# INTRODUCTION TO ALGORITHMS

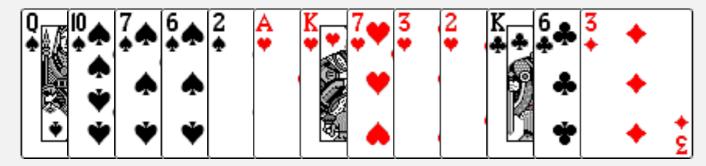
LECTURE 4: ELEMENTARY SORTING ALGORITHM

Yao-Chung Fan yfan@nchu.edu.tw

# Sorting applications



FedEx packages



playing cards



contacts

```
public class Girl {
  private int age = 28;
     public int getAge() {
        return 20;
```



### Selection sort

Algorithm. † scans from left to right.

小到大排列,從陣列中找 出最小的放到左邊第一 個,重複此動作

#### Invariants.

- Entries the left of ↑ (including ↑) fixed and in ascending order.
- No entry to right of ↑ is smaller than any entry to the left of ↑.



# Selection sort inner loop

To maintain algorithm invariants:

Move the pointer to the right.

```
i++;
```

in final order

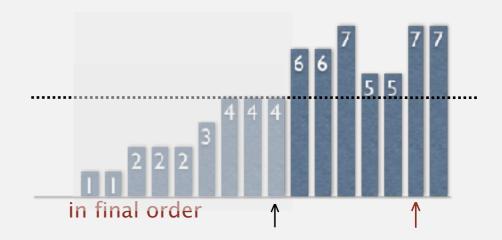
• Identify index of minimum entry on right.

```
int min = i;
for (int j = i+1; j < N; j++)
  if (less(a[j], a[min]))
  min = j;</pre>
```



• Exchange into position.

```
exch(a, i, min);
```



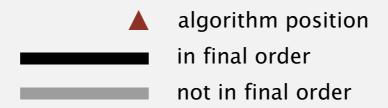
# Selection sort: Java implementation

```
public class Selection
   public static void sort(Comparable[] a)
   {
     int N = a.length;
      for (int i = 0; i < N; i++)
         int min = i;
         for (int j = i+1; j < N; j++)
            if (less(a[j], a[min]))
              min = j;
         exch(a, i, min);
   }
   private static boolean less(Comparable v, Comparable w)
   { /* as before */ }
   private static void exch(Comparable[] a, int i, int j)
   { /* as before */ }
```

# Selection sort: animations

#### 20 random items

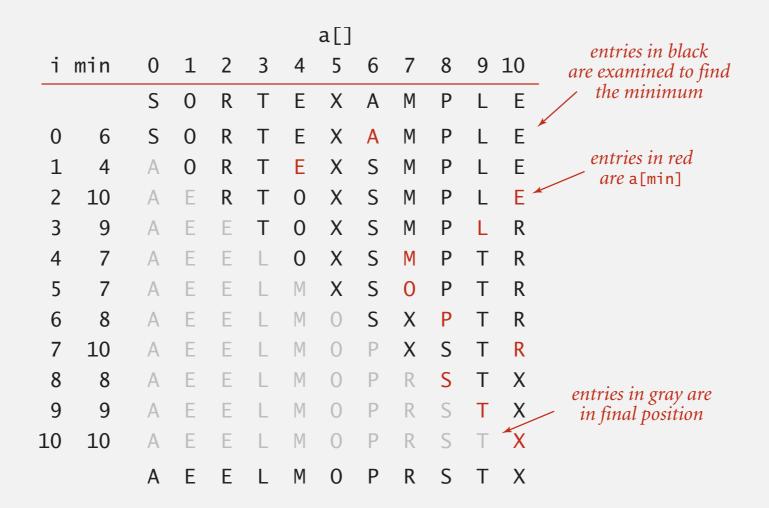




http://www.sorting-algorithms.com/selection-sort

# Selection sort: mathematical analysis

Proposition. Selection sort uses  $(N-1)+(N-2)+...+1+0 \sim N^2/2$  compares and N exchanges.



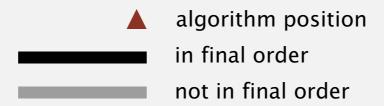
Trace of selection sort (array contents just after each exchange)

Running time insensitive to input. Quadratic time, even if input is sorted. Data movement is minimal. Linear number of exchanges.

# Selection sort: animations

#### 20 partially-sorted items





http://www.sorting-algorithms.com/selection-sort

# Sample sort client 1

- Goal. Sort any type of data.
- Ex 1. Sort random real numbers in ascending order.

seems artificial (stay tuned for an application)

```
public class Experiment
   public static void main(String[] args)
      int N = Integer.parseInt(args[0]);
      Double[] a = new Double[N];
      for (int i = 0; i < N; i++)
         a[i] = StdRandom.uniform();
      Insertion.sort(a);
      for (int i = 0; i < N; i++)
         StdOut.println(a[i]);
```

```
% java Experiment 10
0.08614716385210452
0.09054270895414829
0.10708746304898642
0.21166190071646818
0.363292849257276
0.460954145685913
0.5340026311350087
0.7216129793703496
0.9003500354411443
0.9293994908845686
```

# Sample sort client 2

- Goal. Sort any type of data.
- Ex 2. Sort strings in alphabetical order.

```
public class StringSorter
   public static void main(String[] args)
      String[] a = StdIn.readAllStrings();
      Insertion.sort(a);
      for (int i = 0; i < a.length; i++)
         StdOut.println(a[i]);
       % more words3.txt
        bed bug dad yet zoo ... all bad yes
       % java StringSorter < words3.txt</pre>
        all bad bed bug dad ... yes yet zoo
        [suppressing newlines]
```

# Sample sort client 3

- Goal. Sort any type of data.
- Ex 3. Sort the files in a given directory by filename.

```
import java.io.File;
public class FileSorter
   public static void main(String[] args)
      File directory = new File(args[0]);
      File[] files = directory.listFiles();
      Insertion.sort(files);
      for (int i = 0; i < files.length; i++)
         StdOut.println(files[i].getName());
```

% java FileSorter .
Insertion.class
InsertionX.class
InsertionX.java
Selection.class
Selection.java
Shell.class
Shell.java
ShellX.class
ShellX.java

### Callbacks

Goal. Sort any type of data (for which sorting is well defined).

