Principles of Programming Languages
COMP251: Functional Programming in Scheme (and LISP)

Prof. Dekai Wu

Department of Computer Science and Engineering
The Hong Kong University of Science and Technology
Hong Kong, China

Fall 2006
Part I

Introduction
In the 1930’s, long before digital computers were invented, logicians studied abstract concepts of computation.

- Two simple models of computation:
  1. Alan Turing’s **Turing machines** — based on an updatable store (memory).
  2. Alonzo Church’s **lambda calculus** — based on the mathematical concept of functions.

- The **Turing machine** idea has great influence on the design of **von Neumann computer architecture** — used by most of today’s common CPUs, and imperative programming languages are designed to compute efficiently on the architecture.

- **Lambda calculus** is the foundation of **functional programming paradigm**.
Church-Turing Thesis
Any effective computation can be done in one of the two models.
Functions are first-class objects: They may be (1) passed as arguments (2) returned as results (3) stored in variables. (c.f. variables in imperative languages.)

Basic mode of computation: construction and application of functions. (c.f. assignments in imperative languages.)

Principal control mechanism: recursive function applications. (c.f. for-/while-loops in imperative languages.)

Freed from side-effects (for “pure” FPL).
Side-Effects: Operations which permanently change the value of a variable (by assignments) or other observable objects (e.g. by printing outputs).

Examples: LISP (LISt Processing), Scheme (a dialect of LISP), ML (Meta Language), Haskell, Miranda, etc.
Scheme supports:

- **Static Scope**: All identifier references can be resolved at compile time (unlike LISP, which uses dynamic scoping).
- **Dynamic or Latent Typing**: All variables are dynamically or latently typed. Any variable can hold any type (imagine that all variables are of the type `Object`). There are no explicit type declarations. Many scripting languages take the same approach (e.g., Perl, Python, JavaScript, Tcl, etc.)
- **Proper Tail-Recursion**: Scheme (but not LISP) implementations are required to be properly tail-recursive, supporting an unbounded number of active tail calls.
- **Pure and Impure Functional Programming**: Scheme/LISP encourage pure functional programming, but impure functional programming using some side-effects can also be done.
Scheme supports:

- **List-Oriented Programming**: The main data structure in Scheme/LISP is the list, implemented as a cons data structure or cons tree. This versatile data structure can easily be applied in many ways, and is easy to manipulate.

- **Programs as Data**: The concrete syntax of Scheme/LISP is almost the abstract syntax. All expressions are in prefix notation and fully parenthesized. A big advantage is that programs look just like cons list data structures, which makes it easy to manipulate programs as if they were just data (so programs can write and execute programs).

- **Interpreted**: Traditionally, Scheme/LISP implementations all provide an interpreter (and sometimes also a compiler), making interactive debugging easy.

- **Automatic Garbage Collection**: All implementations all provide this feature, which Java inherits.
Part II

Types and Values
<table>
<thead>
<tr>
<th>TYPE</th>
<th>EXAMPLE</th>
<th>COMMON OPERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>boolean</td>
<td>#t, #f</td>
<td>boolean?, not, and, or</td>
</tr>
<tr>
<td>integer</td>
<td>2, 0, 87</td>
<td>+, -, *, quotient, remainder, modulo</td>
</tr>
<tr>
<td>rational</td>
<td>(/ 4 6)</td>
<td>+, -, *, /, numerator, denominator</td>
</tr>
<tr>
<td>real</td>
<td>1.3, 300.1</td>
<td>real?, +, -, *, /, floor</td>
</tr>
<tr>
<td>complex</td>
<td>3+4i</td>
<td>complex?, +, -, *, /</td>
</tr>
<tr>
<td>character</td>
<td>#\a, #\space, #\newline</td>
<td>char?</td>
</tr>
<tr>
<td>string</td>
<td>&quot;hello&quot;</td>
<td>string?, string-ref, string-set!</td>
</tr>
<tr>
<td>symbol</td>
<td>hello</td>
<td>symbol?, eqv?, symbol-&gt;string</td>
</tr>
<tr>
<td>dotted pair</td>
<td>(1 . &quot;yes&quot;)</td>
<td>list?, null?, cons, car, cdr</td>
</tr>
<tr>
<td>list</td>
<td>() (1 &quot;yes&quot; 2.5) (1 . (&quot;yes&quot; . (2.5 . ())))</td>
<td></td>
</tr>
<tr>
<td>vector</td>
<td>#(1 &quot;yes&quot; 2.5)</td>
<td>vector?, vector-ref, vector-set!, vector</td>
</tr>
</tbody>
</table>
Quoting

