Unit #0: Introduction

CPSC 221: Algorithms and Data Structures

Will Evans and Jan Manuch[1]

2016W1

[1]Thanks to Steve Wolfman for the content of most of these slides with additional material from Alan Hu, Ed Knorr, and Kim Voll.

Unit Outline

- Course logistics
- Course overview
- Fibonacci Fun
- Arrays
- Queues
- Stacks
- Deques

Course Information

Instructors

Will Evans	`will@cs.ubc.ca`	ICCS X841
Jan Manuch	`jmanuch@cs.ubc.ca`	ICCS 247

Tue 10 – 11
Wed 2 – 3p

TAs

Alexander Lim	Chris Hunter	David Zheng	Harman Gakhal
Henry Chee	Jason Zeng	Jin Ziyang	Jordan Coblin
Michael Zhang	Mike Spearman	Muzhi Ou	Nancy Chen
Oliver Zhan	Patience Shyu	Sharon Yang	Xing Zeng

Zheng Dong

Office hours

See `www.ugrad.cs.ubc.ca/~cs221`

Texts

Epp Discrete Math, Koffman Data Structs C++

Course Work

No late work; may be flexible with advance notice

10%	Labs
15%	Programming projects (~ 3)
15%	Written homework (~ 3)
20%	Midterm exam
40%	Final exam

Must pass the final and combo of labs/assignments to pass the course.

Collaboration

You may work in groups of two people on:

- Labs → LIM Wed 5p-7p
- Programming projects (LIN Mon 11-1p)
- Written homework

You may also collaborate with others as long as you follow the rules (see the website) and **acknowledge** their help on your assignment.

Don't violate the collaboration policy.

Course Mechanics

- Web page: `www.ugrad.cs.ubc.ca/~cs221`
- Piazza:
  `https://piazza.com/ubc.ca/winterterm12016/cpsc221/home`
- UBC Connect site: `www.connect.ubc.ca`
- Labs are in ICCS X350
  - Use the `Xshell` program on the lab machines to `ssh` into a undergrad Unix machine (e.g. `lulu.ugrad.cs.ubc.ca`)
- Programming projects will be graded on `UNIX/g++`

What is a Data Structure?

Examples:

- Array
- Tree
- Stack
- Hash table / Map
- Queue
- Heaps
- Graph
- List

A definition:

A method of storing data provides, through a set of operations, a way to manipulate and access the data.

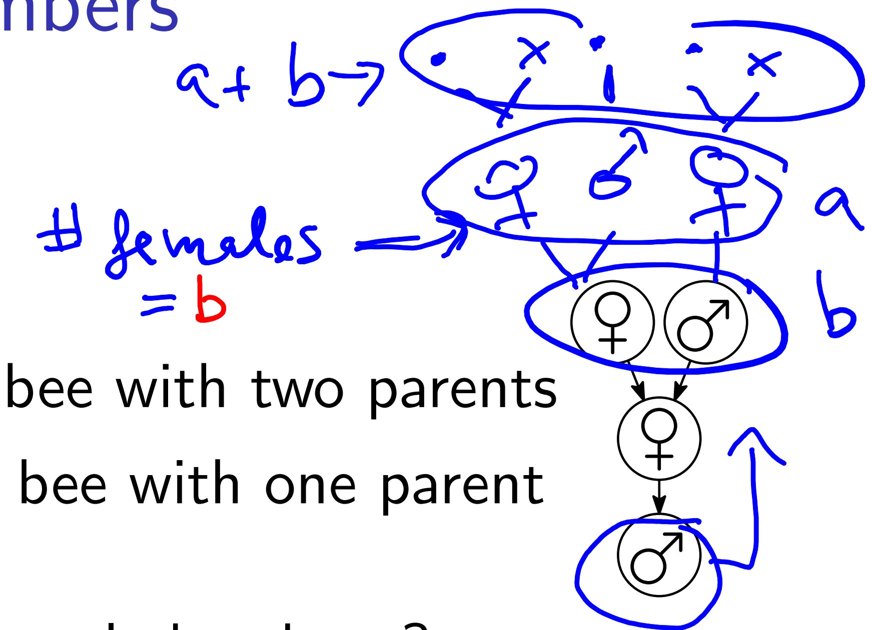
Observation

- All programs manipulate data
  - programs process, store, display, gather data
  - data can be information, numbers, images, sound
- ~~The programmer~~ You must decide how to store and manipulate data
- This choice influences the program in many ways
  - execution speed
  - memory requirements
  - maintenance (debugging, extending, etc.)

Goals of the Course

- Become familiar with some of the fundamental data structures and algorithms in computer science
  - Learn when to use them
- Improve your ability to solve problems abstractly
  - Data structures and algorithms are the building blocks
- Improve your ability to analyze algorithms
  - Prove correctness
  - Gauge, compare, and improve time and space complexity
- Become modestly skilled with C++ and UNIX, but this is largely on your own!

Analysis Example: Fibonacci numbers



Bee ancestory:

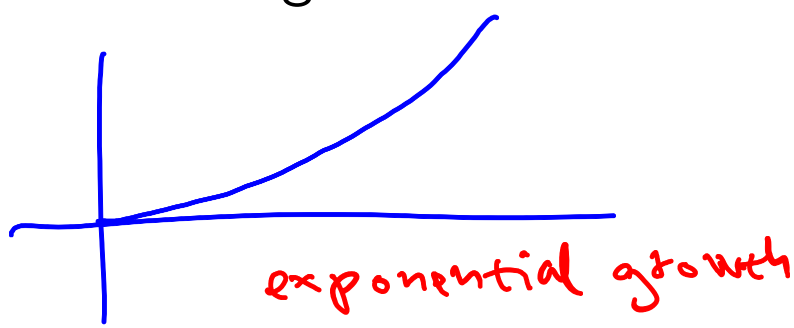
1. Fertilized egg becomes a female bee with two parents
2. Unfertilized egg becomes a male bee with one parent

How many great-grandparents does a male bee have? great-great-grandparents? ...

$$Fib_n = Fib_{n-1} + Fib_{n-2}$$

Fibonacci numbers: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, ...

First two numbers are 1; each succeeding number is the sum of the previous two numbers.