Q. How can sort() know how to compare data of type Double, String, and java.io.File without any information about the type of an item's key?

Callback = reference to executable code.

- Client passes array of objects to sort() function.
- The sort() function calls object's compareTo() method as needed.

### Implementing callbacks.

- Java: interfaces.
- C: function pointers.
- C++: class-type functors.
- C#: delegates.
- Python, Perl, ML, Javascript: first-class functions.

# Callbacks: roadmap

#### client

```
public class StringSorter
{
   public static void main(String[] args)
   {
      String[] a = StdIn.readAllStrings();
      Insertion.sort(a);
      for (int i = 0; i < a.length; i++)
            StdOut.println(a[i]);
   }
}</pre>
```

#### data-type implementation

```
public class String
implements Comparable<String>
{
    ...
    public int compareTo(String b)
    {
        ...
        return -1;
        ...
        return +1;
        ...
        return 0;
    }
}
```

#### Comparable interface (built in to Java)

```
public interface Comparable<Item>
{
    public int compareTo(Item that);
}
```

key point: no dependence on String data type

#### sort implementation

```
public static void sort(Comparable[] a)
{
   int N = a.length;
   for (int i = 0; i < N; i++)
        for (int j = i; j > 0; j--)
            if (a[j].compareTo(a[j-1]) < 0)
            exch(a, j, j-1);
        else break;
}</pre>
```

# Comparable API

Implement compareTo() so that v.compareTo(w)

- Defines a total order.
- Returns a negative integer, zero, or positive integer if v is less than, equal to, or greater than w, respectively.
- Throws an exception if incompatible types (or either is null).



Built-in comparable types. Integer, Double, String, Date, File, ... User-defined comparable types. Implement the Comparable interface.

# Implementing the Comparable interface

Date data type. Simplified version of java.util.Date.

```
public class Date implements Comparable<Date>
{
   private final int month, day, year;
   public Date(int m, int d, int y)
     month = m;
     day = d;
     year = y;
  public int compareTo(Date that)
     if (this.year < that.year ) return -1;
     if (this.year > that.year ) return +1;
     if (this.month < that.month) return -1;
     if (this.month > that.month) return +1;
     if (this.day < that.day ) return -1;
     if (this.day > that.day ) return +1;
     return 0;
```

only compare dates to other dates

# Two useful sorting abstractions

Helper functions. Refer to data through compares and exchanges.

Less. Is item v less than w?

```
private static boolean less(Comparable v, Comparable w)
{ return v.compareTo(w) < 0; }</pre>
```

Exchange. Swap item in array a[] at index i with the one at index j.

```
private static void exch(Comparable[] a, int i, int j)
{
   Comparable swap = a[i];
   a[i] = a[j];
   a[j] = swap;
}
```

# Selection sort: Java implementation

```
public class Selection
   public static void sort(Comparable[] a)
   {
     int N = a.length;
      for (int i = 0; i < N; i++)
         int min = i;
         for (int j = i+1; j < N; j++)
            if (less(a[j], a[min]))
              min = j;
         exch(a, i, min);
   }
   private static boolean less(Comparable v, Comparable w)
   { /* as before */ }
   private static void exch(Comparable[] a, int i, int j)
   { /* as before */ }
```



### Insertion sort

Algorithm. † scans from left to right.

起始兩個數字先排序,依序再加入新的數字,加入後再與先前的 數字比較,較小就往前移

#### Invariants.

- Entries to the left of ↑ (including ↑) are in ascending order.
- Entries to the right of \( \) have not yet been seen.



# Insertion sort inner loop

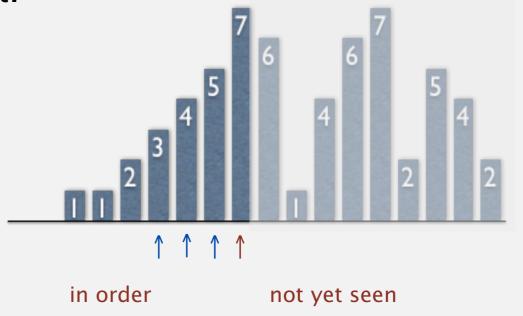
### To maintain algorithm invariants:

Move the pointer to the right.



Moving from right to left, exchange
 a[i] with each larger entry to its left.

```
for (int j = i; j > 0; j--)
  if (less(a[j], a[j-1]))
      exch(a, j, j-1);
  else break;
```

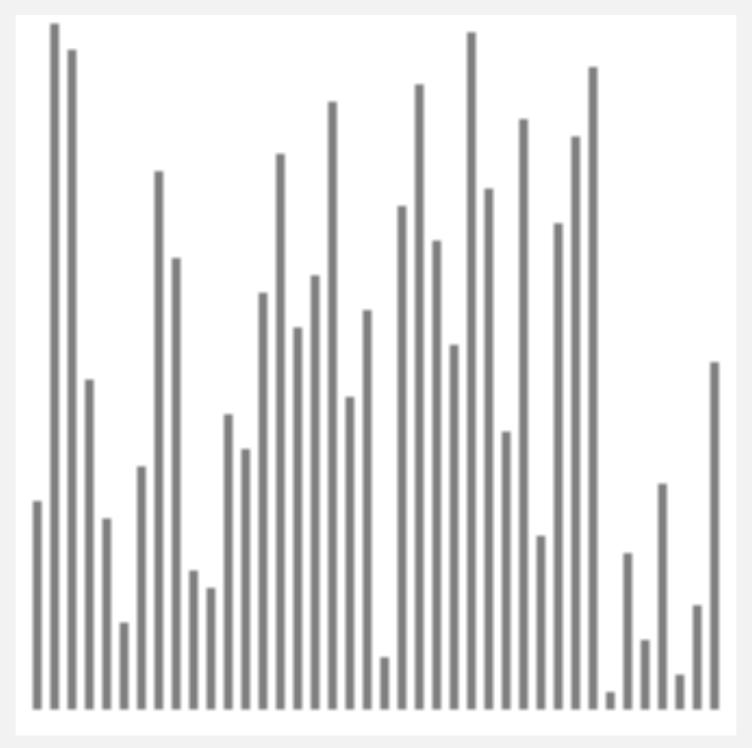


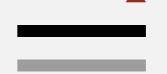
# Insertion sort: Java implementation

```
public class Insertion
   public static void sort(Comparable[] a)
      int N = a.length;
      for (int i = 0; i < N; i++)
         for (int j = i; j > 0; j--)
            if (less(a[j], a[j-1]))
               exch(a, j, j-1);
            else break;
   }
   private static boolean less(Comparable v, Comparable w)
   { /* as before */ }
   private static void exch(Comparable[] a, int i, int j)
   { /* as before */ }
```