Syntax:

\[(\text{quote } < \text{expression } >)\]

\[\text{'} < \text{expression } >\]

- Quoting is needed to treat expressions as data.

\[> (\text{define pi 3.14159})\]
\[> \text{pi}\]
\[3.14159\]

- A quoted item evaluates to itself. \text{quote} prevents \text{eval} from being called:

\[> (\text{quote pi})\]
\[\text{pi}\]

- An equivalent convenient shorthand is provided using \text{'}: 

\[> \text{'pi}\]
\[\text{pi}\]
• Unquoted, * represents the multiplication function:

```
> (define f *)
> (f 2 3)
6
```

• Quoted, ’* represents the symbol with spelling *:

```
> (define f ’*)
> (f 2 3)
ERROR: Wrong type to apply: *
```
Empty list: '()

'(e_1 e_2 \ldots e_n) is an abbreviation for
(cons e_1 (cons e_2 (cons \ldots (cons e_n '())))))

cons is the list constructor

<new-list> = (cons <item> <list>)

(cons 1 (cons 2 (cons 3 '())))
= (cons 1 (cons 2 '(3)))
= (cons 1 '(2 3))
= '(1 2 3)

Equality on 2 lists is item-by-item.
- **cons** operator: creates a dotted pair of the two operands
- **car** operator: returns the first element of a dotted pair (cons cell)
- **cdr** operator: returns the second element of a dotted pair (cons cell)
- **append** operator: returns a list consisting of the elements of the first list followed by the elements of the other lists
List Examples

> (car '(1 2 3 4))
1

> (cdr '(1 2 3 4))
(2 3 4)

> (cons (car '(1 2 3 4)) (cdr '(1 2 3 4)))
(1 2 3 4)

> (append '(5 6) (cdr '(1 2 3 4)))
(5 6 2 3 4)
> (vector 4 true cat)
#(4 true cat)

> (if (equal? 3 8) "X" (vector "Y" 9.5/0.5 (quotient 5 2)))
#("Y" 19.0 2)

> (equal? (vector (modulo 14 3) (not #f))
   (vector (+ 1 1) #t))
#t

> (vector-ref (vector "for" "your" "info") 1)
"your"

- Ordered $n$-tuple: $\#(e_1 e_2 \ldots e_n)$.
- The $n$ expressions may be of mixed types.
- 2 $n$-tuples are equal if their corresponding components are equal.
• “(vector-ref myvector k)” is the item selection operator.
• “(vector-length myvector)” returns $n$.
• “(vector #(e₁ e₂ . . . eₙ))” returns a newly constructed vector containing the given elements.
• “(make-vector n fill)” returns a newly allocated vector of $n$ elements. If the optional second argument is given, then each element is initialized to fill.
Most identifiers allowed by other programming languages are also acceptable to Scheme. The precise rules for forming identifiers vary among implementations of Scheme, but in all implementations a sequence of letters, digits, and “extended alphabetic characters” that begins with a character that cannot begin a number is an identifier. In addition, +, -, and ... are identifiers. Here are some examples of identifiers:

```
lambda          q
list->vector    soup
+                V17a
<=?              a34kTMNs
the-word-recursion-has-many-meanings
```

Extended alphabetic characters may be used within identifiers as if they were letters. The following are extended alphabetic characters:

```
! $ % & * + - . / : < = > ? @ ^ _ ~
```
Identifiers have two uses within Scheme programs:

- Any identifier may be used as a **variable** or as a **syntactic keyword**.
- When an identifier appears as a literal or within a literal, it is being used to denote a **symbol**.
Identifiers: Value Binding

