CS100 Introduction to Programming

Lecture 10. Recursion

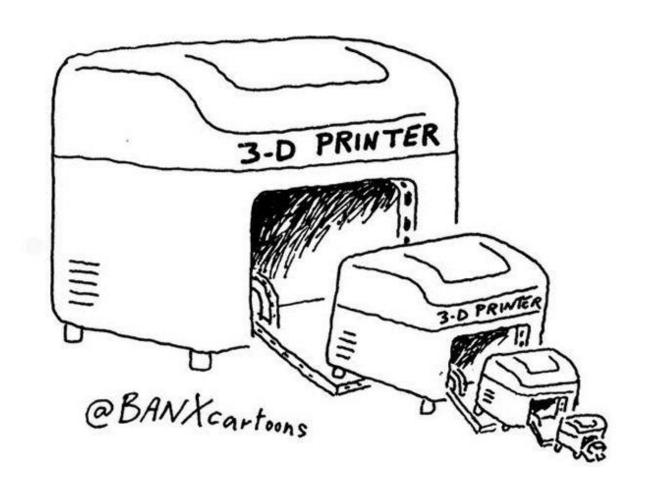
Learning Objectives

- At the end of this lecture, you should be able to understand the following:
 - Recursive Methods
 - Recursive Cheers
 - Recursive PrintSomething
 - Recursive Factorials
 - Recursive Multiplication by Addition
 - Recursive PrintDigits
 - Recursive SumArray
 - Tower of Hanoi

What Is Recursion?

- The method in which a problem is solved by reducing it to smaller cases of the same problem.
- Recursion is the name for the case when
 - a function calls itself, or
 - calls a sequence of other functions, one of which eventually calls the first function again.
- Two parts
 - Base case (with terminating condition)
 - Recursive case (with recursive condition)







Life Mirror Recursive (1909), by Coles Phillips (1880 – 1927)

FROM THE MIRROR.



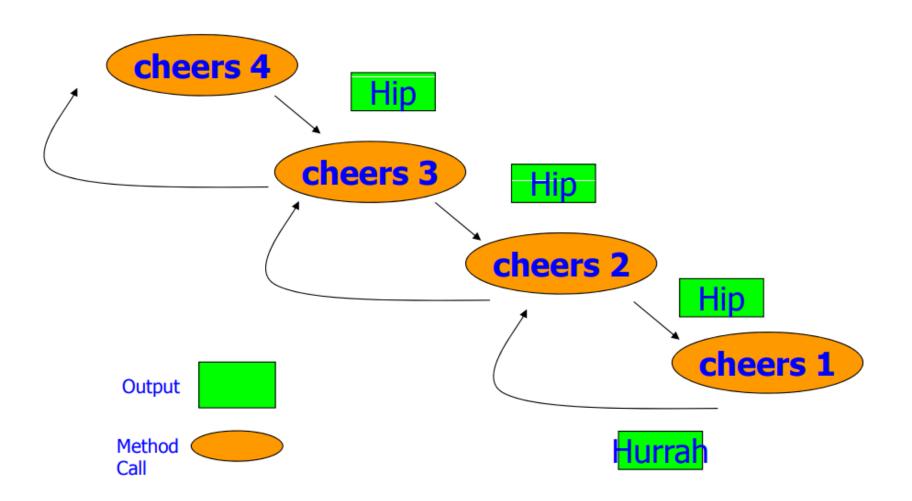
Example 1: Cheers

What does the following recursive method print when called with cheers (4)?

```
void cheers(int n)
{
   if (n <= 1)
     printf("Hurrah \n");
   else {
     printf("Hip \n");
     cheers(n-1);
   }
}</pre>
```

Output: Hip Hip Hip Hurrah

Recursive Cheers – Tracing



Recursive Cheers

```
terminating condition
void cheers(int n)
   if (n <= 1)
      printf("Hurrah \n");
                              recursive condition
  else {
                                    (n > 1)
      printf("Hip \n");
      cheers(n-1);
```

Example 2: PrintSomething

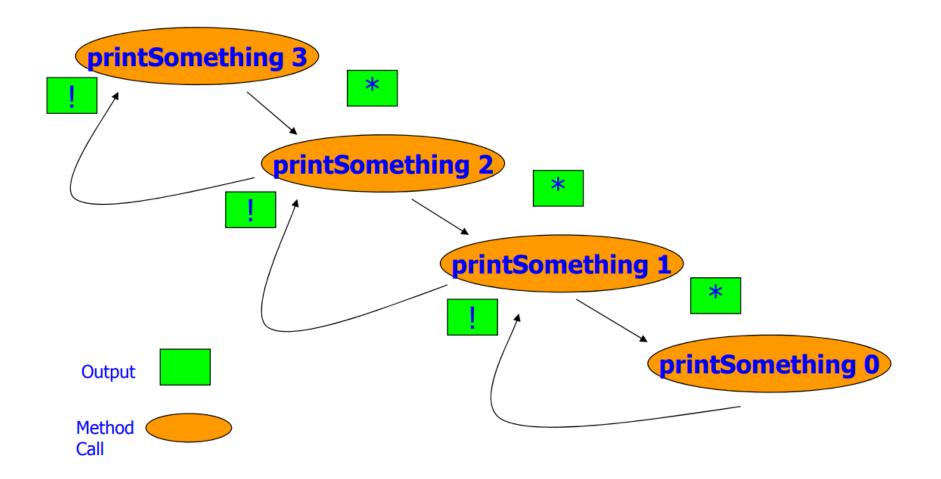
What does the following recursive method print when called with printSomething(3)?

```
void printSomething(int n)
{
    if (n > 0) {
        printf("*");
        printSomething(n-1);
        printf("!");
    }
}
What condition is this?
Continue looping condition

Output:
    ***!!!
    ***!!!
}

Where is the other condition?
```

Recursive PrintSomething – Tracing



Example 3: Factorials

- Problem: Find the factorial of a non-negative integer number.
- The factorial function of a positive integer is usually defined by the formula

$$n! = n \times (n - 1) \times \dots \times 1$$

A more precise definition (recursive definition)

$$n! = 1$$
 if $n = 0$
 $n! = n \times (n - 1)!$ if $n > 0$

Non-recursive Factorials

The following program will do the job, using a

for loop:

```
#include <stdio.h>
int factorial(int n);
int main(void)
   int num;
   printf("Enter an integer:");
   scanf("%d", &num);
   printf("n! = %d\n",
          factorial(num));
   return 0;
```

```
int factorial(int n)
{
    int i;
    int temp = 1;
    for (i=n; i > 0; i--)
        temp *= i;
    return temp;
}
```

```
Output:
Enter an integer: <u>4</u>
n! = 24
```

Recursive Factorials

Suppose we wish to calculate 4!