# Insertion sort: animation

#### 40 random items



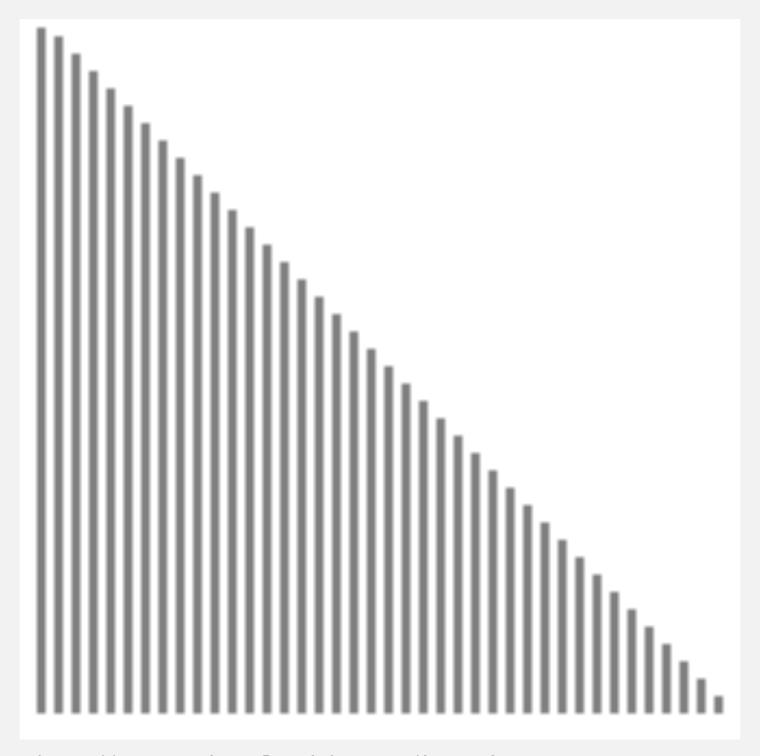


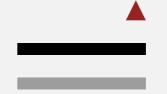
algorithm position in order not yet seen

http://www.sorting-algorithms.com/insertion-sort

# Insertion sort: animation

#### 40 reverse-sorted items





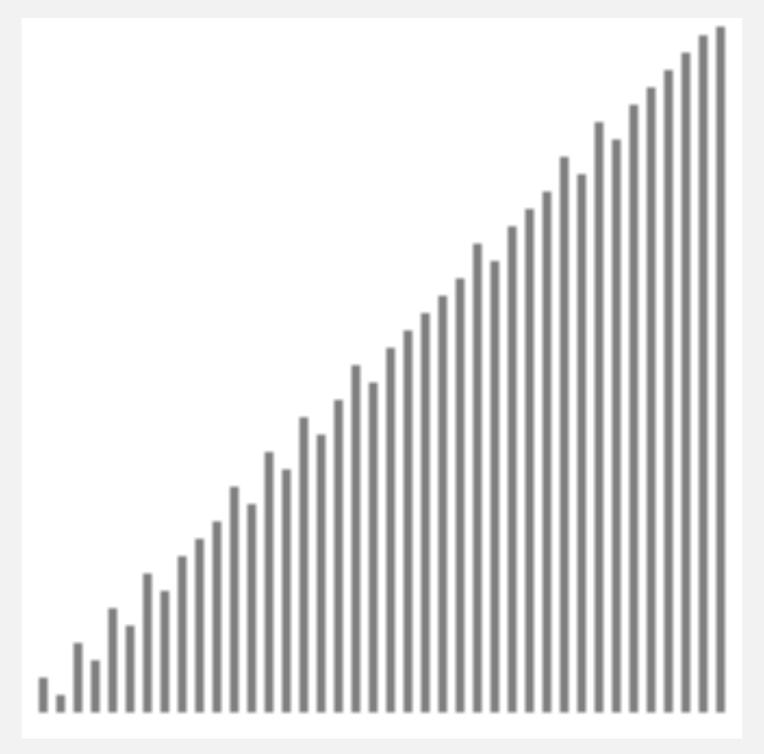
algorithm position in order not yet seen

http://www.sorting-algorithms.com/insertion-sort

Reason it is slow: excessive data movement.

### Insertion sort: animation

#### 40 partially-sorted items



in or

algorithm position in order not yet seen

http://www.sorting-algorithms.com/insertion-sort

Reason it is slow: excessive data movement.

# Insertion sort: mathematical analysis

Proposition. To sort a randomly-ordered array with distinct keys, insertion sort uses  $\sim \frac{1}{4} N^2$  compares and  $\sim \frac{1}{4} N^2$  exchanges on average.

Pf. Expect each entry to move halfway back.

Best case. If the array is in ascending order, insertion sort makes N-1 compares and 0 exchanges.

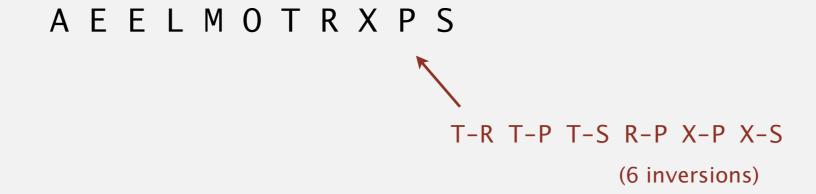
AEELMOPRSTX

Worst case. If the array is in descending order (and no duplicates), insertion sort makes  $\sim \frac{1}{2} N^2$  compares and  $\sim \frac{1}{2} N^2$  exchanges.