Syntax: \( \text{(define < identifier > < expression >)} \)

\[
> \text{(define a\_df (+ 3 2)) \; c.f. int a\_df = 3+2; in C++}
\]
\[
> \text{a\_df}
\]
\[
5
\]
\[
> \text{(define a’a (string-append "Albert" " " "Einstein"))}
\]
\[
> \text{a’a}
\]
\[
"Albert Einstein"
\]
\[
> \text{(define a1b2 2)}
\]
\[
> \text{a1b2}
\]
\[
2
\]
\[
> \text{(define +++$$$(\ast 9 3)) \; may hold integral value}
\]
\[
> \text{+++$$$}
\]
\[
27
\]
\[
> \text{(+ +++$$$ +++$$$)} \; \text{Though you don’t want to do that}
\]
\[
54
\]
Part III

Scheme Functions
Syntax: \( \text{lambda} (\text{<formal parameters>}) \text{<body>} \)

- An \textit{anonymous function} is a function without a name.
- \texttt{lambda} returns a newly constructed anonymous parameterized function \textit{value}.
- \texttt{<formal parameters>} is a sequence of parameter names.
- \texttt{<body>} is an expression, possibly containing occurrences of the parameter names.
- Used when only a locally defined function is needed.

\[
\texttt{> (lambda (x) (* x x)) ; constructing anonymous function}
\]
\[
\texttt{#<procedure #f (x)>}
\]

\[
\texttt{> ((lambda (x) (* x x)) 4) ; applying anonymous function}
\]
\[
\texttt{16}
\]
Defining Named Functions

- To define a named function, you simply have to bind some identifier (which will be the function name) to a function value.

- Parameter passing method: **Call-By-Value**.

```scheme
> (define square (lambda (x) (* x x)))
> square
#<procedure square (x)>

> (square 4)
16

> (define first (lambda (x y) x))
> (first 3 "man")
3
```
Convenient “Syntactic Sugar” for Named Functions

- **Syntax:** `(define (<identifier> <formal-parameters>) <body>)`

```scheme
> (define (square x) (* x x))
> square
#<procedure square (x)>

> (square 4)
16

> (define (first x y) x)
> (first 3 "man")
3
```
Functions taking functions as arguments:

```
> (define (square x) (* x x))
> (define (twice x) (* 2 x))

> (define (apply5 f) (apply f '(5)))
> (apply5 square)
25

> (define (apply-to-twice-x f x) (apply f (list (twice x))))
> (apply-to-twice-x square 3)
36
```
Functions returning function:

> (define (sq_or_twice x) (if (> x 0) square twice))

> (apply (sq_or_twice 2) '(5))
25

> (sq_or_twice 2)
#<procedure square (x)>
Recall this example:

\[
\text{> (lambda (x) (* x x)) ; constructing anonymous function}
\]
\[
\text{#<procedure #f (x)>}
\]

\[
\text{> ((lambda (x) (* x x)) 4) ; applying anonymous function}
\]
\[
\text{16}
\]

In the lambda calculus, which was the origin of lambda in LISP/Scheme, the same two lines would be written:

\[
(\lambda x. (x \times x))
\]

\[
(\lambda x. (x \times x)) 4
\]

We say that applying \( \lambda x \) to the expression \( x \times x \) performs a lambda abstraction.
A couple conventions for the lambda calculus make it much more readable.

First, parentheses may be dropped from \((MN)\) and \((\lambda x. M)\).

Notice that function application is an operator; the operator is left associative. E.g., \(xyz\) is an abbreviation for \(((xy)z)\). Also, function application has higher precedence than lambda abstraction, so

\[
\lambda x. x \times x
\]

is an abbreviation for

\[
(\lambda x. (x \times x))
\]
Quiz: What is the abbreviated syntax for this?

\((\lambda x. (x \times x))\ 4\)

Answer:

\((\lambda x. x \times x)\ 4\)
Second, a sequence of consecutive lambda abstractions, as in

$$\lambda x. \lambda y. \lambda z. M$$

can be written with a single lambda, as in

$$\lambda xyz. M$$
Quiz: Note that $\lambda xyz. M$ corresponds to what we normally think of as a function with three parameters. If you instead write it as $\lambda x. \lambda y. \lambda z. M$, then what is the meaning of this:

$$(\lambda x. \lambda y. \lambda z. M) 8$$

Answer: The value of this expression is a function of two parameters, where $y$ and $z$ remain free variables, but $x$ has already been bound to the value 8.