Recursive Fibonacci

Problem: Calculate the nth Fibonacci number.

Recursive definition:

$$fib_n = \begin{cases} 1 & \text{if } n = 1 \\ 1 & \text{if } n = 2 \\ fib_{n-1} + fib_{n-2} & \text{if } n \geq 3 \end{cases}$$

C++ code:

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

C_n = # of times fib() is called on input n.

$C_n = C_{n-1} + C_{n-2} + 1$

$C_n \geq fib(n)$ ⇒ exponential running time

Too slow!

Iterative Fibonacci

Idea: Use an array

```
int fib(int n) {
  int F[n+1];

  F[0]=0; F[1]=1; F[2]=1;
  for( int i=3; i<=n; ++i ) {
    F[i] = F[i-1] + F[i-2];
  }
  return F[n];
}
```

(We don't really need the array.)

Can we do better?

Fibonacci by formula

Idea: Use a formula (a *closed form solution* to the recursive definition.)

$$fib_n = \frac{\varphi^n - (-\varphi)^{-n}}{\sqrt{5}}$$

where $\varphi = (1 + \sqrt{5})/2 \approx 1.61803$.

```
#include <cmath>
int fib(int n) {
   double phi = (1 + sqrt(5))/2;
   return (pow(phi, n) - pow(-phi,-n))/sqrt(5);
}
```

running time? $O(1)$? NO $O(n)$? PROBABLY NO

PROBABLY YES $O(\log n)$

$\varphi^{16} \rightarrow$ $\varphi^2 = \varphi \times \varphi$, $\varphi^4 = \varphi^2 \times \varphi^2$, $\varphi^8 = \varphi^4 \times \varphi^4$, $\varphi^{16} = \varphi^8 \times \varphi^8$

$\varphi^{20} = \varphi^4 \times \varphi^{16}$

Sadly, it's impossible to represent $\sqrt{5}$ exactly on a digital computer.

if $n = 2^k$, then to compute φ^n we need $k = \log n$ multiplications

Can we do better?

Fibonacci with Matrix Multiplication

$$\begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}1+1\\1\end{bmatrix} = \begin{bmatrix}fib_3\\fib_2\end{bmatrix}$$

$$\begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}\begin{bmatrix}2\\1\end{bmatrix} = \begin{bmatrix}fib_4\\fib_3\end{bmatrix}$$

$$\begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}^{n-2}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}fib_n\\fib_{n-1}\end{bmatrix}$$

How do we calculate $\begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}^{n-2}$? $= \begin{bmatrix}a & b\\c & d\end{bmatrix}$ $a+b=fib_n$

Repeated Squaring

$(11111)_2 = 16+8+4+2+1 = 31$

$100 = (1100100)_2$

$A^{100} = A^{2^6} * A^{2^5} * A^{2^2}$

digits in the binary representation of n is $\sim \log_2 n$

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

$$A \times A = A^2$$
$$A^2 \times A^2 = A^4$$
$$A^4 \times A^4 = A^8$$
$$A^8 \times A^8 = A^{16}$$
$$A^{16} \times A^{16} = A^{32}$$
$$A^{32} \times A^{32} = A^{64}$$
$$\vdots$$

Comparison (# of arithm. operations):

F_n	Iter.	matrix.
F_{20}	18 ops (better)	$12 \times (4+1)+1 = 61$ ops
F_{102}	100 ops	$12 \times (6+2)+1 = 97$ ops (better)
⋮		⋮
F_n	$\sim n$ ops	$\lesssim 24 \log_2 n$ ops (much better for large n)

Example: $A^{100} = A^{64} \times A^{32} \times A^4$. 8 instead of 99 matrix multiplications.
Generally, about $\log_2 n$ multiplications.

$< 2 \log_2 n$ matrix multiplications

Is this better than iterative Fibonacci? $\sim 24 \log_2 n$ int. operations

Abstract Data Type

Abstract Data Type

Mathematical description of an object and the set of operations on the object

what

interface

Example: **Dictionary ADT**

- Stores pairs of strings: (word, definition)
- Operations:
  - Insert(word,definition)
  - Delete(word)
  - Find(word)

Another Example: Array ADT

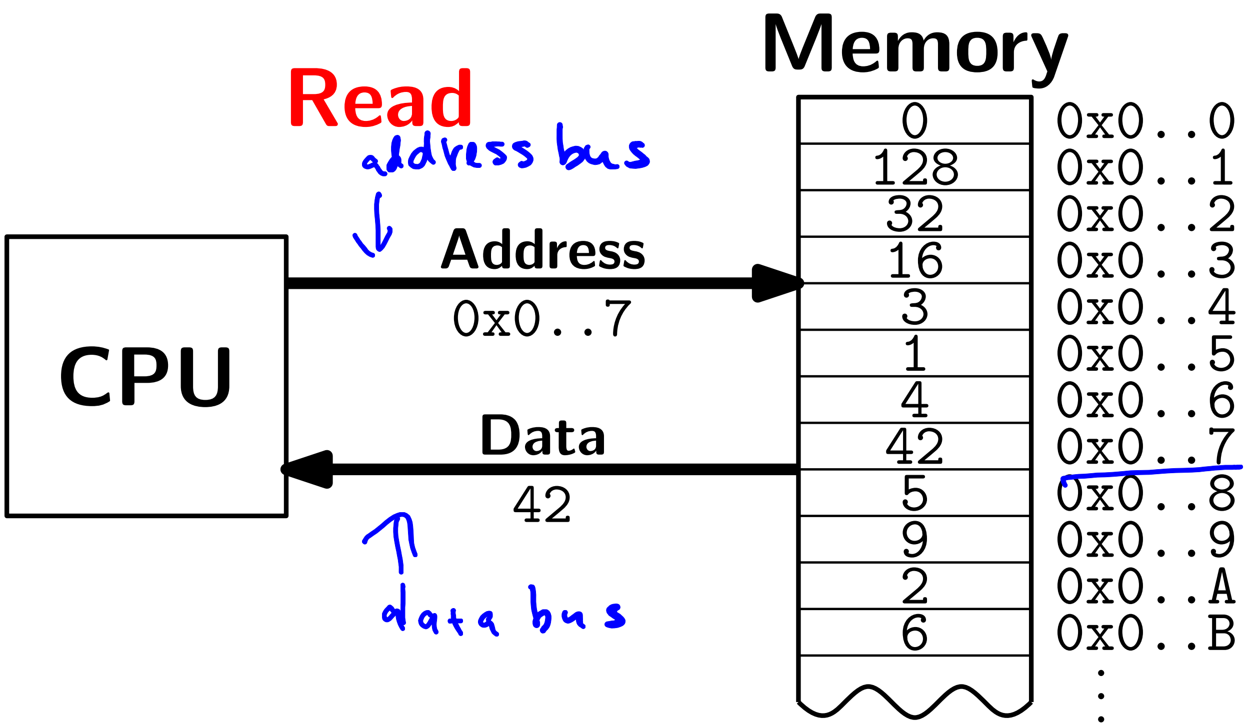
- Store things like integers, (pointers to) strings, etc.
- Operations:
  - Initialize an empty array that can hold n things.