XTSRPOMLFEA

# Insertion sort: partially-sorted arrays

Def. An inversion is a pair of keys that are out of order.



Def. An array is partially sorted if the number of inversions is  $\leq c N$ .

- Ex 1. A sorted array has 0 inversions.
- Ex 2. A subarray of size 10 appended to a sorted subarray of size N.

Proposition. For partially-sorted arrays, insertion sort runs in linear time. Pf. Number of exchanges equals the number of inversions.

number of compares = exchanges + 
$$(N - 1)$$

# Insertion sort: practical improvements

Half exchanges. Shift items over (instead of exchanging).

- Eliminates unnecessary data movement.
- No longer uses only less() and exch() to access data.

ACHHIMNNPQXYKBINARY

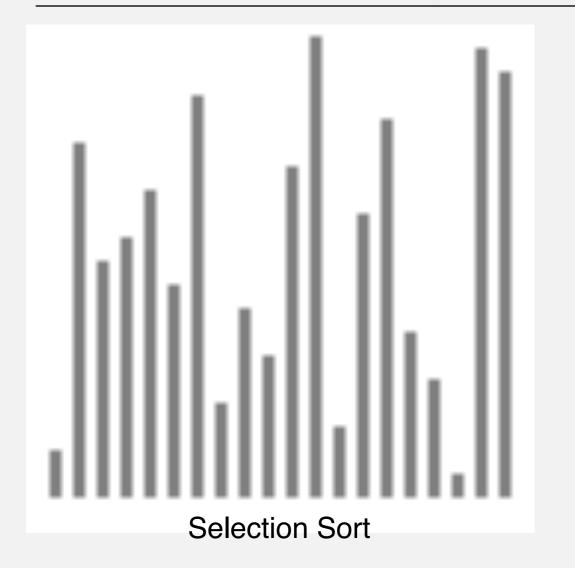
Binary insertion sort. Use binary search to find insertion point.

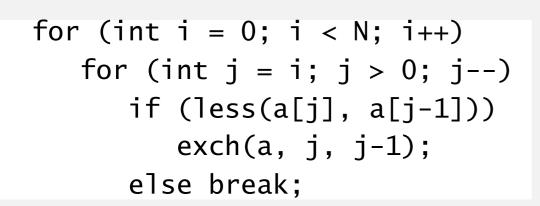
- Number of compares  $\sim N \lg N$ .
- But still a quadratic number of array accesses (worst case).



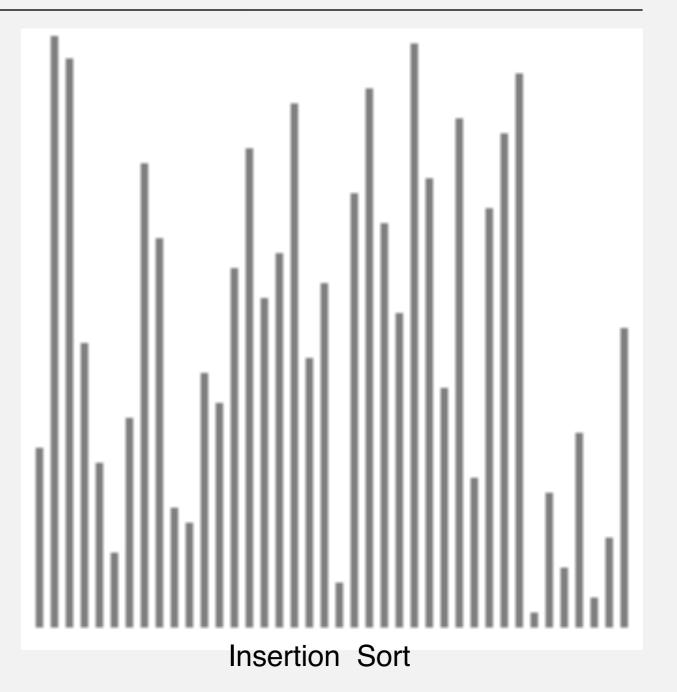


# Remind you that









### Shellsort overview

Idea. Move entries more than one position at a time by h-sorting the array.

an h-sorted array is h interleaved sorted subsequences

Shellsort. [Shell 1959] h-sort array for decreasing sequence of values of h.

```
        input
        S
        H
        E
        L
        L
        S
        O
        R
        T
        E
        X
        A
        M
        P
        L
        E

        13-sort
        P
        H
        E
        L
        L
        S
        O
        R
        T
        E
        X
        A
        M
        S
        L
        E

        4-sort
        L
        E
        E
        A
        M
        H
        L
        E
        P
        S
        O
        L
        T
        S
        X
        R

        1-sort
        A
        E
        E
        E
        H
        L
        L
        L
        M
        O
        P
        R
        S
        S
        T
        X
```

# h-sorting

How to h-sort an array? Insertion sort, with stride length h.

#### 3-sorting an array

```
M O L E E X A S P R T E O L M E X A S P R T E E L M O X A S P R T E E L M O X A S P R T A E L E O X M S P R T A E L E O P M S X R T A E L E O P M S X R T A E L E O P M S X R T A E L E O P M S X R T A E L E O P M S X R T
```

### Why insertion sort?

- Big increments  $\Rightarrow$  small subarray.
- Small increments ⇒ nearly in order. [stay tuned]

