PS4 Written can be turned in during structured office hours.

Quickest Sorting and Double Deltas

Shuttle Rescue Mission

Monday Feb 23 and Wednesday Feb 25

MEC 205 until 5:30pm

http://shuttle.cs.virginia.edu:8080/

Build and program Lego Mindstorms robot to remotely sense and navigate a barren environment and retrieve a life pod from a crater.

Exam 1 Extra Credit: *either* show up and watch one day *or* write paragraph about how to do it

One-Slide Summary

- **Insert-sort** is $\Theta(n^2)$ worst case (reverse list), but is $\Theta(n)$ best case (sorted list).
- A recursive function that divides its input in **half** each time is often in $\Theta(\log n)$.
- If we could divide our input list in half rapidly, we could do a **quicker sort**: $\Theta(n \log n)$.
- **Sorted binary trees** are an efficient data structure for maintaining sorted sets.
- British codebreakers used **cribs** (guesses), brute force, and **analysis** to break the Lorenz cipher. Guessed wheel settings were likely to be correct if they resulted in a message with the right linguistic properties for German (e.g., repeated letters).

Outline

- Insert-sort
- Going half-sies
- Sorted binary trees
- Quicker-sort
- WWII Codebreaking

Pick Up Graded Problem Sets During Structure Hours Today!

How much work is insert-sort?

```
(define (insert-sort lst cf)
  (if (null? lst) null
      (insert-one (car lst) (insert-sort (cdr lst) cf) cf)))

(define (insert-one el lst cf)
  (if (null? lst) (list el)
      (if (cf el (car lst)) (cons el lst)
          (cons (car lst) (insert-one el (cdr lst) cf)))))
```

How many times does insert-sort evaluate insert-one?

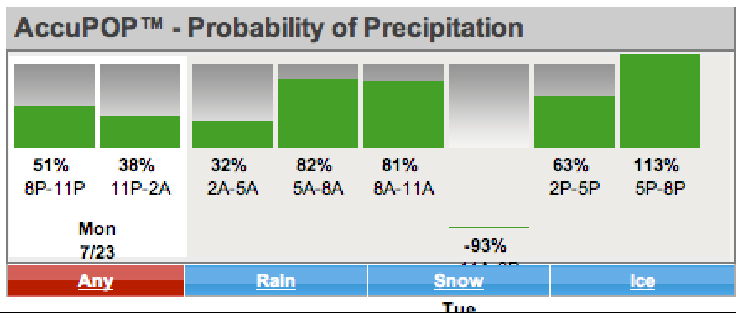
n times (once for each element)

running time of insert-one is in $\Theta(n)$

insert-sort has running time in $\Theta(n^2)$ where n is the number of elements in the input list

Which is better?

- Is insert-sort faster than best-first-sort?



```
> (insert-sort < (revintsto 20))
(1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20)
                Requires 190 applications of <

> (insert-sort < (intsto 20))
(1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20)
                Requires 19 applications of <

> (insert-sort < (rand-int-list 20))
(0 11 16 19 23 26 31 32 32 34 42 45 53 63 64 81 82 84 84 92)
                Requires 104 applications of <
```



```
> (best-first-sort < (intsto 20))

(1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20)
                Requires 210 applications of <

> (best-first-sort < (rand-int-list 20))

(4 4 16 18 19 20 23 32 36 51 53 59 67 69 73 75 82 82 88
89)
                Requires 210 applications of <
```


best-first-sort vs. insert-sort

- Both are $\Theta(n^2)$ worst case (reverse list)
- Both are $\Theta(n^2)$ when sorting a randomly ordered list
  - But insert-sort is about twice as fast
- insert-sort is $\Theta(n)$ best case (ordered input list)

Can we do better?

```
(quicker-insert < 88
          (list 1 2 3 5 6 23 63 77 89 90))
```

Suppose we had procedures
(first-half lst)
(second-half lst)
that quickly divided the list in two halves?

quicker-insert using halves

```
(define (quicker-insert el lst cf)
  (if (null? lst) (list el)      ;; just like insert-one
     (if (null? (cdr lst))
        (if (cf el (car lst)) (cons el lst) (list (car lst) el))
        (let ((front (first-half lst))
              (back (second-half lst)))
          (if (cf el (car back))
             (append (quicker-insert el front cf) back)
             (append front
                     (quicker-insert el back cf)))))))
```


