ITI 1121. Introduction to Computing II *

Marcel Turcotte
School of Electrical Engineering and Computer Science

Version of March 26, 2013

Abstract

- Iterator[1] (part I)

---

*These lecture notes are meant to be looked at on a computer screen. Do not print them unless it is necessary.

[1]A great opportunity to discuss time complexity (asymptotic analysis, big-O), encapsulation, object-oriented programming, interface, and more.

Motivation

Given a (singly) linked list implementation, called **LinkedList**, of the interface **List**, defined as follows.

```
public interface List<E> {
    public abstract boolean add( E obj );
    public abstract E get( int index );
    public abstract boolean remove( E obj );
    public abstract int size();
}
```

Consider writing statements to traverse the list. Compare writing the code inside and outside of the implementation.

⇒ The difficulties would be the same for a doubly linked list.

Traversing a list inside/outside of the class

```
List<String> colors;
colors = new LinkedList<String>();

colors.add( "bleu" );
colors.add( "blanc" );
colors.add( "rouge" );
colors.add( "jaune" );
colors.add( "vert" );
colors.add( "orange" );
```

A — traversing a linked list inside of the class

Being inside of the implementation, the nodes can be accessed directly.

```
Node<E> p = head;
while ( p != null ) {
    System.out.println( p.value );
    p = p.next;
}
```

B — traversing a linked list outside of the class

Traversing a list without having access to the underlying implementation requires using **E get(int pos)**.

```
for ( int i=0; i < colors.size(); i++ ) {
    System.out.println( colors.get( i ) );
}
```

$\Rightarrow$ How efficient, w.r.t. time, this code is compared to writing it within the class **LinkedList**? Much faster, faster, same, slower, much slower?

Timing results

Here are the timing results for $20,000$, $40,000$ and $80,000$ nodes lists, averaged over 5 runs (times in milliseconds).

# nodes	A	B
20,000	2	20,644
40,000	5	94,751
80,000	12	407,059

$\Rightarrow$ For $80,000$ nodes it takes 6.5 minutes traversing the list using **get(pos)** vs 12 ms for the other approach. What can explain this huge difference?

```
for ( int i=0; i< names.size(); i++ ) {
    System.out.println( names.get( i ) );
}
```

Call	Number of nodes visited
get(0)	1
get(1)	2
get(2)	3
get(3)	4
...	...
get(n-1)	n

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2} = \frac{n^2 + n}{2} \approx n^2$$

Implementation A visits n nodes, implemenation B visits n^2 nodes!

# nodes	A	B
20,000	2	20,644
40,000	5	94,751
80,000	12	407,059

Executing the following statements (implementation B).

```
for ( int i=0; i< names.size(); i++ ) {
    System.out.println( names.get( i ) );
}
```

It is as if implementation A had been written like this.

```
for ( int i=0; i< size(); i++ ) {
    Node p = head;
    for ( int j=0; j<i; j++ ) {
        p = p.next;
    }
    System.out.println( p.value );
}
```

Instead of this.

```
Node p = head;
while ( p != null ) {
    System.out.println( p.value );
    p = p.next;
}
```

What mechanism could the implementation provide that would allow to efficiently traverse the list?

We are outside of the class definition, therefore, we do not have access to the implementation (**p.next** and such), yet the need to traverse a data structure is frequent.

The concept that we are introducing today will apply to a specific context, TRAVERSING A LIST.

WE WILL NOT PROVIDE A GENERAL MECHANISM TO MAKE **get(i)** EFFICIENT ALWAYS.

Iterator: introduction

- A uniform mechanism to traverse a variety of data structures, such as lists, but also trees and other data structures to be seen in CSI 2114;
- Gives access to all the elements, one a time.
- It's a mechanism that is used by Java's own collections.

Concept

```
‘‘create an iterator’’ (* positioned before the start of the list *)

while ( ‘‘has more elements?’’ ) {
   ‘‘move the iterator forward and return the value’’
   ‘‘process the value’’
}
```

Concept

Iterator interface

The implementation of the **Iterator** is separated from its usage by creating an interface for it.

```
public interface Iterator<E> {
    public abstract boolean hasNext();
    public abstract E next();
}
```

```java
public interface Iterator<E> {
    /**
     * Returns the next element in the list.  This method may be called
     * repeatedly to iterate through the list.
     *
     * @return the next element in the list.
     * @exception NoSuchElementException if the iteration has no next element.
     */

    public abstract E next();

    /**
     * Returns true if this list iterator has more elements when
     * traversing the list in the forward direction. (In other words, returns
     * true if next would return an element rather than
     * throwing an exception.)
     *
     * @return true if the list iterator has more elements when
     *         traversing the list in the forward direction.
     */

    public abstract boolean hasNext();
}
```

Issues

- Which class will be implementing the interface?
- How to create and initialize a new iterator?
- How to move the iterator forward?
- How to detect the end?

$\Rightarrow$ This process is modeled on traversing an array and should be familiar.

Implementation -1-

In our first implementation, the class **LinkedList** developed earlier is modified to implement the interface **Iterator**. This involves modifying the header of the class, adding "implements Iterator", and to provide an implementation for each of the methods defined in the interface, **next()** and **hasNext()**.

l

head

current

A   B   C
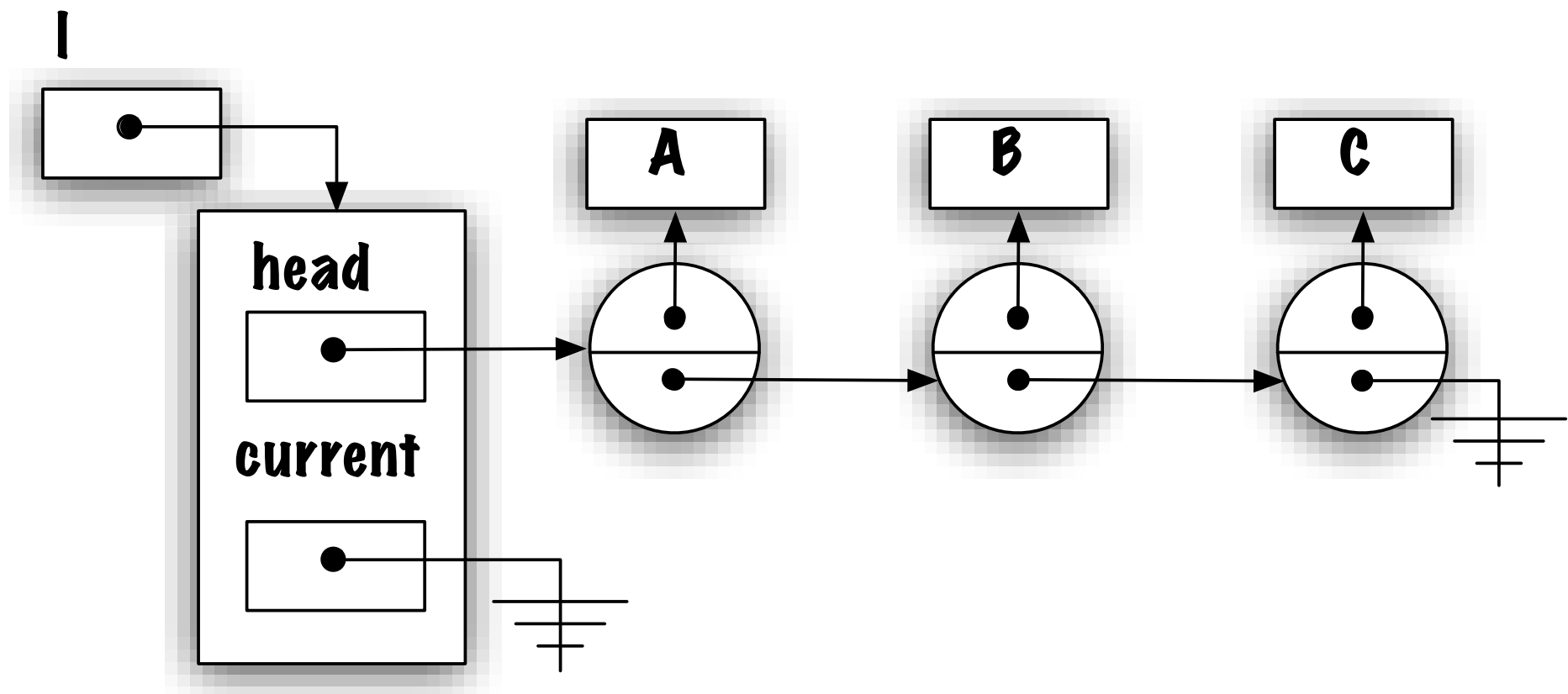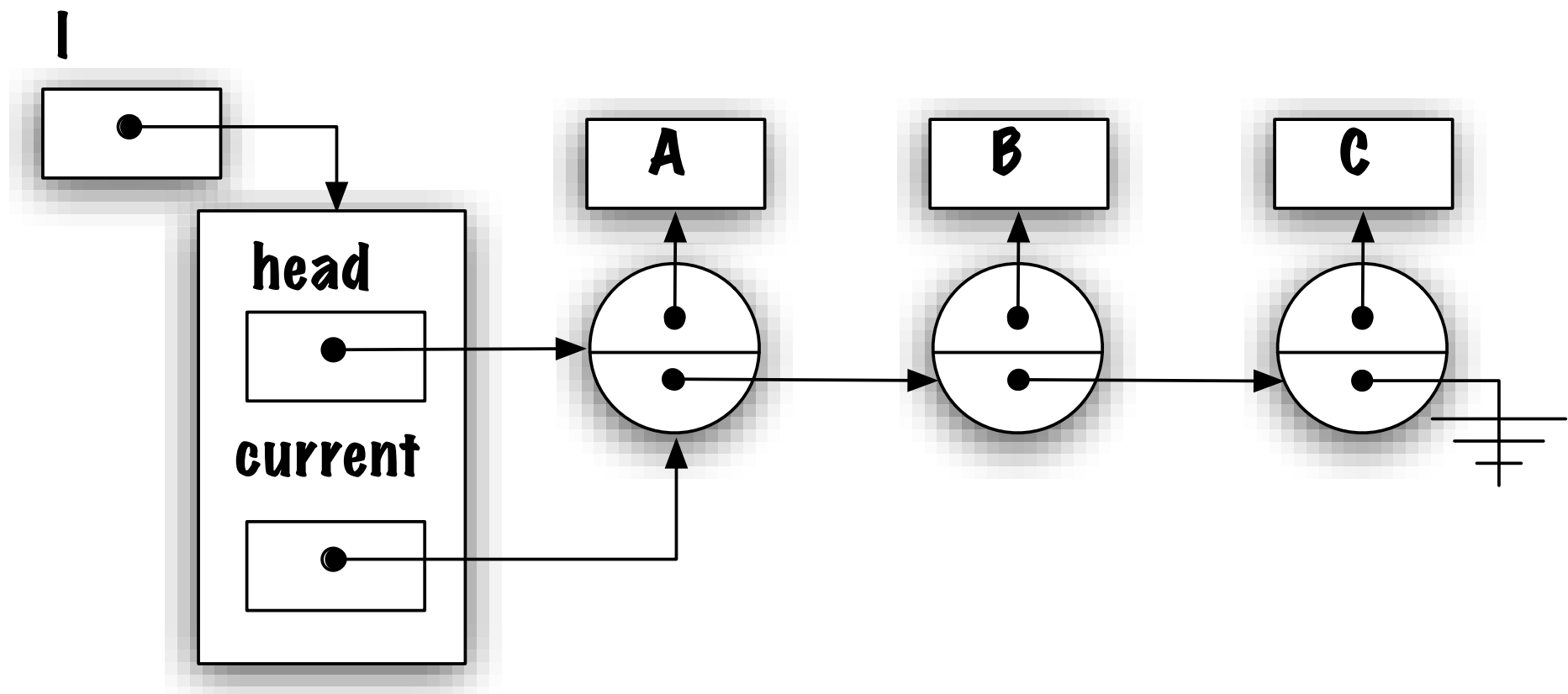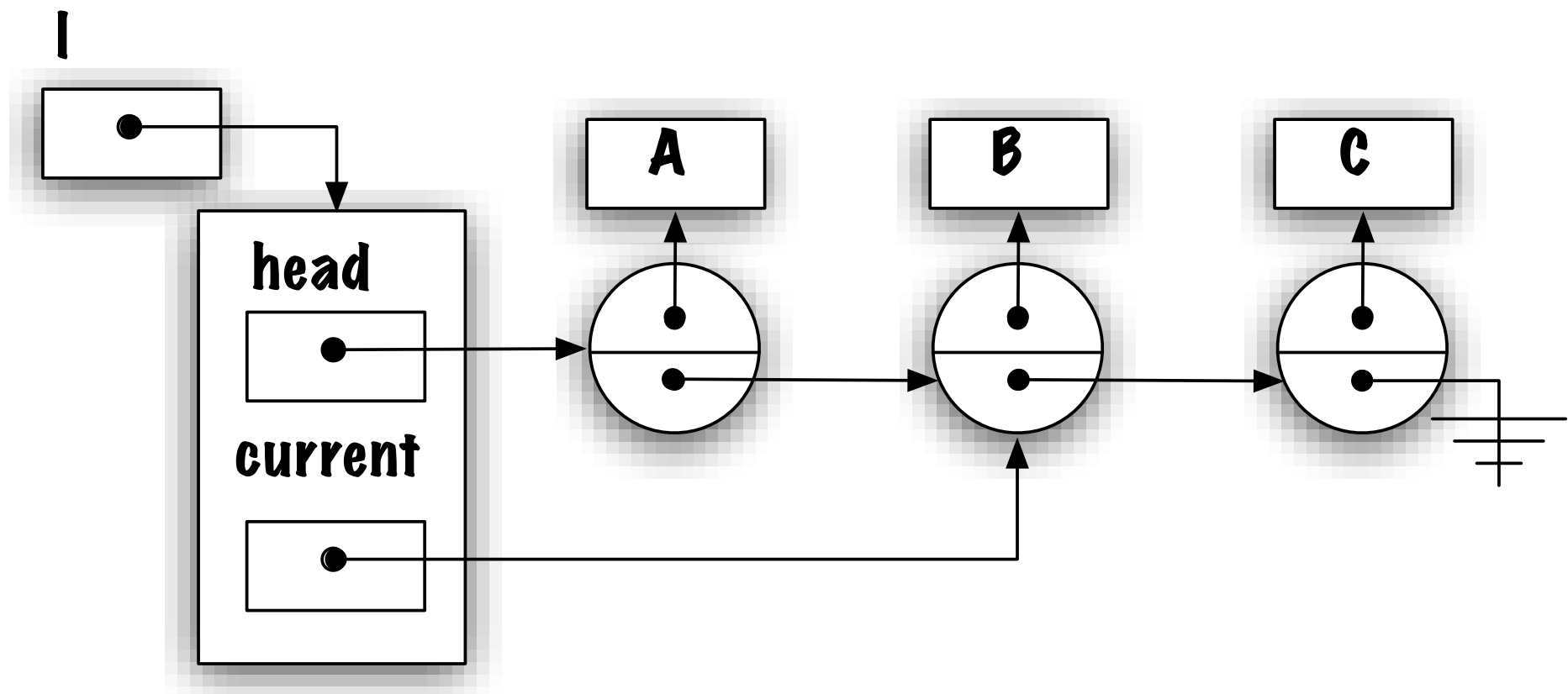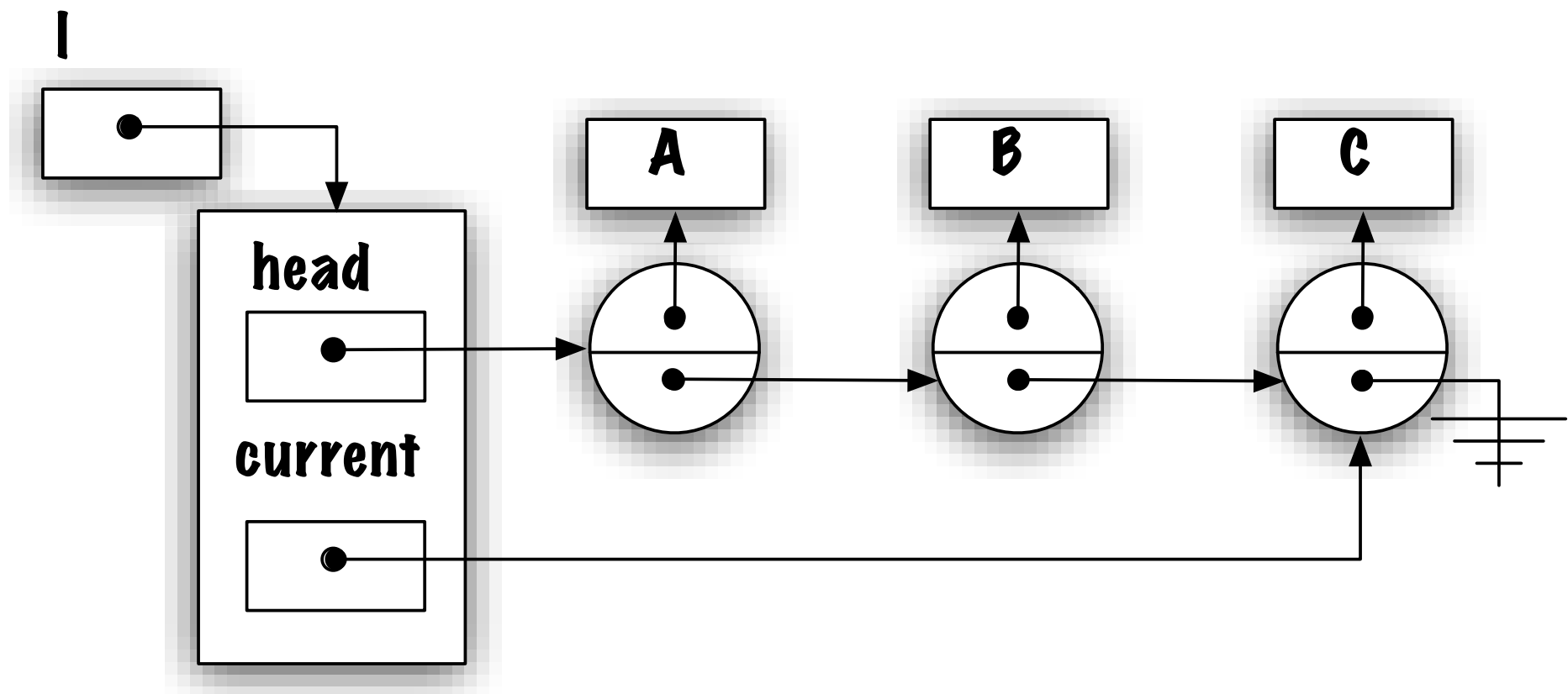
Usage

```java
List<E> l = new LinkedList<E>();

// Operations that add elements to the list ...

// Position the iterator to the left of the list

Iterator<E> i = l.iterator();

while ( i.hasNext() ) {

    E o = i.next();

    // do something with o
}
```