# Shellsort example: increments 7, 3, 1

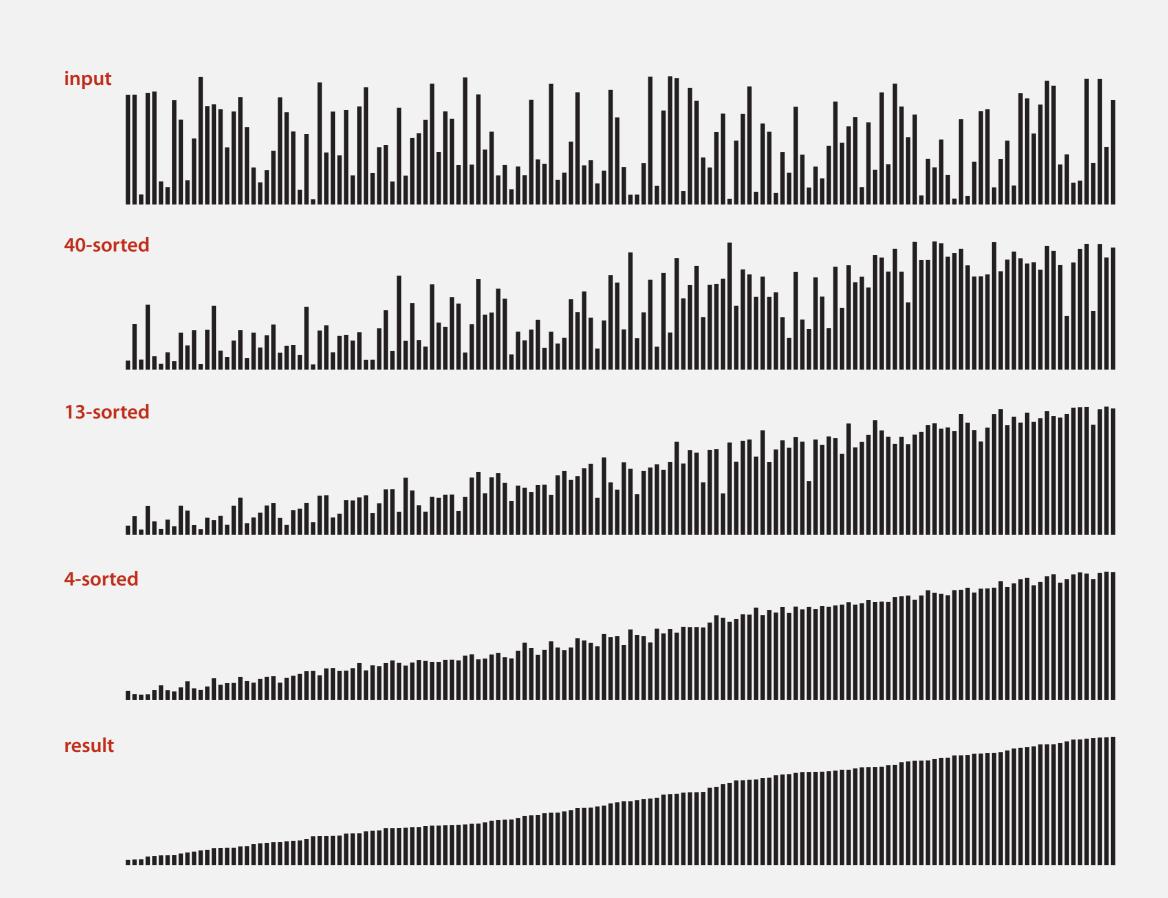
Idea. h-sort the file for a decreasing sequence of values of h.

# input RTEXAMPLE 7-sort Μ Α X 3-sort Ε Ε X A S M 0 X Μ



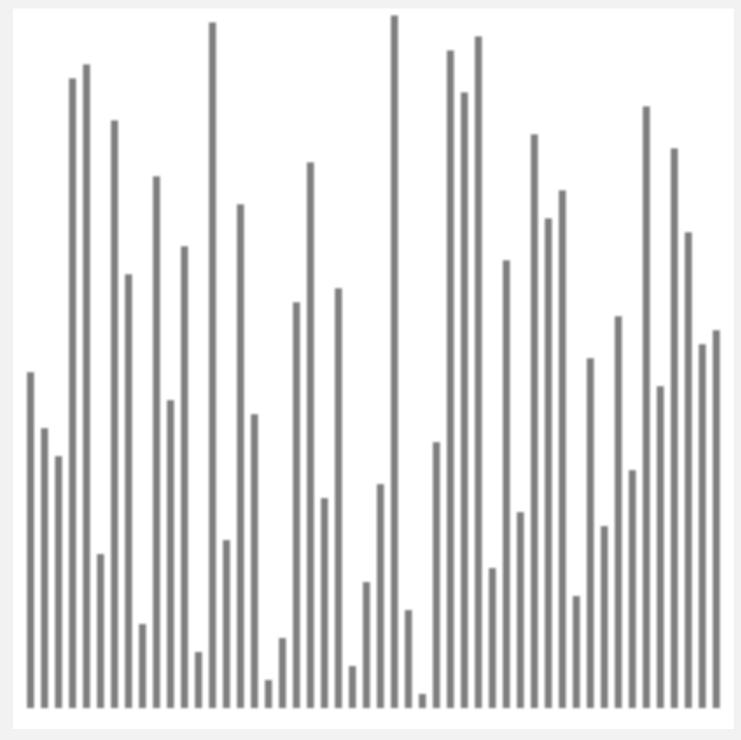
# Shellsort: Java implementation

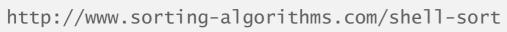
```
To implement: Use insertion sort, modified to h-sort.
public class Shell
    public static void sort(Comparable[] a)
       int N = a.length;
       int h = 1;
                                                                             3x+1 increment
       while (h < N/3) h = 3*h + 1; // 1, 4, 13, 40, 121, 364, ...
                                                                             sequence
       while (h >= 1)
       { // h-sort the array.
          for (int i = h; i < N; i++)
                                                                             insertion sort
             for (int j = i; j >= h && less(a[j], a[j-h]); <math>j -= h)
                exch(a, j, j-h);
                                                                              move to next
          h = h/3;
                                                                             increment
    }
    private static boolean less(Comparable v, Comparable w)
    { /* as before */ }
    private static void exch(Comparable[] a, int i, int j)
    { /* as before */ }
```