    ```
    thing A[n];
    ```
  - Access (read or write) the ith thing in the array ($0 \leq i \leq n-1$).

    ```
    thing1 = A[i]; Read
    A[i] = thing2; Write
    ```

Why Arrays?

- Computer memory is an array.
  Read: CPU provides address i,
  memory unit returns the data stored at i.



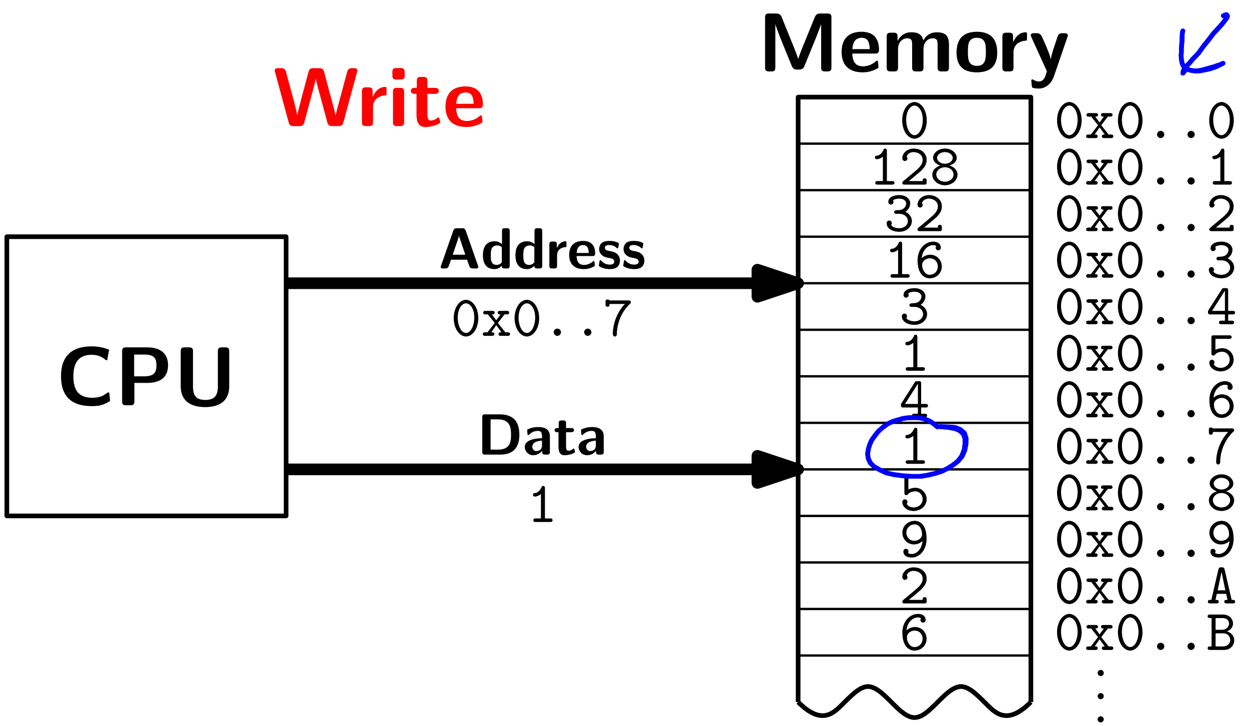
Why Arrays?

$(0100\,1011)_2$ → 4 B

- Computer memory is an array.