Implementation -1-

In our first implementation, the class **LinkedList** developed earlier is modified to implement the interface **Iterator**. This involves modifying the header of the class, adding "implements Iterator", and to provide an implementation for each of the methods defined in the interface.

```java
public class LinkedList<E> implements List<E>, Iterator<E> {

    private static class Node<E> {
        ...
    }

    private Node head<E>;

    private Node current<E>;

    public E next() { ... }

    public boolean hasNext() { ... }

    // ...
}
```

Remarks

1. Why is the iterator positioned to the left of the list and not on the first element?
   - To correctly process the case of the empty list; positioning the iterator onto the first element implies there will always be a first element.
2. Functional definition of "**next()**":
   (a) **Iterator moves one position forward**;
   (b) **then returns the value of the current element**.
3. A danger awaits you when calling "**next**", what is it?
   - Look at the functional definition of next, the iterator moves forward one position then returns a value, it's possible that there is no next element!
   - Give two examples where that would be the case:
     - The empty list;
     - The iterator was already positioned on the last element of the list.

Current limitation

Our current implementation does not allow for multiple iterators!

```
while ( i.hasNext() ) {
    oi = i.next();
    while ( j.hasNext() ) {
        oj = i.next();
        // process oi and oj
    }
}
```

What's needed then?

- We need as many references (variables) as iterators;

- **An iterator has to have access to the implementation (Node);**

- **An iterator needs access to the instance variables of the class LinkedList.**
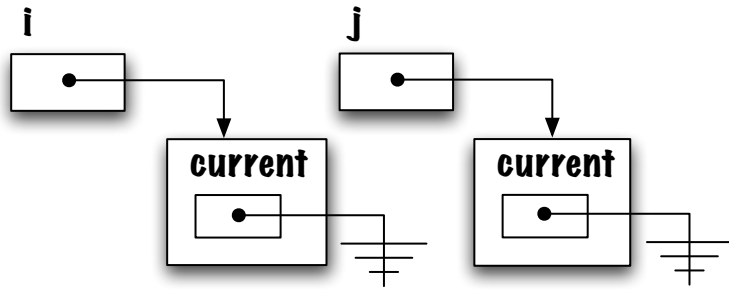
Allowing for multiple iterators