Evaluating quicker-sort

```
> (quicker-insert < 3 (list 1 2 4 5 7))
|(quicker-insert #<procedure:traced-<> 3 (1 2 4 5 7))
| (< 3 1)
| #f
| (< 3 5)
| #t
| (quicker-insert #<procedure:traced-<> 3 (1 2 4))
| |(< 3 1)
| |#f
| |(< 3 4)
| |#t
| |(quicker-insert #<procedure:traced-<> 3 (1 2))
| | (< 3 1)
| | #f
| | (< 3 2)
| | #f
| | (quicker-insert #<procedure:traced-<> 3 (2))
| | |(< 3 2)
| | |#f
| | (2 3)
| |(1 2 3)
| (1 2 3 4)
|(1 2 3 4 5 7)
(1 2 3 4 5 7)
```

```
(define (quicker-insert el lst cf)
  (if (null? lst) (list el)
     (if (null? (cdr lst))
        (if (cf el (car lst))
           (cons el lst)
           (list (car lst) el))
        (let ((front (first-half lst))
              (back (second-half lst)))
          (if (cf el (car back))
             (append (quicker-insert el front cf) back)
             (append front
                     (quicker-insert el back cf)))))))
```

Every time we call quicker-insert, the length of the list is approximately **halved**!

How much work is quicker-sort?

Each time we call quicker-insert, the size of lst halves. So doubling the size of the list only increases the number of calls by 1.

```
(define (quicker-insert el lst cf)
  (if (null? lst) (list el)
    (if (null? (cdr lst))
      (if (cf el (car lst))
        (cons el lst)
        (list (car lst) el))
      (let ((front (first-half lst))
            (back (second-half lst)))
       (if (cf el (car back))
          (append (quicker-insert el front cf) back)
          (append front
                  (quicker-insert el back cf)))))))
```

List Size	# quicker-insert applications
1	1
2	2
4	3
8	4
16	5

Liberal Arts Trivia:

- The argan tree, found primarily in Morocco, has a knobby, twisted trunk that allows these animals to climb it easily. The animals eat the fruit, which has an indigestible nut inside, which is collected by farmers and used to make argan oil: handy in cooking and cosmetics, but pricey at $45 per 500 ml.

Liberal Arts Trivia: Scandinavian Studies

- This capital of and largest city in Denmark is situated on the islands of Zealand and Amager. It is the birthplace of Neils Bohr, Søren Kierkegaard, and Victor Borge. The city's origin as a harbor and a place of commerce is reflected in its name. Its original designation, from which the contemporary Danish name is derived, was Køpmannæhafn, "merchants' harbor". The English name for the city is derived from its (similar) Low German name.

Remembering Logarithms

$\log_b n = x$ means $b^x = n$

What is $\log_2 1024$?

What is $\log_{10} 1024$?

Is $\log_{10} n$ in $\Theta(\log_2 n)$?

Changing Bases

$$\log_b n = \underbrace{(1/\log_k b)}_{\text{If } k \text{ and } b \text{ are constants, this is constant}} \log_k n$$

$$\Theta(\log_2 n) \equiv \Theta(\log_{10} n) \equiv \Theta(\log n)$$

No need to include a constant base within asymptotic operators.

Number of Applications

Assuming the list is well-balanced, the number of applications of quicker-insert is in $\Theta(\log n)$ where n is the number of elements in the input list.

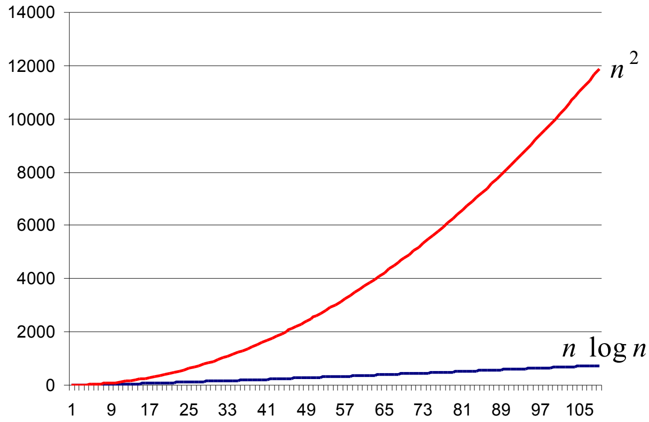
quicker-sort ?

```
(define (quicker-sort lst cf)
  (if (null? lst) null
      (quicker-insert
        (car lst)
        (quicker-sort (cdr lst) cf)
      cf)))
```

```
(define (quicker-insert el lst cf)
  (if (null? lst) (list el)
      (if (null? (cdr lst))
          (if (cf el (car lst))
              (cons el lst)
              (list (car lst) el))
          (let ((front (first-half lst))
                (back (second-half lst)))
            (if (cf el (car back))
                (append (quicker-insert el front cf) back)
                (append front
                        (quicker-insert el back cf)))))))
```

quicker-sort using halves would have running time in $\Theta(n \log n)$ ***if*** we have first-half, second-half, and append procedures that run in constant time

Orders of Growth



Is there a fast first-half procedure?