As
$$4 > 0$$
, $4! = 4 \times 3!$

But we do not know what is 3!

As
$$3 > 0$$
, $3! = 3 \times 2!$
As $2 > 0$, $2! = 2 \times 1!$
As $1 > 0$, $1! = 1 \times 0!$

What is 0! equal to?

$$4! = 4 \times 3!$$

 $= 4 \times (3 \times 2!)$
 $= 4 \times (3 \times (2 \times 1!))$
 $= 4 \times (3 \times (2 \times (1 \times 0!)))$
 $= 4 \times (3 \times (2 \times (1 \times 1)))$
 $= 4 \times (3 \times (2 \times 1))$
 $= 4 \times (3 \times 2)$
 $= 4 \times 6$
 $= 24$

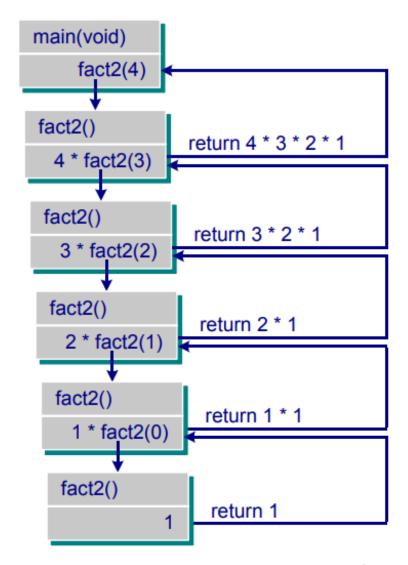
Recursive Factorials

```
int factorial(int n)
{
  if (n == 0) {
  // teminating condition
     return 1;
  } else {
  // recursive condition
     return n*factorial(n-1);
```

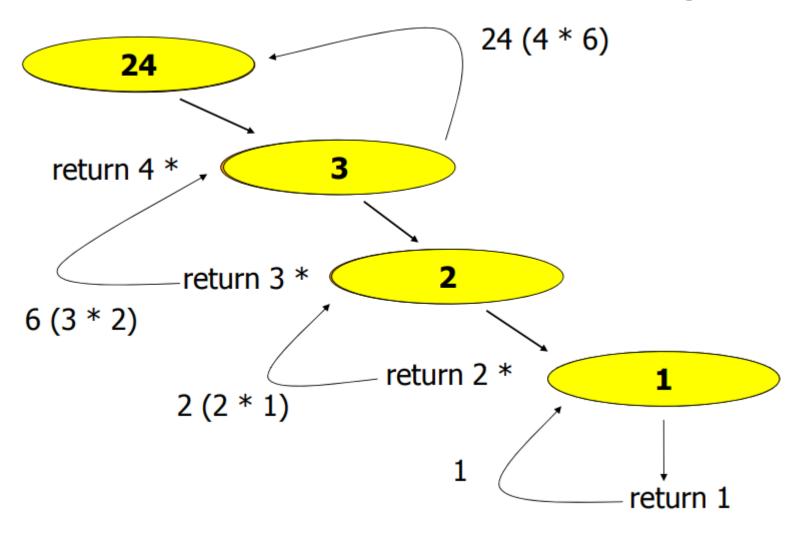
Output:

Enter an integer: 4

n! = 24



Recursive Factorials: Tracing



Example 4: Multiplication by Addition

The multiplication operation

$$multi(m, n) = m \times n$$

can be defined **recursively** as

```
multi(m, n) = m if n = 1

multi(m, n) = m + multi(m, n-1) if n > 1
```

Using Call by Value

```
/* multiplication by addition, pass parameter using
call by value */
#include <stdio.h>
int multi1(int, int);
int main(void)
   printf("5 * 3 = %d\n", multi1(5, 3));
   return 0;
                                              Output:
                                              5 * 3 = 15
int multi1(int m, int n)
   if (n == 1)
                           // terminating condition
      return m;
   else
                           // recursive condition
      return m + multi1(m, n-1);
}
```

Using Call by Reference

```
/* multiplication by addition, pass parameter using call by
reference */
#include <stdio.h>
void multi2(int, int, int*);
int main(void)
   int result;
   multi2(5, 3, &result);
   printf("5 * 3 = %d\n", result);
                                                   Output:
   return 0;
}
                                                   5 * 3 = 15
void multi2(int m, int n, int *product)
{
   if (n == 1)
                              // terminating condition
       *product = m;
                              // recursive condition
   else {
       multi2(m, n-1, product);
       *product = *product + m;
```

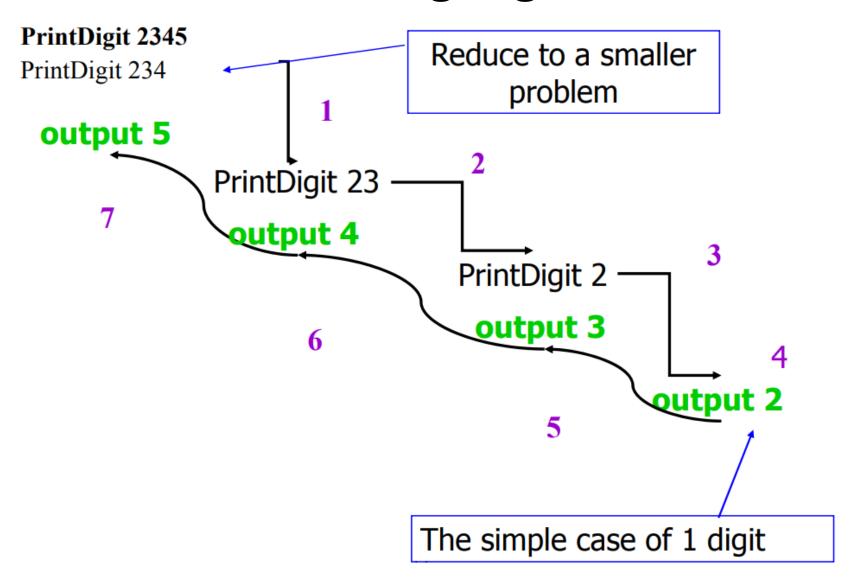
Recursive Call by Reference

```
#include <stdio.h>
void fn(int x, int y, int *z);
int main(void)
    int n1=20, n2=15, n3=0;
    fn(n1, n2, &n3);
    printf("n1 = %d, n2 = %d, n3 = %d\n", n1, n2, n3);
    return 0;
                                        Output:
                                        x = 10, y = 10, *z = 100
void fn(int x, int y, int *z)
                                        x = 10, y = 10, *z = 100
    if (x > y) {
                                        x = 12, y = 11, *z = 100
       x = x - 2;
                                        x = 14, y = 12, *z = 100
       y = y - 1;
                                        x = 16, y = 13, *z = 100
        *z = x * y;
                                        x = 18, y = 14, *z = 100
       fn(x, y, z);
                                        n1 = 20. n2 = 15. n3 = 100
    printf("x = %d, y = %d, *z = %d\n", x, y, *z);
```

Example 5: Printing Digits

- Problem: Given a number, say 2345, print each digit of the number, in the order from left to right, one per line.
- Recursive Solution:
 - Look for the <u>simplest case</u> (terminating condition). If the number is a single digit, just print that digit.
 - Look into <u>reducing</u> the problem into a <u>smaller</u> but same problem (<u>recursive</u> <u>condition</u>).

