \ 1 \ (* \ n \ (FAC \ (- \ n \ 1))) \right)
\]

This does not work: functions are unnamed in the lambda calculus. But it is possible to express recursion as a function.

\[
FAC = (\lambda n . \ldots \ FAC \ldots)
\leftarrow \beta \ (\lambda f . \ (\lambda n . \ldots f \ldots)) \ FAC
= \ H \ FAC
\]

That is, the factorial function, \(FAC\), is a fixed point of the (non-recursive) function \(H\):

\[
H = \lambda f . \lambda n . \ IF \ (= \ n \ 0) \ 1 \ (* \ n \ (f \ (- \ n \ 1)))
\]
Recursion

Let’s invent a function $Y$ that computes $FAC$ from $H$, i.e., $FAC = Y H$:

$$
FAC = H \, FAC \\
Y \, H = H \, (Y \, H)
$$

\[
\begin{align*}
FAC \, 1 &= Y \, H \, 1 \\
&= H \, (Y \, H) \, 1 \\
&= (\lambda f.\, \lambda n.\, IF \,(= \, n \, 0) \, 1 \,(\ast \, n \,(f \,(- \, n \, 1)))) \, (Y \, H) \, 1 \\
&\rightarrow (\lambda n.\, IF \,(= \, n \, 0) \, 1 \,(\ast \, n \,(((Y \, H) \,(- \, n \, 1)))) \, 1 \\
&\rightarrow IF \,(= \, 1 \, 0) \, 1 \,(\ast \, 1 \,((Y \, H) \,(- \, 1 \, 1)))) \\
&\rightarrow * \, 1 \,(Y \, H \, 0) \\
&= * \, 1 \,(H \, (Y \, H) \, 0) \\
&= * \, 1 \,((\lambda f.\, \lambda n.\, IF \,(= \, n \, 0) \, 1 \,(\ast \, n \,(f \,(- \, n \, 1)))) \, (Y \, H) \, 0) \\
&\rightarrow * \, 1 \,((\lambda n.\, IF \,(= \, n \, 0) \, 1 \,(\ast \, n \,(Y \, H \,(- \, n \, 1)))) \, 0) \\
&\rightarrow * \, 1 \,(IF \,(= \, 0 \, 0) \, 1 \,(\ast \, 0 \,(Y \, H \,(- \, 0 \, 1)))) \\
&\rightarrow * \, 1 \, 1 \\
&\rightarrow 1
\end{align*}
\]
The \textit{Y} Combinator

Here’s the eye-popping part: \textit{Y} can be a simple lambda expression.

\[
Y = \lambda f. (\lambda x. f(x)) (\lambda x. f(x))
\]

\[
Y H = \left( \lambda f. (\lambda x. f(x)) (\lambda x. f(x)) \right) H \\
\quad \to (\lambda x. H(x)) (\lambda x. H(x)) \\
\quad \to H \left( (\lambda x. H(x)) (\lambda x. H(x)) \right) \\
\quad \leftrightarrow H \left( \left( \lambda f. (\lambda x. f(x)) (\lambda x. f(x)) \right) H \right) \\
\quad = H (Y H)
\]

“\textit{Y}: The function that takes a function \(f\) and returns \(f(f(f(f(\cdots))))\)”
Prolog Execution

Facts
nerd(X) :- techer(X).
techer(stephen).

Query
?- nerd(stephen).

Search (Execution)

Result
yes
Simple Searching

Starts with the query:

\(?- \text{nerd}(\text{stephen}).\)

*Can we convince ourselves that \text{nerd}(\text{stephen}) is true given the facts we have?*

\text{techer}(\text{stephen}).
\text{nerd}(X) :- \text{techer}(X).$

First says \text{techer}(\text{stephen}) is true. Not helpful.

Second says that we can conclude \text{nerd}(X) is true if we can conclude \text{techer}(X) is true. More promising.
Unifying \texttt{nerd(stephen)} with the head of the second rule, \texttt{nerd(X)}, we conclude that \(X = \text{stephen}\).

We’re not done: for the rule to be true, we must find that all its conditions are true. \(X = \text{stephen}\), so we want \texttt{techer(stephen)} to hold.

This is exactly the first clause in the database; we’re satisfied. The query is simply true.
More Clever Searching

```prolog
techer(stephen).
techer(todd).
nerd(X) :- techer(X).
```

```prolog
?- nerd(X).
```

“Tell me about everybody who’s provably a nerd.”

As before, start with query. Rule only interesting thing.

Unifying `nerd(X)` with `nerd(X)` is vacuously true, so we need to establish `techer(X)`.

Unifying `techer(X)` with `techer(stephen)` succeeds, setting `X = stephen`, but we’re not done yet.

Unifying `techer(X)` with `techer(todd)` also succeeds, setting `X = todd`, but we’re still not done.

Unifying `techer(X)` with `nerd(X)` fails, returning no.
Database consists of Horn clauses. (“If a is true and b is true and ... and y is true then z is true”.)

Each clause consists of terms, which may be constants, variables, or structures.