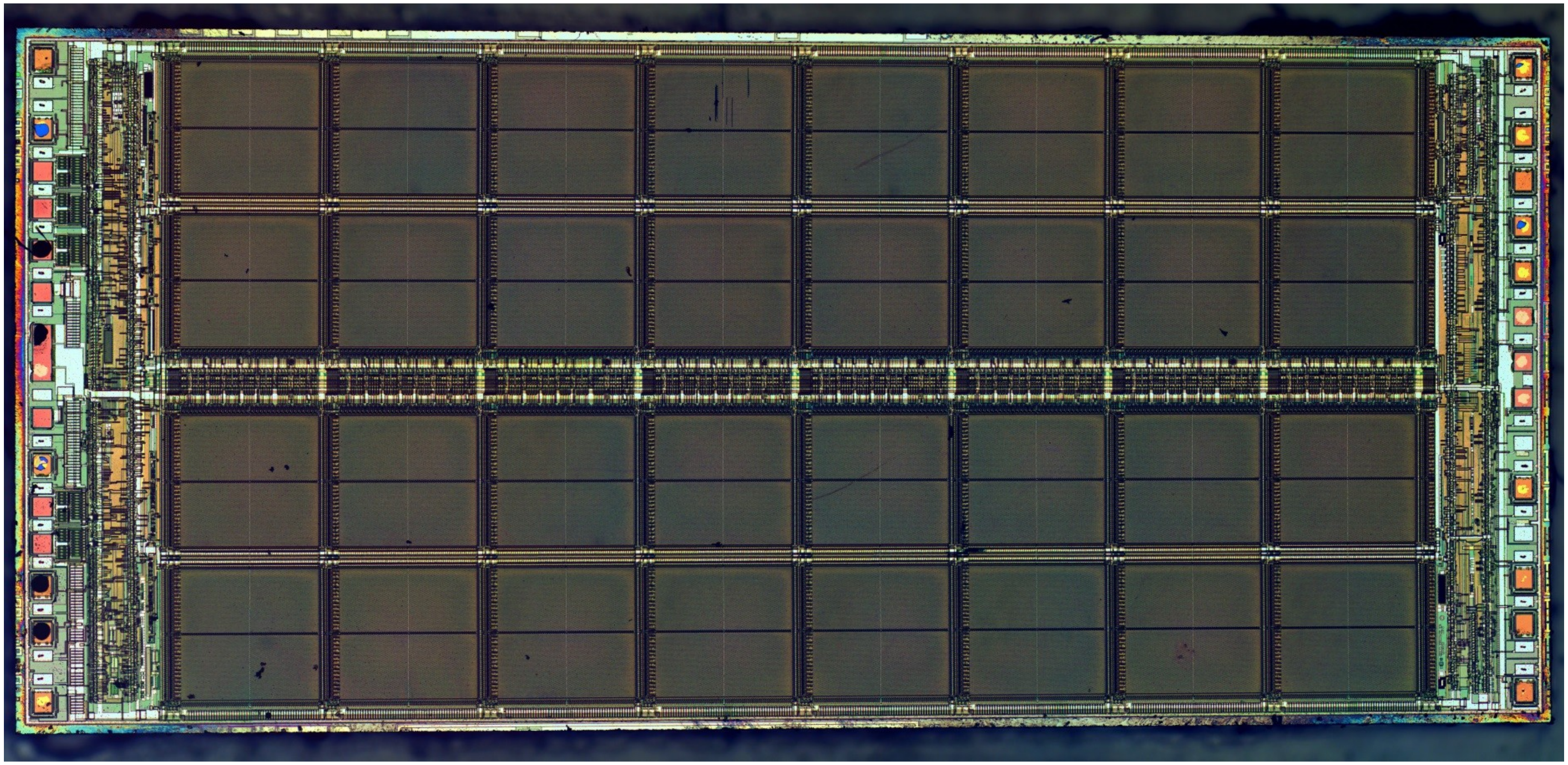
  Write: CPU provides address *i* and data *d*,
  memory unit stores data *d* at *i*.

hexadecimal #s



Why Arrays?

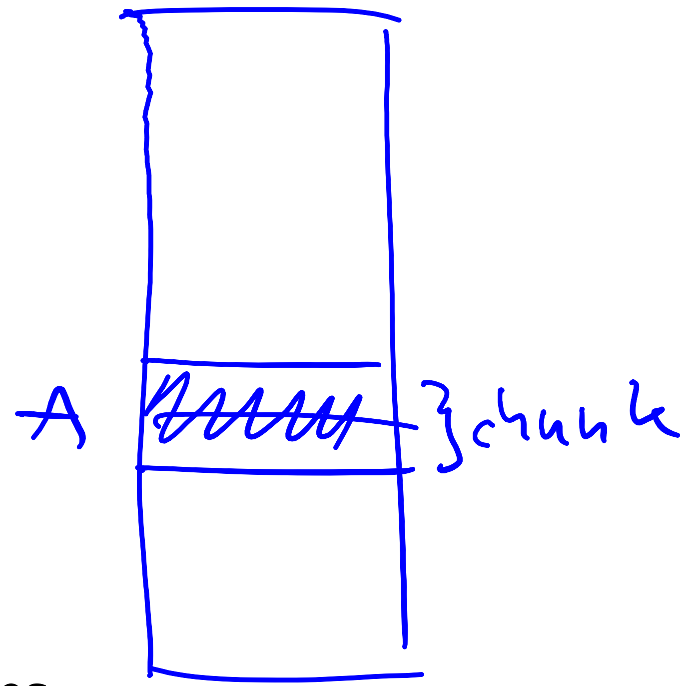
- Computer memory is an array. Every bit has a physical location.



http://zeptobars.ru/en/read/how-to-open-microchip-asic-what-inside licensed under Creative Commons Attribution 3.0 Unported.

Why Arrays?

- Computer memory is an array.
- Simple and fast.
- Used in almost every program.
- Used to implement other data structures.



Array limitations

- Need to know size when array is created.

  Fix: Resizeable arrays.
  If the array fills up, allocate a new, bigger array and copy the old contents to the new array.

  vector in C++

  2.5GB array A .. 1GB

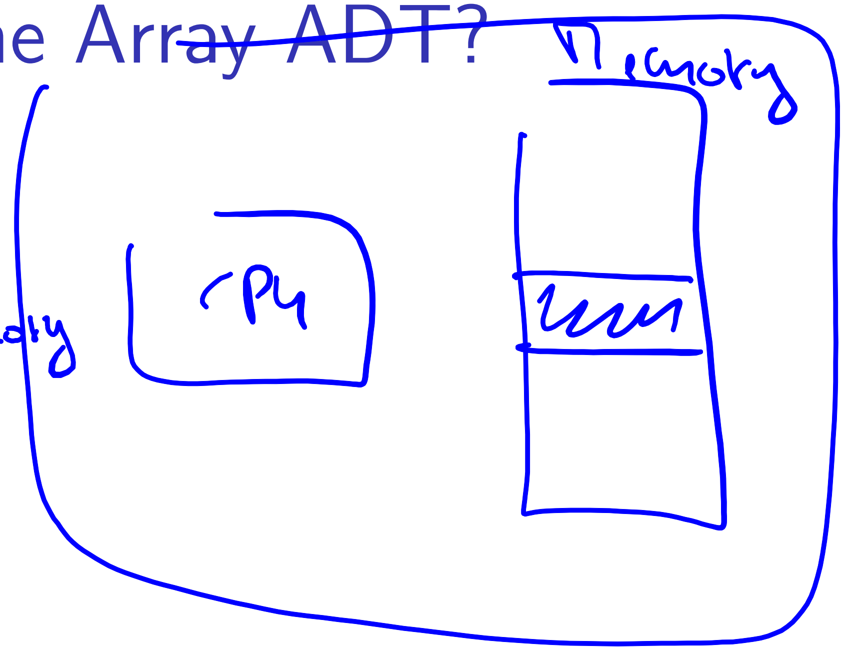
- Indices are integers 0,1,2,...