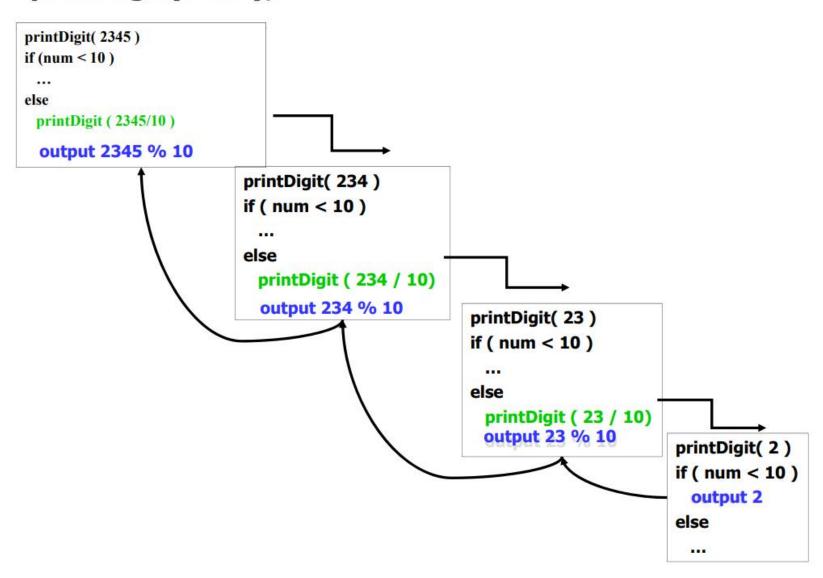
Printing Digits



Printing Digits: Recursive Solution

```
#include <stdio.h>
                                           Output:
void printDigit(int);
                                           Enter a number: 2345
int main(void) {
   int num;
                                           3
   printf("Enter a number: ");
   scanf("%d", &num);
   printDigit(num);
                                           5
   return 0;
void printDigit(int num) {
   if (num < 10)
                              // terminating condition
       printf("%d\n", num);
                              // recursive condition
   else {
       printDigit(num/10);
       printf("%d\n", num%10);
```

printDigit(2345);



Recursion: Summary

- To obtain the answer to a larger problem, a general method is used that
 - reduces the large problem to one or more problems of a similar nature but a smaller size.
- Recursion continues until the size of the problem is reduced to the smallest, i.e. base case
 - where the solution is given directly without using further recursion.
- As such, recursive methods consist of two parts:
 - A smallest, base case that is processed without recursion (terminating condition).
 - A general method that reduces a particular case to one or more of the smaller cases (recursive condition).

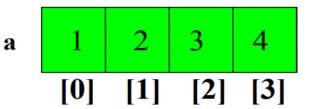
Recursion: Summary

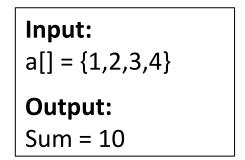
- Each function makes a call to itself with an argument which is <u>closer</u> to the terminating condition.
- Each call to a function has its own set of values/arguments for the formal arguments and local variables.
- When a recursive call is made, control is transferred from the calling point to the first statement of the recursive function. When a call at a certain level is finished, control returns to the calling point one level up.

Example 6: Summing Array of Integers

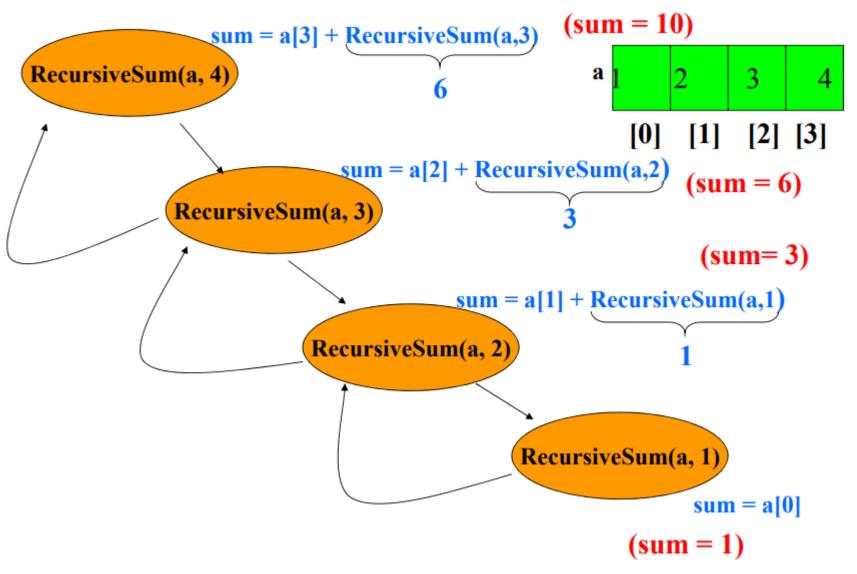
```
int sumArray(int a[], int size)
   int sum = 0;
   for (int i = 0; i < size; i++)
       sum = sum + a[i];
   return sum;
int recursiveSum(int a[], int size)
   if (size == 1)
       return a[0];
   else
       return a[size - 1] +
           recursiveSum(a, size - 1);
```

 Given an array of integers, write a method to calculate the sum of all the integers.





Recursive Sum – Tracing

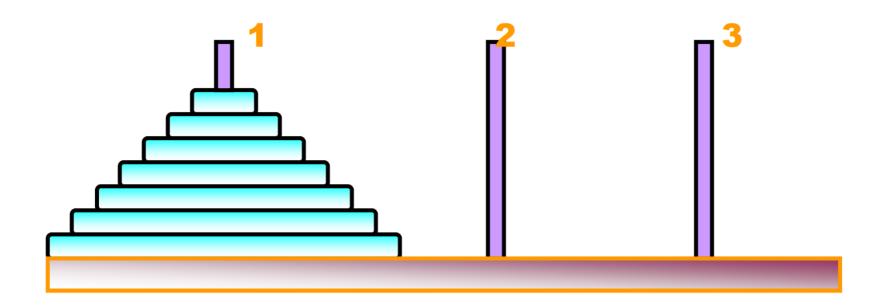


Example 7: Tower of Hanoi

- In the 19th century, a game called the **Towers of Hanoi** appeared in Europe.
- The game represents a task underway in the Temple of Brahma.
- At the creation of the world, the priests were given a brass platform on which were 3 diamond needles.
- On the first needle were stacked 64 golden disks.
- Each one slightly smaller than the one under it.
- The priest were assigned the task of moving all the golden disks from the first needle to the third
 - Condition: Only one disk can be moved at a time, and no disk is allowed to be placed on top of a smaller disk.