- No! (at least not on lists)
- To produce the first half of a list length n, we need to cdr down the first $n/2$ elements
- So, first-half on lists has running time in $\Theta(n)$

Making it faster

We need to either:

1. Reduce the number of applications of insert-one in insert-sort

   Impossible – need to consider each element

3. Reduce the number of applications of quicker-insert in quicker-insert

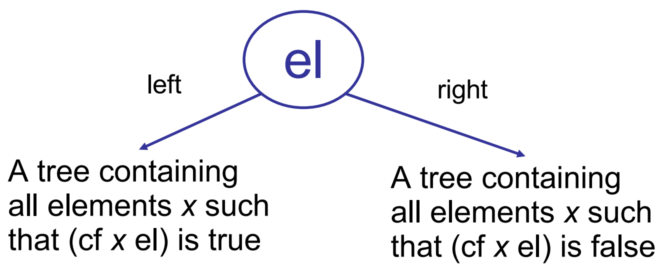
   Unlikely… each application already halves the list

5. Reduce the time for each application of quicker-insert

   Need to make first-half, second-half and append faster than $\Theta(n)$

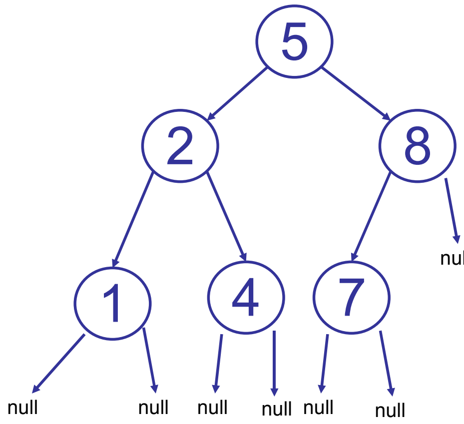
"Nothing yet. ...How about you, Newton?"

Sorted Binary Trees



Tree Example

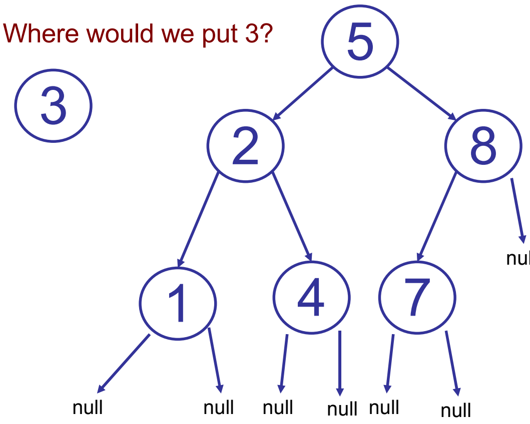
cf is **<**



Tree Example

Where would we put 3?

cf is **<**



Representing Trees

```
(define (make-tree left el right)
  (cons el (cons left right))
```

left and right are trees
(**null** is a tree)

```
(define (tree-element tree)
  (car tree))
```

tree must be a non-null tree

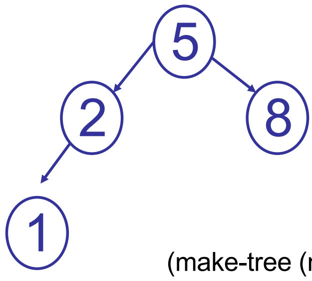
```
(define (tree-left tree)
  (car (cdr tree)))
```

tree must be a non-null tree

```
(define (tree-right tree)
  (cdr (cdr tree)))
```

tree must be a non-null tree

Representing Trees




```
(make-tree (make-tree (make-tree null 1 null)
                      2
                      null)
           5
           (make-tree null 8 null))
```


insert-one-tree

```
(define (insert-one-tree cf el tree)
  (if (null? tree)
      (make-tree null el null)
      (if (cf el (get-element tree))
          (make-tree
            (insert-one-tree cf el (get-left tree))
            (get-element tree)
            (get-right tree))
          (make-tree
            (get-left tree)
            (get-element tree)
            (insert-one-tree cf el (get-right tree))))))
```

If the tree is null, make a new tree with el as its element and no left or right trees.

Otherwise, decide if el should be in the left or right subtree. insert it into that subtree, but leave the other subtree unchanged.