  Fix: Hashing.
  (more later)

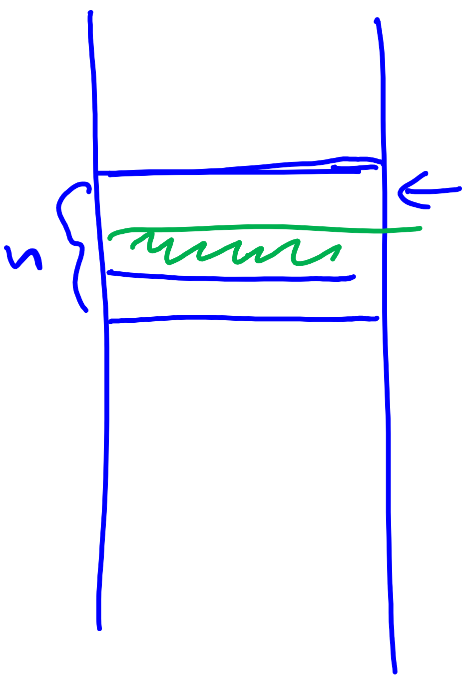
  associative arrays (javascript, Perl, Python, Awk)

  Example:
  A["cpsc211"] = "course";

How would you implement the Array ADT?



- memory management
  - list of free chunks of memory (freelist)



pointer

create A of size n objects ⇒ we get address of the beginning of this chunk

A .. address

$A = 0x0007$

$A + 3 = 0x000A$

A[3]

$A + 3 * size(object)$

address of the "3-rd" object

in C++: $*(A+3) \approx A[3]$

dereferencing — pointer arithmetic

How would you implement the Array ADT?

Arrays in C++

Create `int A[100];`

Access
```
for( int i=0; i<100; i++ )
    A[i] = (i+1) * A[i-1];
```

g++ … not initialized

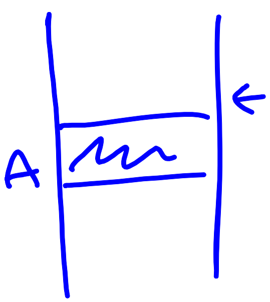
MSVS … might initialize to all zeros

$*(A+i)$

$*(A+i-1)$

A[-1] → $*(A-1)$

probably crashes



How would you implement the Array ADT?

Arrays in C++

Create `int A[100];`

Access
```
for( int i=0; i<100; i++ )
    A[i] = (i+1) * A[i-1];
```

Warning No bounds checking!

Data Structures as Algorithms

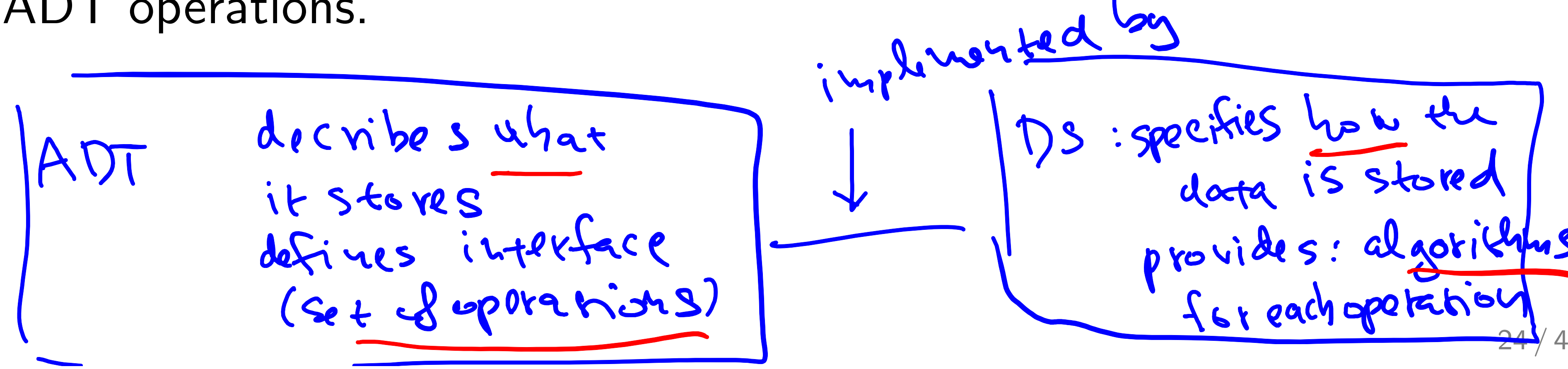
Algorithm

a high level, language independent description of a step-by-step process for solving a problem

Data Structure

a way of storing and organizing data so that it can be manipulated as described by an ADT

A data structure is defined by the algorithms that implement the ADT operations.



Why so many data structures?

Ideal data structure

fast, elegant, memory efficient

Trade-offs

- time vs. space
- performance vs. elegance
- generality vs. simplicity
- one operation's performance vs. another's

Data structures for Dictionary ADT

- List
- Skip list
- Binary search tree
- AVL tree
- Splay tree
- B-tree
- Red-Black tree
- Hash table

...

why so many?

Code Implementation

Theory

- abstract base class (interface) describes ADT
- descendents implement data structures for the ADT
- data structures can change without affecting client code

Practice

- different implementations sometimes suggest different interfaces (generality vs. simplicity)
- performance of a data structure may influence the form of the client code (time vs. space, one operation vs. another)

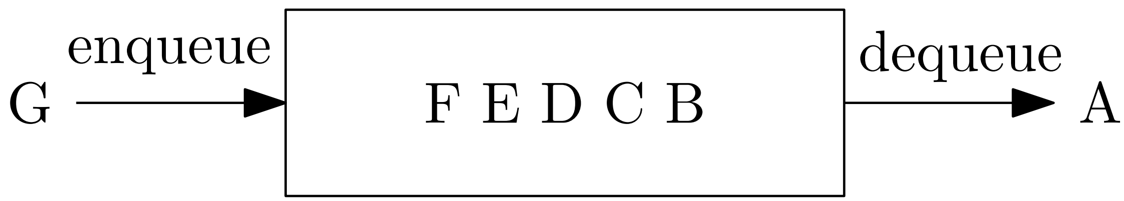
ADT Presentation Algorithm

1. Present an ADT
2. Motivate with some applications
3. Repeat
   3.1 develop a data structure for the ADT
   3.2 analyze its properties
      - efficiency
      - correctness
      - limitations
      - ease of programming
4. Contrast data structure's strengths and weaknesses
   - understand when to use each one

Queue ADT

Queue operations

- create
- destroy
- enqueue
- dequeue
- is_empty



Queue property

If x is enqueued before y is enqueued, then x will be dequeued before y is dequeued.