The practice of breaking up multi-parameter functions into single-parameter functions like this in programming is called **currying** a function (after Tim Curry who promoted this practice, widely found in functional languages like ML and Haskell).
The result of applying an abstraction \((\lambda x. M)\) to an argument \(N\) is formalized as \textit{substitution} of \(N\) for \(x\) in \(M\), written \(\{N/x\} M\).

Informally, \(N\) replaces all free occurrences of \(x\) in \(M\).

Caution! A fully correct definition of substitution is tricky (there was a long history of not-quite-adequate definitions). This is beyond our scope here.

Here is an almost-correct definition:

1. Suppose the free variables of \(N\) have no bound occurrences in \(M\). Then the term \(\{N/x\} M\) is formed by replacing all free occurrences of \(x\) in \(M\) by \(N\).

2. Otherwise, suppose variable \(y\) is free in \(N\) and bound in \(M\). Consistently replace the binding and corresponding bound occurrences of \(y\) in \(M\) by some fresh variable \(z\). Repeat the renaming of bound variables in \(M\) until case 1 applies, then proceed as in case 1.
In the following examples, $M$ has no bound occurrences, so $N$ replaces all occurrences of $x$ in $M$ to form $\{N/x\} M$:

\[
\begin{align*}
\{u/x\} x &= u \\
\{u/x\} (x x) &= (u u) \\
\{u/x\} (x y) &= (u y) \\
\{u/x\} (x u) &= (u u) \\
\{(\lambda x. x)/x\} (x u) &= (\lambda x. u)
\end{align*}
\]
In the following examples, \( M \) has no free occurrences, so \( \{ N/x \} M \) is \( M \) itself:

\[
\begin{align*}
\{u/x\} y &= y \\
\{u/x\} (y z) &= (y x) \\
\{u/x\} (\lambda y. y) &= (\lambda y. y) \\
\{u/x\} (\lambda x. x) &= (\lambda x. x) \\
\{(\lambda x. x)/x\} y &= y
\end{align*}
\]
In the following examples, free variable $u$ in $M$ has bound occurrences in $M$, so $\{N/x\} M$ is formed by first renaming the bound occurrences of $u$ in $M$:

\[
\{u/x\} (\lambda u. x) = \{u/x\} (\lambda z. x) = (\lambda z. u) \\
\{u/x\} (\lambda u. u) = \{u/x\} (\lambda z. z) = (\lambda z. z)
\]
List Function: \texttt{map}

- The built-in library function \texttt{map()} has 2 or more arguments: a function \texttt{<func>} and one or more lists.
- It applies function \texttt{<func>} to the elements of the lists as follows.

\begin{verbatim}
(map <func> <list_1> <list_2> ...)
\end{verbatim}

\texttt{<func>} must be a function taking as many arguments as there are lists and returning a single value. If more than one list is given, then they must all be the same length. Map applies \texttt{<func>} element-wise to the elements of the lists and returns a list of the results, in order. The dynamic order in which \texttt{<func>} is applied to the elements of the lists is unspecified.
map: Examples

> (define (odd x) (equal? (modulo x 2) 1))

> (map odd '(1 2 3))
(#t #f #t)

> (map cadr '((a b) (d e) (g h)))
(b e h)

> (map (lambda (n) (expt n n)) '(1 2 3 4 5))
(1 4 27 256 3125)

> (map + '(1 2 3) '(4 5 6))
(5 7 9)
(let
  ((<1st-identifier> < E₁ >)
   (<2nd-identifier> < E₂ >)
   ...
   <body-expression>))

- c.f. Declaration of local variables in C++
let Example

> (define z
  (let ((x 3)
        (y 5))
    (+ (* x x) (* 3 y))))

> z
24

• As spaces are immaterial, the statement may as well be written all in one single line as follows:

> (define z (let ((x 3) (y 5)) (+ (* x x) (* 3 y))))

• Quiz: What is the relationship between let and lambda?