How much work is insert-one-tree?

```
(define (insert-one-tree cf el tree)
  (if (null? tree)
      (make-tree null el null)
      (if (cf el (get-element tree))
          (make-tree
           (insertel-tree cf el (get-left tree))
           (get-element tree) (get-right tree))
          (make-tree (get-left tree)
           (get-element tree)
           (insertel-tree cf el (get-right tree))))))
```

Each time we call insert-one-tree, the size of the tree approximately halves (if it is well balanced).

Each application is constant time.

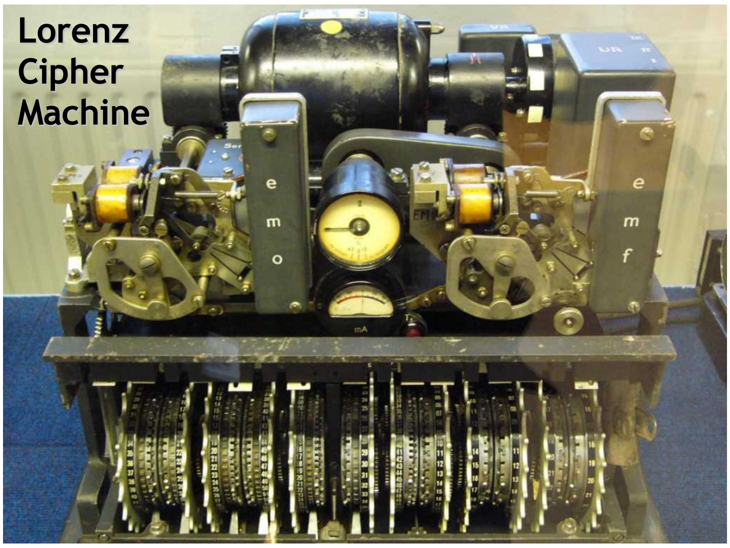
The running time of insertel-tree is in $\Theta(\log n)$ where n is the number of elements in the input tree, which must be well-balanced.

quicker-insert-one

```
(define (quicker-insert-one cf lst)
  (if (null? lst) null
      (insert-one-tree
       cf (car lst)
       (quicker-insert-one cf (cdr lst)))))
```

No change (other than using insert-one-tree)…but evaluates to a tree not a list!

(((() 1 ()) 2 ()) 5 (() 8 ()))



Liberal Arts Trivia: Classics

- This ancient Greek epic poem, traditionally attributed to Homer, is widely believed to be the oldest extant work of Western literature. It describes the events of the final year of the Trojan War. The plot follows Achilles and his anger at Agamemnon, king of Mycenae. It is written in dactylic hexameter and comprises 15,693 lines of verse. It begins:
  - μῆνιν ἄειδε θεὰ Πηληϊάδεω Ἀχιλῆος
  - οὐλομένην, ἣ μυρί' Ἀχαιοῖς ἄλγε' ἔθηκεν

Liberal Arts Trivia: Literature

- Name the author of the Age of Innocence (1920). The novel describes the upper class in New York city in the 1870s and questions the mores and assumptions of society. The title is an ironic comment on the polished outward manners of New York society, when compared to its inward machinations. The authors was the first woman to win the Pulitzer Prize for Literature.

Lorenz Wheels

12 wheels
501 pins total (set to control wheels)

Work to break in $\Theta(p^w)$ so real Lorenz is $41^{12}/5^3 \sim$ 1 quintillion (10^{18}) times harder!

Code Breaking Intuition

- Suppose we are using a simple letter substitution cipher (i.e., replace every A with Q, etc.)
- You intercept these two messages:
  - pf150: Pbzchgre Fpvrapr sebz Nqn naq Rhpyvq gb Dhnaghz Pbzchgvat naq gur Jbeyq Jvqr Jro.
  - pf150: Pbzchgre Fpvrapr sebz Nqn gb gur Jbeyq Jvqr Jro.
- What does the first one say? What hints did you have?