$\Rightarrow$ Suggestions?
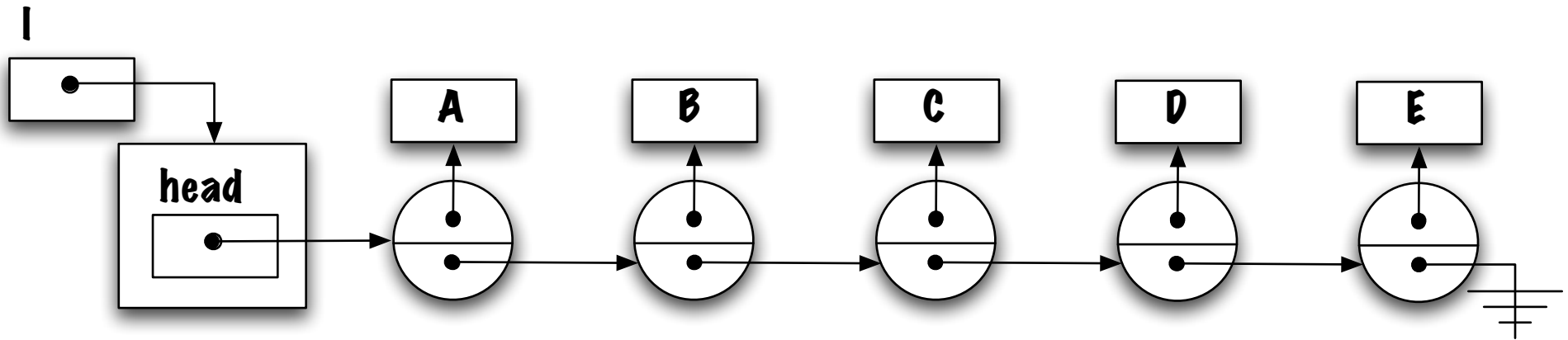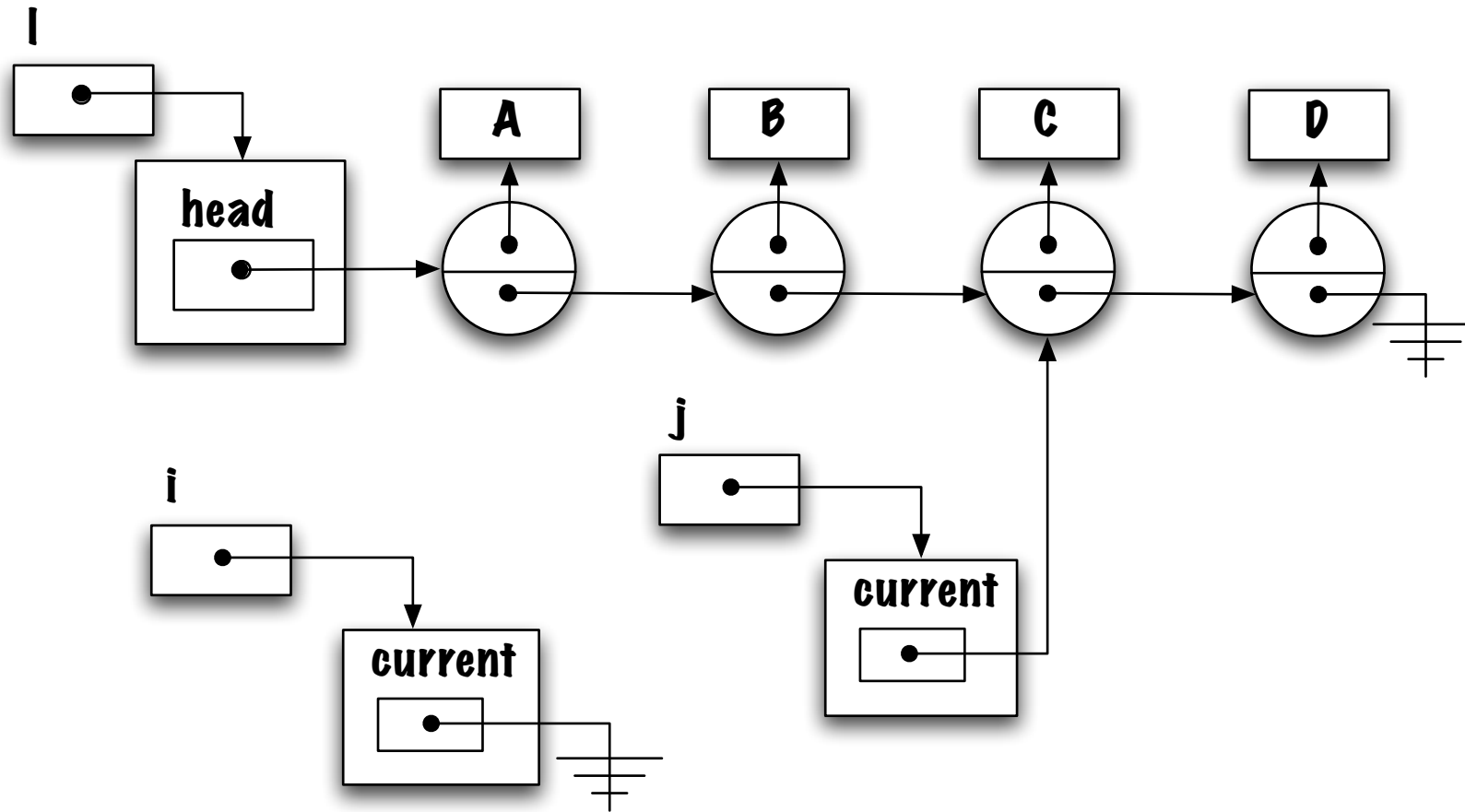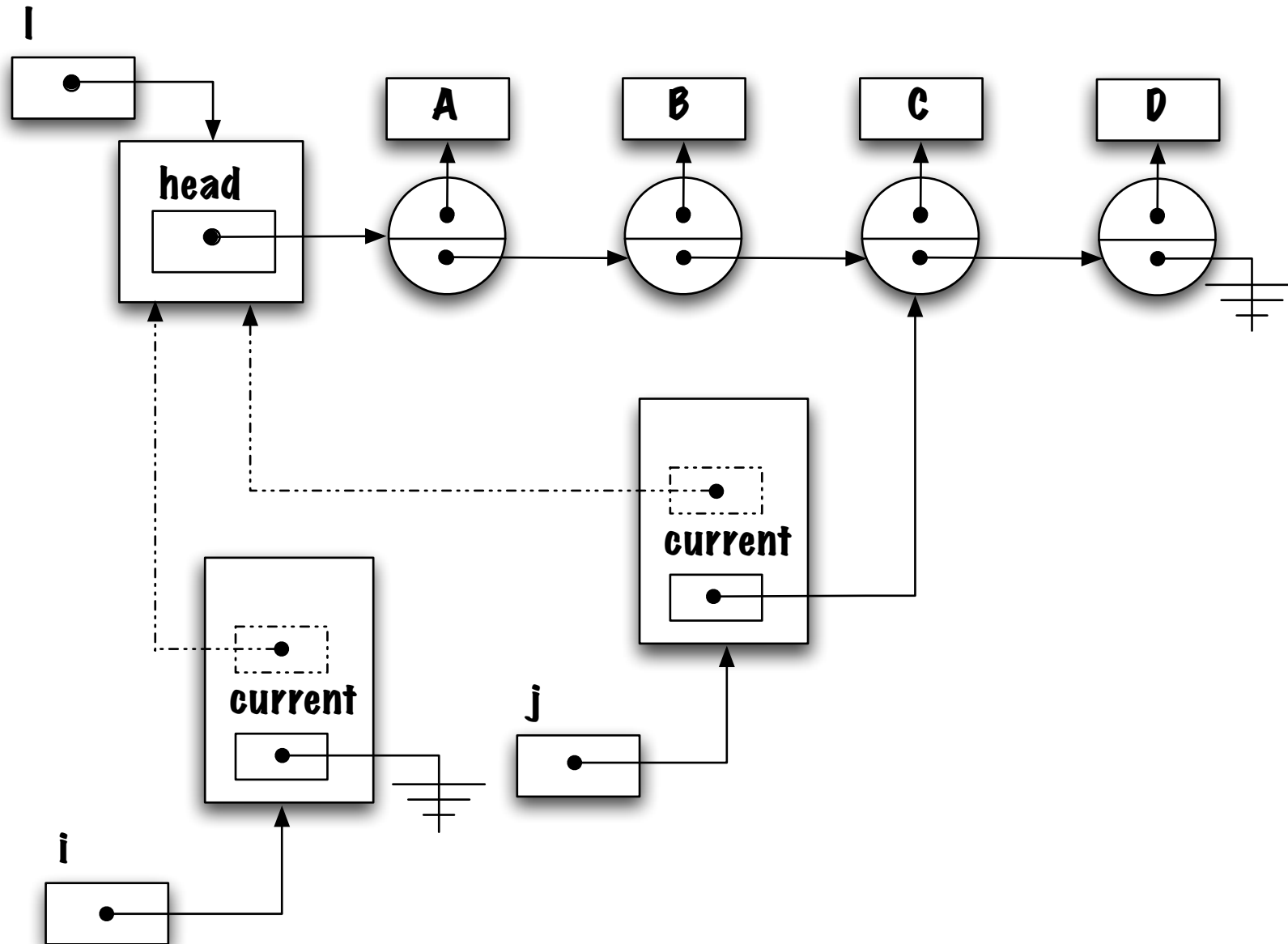
```
LinkedList<String> l1 = new LinkedList<String>();
l1.addLast( "A" );
l1.addLast( "B" );
l1.addLast( "C" );
l1.addLast( "D" );

LinkedList<Integer> l2 = new LinkedList<Integer>();
l2.addLast( new Integer( 1 ) );
l2.addLast( new Integer( 2 ) );
l2.addLast( new Integer( 3 ) );

Iterator<String> i, j;
Iterator<Integer> k;

i = l1.iterator();
j = l1.iterator();
k = l2.iterator();
```
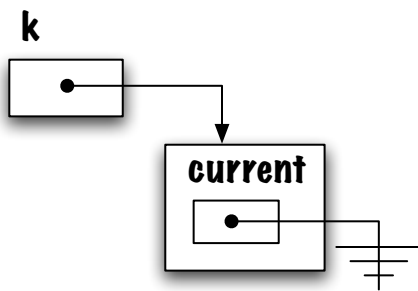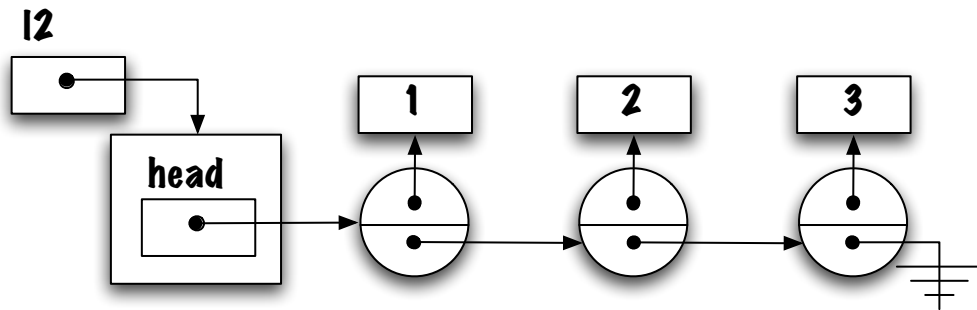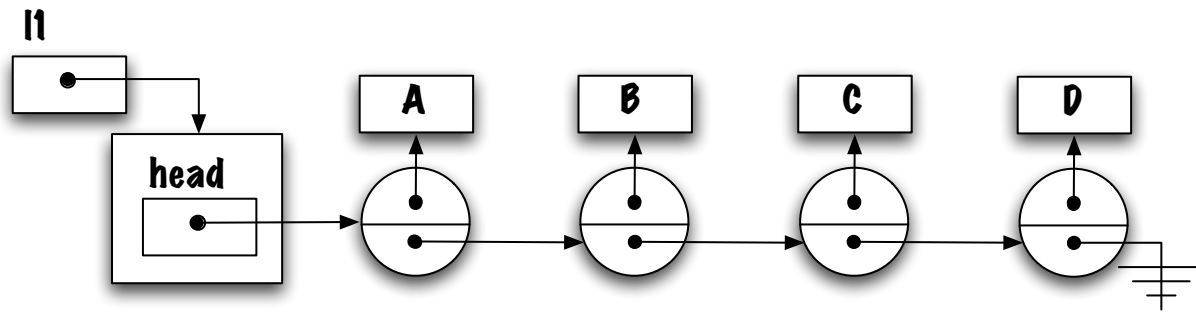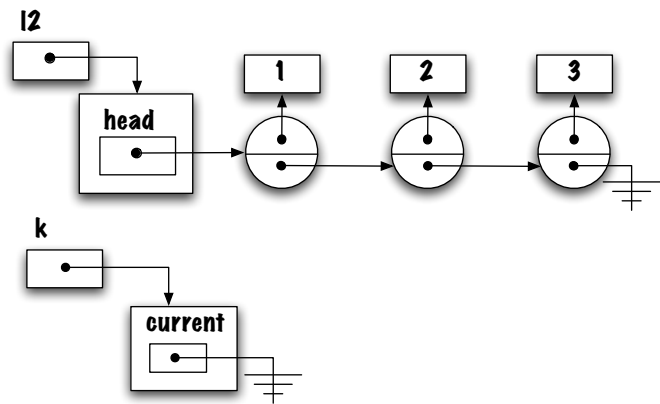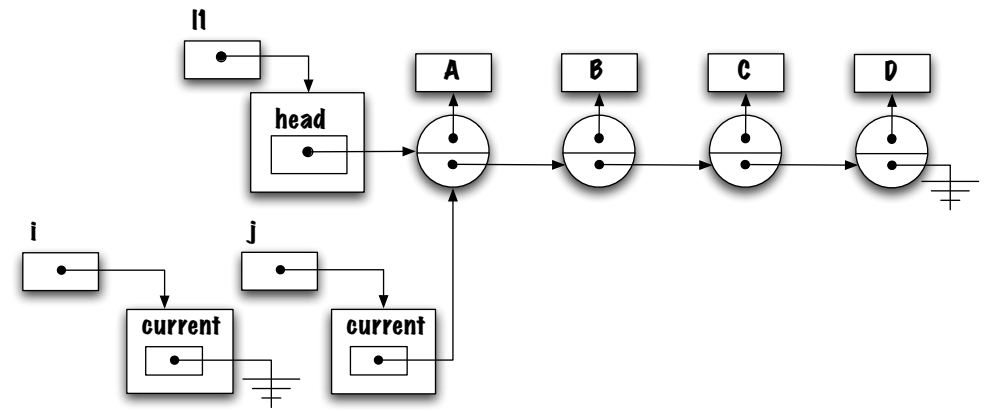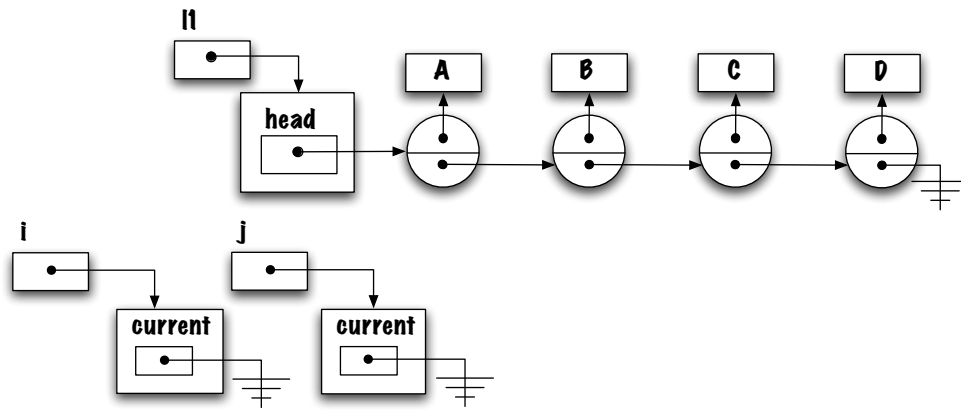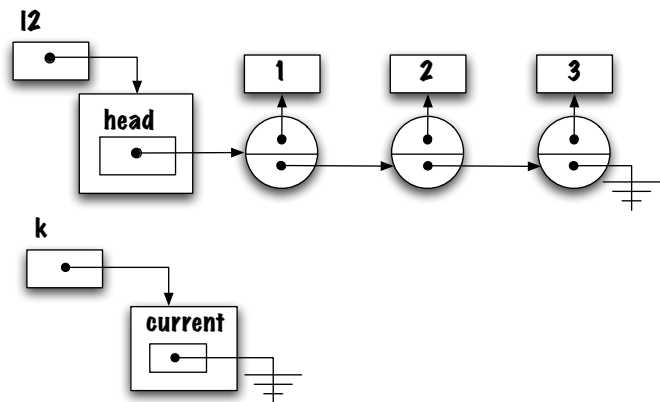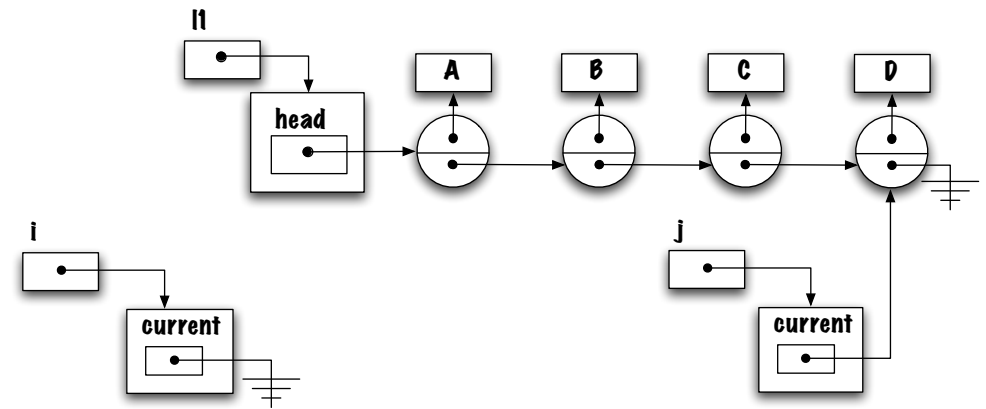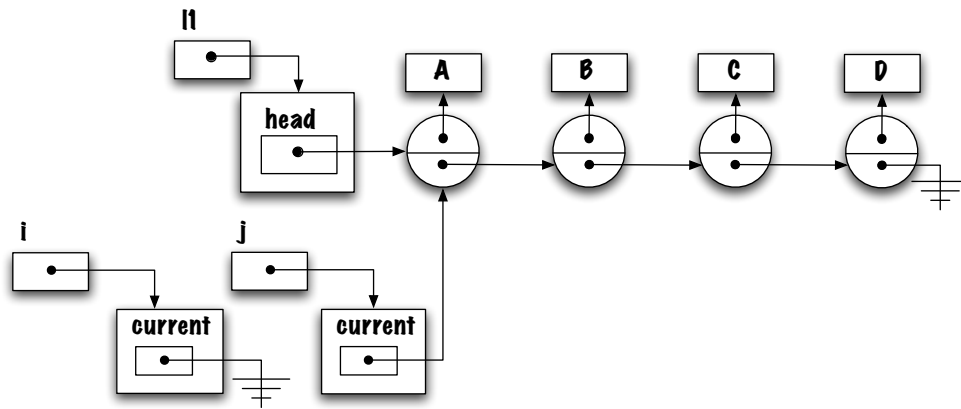