> (define z
  ((lambda (x y) (+ (* x x) (* 3 y)))
   3 5))

> z
24
Nested let Example

> (let ((x 3.0)
    (y 5.0))
  (let ((a (+ x y))
    (b (- x y)))
  (let ((f (* a b x))
    (g (/ a b y)))
  (/ f g)))

Quiz: What is the output?
Part V

Misc: Different Notions of Equality and Equivalence
Scheme (like LISP) offers various notions of equality and equivalence, to support reference vs. value comparisons over different types, with different efficiency tradeoffs.

- $\equiv$: Only applies to numeric values. (Very efficient.)
- $\text{char}\equiv$: Only applies to character values. (Very efficient.)
- $\text{string}\equiv$: Only applies to string values.
- $\text{eq?}$: Merely compares references and booleans. (Very efficient; essentially just pointer comparison.)
- $\text{eqv?}$: Like $\text{eq?}$ combined with $\equiv$ plus handles numeric and character values. (Still efficient but slightly less so.)
- $\text{equal?}$: Recursively compares the contents of pairs, vectors, and strings, applying $\text{eqv?}$ on other objects such as numbers and symbols. A rule of thumb is that objects are generally equal? if they print the same. (Least efficient. Why?)
### eq? vs. eqv? vs. equal?

<table>
<thead>
<tr>
<th>Expression</th>
<th>eq?</th>
<th>eqv?</th>
<th>equal?</th>
</tr>
</thead>
<tbody>
<tr>
<td>(eq? '(()) #f) ; cf LISP's nil</td>
<td>#f</td>
<td>#f</td>
<td>#f</td>
</tr>
<tr>
<td>(eq? 'a 'a)</td>
<td>#t</td>
<td>#t</td>
<td>#t</td>
</tr>
<tr>
<td>(eq? 'a 'b)</td>
<td>#f</td>
<td>#f</td>
<td>#f</td>
</tr>
<tr>
<td>(eq? 2 2)</td>
<td>#t only if eqv?</td>
<td>#t</td>
<td>#t</td>
</tr>
<tr>
<td>(eq? #\a #\a)</td>
<td>#t only if eqv?</td>
<td>#t</td>
<td>#t</td>
</tr>
<tr>
<td>(eq? &quot;&quot; &quot;&quot;)</td>
<td>unspecified</td>
<td>unspecified</td>
<td>#t</td>
</tr>
<tr>
<td>(eq? &quot;a&quot; &quot;a&quot;)</td>
<td>#t iff eqv?</td>
<td>unspecified</td>
<td>#t</td>
</tr>
<tr>
<td>(eq? '#() '#())</td>
<td>unspecified</td>
<td>unspecified</td>
<td>#t</td>
</tr>
<tr>
<td>(eq? '#(1 2) '#(1 2))</td>
<td>#t iff eqv?</td>
<td>unspecified</td>
<td>#t</td>
</tr>
<tr>
<td>(eq? '() ')()</td>
<td>#t</td>
<td>#t</td>
<td>#t</td>
</tr>
<tr>
<td>(eq? '(a) '(a))</td>
<td>unspecified</td>
<td>unspecified</td>
<td>#t</td>
</tr>
<tr>
<td>(eq? '(b) (cdr '(a b)))</td>
<td>unspecified</td>
<td>unspecified</td>
<td>#t</td>
</tr>
<tr>
<td>(eq? (cons 1 2) (cons 1 2))</td>
<td>#f</td>
<td>#f</td>
<td>#t</td>
</tr>
<tr>
<td>(let ((x '(a))) (eq? x x))</td>
<td>#t</td>
<td>#t</td>
<td>#t</td>
</tr>
<tr>
<td>(eq? (lambda () 'a) (lambda () 'b))</td>
<td>#f</td>
<td>#f</td>
<td>#f</td>
</tr>
<tr>
<td>(eq? (lambda (x) x) (lambda (x) x))</td>
<td>unspecified</td>
<td>unspecified</td>
<td>unspecified</td>
</tr>
<tr>
<td>(let ((p (lambda (x) x))) (eq? p p))</td>
<td>#t</td>
<td>#t</td>
<td>#t</td>
</tr>
</tbody>
</table>
Part VI

Misc: Value Binding
Impure FP: Imperative-style Side-effects

- `(set! <variable> <expression>)`
  `<expression>` is evaluated, and the resulting value is stored in the location to which `<variable>` is bound.

- `(set-car! <pair> <expression>)`
  Stores `<expression>` in the car field of `<pair>`.