**Constants:** foo my_Const + 1.43

**Variables:** X Y Everybody My_var

**Structures:** rainy(rochester)

    teaches(edwards, cs4115)
Structures and Functors

A structure consists of a functor followed by an open parenthesis, a list of comma-separated terms, and a close parenthesis:

```
“Functor”
paren must follow immediately
bin_tree( foo, bin_tree(bar, glarch) )
```

What’s a structure? Whatever you like.

A predicate `nerd(stephen)`
A relationship `teaches(edwards, cs4115)`
A data structure `bin(+, bin(-, 1, 3), 4)`
Unification

Part of the search procedure that matches patterns. The search attempts to match a goal with a rule in the database by unifying them.

Recursive rules:

- A constant only unifies with itself
- Two structures unify if they have the same functor, the same number of arguments, and the corresponding arguments unify
- A variable unifies with anything but forces an equivalence
Unification Examples

The = operator checks whether two structures unify:

| ?- a = a. | yes | % Constant unifies with itself |
| ?- a = b. | no  | % Mismatched constants          |
| ?- 5.3 = a. | no  | % Mismatched constants          |
| ?- 5.3 = X. |       |                               |
| X = 5.3 ? ; | yes | % Variables unify              |
| ?- foo(a,X) = foo(X,b). | no  | % X=a required, but inconsistent |
| ?- foo(a,X) = foo(X,a). |       | % X=a is consistent            |
| X = a       | yes |                               |
| ?- foo(X,b) = foo(a,Y). |       | % X=a, then b=Y                |
| X = a       | yes |                               |
| Y = b       |     |                               |
| ?- foo(X,a,X) = foo(b,a,c). | no | % X=b required, but inconsistent |
The Searching Algorithm

search(goal \( g \), variables \( e \))
for each clause \( h \leftarrow t_1, \ldots, t_n \) in the database
\[ e = \text{unify}(g, h, e) \]
if successful,
for each term \( t_1, \ldots, t_n, \)
\[ e = \text{search}(t_k, e) \]
if all successful, return \( e \)
return no

Note: This pseudo-code ignores one very important part of the searching process!
Order Affects Efficiency

```
edge(a, b). edge(b, c).
edge(c, d). edge(d, e).
edge(b, e). edge(d, f).

path(X, X).

path(X, Y) :-
    edge(X, Z), path(Z, Y).
```

Consider the query

```
?- path(a, a).
```

```
path(a,a)
| path(a,a)=path(X,X)
|   X=a
|   yes
```

Good programming practice: Put the easily-satisfied clauses first.
Order Affects Efficiency

\[ \text{edge}(a, b). \text{edge}(b, c). \]
\[ \text{edge}(c, d). \text{edge}(d, e). \]
\[ \text{edge}(b, e). \text{edge}(d, f). \]

\[ \text{path}(X, Y) :\]
\[ \text{edge}(X, Z), \text{path}(Z, Y). \]

\[ \text{path}(X, X). \]

Consider the query

\[ ?- \text{path}(a, a). \]

Will eventually produce the right answer, but will spend much more time doing so.
Order Can Cause Infinite Recursion

\[\text{edge}(a, b). \text{edge}(b, c). \]
\[\text{edge}(c, d). \text{edge}(d, e). \]
\[\text{edge}(b, e). \text{edge}(d, f). \]

\[\text{path}(X, Y) \leftarrow \]
\[\text{path}(X, Z), \text{edge}(Z, Y). \]
\[\text{path}(X, X). \]

Consider the query
\[\text{?- path}(a, a). \]

\[\text{path}(a, a) \]
\[\text{path}(a, a) = \text{path}(X, Y) \]
\[\text{X} = a \quad \text{Y} = a \]

Subgoal
\[\text{path}(a, Z) \quad \text{edge}(Z, a) \]
\[\text{path}(a, Z) = \text{path}(X, Y) \]
\[\text{X} = a \quad \text{Y} = Z \]
\[\text{path}(a, Z) \quad \text{edge}(Z, Z) \]
\[\text{path}(a, Z) = \text{path}(X, Y) \]
\[\text{X} = a \quad \text{Y} = Z \]
\[\vdots \]
Prolog as an Imperative Language

A declarative statement such as

\[
\text{P if Q and R and S}
\]

can also be interpreted procedurally as

\[
\text{To solve P, solve Q, then R, then S.}
\]

This is the problem with the last path example.

\[
\text{path}(X, Y) \ :- \ \\
\text{path}(X, Z), \text{edge}(Z, Y).
\]

“To solve P, solve P…”
Cuts
Ways to shape the behavior of the search:

- Modify clause and term order.
  Can affect efficiency, termination.

- “Cuts”
  Explicitly forbidding further backtracking.

When the search reaches a cut (!), it does no more backtracking.

```
teacher(stephen) :- !.
teacher(todd).
nerd(X) :- teacher(X).
```

<table>
<thead>
<tr>
<th>?- nerd(X).</th>
</tr>
</thead>
<tbody>
<tr>
<td>X = stephen</td>
</tr>
<tr>
<td>yes</td>
</tr>
</tbody>
</table>