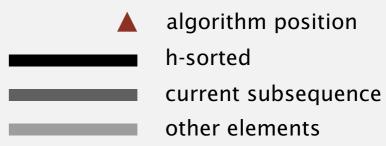


### Shellsort: animation

#### 50 random items

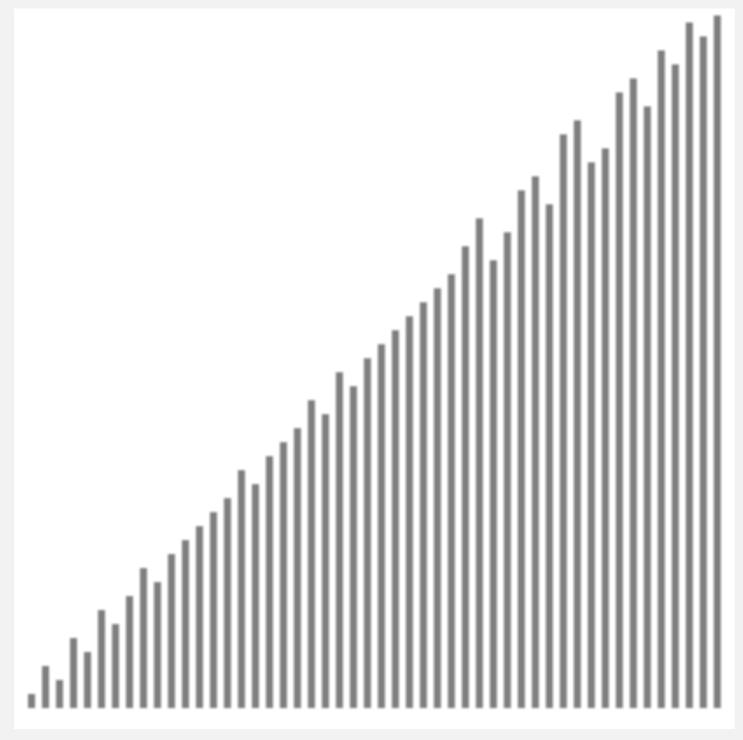


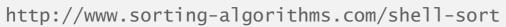




### Shellsort: animation

#### 50 partially-sorted items







algorithm position h-sorted current subsequence other elements

## Shellsort: which increment sequence to use?

Powers of two. 1, 2, 4, 8, 16, 32, ...
No.

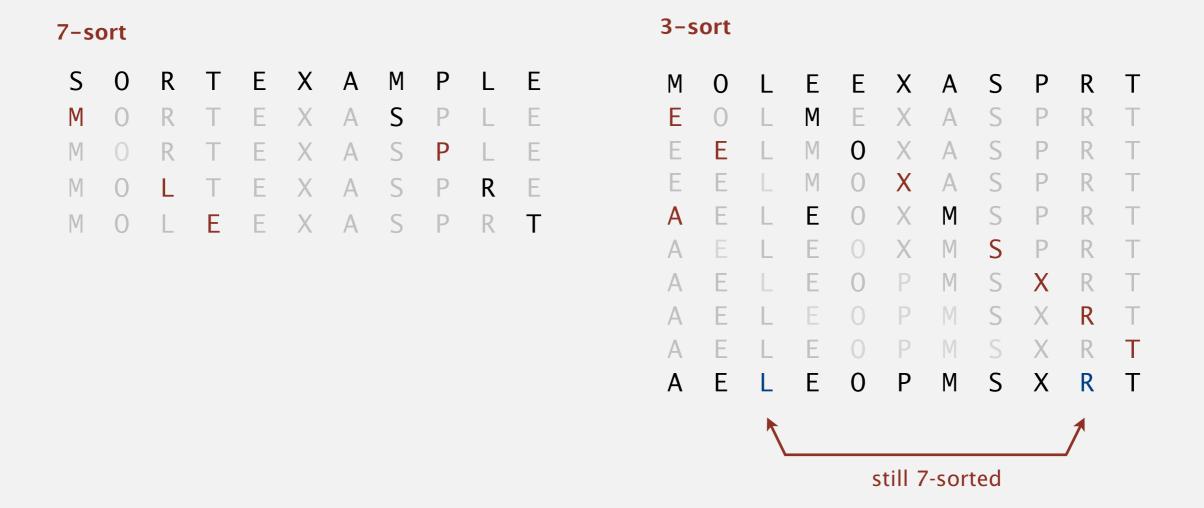
期中會考

Powers of two minus one. 1, 3, 7, 15, 31, 63, ... Maybe.

→ 3x + 1. 1, 4, 13, 40, 121, 364, ...
OK. Easy to compute.

### Shellsort: intuition

Proposition. An h-sorted array remains h-sorted after g-sorting it.



Challenge. Prove this fact—it's more subtle than you'd think!

### Shellsort: analysis

Proposition. The order of growth of the worst-case number of compares used by shellsort with the 3x+1 increments is  $N^{3/2}$ .