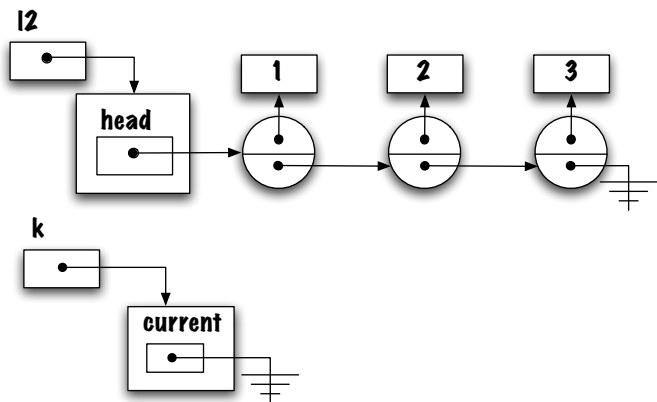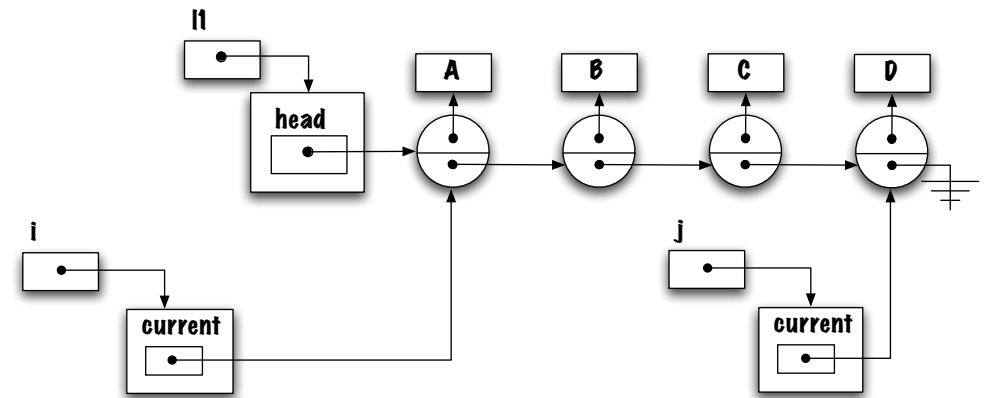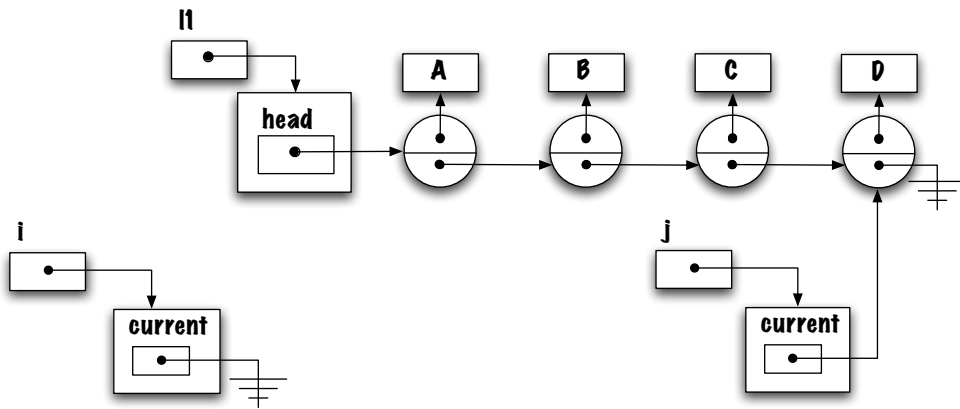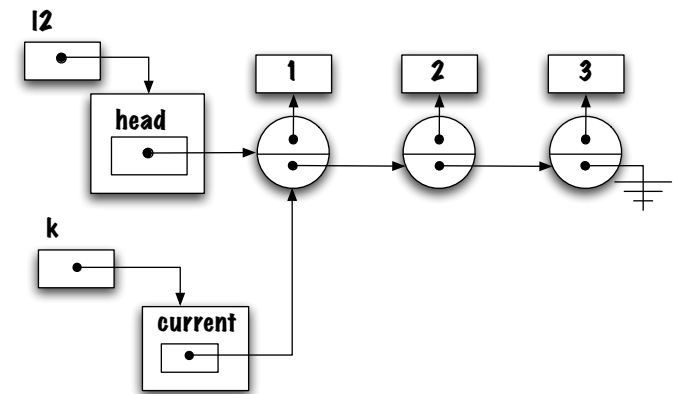
```
if ( j.hasNext() ) {
    String o = j.next();
}
```
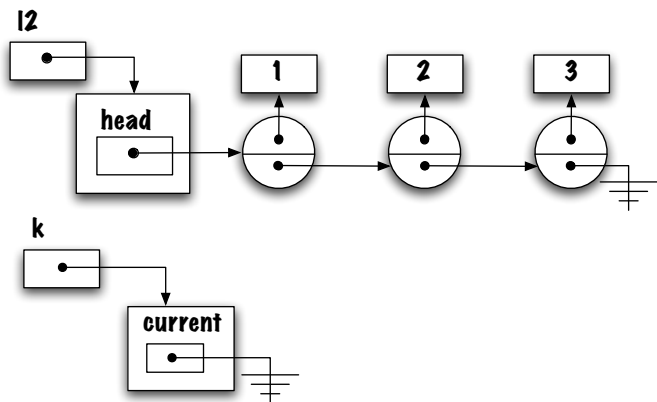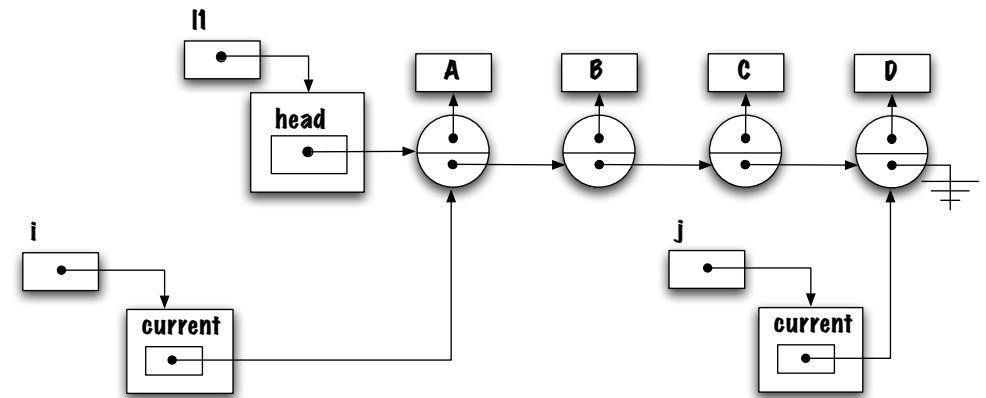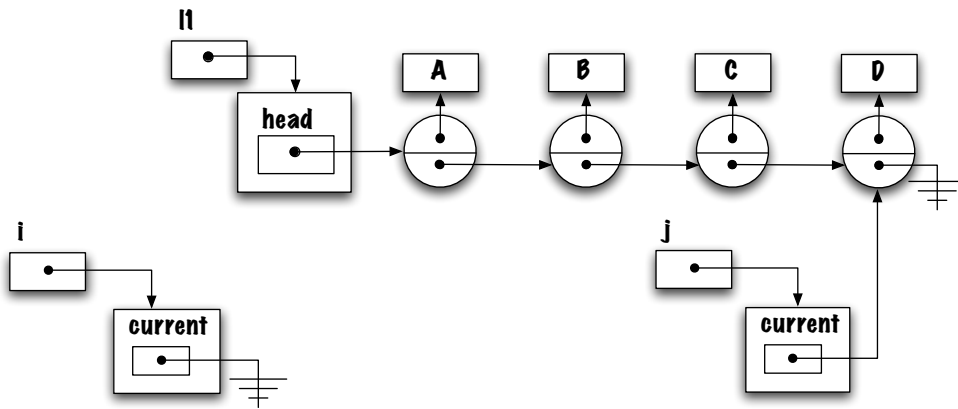
```
while ( j.hasNext() ) {
    String o = j.next();
}
```

```
if ( i.hasNext() ) {
    String o = i.next();
}
```

```
if ( k.hasNext() ) {
    Integer o = k.next();
}
```
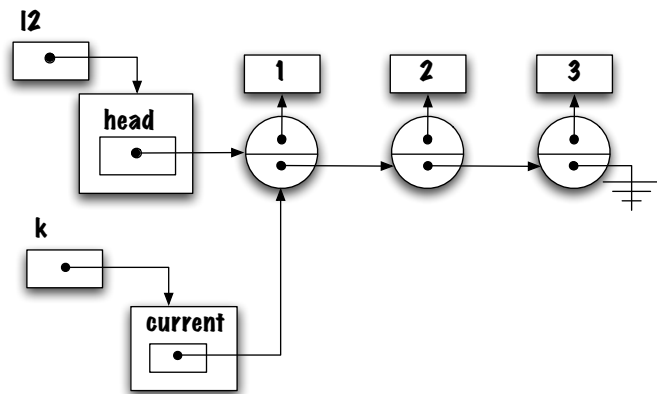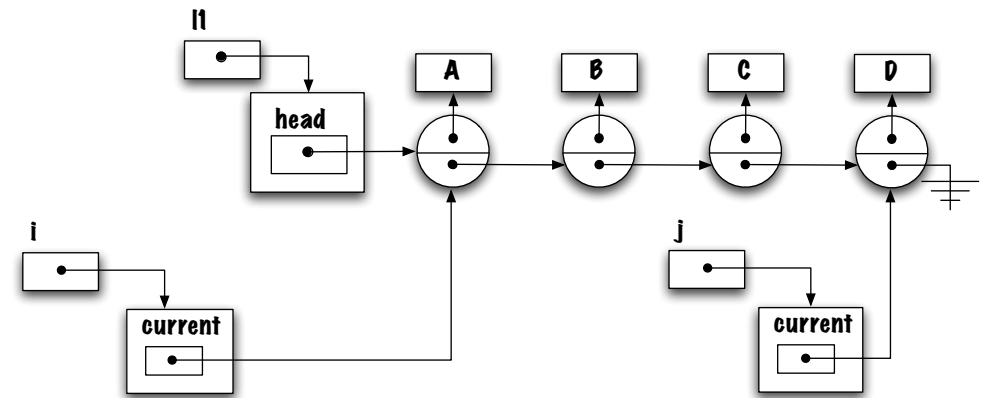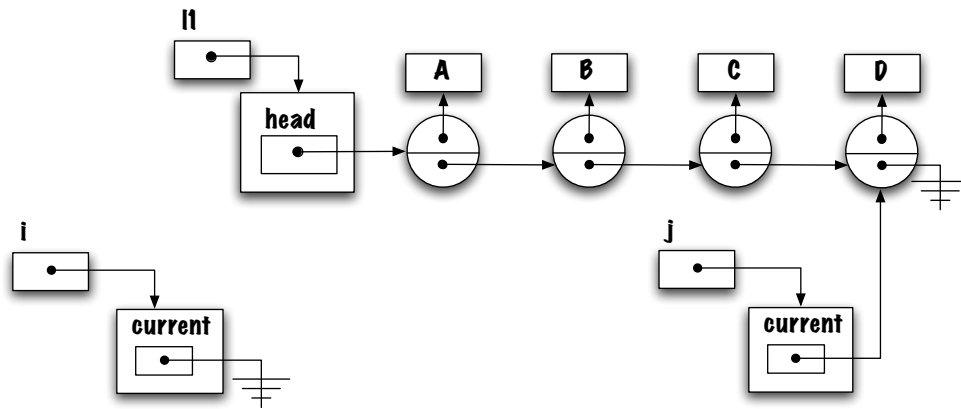
```
while ( k.hasNext() ) {
    Integer o = k.next();
}
```
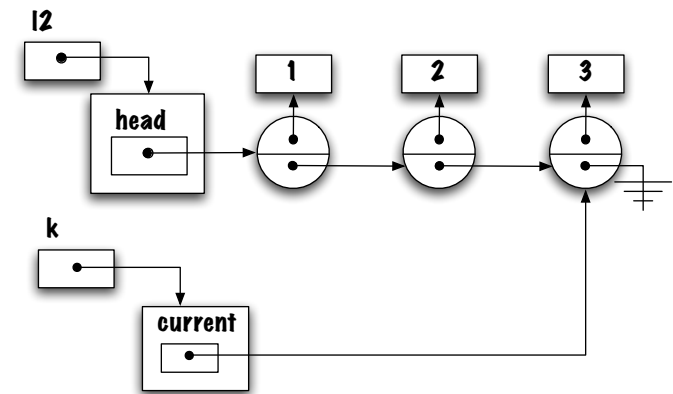
```
public class LinkedList<E> implements List<E> {
    private static class Node<E> { ... }

    Node<E> head;

    private class ListIterator implements Iterator<E> {
        // ...
    }
}
```