FIFO: First In First Out

Applications of the Q

- Hold jobs for a printer
- Store packets on network routers
- Hold memory “freelists”
- Make waitlists fair
- Breadth first search

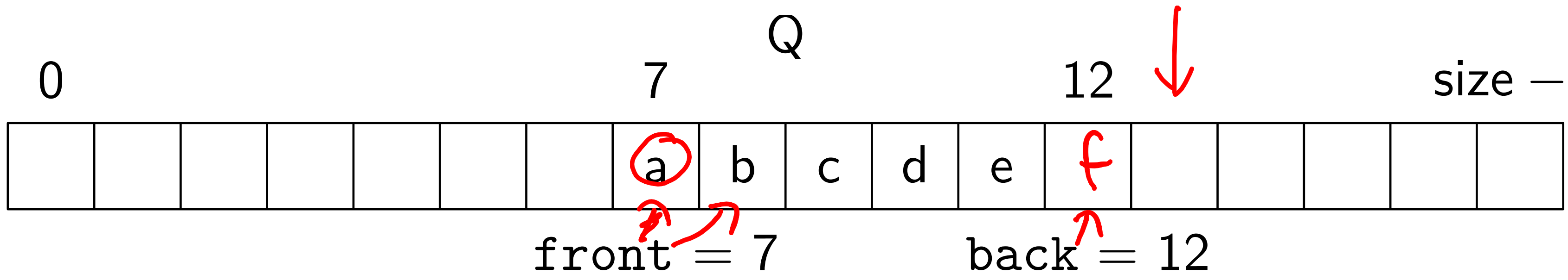
Abstract Q Example

enqueue R
enqueue O
dequeue
enqueue T
enqueue A
enqueue T
dequeue
dequeue
enqueue E
dequeue

In order, what letters are dequeued?

a. OATE
b. ROTA
c. OTAE
d. None of these, but it **can** be determined from just the ADT.
e. None of these, and it **cannot** be determined from just the ADT.

Circular Array Q Data Structure




```
void enqueue(Object x) {
  Q[back] = x;
  back = (back + 1) % size;
}

bool is_empty() {
  return (front == back);
}
```

```
Object dequeue() {
  x = Q[front];
  front = (front + 1) % size;
  return x;
}

bool is_full() {
  return (front ==
    (back + 1) % size);
}
```

Circular Array Q Example

red .. elements of the Q
green .. leftover elements

	front	back
(initial)	0	0
enqueue R	0	1
enqueue O	0	2
dequeue	1	2
enqueue T	1	3
enqueue A	1	0
enqueue T	1	1
dequeue	2	1
dequeue	3	1
enqueue E	3	2
dequeue	0	2

calls explode() *

What are the final contents of the array?

a. RTE

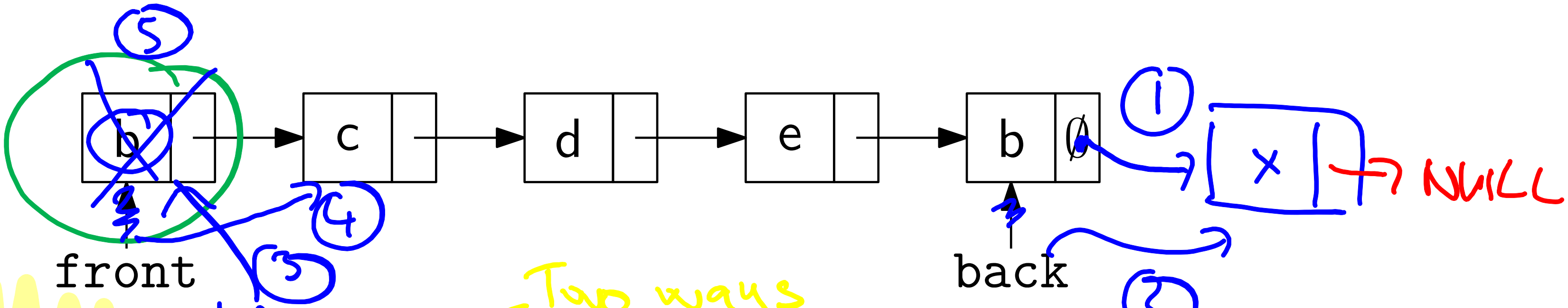
b. RTET

c. TETA * correct

d. TE

e. None

Linked List Q Data Structure




```
void enqueue(Object x) {
  if (is_empty())
    front = back = new Node(x);
  else {
    back->next = new Node(x);
    back = back->next;
  }
}

bool is_empty() {
  return (front == NULL);
}
```

```
Object dequeue()
  assert(!is_empty());
  Object ret = front->data;
  Node *temp = front;
  front = front->next;
  delete temp;
  return ret;
```

DIY memory management

Circular Array vs. Linked List

Ease of implementation — same

Generality — CA .. limited # of elements
.. alt. CA could use dynamic arrays

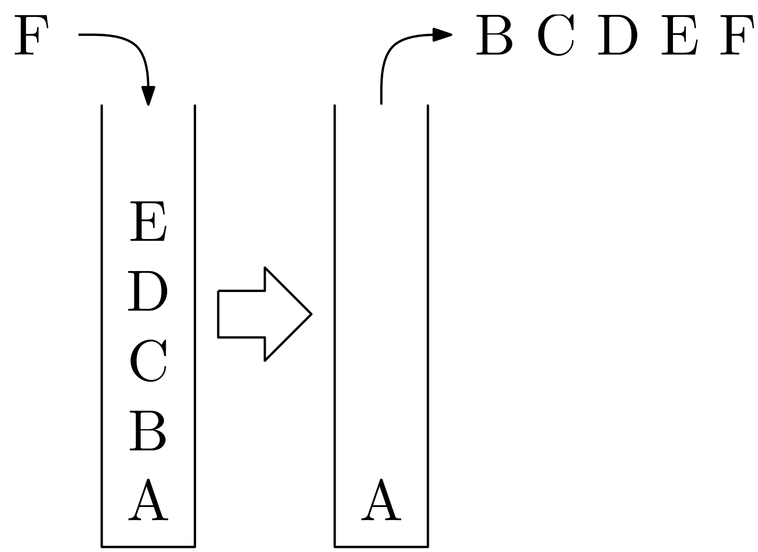
Speed — same, but subtle differences:
- new/delete takes time
- cache performance worse for LL
- time for resizing CA

Memory use — LL uses more memory (to store pointers)

Stack ADT

Stack operations

- create
- destroy
- push
- pop
- top
- is_empty



Stack property

if x is pushed before y is pushed, then x will be popped after y is popped.