- `(set-cdr! <pair> <expression>)`
  Stores `<expression>` in the cdr field of `<pair>`.

```
> (define x '(b c)) | > (set! x '(b c))
> (define y x) | > (eq? x y)
> (cadr x) | #f
  c | > (equal? x y)
> (eq? x y) | #t
  #t | > (set! y x)
> (set! x '(d e f)) | > (eq? x y)
> (cadr x) | #t
  e | > (set-cdr! x '(g h))
> (eq? x y) | #f
  #f | > y
```

Prof. Dekai Wu, HKUST (dekai@cs.ust.hk)
The phrase: “(define x '(b c))” is called a binding; it binds the variable x to a value '(b c).

Can occur at top-level (global) or at the beginning of a lambda or let body (local).

Don’t re-define variables; think of them as aliases or constants. You can re-define existing variables at the top-level, for convenience. Whenever an identifier is defined, it’s as if a new identifier is “created” — it has nothing whatever to do with any previously existing identifier of the same name.

The phrase: “(set! x '(d e f))” rebinds the variable x to another value '(d e f).

Don’t use set! unless you are intentionally violating pure functional programming.

Environment: the current set of ordered pairs (identifier, value) that are visible.
env:

- val x = 17;
val x = 17 : int

- val y = x;
val y = 17 : int

- val x = true;
val x = true : bool

- val z = x;
val z = true : bool
Alias: When a data object is visible through more than one name in a single referencing environment, each name is termed an alias.

- **Examples**: passed parameters by reference in a function, several pointers to the same object.
- **Pitfall**: programs are harder to understand.

Side Effects: An operation has side effects if it makes changes which persist after it returns.

- **Examples**: A function changes its parameters or modifies global variables (through assignments); printouts.
- **Pitfall**: programs are harder to understand, evaluation order of expressions becomes important.
**Alias and Side Effects: Example**

```c
int x = 2, y = 5;
int Bad(int m) { return x+=m; }

void Swap(int* a, int* b)
{
    int temp = *a; *a = *b; *b = temp;
    x = 4;
}

int main()
{
    int* z = &x;
    int k = x * Bad(7) + x;

    printf("k = %d\n", k);
    Swap(&x, &y);
    printf("(x,y) = (%d,%d)\n", x, y);
}
```
Part VII

Summary
A task is achieved through applications of functions.

No pointers!

No coercion!

No side-effects!

Assignment is replaced by value binding.

Implicit type inference.

Implicit memory management: Objects are allocated as needed, and deallocated when they become inaccessible.

Pattern matching $\Rightarrow$ program by examples.

Allow recursive definition of polymorphic datatypes.

Simple exception handling.
<table>
<thead>
<tr>
<th><strong>IP:</strong></th>
<th>Since IP languages are based on the von Neumann architecture, programmers must deal with the management of variables, assignment of values to them, memory locations, and sometimes even memory allocations.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adv:</strong></td>
<td>efficient computation</td>
</tr>
<tr>
<td><strong>Disadv:</strong></td>
<td>laborious construction of programs</td>
</tr>
<tr>
<td><strong>FP:</strong></td>
<td>Do not manipulate memory directly; no variables, no assignments. Instead they work on values that are independent of an underlying machine.</td>
</tr>
<tr>
<td><strong>Adv:</strong></td>
<td>compact language, simple syntax, higher level of programming</td>
</tr>
<tr>
<td><strong>Disadv:</strong></td>
<td>efficiency is sacrificed</td>
</tr>
<tr>
<td><strong>IP:</strong></td>
<td>Due to aliases and side effects, the effects of a subprogram or a block cannot be determined in isolation from the entire program.</td>
</tr>
<tr>
<td><strong>FP:</strong></td>
<td>Since they only manipulate values, there are no aliases nor side effects.</td>
</tr>
</tbody>
</table>

| **IP:** | Explicit memory management. |
| **FP:** | Storage is allocated as necessary; and storage that becomes inaccessible is automatically deallocated and reclaimed during garbage collection. |

| **IP:** | The power comes from mimicking operations on the underlying computer architecture with assignments, loops, and jumps. |
| **FP:** | The power comes from recursion and treating functions as “first-class” values. |