- **LinkedListIterator** is an inner class;
- **LinkedListIterator** implements **Iterator**;
- **LinkedList** does not implement **Iterator**.

Getting in Touch with your Inner Class (web resource)

www.javaranch.com/campfire/StoryInner.jsp

attractive object seeks
that special someone...
for sharing private thoughts,
walks on the beach,
drinking wine from a glass,
subclasses and pets OK.
NO STATICS!!

Inner class: summary

- "An *inner class* is a nested class that is not explicitly or implicitly declared static." The Java Language Specification;

- An object of the outer class creates an instance of an inner class:

  – An object of an inner class has **one** outer object;
  – An object of the outer class may have **zero**, **one** or **several** instances of its inner class;
  – **An object of an inner class has access to the instance variables and methods of its outer object**, we say that their **name space** is shared.

```
public class LinkedList<E> implements List<E> {
    private static class Node<E> { ... }

    Node<E> head;

    private class ListIterator implements Iterator<E> {
        // ...
    }
}
```
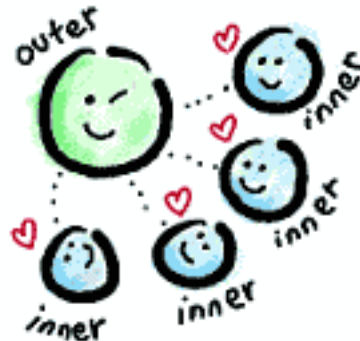
- **LinkedListIterator** is an inner class;
- **LinkedListIterator** implements **Iterator**;
- **LinkedList**
  underlinedoes not implement **Iterator**.

$\Rightarrow$ What are the instance variable of a **LinkedListIterator**?

Each object must have a variable (instance), a reference to the current element that it designates, **current**:

```
public class LinkedList<E> implements List<E> {

    private static class Node<E> { ... }

    private Node<E> head;

    private class ListIterator implements Iterator<E> {
        private Node<E> current; // <---
        ...
    }
}
```

The new implementation allows for several iterators to co-exist.

Its constructor initialises **current** to **null**, a value which, symbolically, designates being before the start of the list.

```java
public class LinkedList<E> implements List<E> {
    private static class Node<E> { ... }

    private Node<E> head;

    private class ListIterator implements Iterator<E> {
        private Node<E> current;
        private ListIterator() {
            current = null; // <---
        }
        ...
    }
    ...
}
```
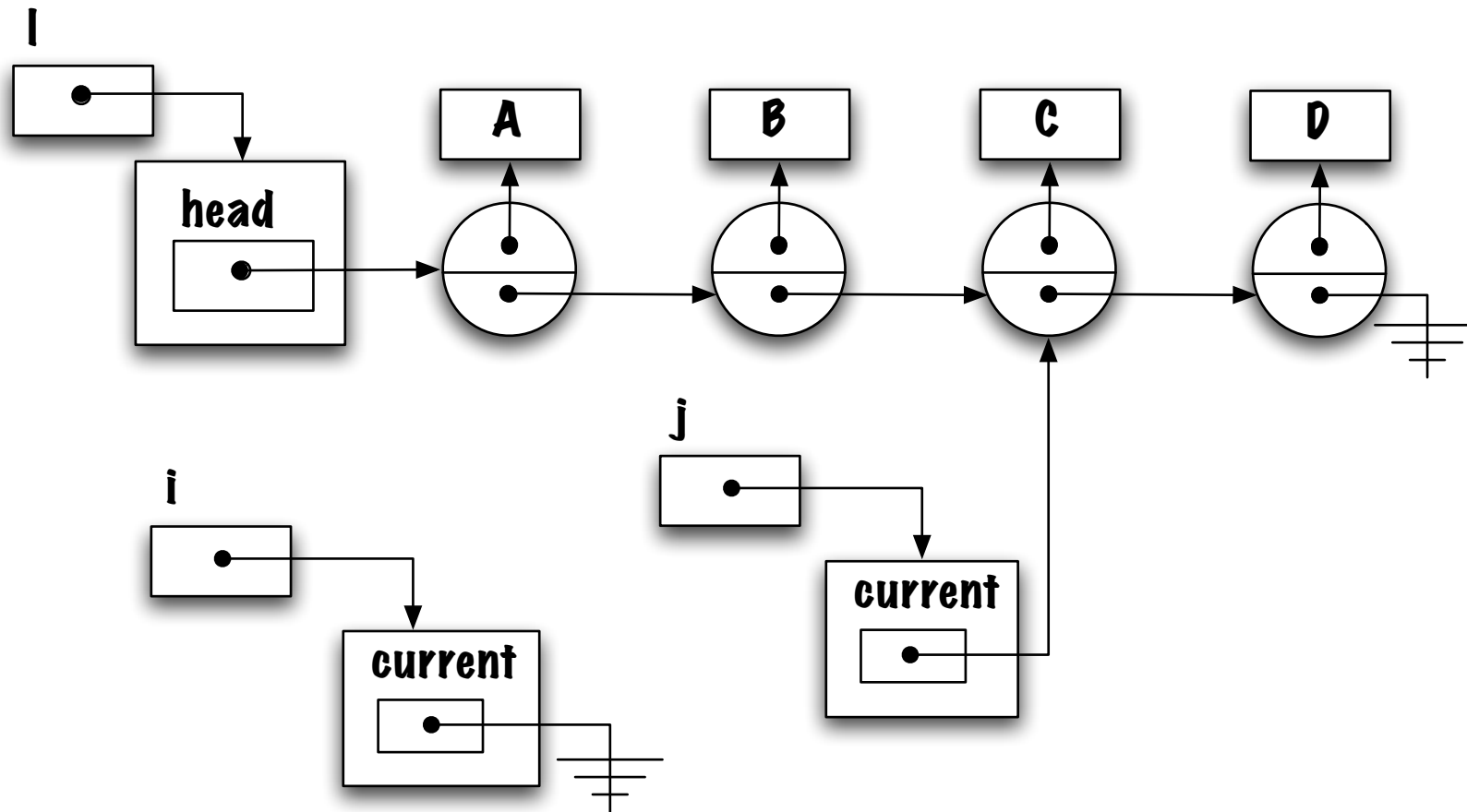
How to create an iterator?

```
public class LinkedList<E> implements List<E> {
    private static class Node<E> { ... }

    private Node<E> head;

    private class ListIterator implements Iterator<E> {
        private Node<E> current;
        private ListIterator() {
            current = null;
        } ...
    }
    public Iterator<E> iterator() {
        return new ListIterator(); // <---
    } ...
}
```

⇒ **iterator()**: is an instance method of an object of the class **LinkedList**.

```java
public class LinkedList<E> implements List<E> {
    private static class Node<E> { ... }
    private Node<E> head;
    private class ListIterator implements Iterator<E> {
        private Node<E> current;
        private ListIterator() {
            current = null;
        }
        public E next() {
            if ( current == null ) {
                current = head;
            } else {
                current = current.next;
            }
            return current.value;
        }
        public boolean hasNext() { ... }
    }
    public Iterator<E> iterator() {
        return new ListIterator<E>();
    } ...
}
```

Usage

```
List<Double> doubles = new LinkedList<Double>();

doubles.add( new Double( 5.1 ) );
doubles.add( new Double( 3.2 ) );

double sum = 0.0;

Iterator<Double> i = doubles.iterator();

while ( i.hasNext() ) {
    Double v = i.next();
    sum = sum + v.doubleValue();
}
```

⇒ Exercise: trace the execution of the above statements.

We've worked hard to create a mechanism to traverse the list.
What is worth it?

Intel Pentium III 750.0 MHz (2000)

# nodes	inside (ms)	iterator (ms)
10,000	2	3
20,000	4	5
40,000	8	11
80,000	14	19
160,000	23	48

2.66 GHz Intel Core i5 (2009)

# nodes	inside (ms)	iterator (ms)	get (ms)
200,000	2	1	54,896
400,000	3	3	336,844
800,000	6	5	1,715,142
1,600,000	11	9	7,209,248

Many iterators

```java
List<Double> doubles = new LinkedList<Double>();

for ( int c=0; c<5; c++ ) {
    doubles.add( new Double( c ) );
}
Iterator<Double> i = doubles.iterator();
while ( i.hasNext() ) {
    Double iVal = i.next();
    Iterator<Double> j = doubles.iterator();
    while ( j.hasNext() ) {
        Double jVal = j.next();
        System.out.println( "("+iVal+","+jVal+")" );
    }
}
```

$\Rightarrow$ (0.0,0.0), (0.0,1.0), . . . , (0.0,4.0), . . . , (4.0,4.0).

Iterator and new for loop

```
List<Integer> ints = Arrays.asList( 1, 2, 3, 4, 5 );

int s = 0;

for ( Iterator<Integer> it = ints.iterator(); it.hasNext(); ) {
    Integer i = it.next();
    s += i.intValue();
}

System.out.println( s );
```

Iterator and new for loop

Java 5 introduces a new syntax for the **for** loop that automagically creates and uses an iterator to traverse a list, and unboxes the elements returned by the iteration.

```
List<Integer> ints = Arrays.asList( 1, 2, 3, 4, 5 );

int s = 0;
for ( int i : ints ) {
    s += i;
}

System.out.println( s );
```

Iterator and new for loop

```
List<Integer> ints = Arrays.asList( 1, 2, 3, 4, 5 );

int s = 0;
for ( int i : ints ) {
    s += i;
}

System.out.println( s );
```

What type can be used with the new for loop? Any type that has an iterator. What is the Java way to enforce this? Any class that implements the interface **Iterable<E>**.

Iterable$<$E$>$

```
public interface Iterable<E> {
    public Iterator<E> iterator();
}
```

where

```
public interface Iterator<E> {
    public boolean hasNext();
    public E next();
    public void remove();
}
```

Iterable

```
public class LinkedList<E> implements List<E>, Iterable<E> {

    private static class Node<T> { ... }

    private Node<E> head;

    private class ListIterator implements Iterator<E> {
        ...
    }

    ...
}
```

Iterable

```java
public interface List<E> extends Iterable<E> {
    ...
}

public class LinkedList<E> implements List<E> {

    private static class Node<T> { ... }

    private Node<E> head;

    private class ListIterator implements Iterator<E> {
        ...
    }

    ...
}
```

Arrays and new for loop

For the sake of completeness, the new for loop can also be applied to arrays.

```
int[] xs = new int[] { 1, 2, 3 };

int sum = 0;

for ( int x : xs ) {
    sum += x;
}
```

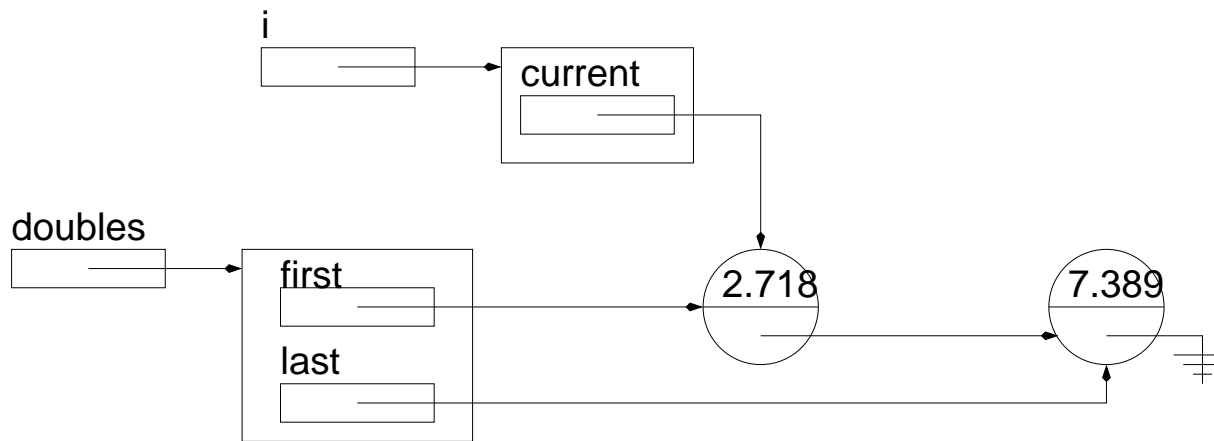# Concurrent modifications: "Fail-fast"

add(E o) and remove()