Breaking Fish

- Gov't Communications HQ learned about first Fish link (Tunny) in May 1941
  - British codebreakers used "Fish" to refer to German teleprinter traffic
  - Intercepted unencrypted Baudot-encoded test messages
- August 30, 1941: Big Break!
  - Operator retransmits failed message with same starting configuration
  - Gets lazy and uses some abbreviations, makes some mistakes
    - SPRUCHNUMMER/SPRUCHNR (Serial Number)

"Two Time" Pad

- Allies have intercepted:

  $C1 = M1 \oplus \mathbf{K1}$

  $C2 = M2 \oplus \mathbf{K1}$

  Same key used for both (same starting configuration)

$\oplus$ means XOR

- Breaking message:

$$C1 \oplus C2 = (M1 \oplus \mathbf{K1}) \oplus (M2 \oplus \mathbf{K1})$$
$$= (M1 \oplus M2) \oplus (\mathbf{K1} \oplus \mathbf{K1})$$
$$= M1 \oplus M2$$

"Cribs"

- Know: C1, C2 (intercepted ciphertext)

  $C1 \oplus C2 = M1 \oplus M2$
- Don't know M1 or M2
  - But, can make some guesses (cribs)
    - SPRUCHNUMMER
    - Sometimes allies moved ships, sent out bombers to help the cryptographers get good cribs
- Given guess for M1, calculate M2

  $M2 = C1 \oplus C2 \oplus M1$
- Once guesses that work for M1 and M2

  $K1 = M1 \oplus C1 = M2 \oplus C2$

Reverse Engineering Lorenz

- From the 2 intercepted messages, Col. John Tiltman worked on guessing cribs to find M1 and M2: 4000 letter messages, found 4000 letter key K1
- Bill Tutte (recent Chemistry graduate) given task of determining machine structure
  - Already knew it was 2 sets of 5 wheels and 2 wheels of unknown function
  - Six months later new machine structure likely to generate K1

Intercepting Traffic

- Set up listening post to intercept traffic from 12 Lorenz (Fish) links
  - Different links between conquered capitals
  - Slightly different coding procedures, and different configurations
- 600 people worked on intercepting traffic

Breaking Traffic

- Knew machine structure, but a different initial configuration was used for each message
- Need to determine wheel setting:
  - Initial position of each of the 12 wheels
  - 1271 possible starting positions
  - Needed to try them fast enough to decrypt message while it was still strategically valuable

This is what you did for PS4 (except with fewer wheels)

Recognizing a Good Guess

- Intercepted Message (divided into 5 channels for each Baudot code bit)

  $Z_c = z_0 z_1 z_2 z_3 z_4 z_5 z_6 z_7 \ldots$

  $z_{c,i} = m_{c,i} \oplus x_{c,i} \oplus s_{c,i}$

  Message Key (parts from S-wheels and rest)
- Look for statistical properties
  - How many of the $z_{c,i}$'s are 0? ½ (not useful)
  - How many of $(z_{c,i+1} \oplus z_{c,i})$ are 0? ½

Double Delta

$$\Delta Z_{c,i} = Z_{c,i} \oplus Z_{c,i+1}$$

Combine two channels:

$\Delta Z_{1,i} \oplus \Delta Z_{2,i} = \Delta M_{1,i} \oplus \Delta M_{2,i}$ > ½ Yippee!

$\oplus \Delta X_{1,i} \oplus \Delta X_{2,i}$ = ½ (key)

$\oplus \Delta S_{1,i} \oplus \Delta S_{2,i}$ > ½ Yippee!

Why is $\Delta M_{1,i} \oplus \Delta M_{2,i} > ½$

Message is in German, more likely following letter is a repetition than random

Why is $\Delta S_{1,i} \oplus \Delta S_{2,i} > ½$

S-wheels only turn when M-wheel is 1

Actual Advantage

- Probability of repeating letters

  $\text{Prob}[\Delta M_{1,i} \oplus \Delta M_{2,i} = 0]$ ~ **0.614**

  3.3% of German digraphs are repeating
- Probability of repeating S-keys

  $\text{Prob}[\Delta S_{1,i} \oplus \Delta S_{2,i} = 0]$ ~ **0.73**

$\text{Prob}[\Delta Z_{1,i} \oplus \Delta Z_{2,i} \oplus \Delta X_{1,i} \oplus \Delta X_{2,i} = 0]$

$= 0.614 * 0.73 \quad + (1\text{-}0.614) * (1\text{-}0.73)$

Δ M and S are 0 Δ M and S are 1

= 0.55 **if** the wheel settings guess is correct (0.5 otherwise)

Using the Advantage

- If the guess of **X** is correct, should see higher than ½ of the double deltas are 0
- Try guessing different configurations to find highest number of 0 double deltas
- Problem:

  \# of double delta operations to try one config

  = length of Z * length of X

  = for 10,000 letter message = 12 M for each setting * 7 ⊕ per double delta

  = 89 M ⊕ operations

Need a fast way to compute XOR!

Homework

- **Problem Set 4 Due Today**
- **Study for Exam 1**
  - **Out on Monday**