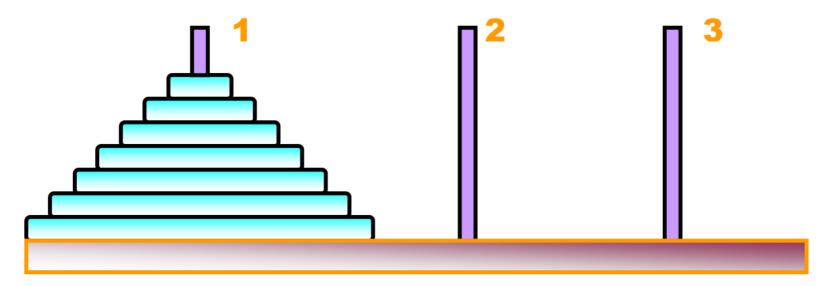
Tower of Hanoi

 Problem: The priests were told that, when they had finished moving the 64 disks from tower 1 to tower 3, it would signify the end of the world!!!

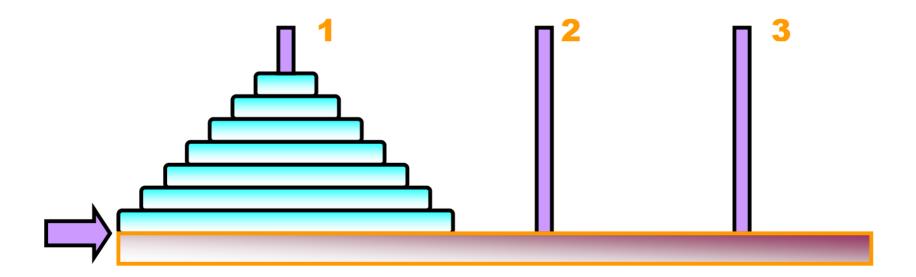


The Problem

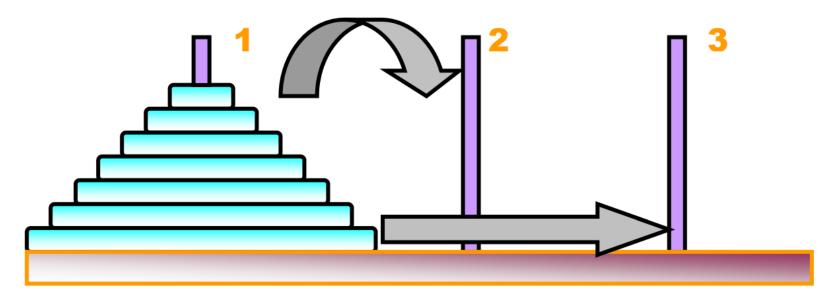
- Write a computer program that will type out a list of instructions for the priests
- move(64, 1, 3, 2)
 - Move 64 disks from tower 1 to tower 3 using tower 2 as temporary storage



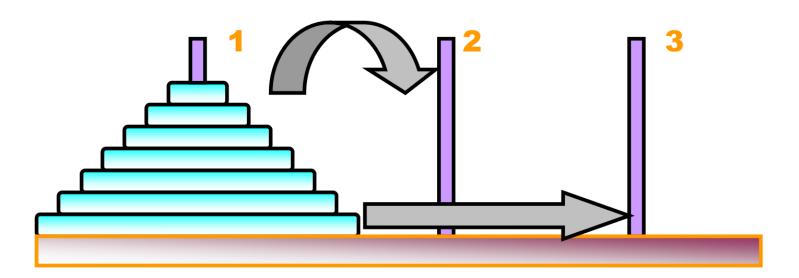
- Concentrate our attention not on the first step
 - Move the top disk anywhere
- Rather, on the hardest step
 - Moving the bottom disk



- How to reach the bottom disk?
 - The top 63 disks have to be moved to tower 2
 - The bottom disk can then be moved from tower 1 to 3
- There cannot be any other disk in tower 3

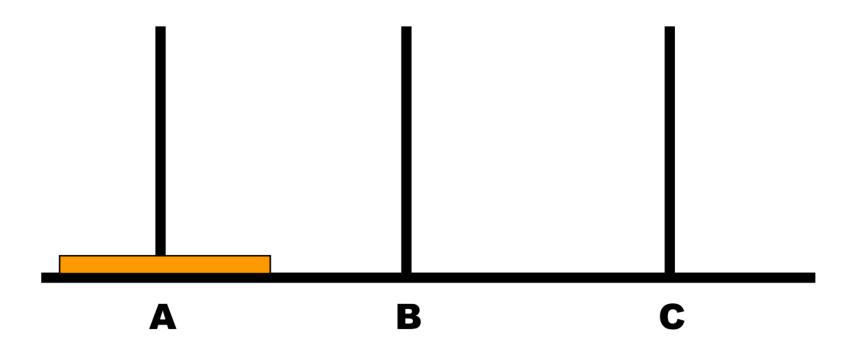


- move(63, 1, 2, 3); // Move 63 disks from tower 1 to tower 2 (using tower 3 as temporary)
- Display "Move disk 64 from tower 1 to 3"
- move (63, 2, 3, 1); // Move 63 disks from tower 2 to tower 3 (using tower 1 as temporary)

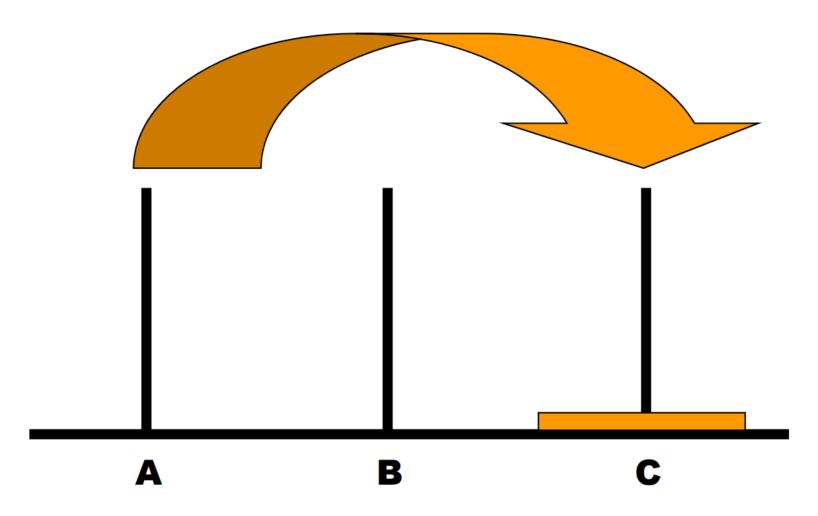


- The remaining 63 disks can be moved in the same way
- Divide and conquer
 - To solve a problem we split it into smaller parts which are easier to handle.