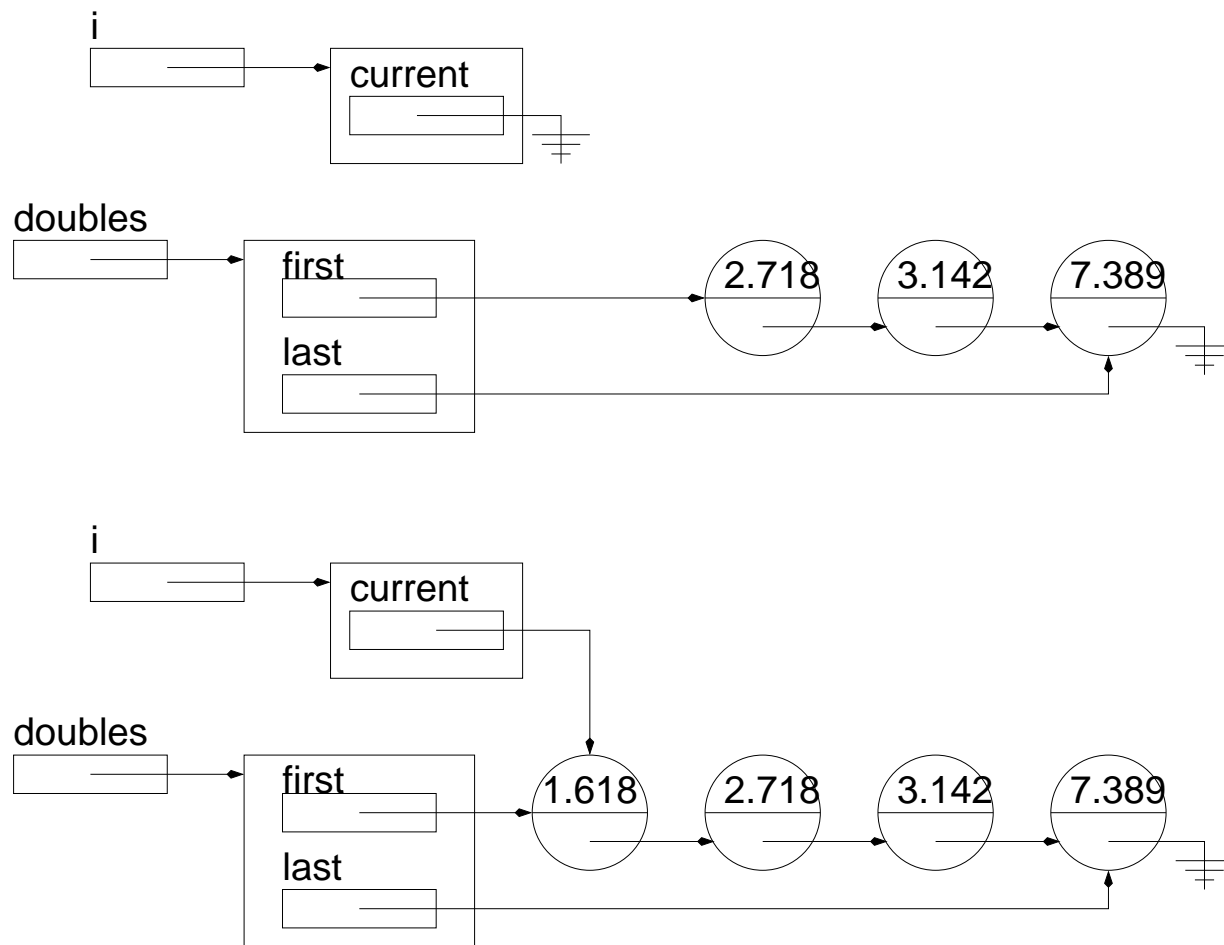Let's add two methods to the interface **Iterator**, **add(E o)** and **remove()**.

add()

By definition, **i.add(value)** will insert the new value immediately before the next element that would be returned by **next()**.

Therefore, the next call to **next()** is unaffected!
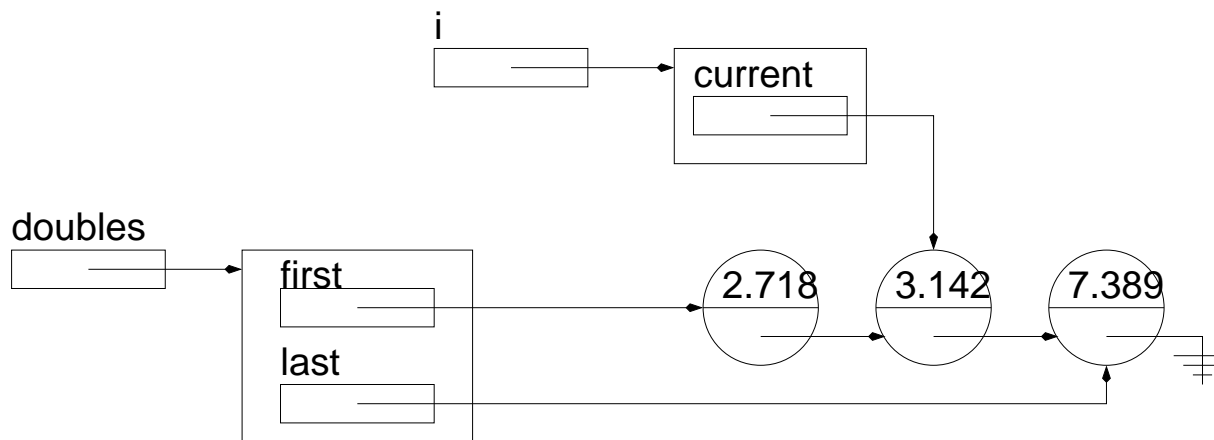
$\Rightarrow$ Call to **i.add(new Double(3.142))**, adding an element at an intermediate position in the list, notice that **i.next()** returns the same value with or without the insertion!

i

current

doubles

first
2.718   3.142   7.389

last

i

current

doubles

first
1.618   2.718   3.142   7.389

last

⇒ Adding an element when the iterator is positioned to the left of the list, **i.add(new Double(1.618))**, again **i.next()** returns the same value with or without insertion.
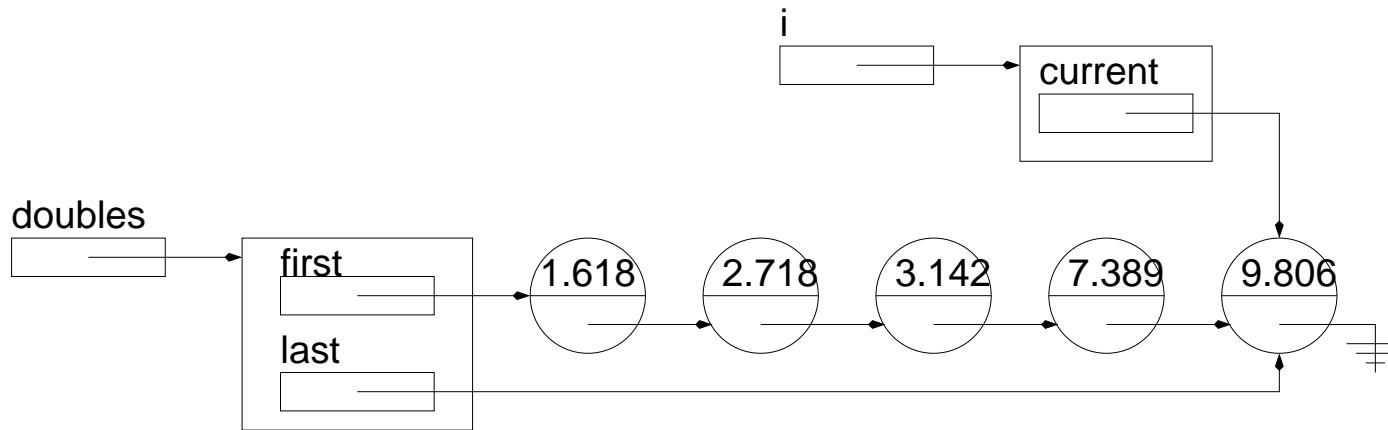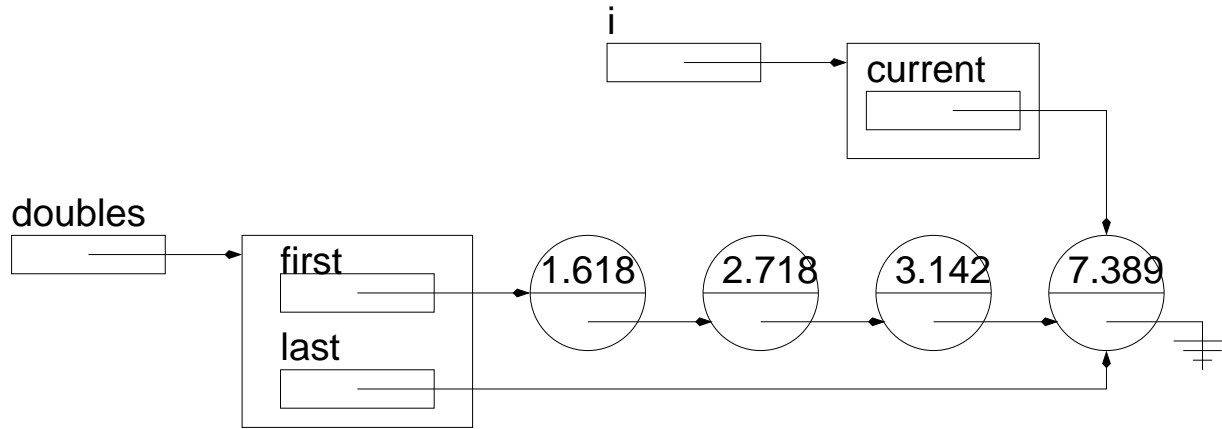
⇒ Adding an element when positioned at the end of the list, **i.add(new Double(1.618))**, notice that **hasNext()** returns the same value with or without insertion!

```java
public class LinkedList<E> implements List<E> {
    private static class Node<E> { ... }
    private Node<E> first;
    private class ListIterator implements Iterator<E> {
        private Node<E> current;
        private ListIterator() { current = null;  }
        public E next() { ... }
        public boolean hasNext() { ... }
        public boolean add(E o) {
            if ( o == null )
                return false;
            if ( current == null ) {
                first = new Node<E>( o, first );
                current = first;
            } else {
                current.next = new Node<E>( o, current.next );
                current = current.next;
            }
            return true;
        }
        public void remove() { ... }
    }
```

```
    public Iterator<E> iterator() { return new ListIterator(); }
    // ... all the usual methods of LinkedList
}
```

What would the following test program display?

```
public class Test {
    public static void main( String[] args ) {
        LinkedList<String> l = new LinkedList<String>();
        Iterator<String> i = l.iterator();

        for ( int c=0; c<5; c++ ) {
            i.add( "element-" + c );
        }
        i = l.iterator();
        while ( i.hasNext() ) {
            System.out.println( i.next() );
        }
    }
}
```

```
element-0
element-1
element-2
element-3
element-4
```
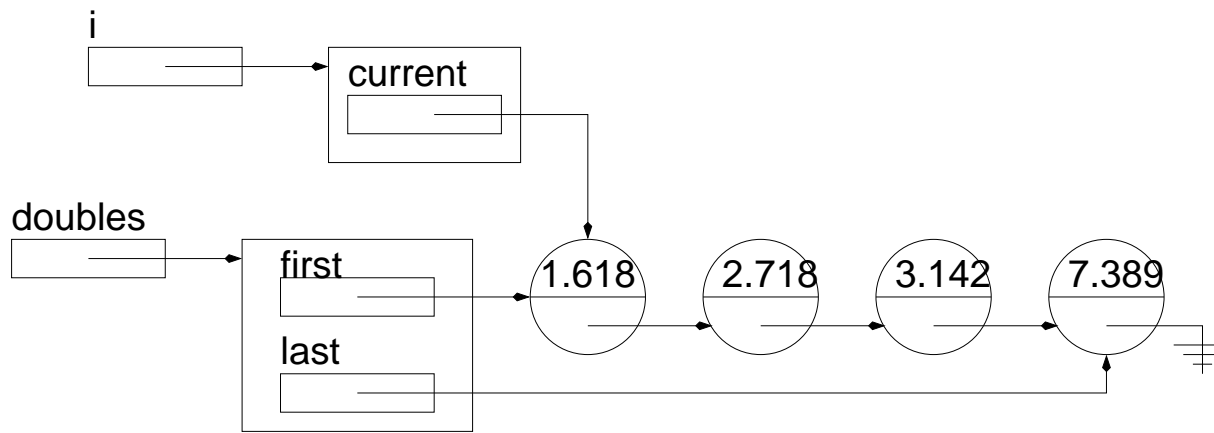
remove()

Removes from the list the last element that was returned by **next()**.