LIFO: Last In First Out

Stacks in Practice

- Function call stack
- Removing recursion
- Balancing symbols (parentheses)
- evaluating arithmetic expressions
- Evaluating Reverse Polish Notation
  = postfix
- Depth first search

in arithmetic expression:

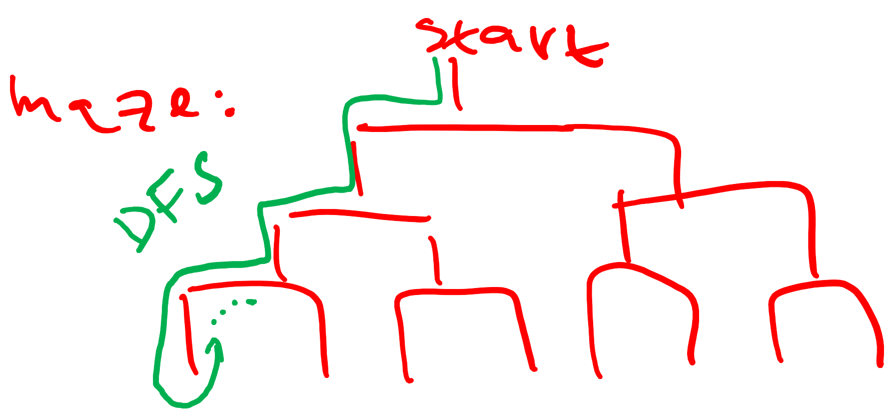
$a + b * (c + d)$

$a\ b\ c\ d\ +\ *\ +$

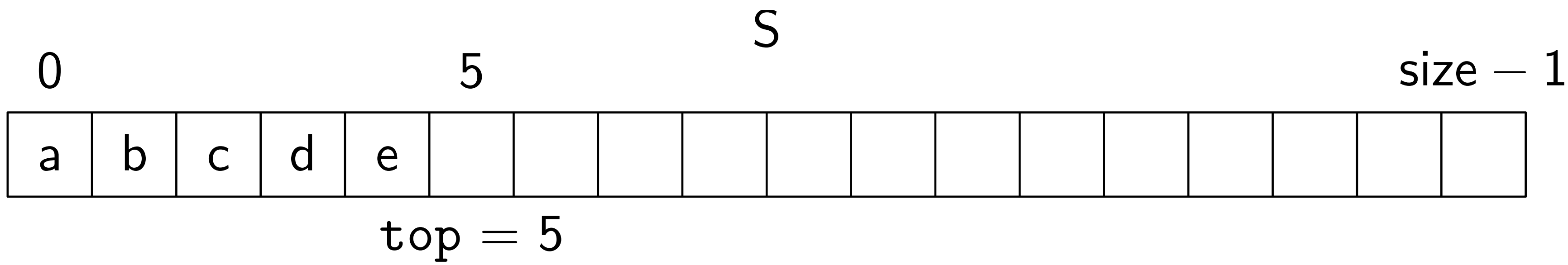
used by:

PS (PostScript) .. Laser printers

Java VM



Array Stack Data Structure

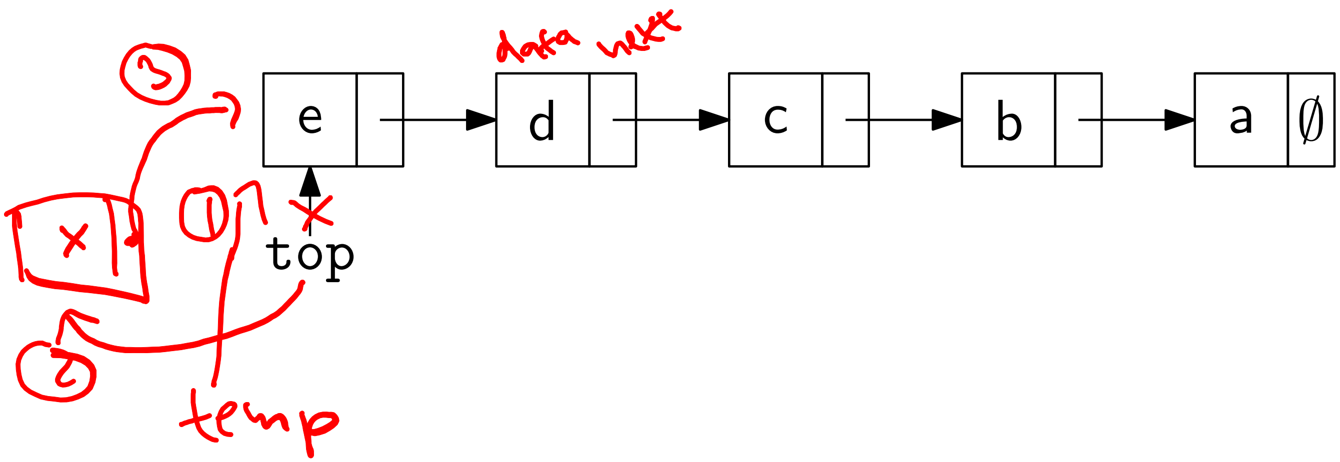



```
void push(Object x) {
  assert(!is_full());
  S[top] = x;
  top++;
}

Object top() {
  assert(!is_empty());
  return S[top-1];
}
```

```
Object pop() {
  assert(!is_empty());
  top--;
  return S[top];
}
bool is_empty() {
  return( top == 0 );
}
bool is_full() {
  return( top == size);
}
```