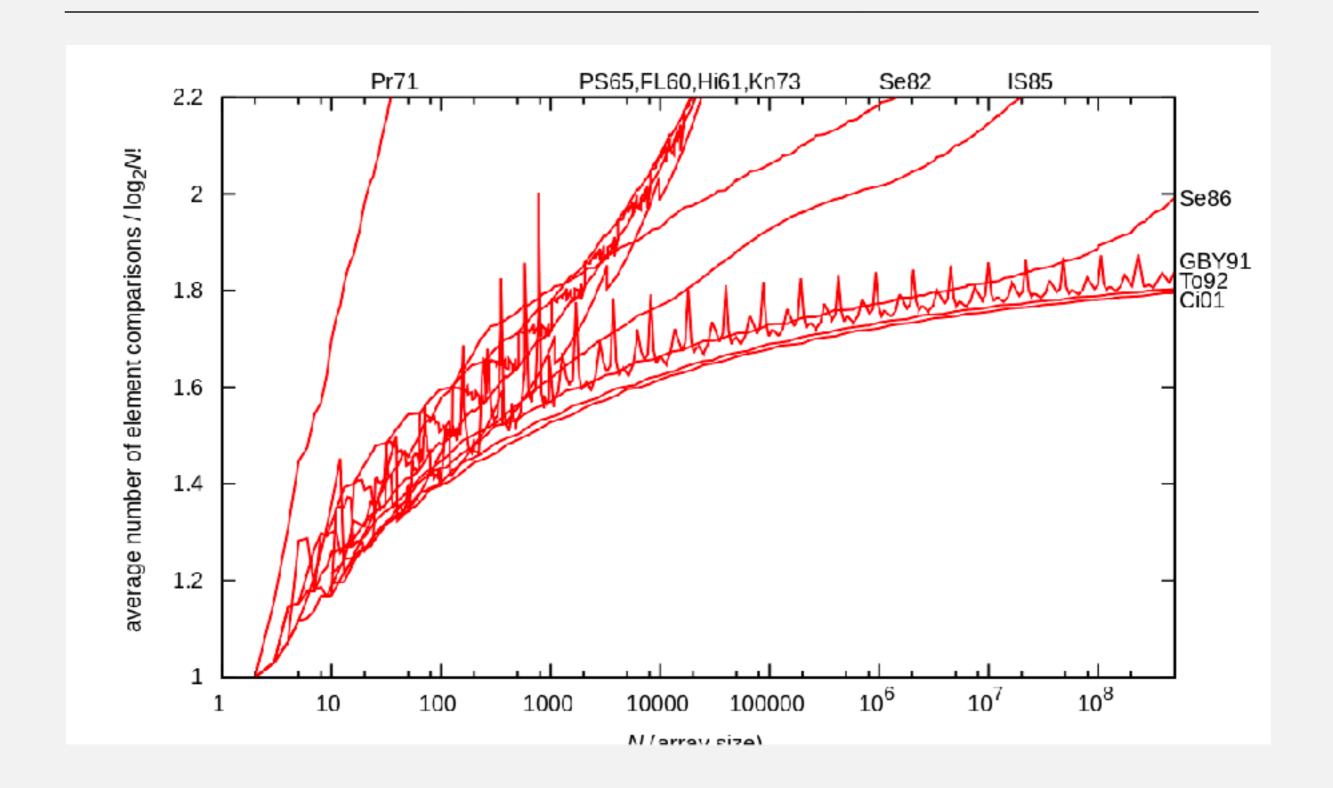
Property. The expected number of compares to shellsort a randomly-ordered array using 3x+1 increments is....

N	compares	2.5 N ln N	0.25 N ln <sup>2</sup> N	<b>N</b> 1.5
5,000	93K	106K	91K	64K
10,000	209K	230K	213K	158K
20,000	467K	495K	490K	390K
40,000	1022K	1059K	1122K	960K
80,000	2266K	2258K	2549K	2366K

Remark. Accurate model has not yet been discovered (!)

## About this research

General term (k ≥ 1)	Concrete gaps	worst-case time complexity	Author and year of publication
$\left\lfloor \frac{N}{2^k} \right\rfloor$	$\left\lfloor \frac{N}{2} \right\rfloor, \left\lfloor \frac{N}{4} \right\rfloor, \dots, 1$	$\Theta(N^2)$ [e.g. when $N=2P$ ]	Shell, 1959 <sup>[4]</sup>
$2\left\lfloor rac{N}{2^{k+1}} ight floor+1$	$2\left\lfloor \frac{N}{4} \right\rfloor + 1,\ldots,3,1$	$\Theta\left(N^{3/2}\right)$	Frank & Lazarus, 1960 <sup>[8]</sup>
$2^k - 1$	1, 3, 7, 15, 31, 63,	$\Theta\left(N^{3/2}\right)$	Hibbard, 1963 <sup>[9]</sup>
$2^k+1$ , prefixed with 1	1, 3, 5, 9, 17, 33, 65,	$\Theta\left(N^{3/2}\right)$	Papernov & Stasevich, 1965 <sup>[10]</sup>
Successive numbers of the form $2^p 3^q$	1, 2, 3, 4, 6, 8, 9, 12,	$\Theta\left(N\log^2N\right)$	Pratt, 1971 <sup>[1]</sup>
$\left\lceil rac{3^k-1}{2}  ight ceil$ , not greater than $\left\lceil rac{N}{3}  ight ceil$	1, 4, 13, 40, 121,	$\Theta\left(N^{3/2}\right)$	Pratt, 1971 <sup>[1]</sup>
$egin{align} &\prod_I a_q,  ext{where} \ &a_q = \min \left\{ n \in \mathbb{N} : n \geq \left(rac{5}{2} ight)^{q+1}, orall p : 0 \leq p < q \Rightarrow \gcd(a_p,n) = 1  ight\} \ &I = \left\{ 0 \leq q < r \mid q  eq rac{1}{2} \left(r^2 + r  ight) - k  ight\} \ &r = \left\lfloor \sqrt{2k + \sqrt{2k}}  ight floor \end{aligned}$	1, 3, 7, 21, 48, 112,	$O\left(N^{1+\sqrt{rac{8\ln(5/2)}{\ln(N)}}} ight)$	Incerpi & Sedgewick, 1985,[11] Knuth[3]
$4^k + 3 \cdot 2^{k-1} + 1$ , prefixed with 1	1, 8, 23, 77, 281,	$O\left(N^{4/3}\right)$	Sedgewick, 1986 <sup>[6]</sup>
$9\left(4^{k-1}-2^{k/2} ight)+1,4^{k+1}-6\cdot 2^{(k+1)/2}+1$	1, 5, 19, 41, 109,	$O\left(N^{4/3}\right)$	Sedgewick, 1986 <sup>[12]</sup>
$h_k = \max\left\{\left\lfloor \frac{5h_{k-1}}{11} \right floor, 1\right\}, h_0 = N$	$\left\lfloor \frac{5N}{11} \right\rfloor, \left\lfloor \frac{5}{11} \left\lfloor \frac{5N}{11} \right\rfloor \right\rfloor, \dots, 1$	Unknown	Gonnet & Baeza-Yates, 1991 <sup>[13]</sup>
$\left\lceil rac{9^k - 4^k}{5 \cdot 4^{k-1}}  ight ceil$	1, 4, 9, 20, 46, 103,	Unknown	Tokuda, 1992 <sup>[14]</sup>
Unknown (experimentally derived)	1, 4, 10, 23, 57, 132, 301, 701	Unknown	Ciura, 2001 <sup>[15]</sup>



### Why are we interested in shellsort?

#### Example of simple idea leading to substantial performance gains.

### Useful in practice.

- Fast unless array size is huge (used for small subarrays).
- Tiny, fixed footprint for code (used in some embedded systems).