A call **i.add(value)** immediately followed by **i.remove()** keeps the list unchanged.
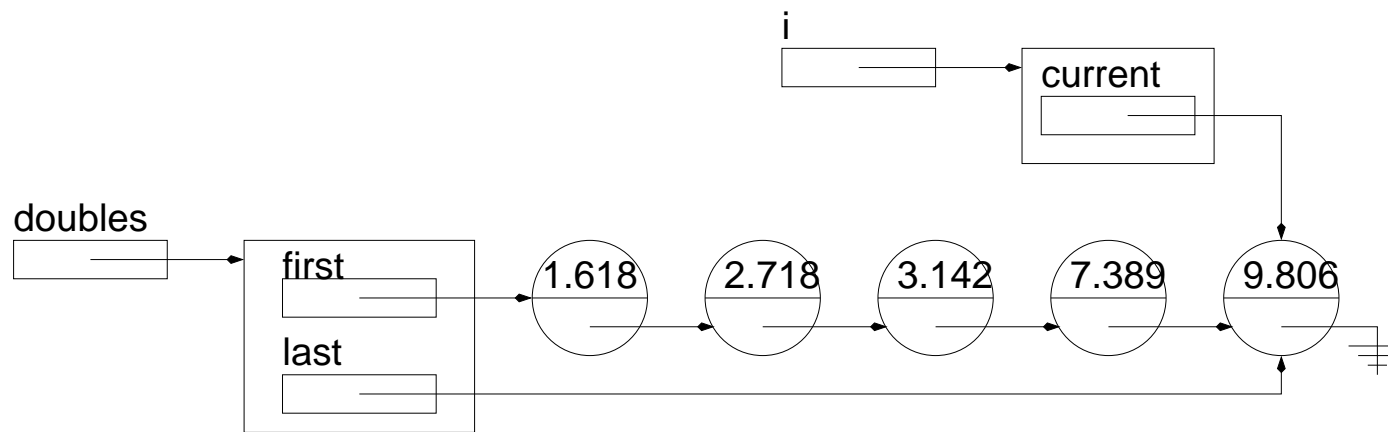
$\Rightarrow$ Notice that **remove()** does return the element that was removed, it is intended that a previous call to **next()** has saved the value.
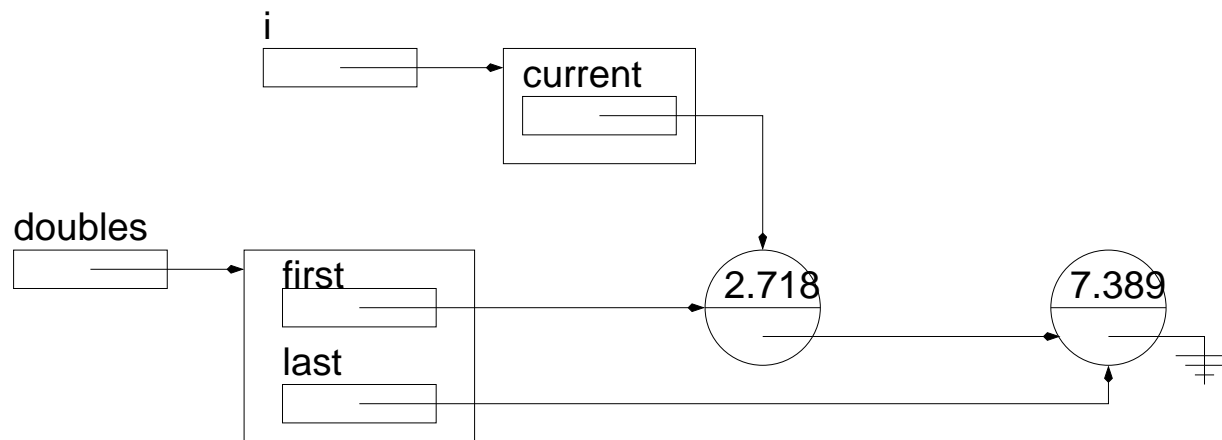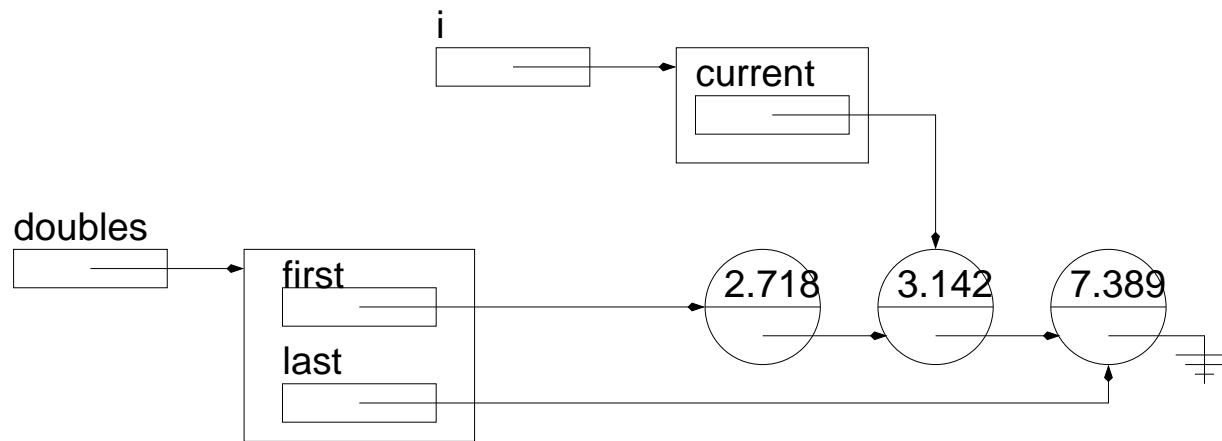
⇒ Removing an element at an intermediate position, **i.remove()**, notice that the value of **i.next()** is the same with or without removal.
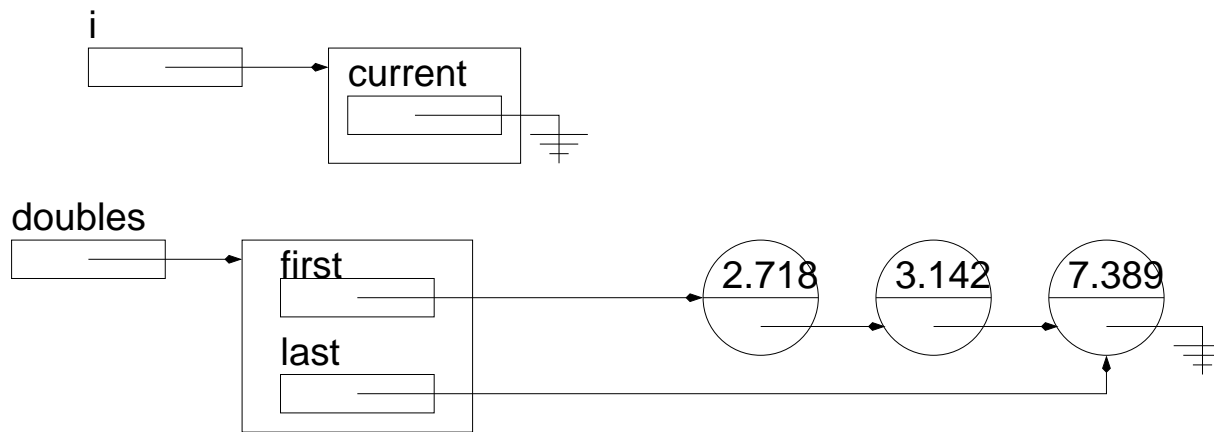
⇒ Removing the first element, **i.remove()**, the value of **i.next()** remains unchanged.

⇒ Removing the last element, **i.remove()**, the value of **hasNext()** is the same with or without removal.

Iterator

```
public interface Iterator<E> {
    public abstract E next();
    public abstract boolean hasNext();
    public abstract void add( E o );
    public abstract void remove();
    public abstract boolean hasPrevious();
}
```

Public methods of LinkedList

```
int size();
void add( E o );
E get( int pos );
E remove( int pos );
Iterator<E> iterator();
```

Limitation of the current implementation?

Concurrent modifications to the list! Having many iterators on the same list poses no problems as long as long no concurrent modifications occur.

Many iterators and concurrent modifications

Consider:

```
[1.618, 2.718, 3.142, 7.389, 9.806]
          i
          j
```

What would occur if a program calls **i.remove()**? What should occur of **j**?
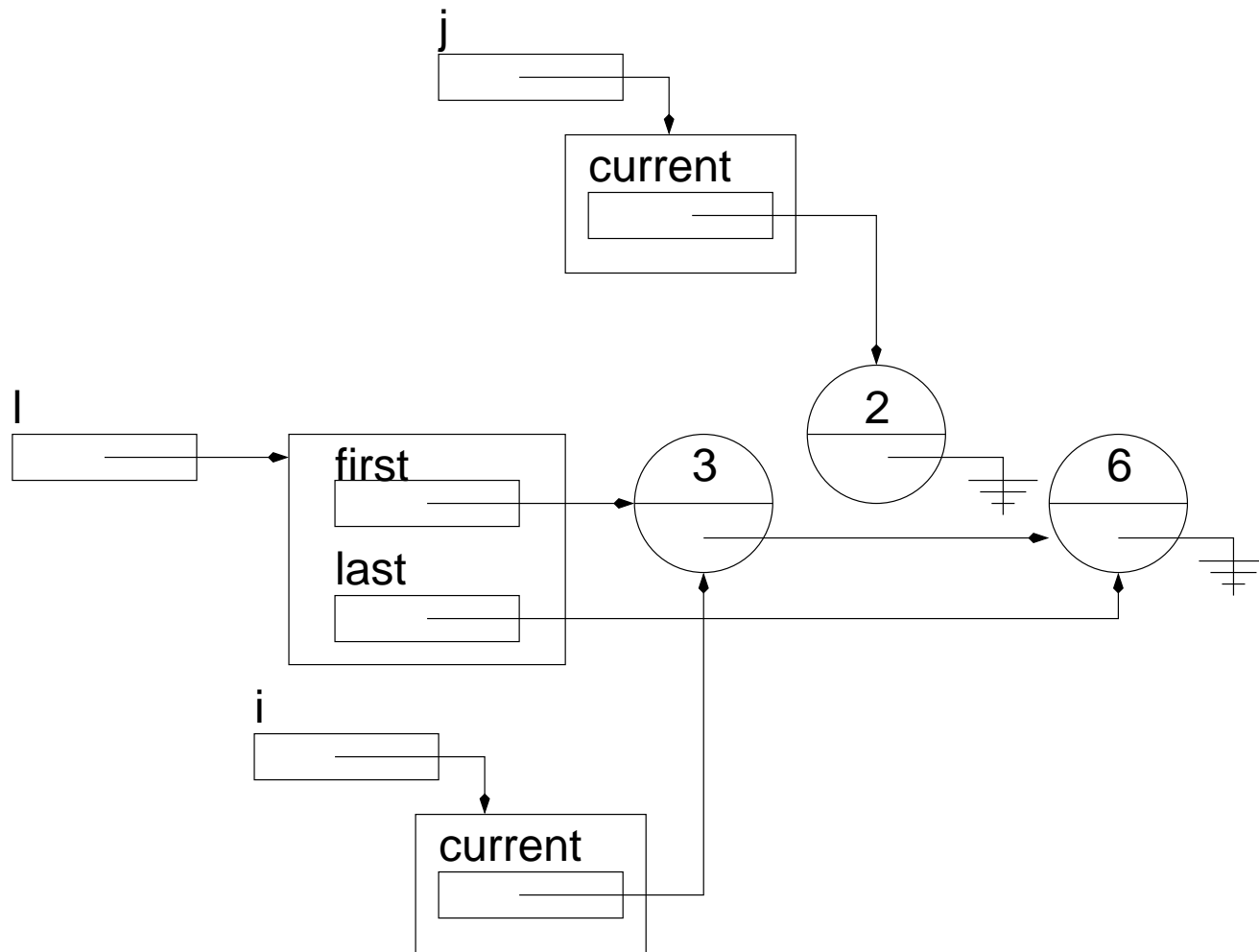
Consider:

```
[1.618, 2.718, 3.142, 7.389, 9.806]
          i
                    j
```

After the call **j.remove()** what should be the value of **i.next()**?
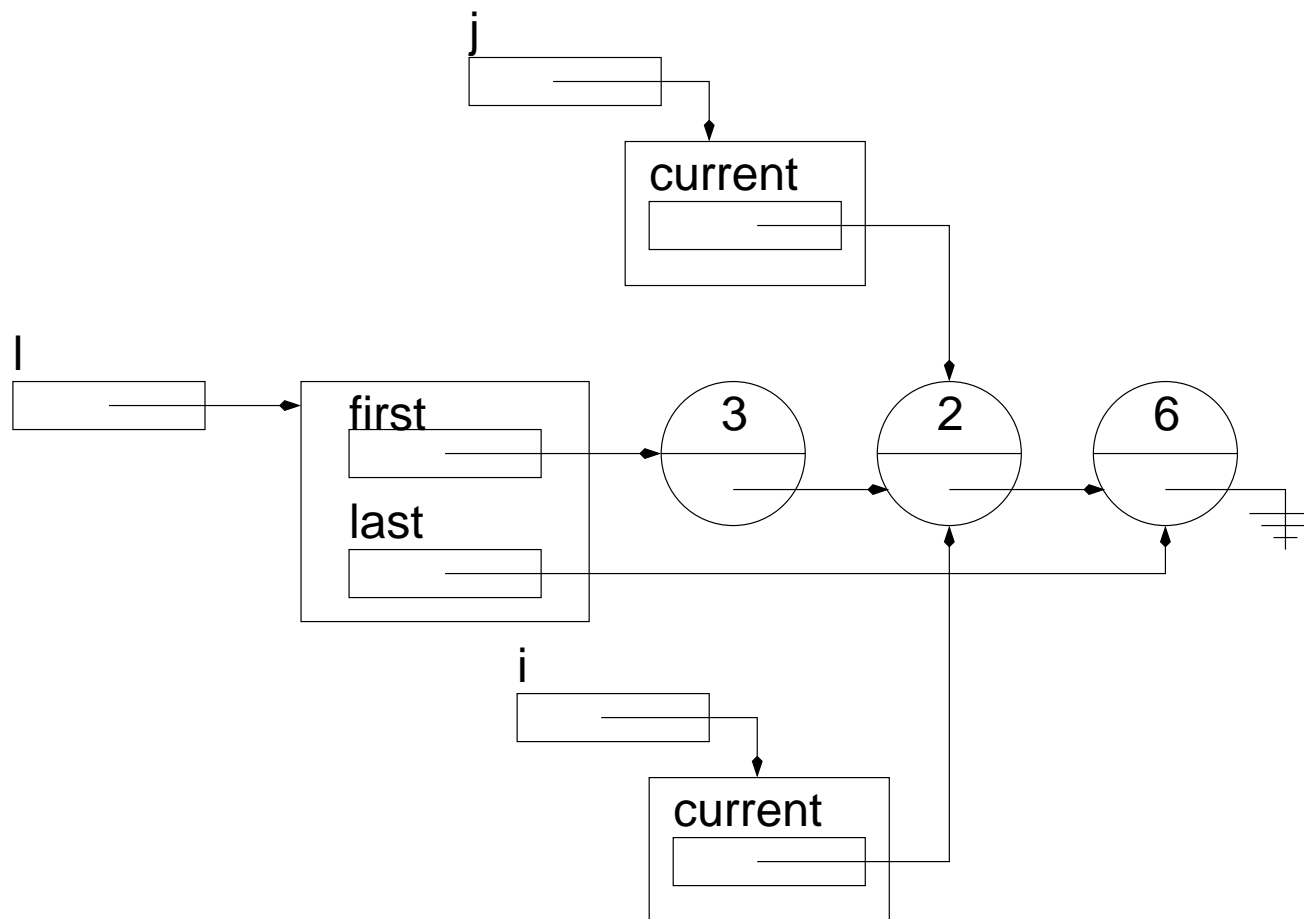
$\Rightarrow$ Consider what would occur after **i.remove()**.