Linked List Stack Data Structure




```
void push(Object x) {
  Node *temp = top;
  top = new Node(x);
  top->next = temp;
}
```

(*top).next

```
Object top() {
  assert(!is_empty());
  return top->data;
}
```

```
Object pop() {
  assert(!is_empty());
  Object ret = top->data;
  Node *temp = top;
  top = top->next;
  delete temp;
  return ret;
}
```

```
bool is_empty() {
  return( top == NULL );
}
```

Deque ADT

[:deck:]

Deque (Double-ended queue) operations

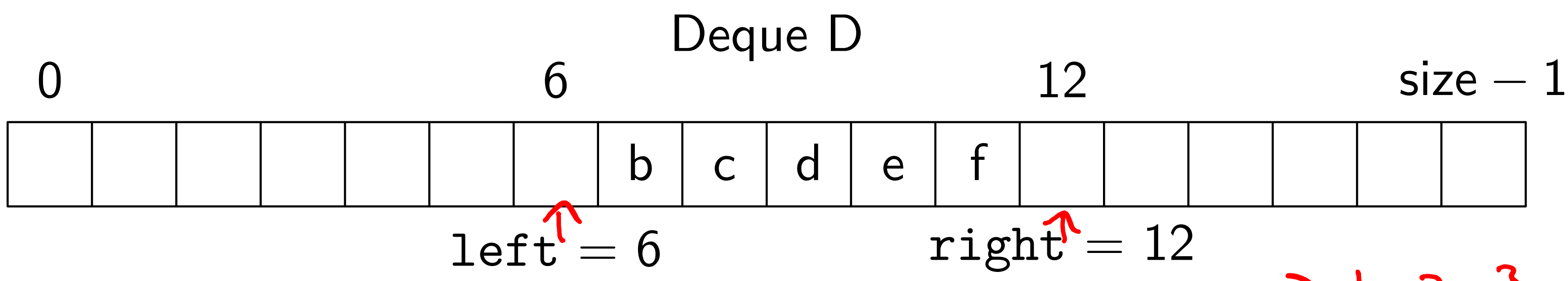
- create/destroy
- pushL/pushR
- popL/popR
- is_empty

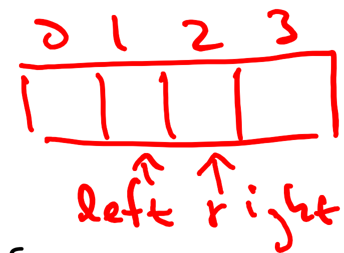
Deque property

Deque maintains a list of items.
push/pop adds to/removes from front(L)/back(R) of list.

undo "stack"

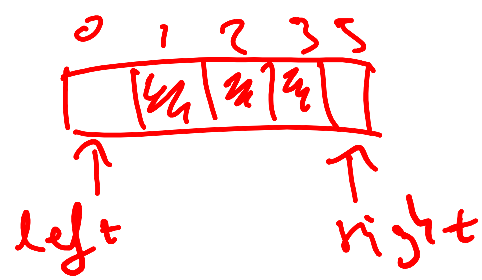
Circular Array Deque Data Structure





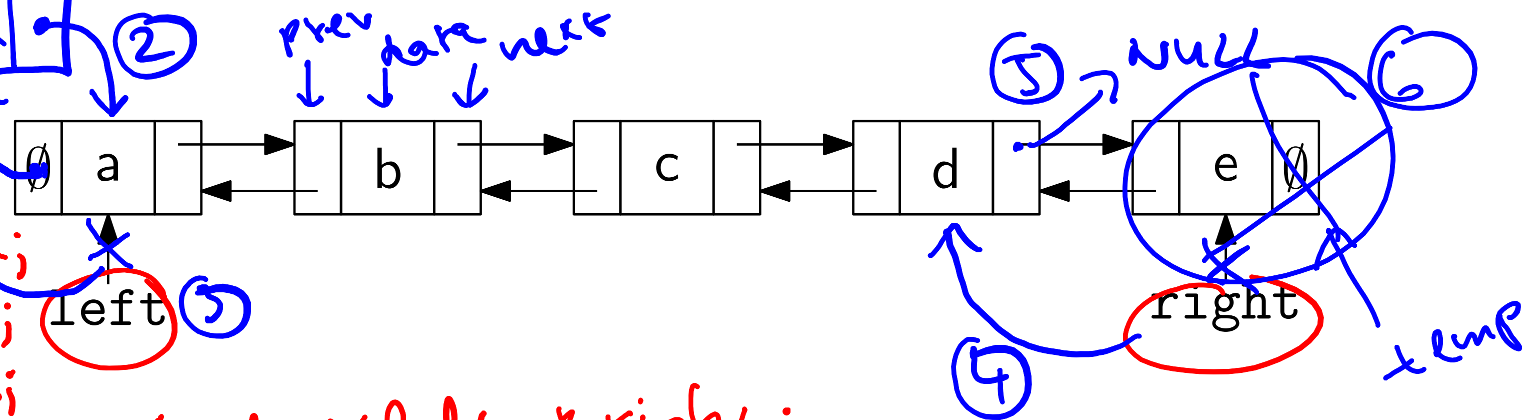

```
void pushL(Object x) {
  assert(!is_full());
  D[left] = x;
  left = (left - 1) % size;
}
Object popR() {
  assert(!is_empty());
  right = (right - 1) % size;
  return D[right];
}
...
```

```
bool is_empty() {
  return( left ==
      (right - 1) % size);
}
bool is_full() {
  return( left ==
      (right + 1) % size);
}
```



Linked List Deque Data Structure

struct Node {
object data;
Node * prev;
Node * next;
}



Node *left, *right;

```
void pushL(Object x) {
  if( is_empty() )
    left = right = new Node(x);
  else {
    left->prev = new Node(x);
    left->prev->next = left;
    left = left->prev;
  }
}

bool is_empty() {return left==NULL;}
```

```
Object popR() {
  assert(!is_empty());
  Object ret = right->data;
  Node *temp = right;
  right = right->prev;
  if( right ) right->next = NULL;
  else left = NULL;
  delete temp;
  return ret;
}
```

test if right ≠ NULL

Data structures you should already know (a bit)

- Arrays
- Linked lists
- Trees
- Queues
- Stacks