A Single Disk Tower

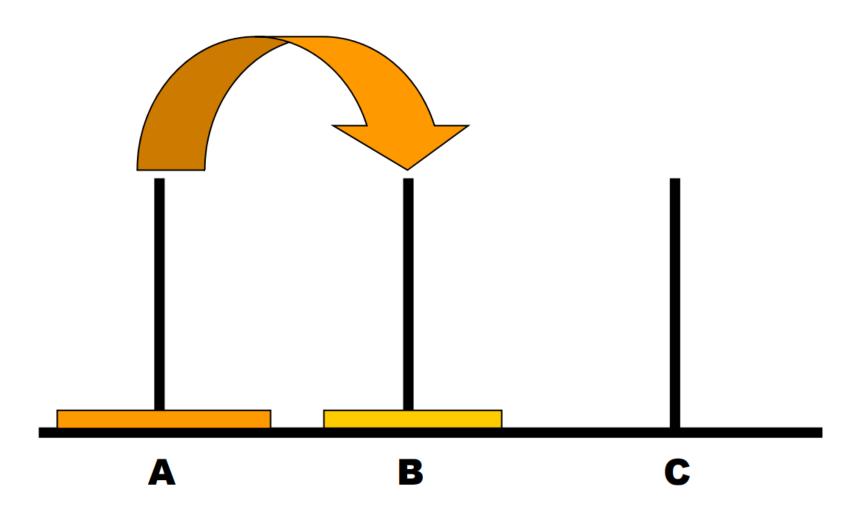


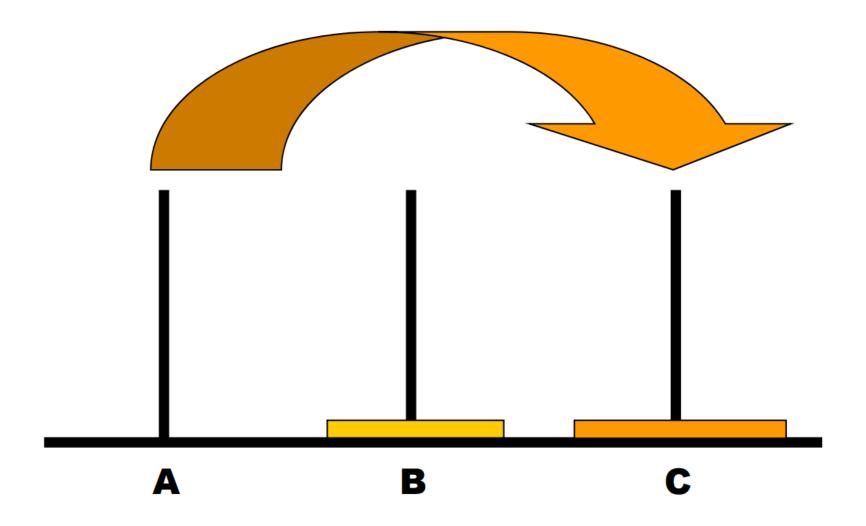
A Single Disk Tower



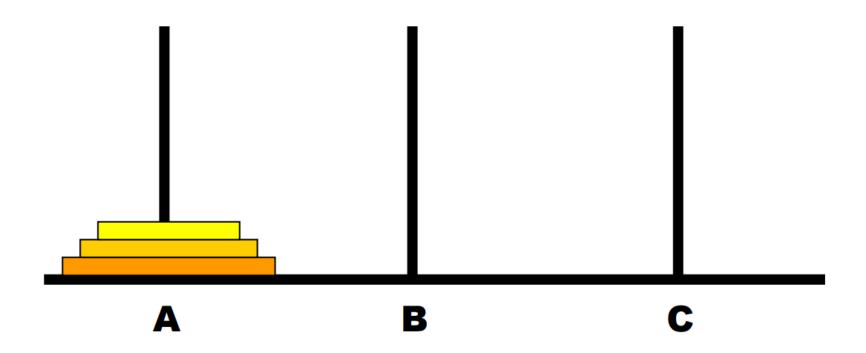
A 2-Disk Tower

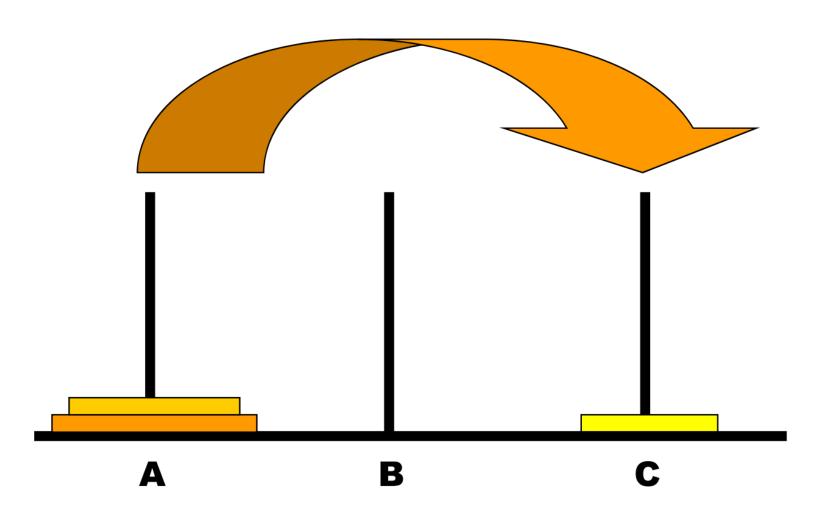


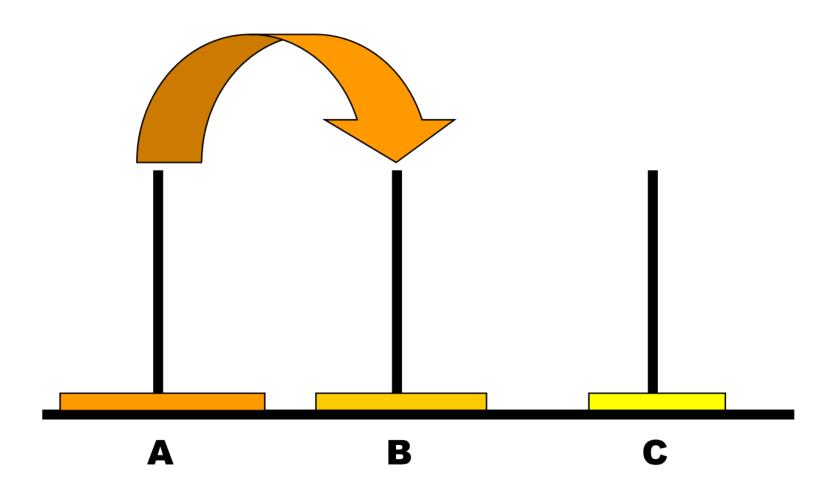


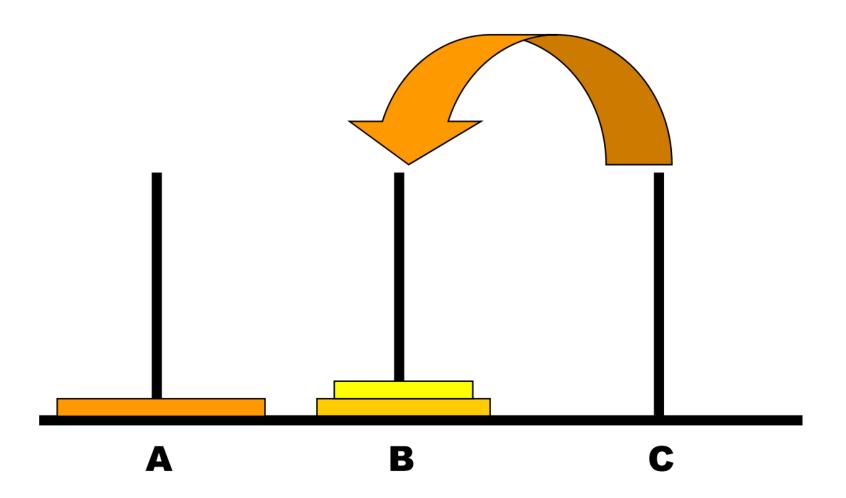


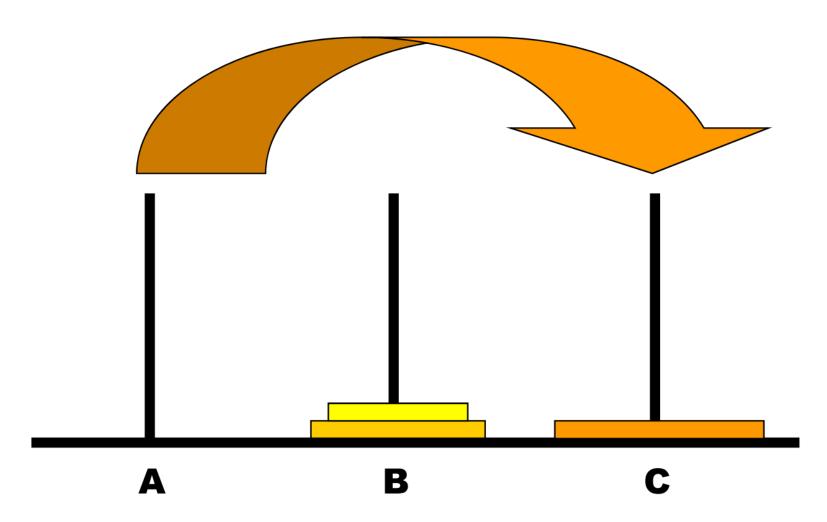
A 3-Disk Tower