$\Rightarrow$ **j** is now invalid!

# Implementation 3: *fail-fast*

The solution that we will adopt is called **fail-fast** and consists of making an iterator invalid as soon as a modification to the list has occurred that was not made by the iterator itself; i.e. the modification was made by another iterator or an instance method of the class **LinkedList**.
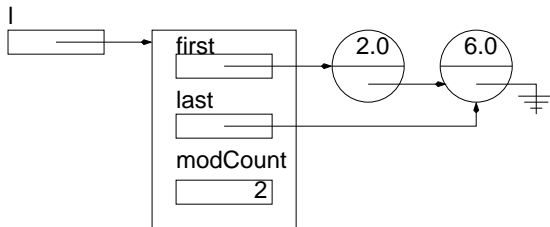
Implementing this solution requires:

1. Adding a counter of modifications to the header of the list (**modCount**);

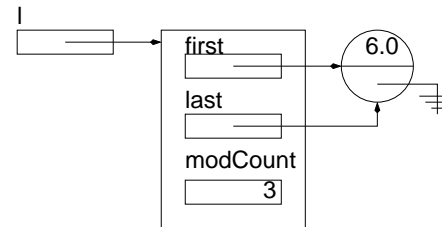2. Adding a counter of modifications to the iterators (**expectedModCount**);
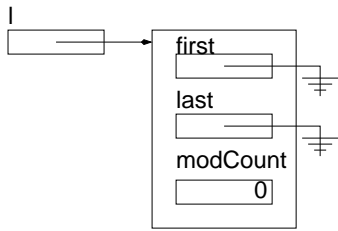
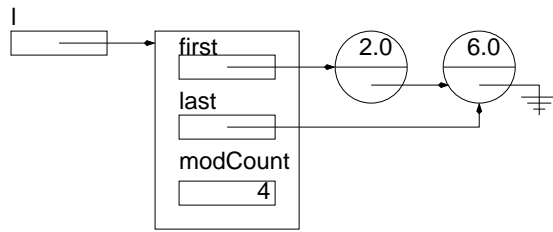l = new LinkedList();
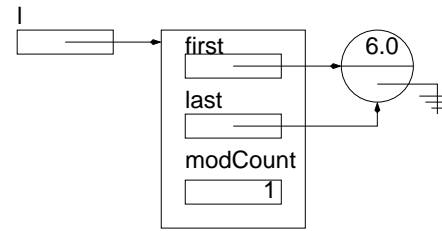
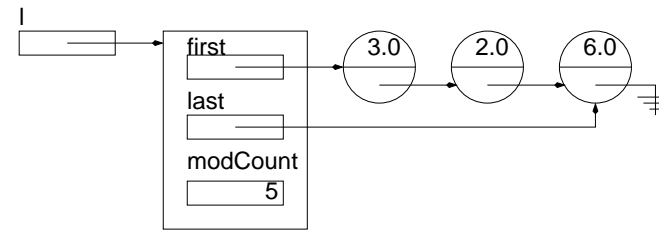l.addFirst(new Double(6.0));

l.addFirst(new Double(2.0));
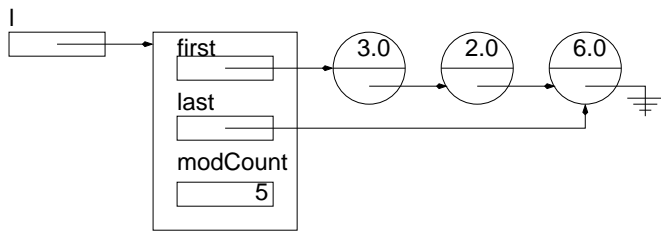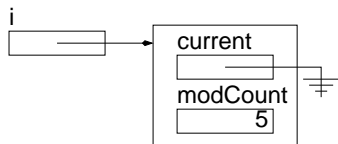

l.deleteFirst();

1. When creating a new list, the modification counter is set to 0;
2. Notice that **modCount** counts the number of modifications and **not** the number of elements in the list. Following the call to **deleteFirst()**, **modCount** is 3 and not 1.
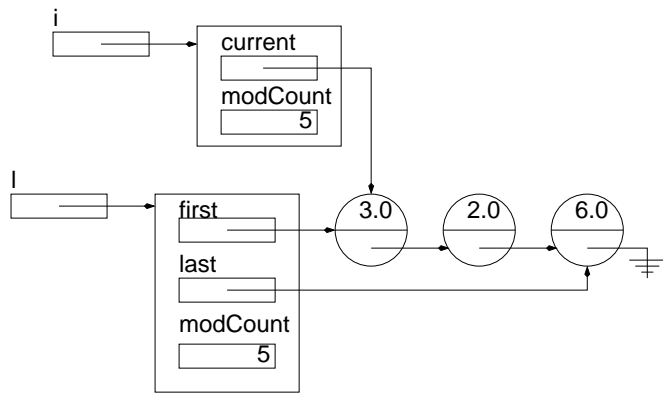
l.addFirst(2.0);

l.addFirst(3.0);
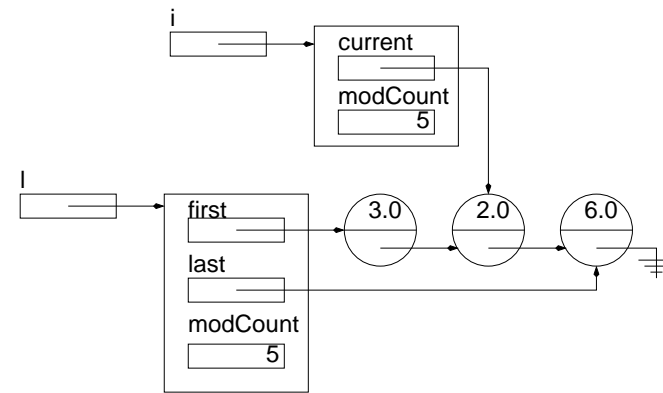
Iterator i = iterator();

⇒ Creating a new **Iterator** does not change **modCount**!

i.next();

i.next();
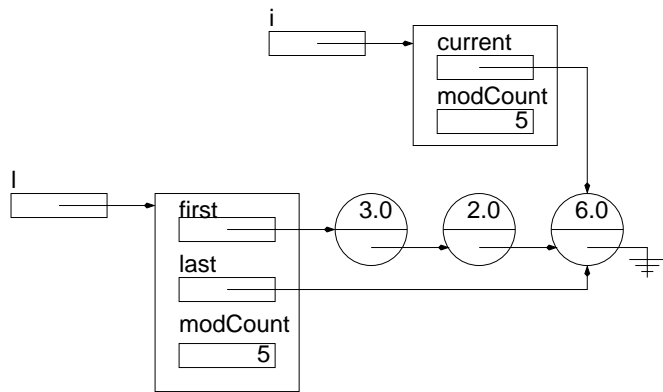
i.next();

$\Rightarrow$ Similarly, traversing the list does not change **modCount**.
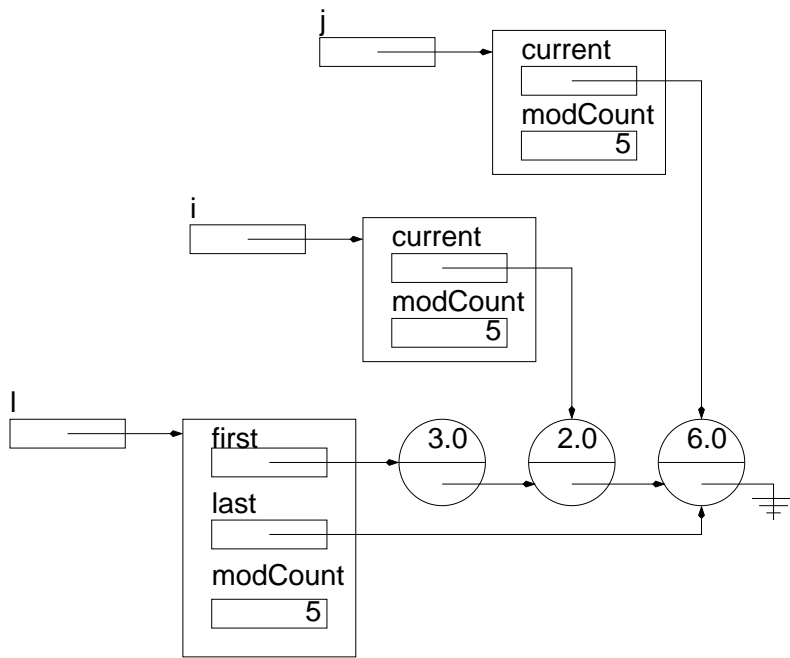
Both iterators are valid
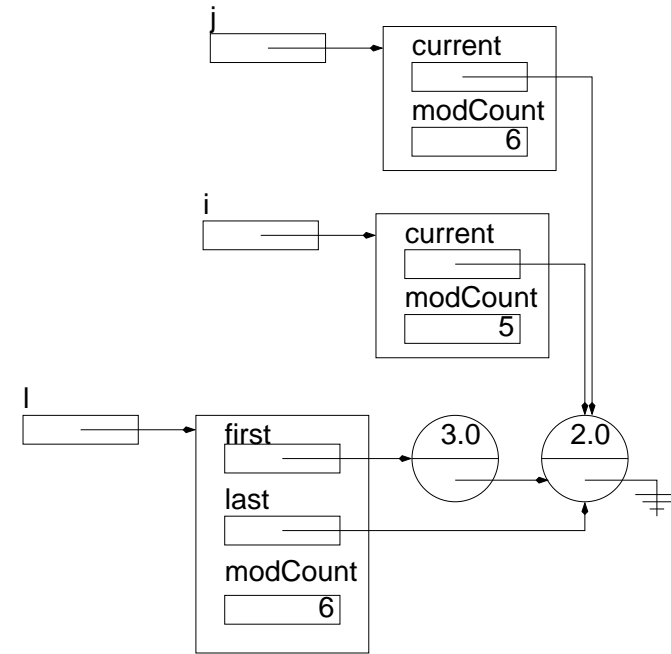
j.remove();

⇒ The iterator whose **expectedModCount** has the same value as the one in the header of the list remains valid; all the other ones are not.

Changes to the class LinkedList

The list operations that modify the list, **addFist**, **deleteFirst**, etc. must now change the value of **modCount**.

When creating a new list, **modCount** must be set to zero.

Modifications to ListIterator

When creating a new iterator, the value **modCount** from the header of the list must be copied to the instance variable **expectedModCount**.

The methods for adding and removing elements must update both, their own **expectedModCount** and **modCount** in the header of the list.

All the iterator's methods must be changed so that the precondition "is valid?" is true, otherwise they must throw the appropriate exception, **ConcurrentModificationException**.

isValid()

To facilitate writing the methods of the inner class **ListIterator**, we create a new method called **isValid()**.

The method checks that **expectedModCount == modCount**.

Appendix : Accessing Members of the Outer Object

In the previous example, the modification counter of the inner class (**expectedModCount**) was given a different name than that of the outer class (**modCount**) to avoid a name conflict.

The syntax **LinkedList.this.modCount** could also be used in the inner class (**ListIterator**) to refer to the instance variable of the outer object (**LinkedList**).

```java
public class Outer {
    private int value = 99;
    public class Inner {
        private int value;
        public Inner() {
            this.value = Outer.this.value + 1;
        }
        public int getValue() {
            return value;
        }
    }
    public Inner newInner() {
        return new Inner();
    }
    public int getValue() {
        return value;
    }
}
```

Accessing Members of the Outer Object

```
class Test {
    public static void main( String[] args ) {

        Outer o = new Outer();
        System.out.println( "o.getValue() -> " + o.getValue() );

        Outer.Inner i = o.newInner();
        System.out.println( "i.getValue() -> " + i.getValue() );

        System.out.println( "o.getValue() -> " + o.getValue() );
    }
}
```

Accessing Members of the Outer Object

```
// > java Test
// o.getValue() -> 99
// i.getValue() -> 100
// o.getValue() -> 99
```

Advanced (optional) example

```
public class A {
    private int value = 99;
    public class B {
        private int value;
        public B() { this.value = A.this.value + 1; }
        public class C {
            private int value;
            public C() {
                this.value = B.this.value + 1;
            }
            public int getValue() {
                System.out.println( "A.this.value = " + A.this.value );
                System.out.println( "B.this.value = " + B.this.value );
                System.out.println( "this.value = " + this.value );
                return value;
            }
        }
    }
}
```

```
class D {
    public static void main( String[] args ) {
        A.B.C abc = new A().new B().new C();
        System.out.println( "abc.getValue() -> " + abc.getValue() );
    }
}


// > java D
// A.this.value = 99
// B.this.value = 100
// this.value = 101
// abc.getValue() -> 101
```