R, bzip2, /linux/kernel/groups.c

Hardware sort prototype.

uClibc

#### Simple algorithm, nontrivial performance, interesting questions.

- Asymptotic growth rate?
- Best sequence of increments? open problem: find a better increment sequence
- Average-case performance?

Lesson. Some good algorithms are still waiting discovery.

## Elementary sorts summary

Today. Elementary sorting algorithms.

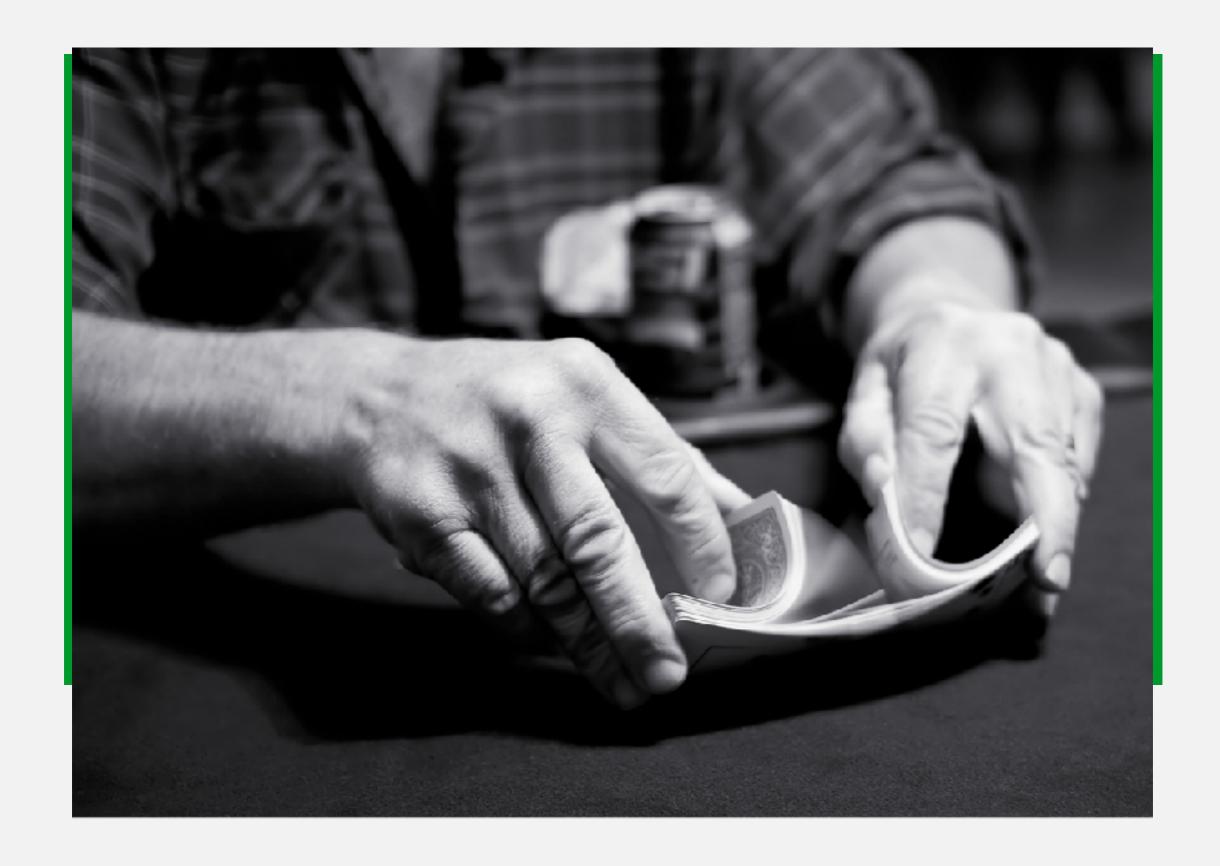
algorithm	best	average	worst
selection sort	<i>N</i> <sup>2</sup>	<i>N</i> <sup>2</sup>	$N^2$
insertion sort	N	<i>N</i> <sup>2</sup>	$N^2$
Shellsort (3x+1)	N log N	?	$N^{3/2}$
goal	N	$N \log N$	$N \log N$

order of growth of running time to sort an array of N items





# Online Gambling System?



## War story (Microsoft)

Microsoft antitrust probe by EU. Microsoft agreed to provide a randomized ballot screen for users to select browser in Windows 7.

http://www.browserchoice.eu

## Select your web browser(s)



A fast new browser from Google. Try it now!



Safari for Windows from Apple, the world's most innovative browser.



Your online security is Firefox's top priority. Firefox is free, and made to help you get the most out of the



The fastest browser on Earth. Secure, powerful and easy to use, with excellent privacy protection.



Designed to help you take control of your privacy and browse with confidence. Free from Microsoft.

appeared last 50% of the time

## War story (Microsoft)

Microsoft antitrust probe by EU. Microsoft agreed to provide a randomized ballot screen for users to select browser in Windows 7.

Solution? Implement shuffle sort by making comparator always return a random answer.

#### Microsoft's implementation in Javascript

```
public int compareTo(Browser that)
{
  double r = Math.random();
  if (r < 0.5) return -1;
  if (r > 0.5) return +1;
  return 0;
}
browser comparator
(should implement a total order)
```

### Knuth shuffle

- In iteration i, pick integer r between 0 and i uniformly at random.
- Swap a[i] and a[r].

common bug: between 0 and N – 1 correct variant: between i and N – 1

```
不必要再sorting
                                                           -次
public class StdRandom
   public static void shuffle(Object[] a)
   {
      int N = a.length;
      for (int i = 0; i < N; i++)
      {
                                                          between 0 and i
         int r = StdRandom.uniform(i + 1);
         exch(a, i, r);
```

### Broken Knuth shuffle

- Q. What happens if integer is chosen between 0 and N-1?
- A. Not uniformly random!



permutation	Knuth shuffle	broken shuffle
ABC	1/6	4/27
A C B	1/6	5/27
ВАС	1/6	5/27
BCA	1/6	5/27
C A B	1/6	4/27
СВА	1/6	4/27

probability of each result when shuffling { A, B, C }