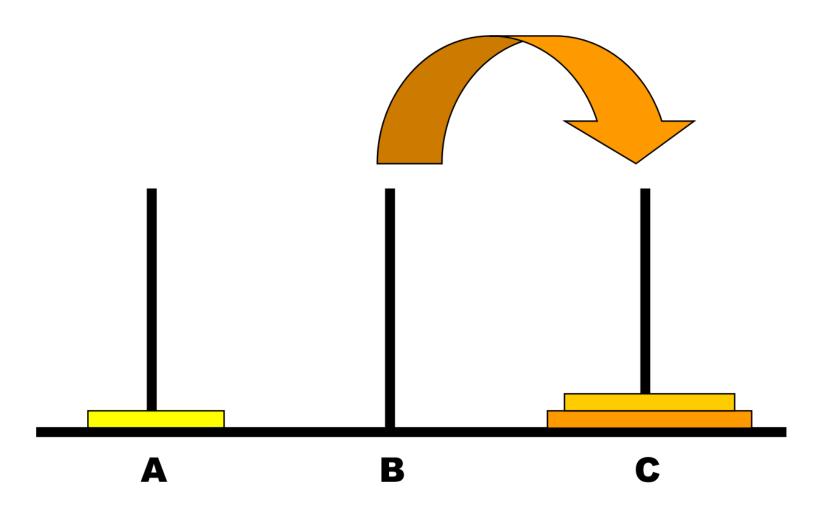


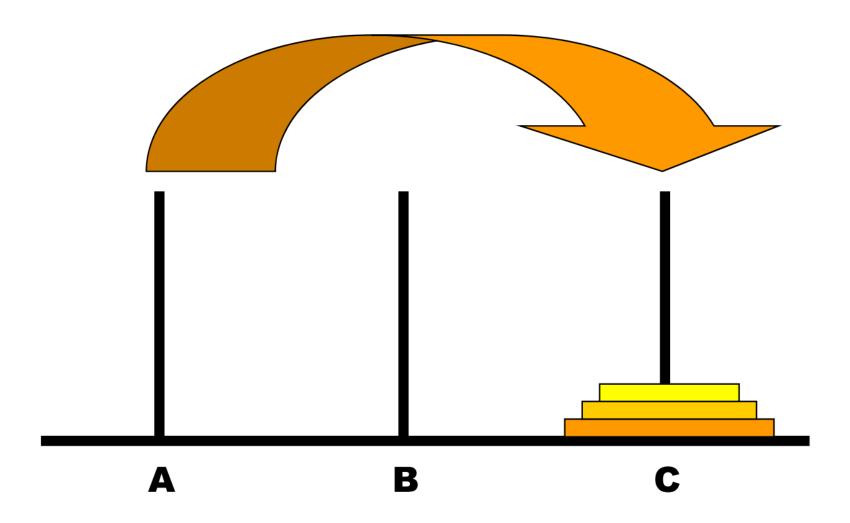






Move 5 B





The Algorithm

- void move(int count, int start, int finish, int temp)
- Precondition
 - There are at least count disks on the tower start
- Postcondition
 - The top count disks on start have been moved to tower finish;
 - Tower temp (used for temporary storage) has been returned to its starting position

The Algorithm

```
void move(int count, int start, int finish, int temp);
int main() {
   int disks = 3;
   printf("Disks: %d\n", disks);
   move(disks, 1, 3, 2);
   return 0;
void move(int count, int start, int finish, int temp) {
   if (count == 1) {
      printf("%s %d %s %d %s %d\n", "move disk", count,
               "from tower", start, "to tower", finish);
      return;
   move(count - 1, start, temp, finish);
   printf("%s %d %s %d %s %d\n", "move disk", count,
            "from tower", start, "to tower", finish);
   move(count - 1, temp, finish, start);
```

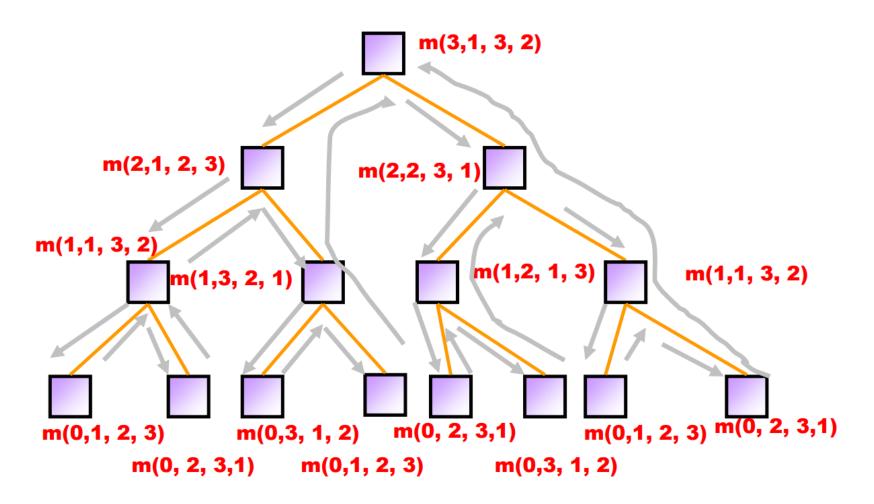
The Algorithm

Output:

Disks: 3

Move disk 1 from tower 1 to tower 3 Move disk 2 from tower 1 to tower 2 Move disk 1 from tower 3 to tower 2 Move disk 3 from tower 1 to tower 3 Move disk 1 from tower 2 to tower 1 Move disk 1 from tower 2 to tower 3 Move disk 1 from tower 1 to tower 3

The Recursion Tree



Recursion: Summary

- To obtain the answer to a larger problem, a general method is used that
 - reduces the large problem to one or more problems of a similar nature but a smaller size.
- Recursion continues until the size of the problem is reduced to the smallest, i.e. base case
 - where the solution is given directly without using further recursion.
- As such, recursive methods consist of two parts:
 - A smallest, base case that is processed without recursion (terminating condition).
 - A general method that reduces a particular case to one or more of the smaller cases (recursive condition).

Recursion: Summary

- Each function makes a call to itself with an argument which is <u>closer</u> to the terminating condition.
- Each call to a function has its own set of values/arguments for the formal arguments and local variables.
- When a recursive call is made, control is transferred from the calling point to the first statement of the recursive function. When a call at a certain level is finished, control returns to the calling point one level up.

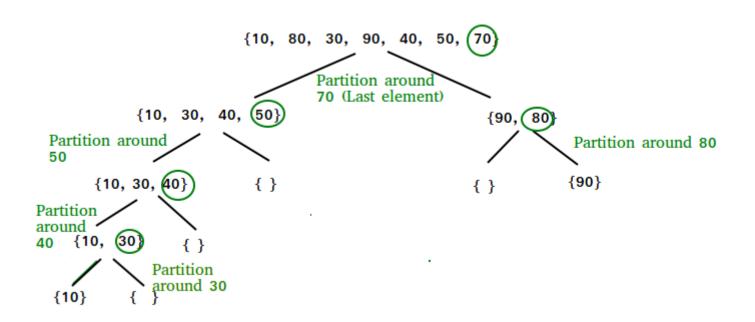
Sorting a Sequence of Data

Quick sort

- a divide-and-conquer algorithm
- pick an element as pivot; partition the data sequence into two sequences around the pivot
 - The left sequence contains element smaller than pivot
 - The right sequence contains element larger than pivot
- This partition is recursively applied to the sequence until no partition can be done

Sorting a Sequence of Data

An example of quick sort



The algorithm

Key part: partition of the data sequence

```
partition(arr[], low, high)
    // pivot (Element to be placed at right position)
    pivot = arr[high];
    i = (low - 1) // Index of smaller element
    for (j = low; j <= high - 1; j++)
         // If current element is smaller than the pivot
         if (arr[i] < pivot)</pre>
              i++; // increment index of smaller element
              if(i != j) swap arr[i] and arr[j]
         }
    swap arr[i + 1] and arr[high])
    return (i + 1)
```

The algorithm

Key part: partition of the data sequence

```
arr[] = \{10, 80, 30, 90, 40, 50, 70\}
Indexes: 0 1 2 3 4 5 6
low = 0, high = 6, pivot = arr[h] = 70
Initialize index of smaller element, i = -1
Traverse elements from j = low to high-1
j = 0 : Since arr[j] <= pivot, do i++</pre>
i = 0
arr[] = \{10, 80, 30, 90, 40, 50, 70\} // No change as i and j
                                     // are same
j = 1 : Since arr[j] > pivot, do nothing
// No change in i and arr[]
```

The algorithm

Key part: partition of the data sequence

```
j = 2 : Since arr[j] <= pivot, do i++ and swap(arr[i], arr[j])</pre>
i = 1
arr[] = \{10, 30, 80, 90, 40, 50, 70\} // We swap 80 and 30
j = 3 : Since arr[j] > pivot, do nothing
// No change in i and arr[]
j = 4 : Since arr[j] <= pivot, do i++ and swap(arr[i], arr[j])</pre>
i = 2
arr[] = \{10, 30, 40, 90, 80, 50, 70\} // 80  and 40 Swapped
j = 5 : Since arr[j] <= pivot, do i++ and swap arr[i] with arr[j]</pre>
i = 3
arr[] = \{10, 30, 40, 50, 80, 90, 70\} // 90  and 50 Swapped
We come out of loop because j is now equal to high-1.
Finally we place pivot at correct position by swapping
arr[i+1] and arr[high] (or pivot)
arr[] = \{10, 30, 40, 50, 70, 90, 80\} // 80  and 70 Swapped
```

Partition around

The algorithm

The recursive part

```
quickSort(arr[], low, high)
{
                                         40 {10, (30)
    if (low < high)
    {
        /* pi is partitioning index, arr[pi] is now
           at right place */
        pi = partition(arr, low, high);
        quickSort(arr, low, pi - 1); // Before pi
        quickSort(arr, pi + 1, high); // After pi
```

 $\{90, (80)\}$

{ }

Partition around 80

{90}

{10, 80, 30, 90, 40, 50, 70}

{ }

70 (Last element)

- Constructing a dynamic array
 - Declaring dynamic array structure

```
struct ARRAY
{
    int size;
    int count;
    float* data;
};
```

Note that the size of the array (for storage) could be larger than the real number (count) of the array

- Constructing a dynamic array
 - What functions we need (by design)?

```
bool create_array(int, ARRAY*);
void destroy_array(ARRAY*);

bool array_add_element(ARRAY*, float);
bool array_remove_element(ARRAY*, int);

int get_array_count(const ARRAY*);
int get_array_size(const ARRAY*);

float& get_array_element(int, const ARRAY*);
void set_array_element(int, float, ARRAY*);

void print_array(const ARRAY*);
```

- Constructing a dynamic array
 - Create the array

```
bool create array(int size, ARRAY* p array out)
     if (size <= 0)
          return false;
     if (p array out != NULL)
          p array out->data = (float*)malloc(sizeof(float) * size);
          if (p array out->data == NULL)
                    return false;
          p array out->size = size;
          p array out->count = 0;
          return true;
     else
          return true;
```

- Constructing a dynamic array
 - Destroy the array

Constructing a dynamic array

Add array element (dynamically)

```
bool array_add_element(ARRAY* p_array, float element value)
    if (p_array == NULL || p_array->data == NULL
                   || p array->count < 0 || p array->size <= 0)
         return false;
    if (p array->count + 1 <= p array->size)
         p array->count++;
         p_array->data[p_array->count - 1] = element_value;
         return true;
    else
    return true;
```

- Constructing a dynamic array
 - Add array element (dynamically)

```
bool array add element(ARRAY* p array, float element value)
     else
          float* data prev = p array->data;
          p array->data = (float*)realloc(data prev, p array->size
                                                    + ARRAY INCREMENT SIZE);
          if (p array->data == NULL)
                    return false;
          p array->size += ARRAY INCREMENT SIZE;
          p array->count++;
          p array->data[p array->count - 1] = element value;
          return true;
     return true;
```

Constructing a dynamic array

Remove array element

```
bool array remove element(ARRAY* p array, int i)
     if (i < 0 \mid | i >= p array -> count)
          return false;
     int rest count = p array->count - i - 1;
     int copy size = sizeof(float) * rest_count;
     float* p temp data = (float*)malloc(copy size);
     if (p temp_data == NULL)
          return false;
     memcpy(p temp data, &p array->data[i + 1], copy size);
     memcpy(&p array->data[i], p temp data, copy size);
     p array->data[p array->count - 1] = 0.0f;
     p array->count--;
```

- Constructing a dynamic array
 - Remove array element

Constructing a dynamic array

Remove array element

```
bool array remove element2(ARRAY* p_array, int i)
    if (i < 0 || i >= p_array->count)
         return false:
     for (int j = i; j < p_array->count - 1; j++)
          p_array->data[j] = p_array->data[j + 1];
     p array->data[p array->count - 1] = 0.0f;
     p_array->count--;
     if (p array->count size - 2 * ARRAY INCREMENT SIZE)
         float* p temp data = p array->data;
          p_array->data = (float*)realloc(
                    p temp data, p array->size - ARRAY INCREMENT SIZE);
          if (p array->data == NULL)
               return false;
     return true;
```

Constructing a dynamic array

Other array functions

```
int get array count(const ARRAY* p array)
      return p array->count;
int get array size(const ARRAY* p array)
      return p array->size;
float& get array element(int i, const ARRAY* p array)
      return p array->data[i];
void set array element(int i, float data value, ARRAY* p array)
      p array->data[i] = data value;
```

- Constructing a dynamic array
 - Print(display) the array on the screen

Implementing quicksort with dynamic array

Partition function

return i+1;

```
int quick sort array partition
           (ARRAY*p array , int low, int high)
     float pivot = get array element(high,p array);
     int i = low-1;
     for (int j = low; j < high; j++)</pre>
           if (get_array_element(j, p_array) < pivot)</pre>
                 i++:
                 if (i != j)
                       swap data(&get array element(i, p array),
                                   &get array element(j, p array));
           }
     swap data(&get array element(i+1, p array),
            &get array element(high, p array));
```

```
void swap_data(float* a, float* b)
{
    float t = *a;
    *a = *b;
    *b = t;
}
```

- Implementing quicksort with dynamic array
 - Recursive function call for partitioning

```
bool quick sort array(ARRAY* p array in out, int low, int high)
    if (low < high)</pre>
        int partition index =
                 quick_sort_array_partition(p_array_in_out, low, high);
        quick_sort_array(p_array_in_out, low, partition_index - 1);
        quick sort array(p array in out, partition index + 1, high);
        return true;
    else
        return false;
}
```

How to use the quicksort for dynamic array?

```
void test array sort()
    ARRAY a;
     if (!create_array(20, &a))
         printf("unable to create the array!\n");
         return;
     printf("Please input the array elements:\n");
    while (1)
         char str[32];
         scanf("%s", str);
         if (strcmp(str, "quit") == 0)
         break;
         array add element(&a, (float)atof(str));
```

How to use the quicksort for dynamic array?

```
void test_array_sort()
{
    ...
    printf("Your input ARRAY is:\n");
    print_array(&a);

    printf("\nSorting the whole array...\n");
    quick_sort_array(&a, 0, get_array_count(&a) - 1);
    printf("Your sorted array is:\n");
    print_array(&a);

    destroy_array(&a);//don't forget to destroy the array
}
```

Designing Recursive Algorithms

- Find the key step
 - How can this problem be divided into parts?
 - How will the key step in the middle be done?
- Find a stopping rule
 - Small, special case that is trivial or easy to handle without recursion
- Outline your algorithm
 - Combine the stopping rule and the key step, using an if statement to select between them
- Check termination
 - Verify that the recursion always terminates

Recursion or Iteration?

- The main <u>advantage</u> of using recursive functions:
 When the problem is recursive in nature, a recursive function results in <u>short</u>, clear code.
- The main <u>disadvantage</u> of using recursive functions: recursion is more <u>expensive</u> than iteration.
- Any problem that can be solved recursively can also be solved iteratively (by using loops).
- A recursive approach is normally chosen in preference to an iterative approach when the recursive approach more naturally mirrors the problem and results in a program that is easier to understand and debug.

Summary

- Study several simple examples to see whether recursion should be used and how it will work.
- Attempt to formulate a method that will work more generally
 - How can this problem be divided into parts?
- Ask whether the remainder of the problem can be done in the same or simple way.
- Find a stopping rule.
- Algorithm must terminate.