

# Java Threads

---

Università di Modena e Reggio Emilia

*Prof. Nicola Bicocchi ([nicola.bicocchi@unimore.it](mailto:nicola.bicocchi@unimore.it))*



# Processes

---

- In operating systems (OS), a process is an instance of a running application
- A process has its own private address space, code, data, opened files, PID, etc..
- Processes do not share memory (separate address spaces), thus they have to communicate through IPC mechanisms offered by the OS (i.e., pipes, signals)
- A process might contain one or more threads running within the context of the process



# Threads

---

- Threads are sometimes called **lightweight processes**. Like processes, **each thread has its own stack, program counter, and local variables**
- However, **threads within the same process are less insulated from each other** than processes. They share the same address space and, consequently, can share variables and objects
- Sharing variables is a simple and fast way threads use for communicating but frequently causes bugs unseen in single-thread programs
- OOP principle of *separation of concerns* can be broken!

# Threads

---

- Every Java program must have at least one thread, the `main()` thread. When a program starts running, the JVM creates a new thread and calls the `main()` method
- There are other threads created by the JVM that users usually don't interact with explicitly (e.g., garbage collector)



# JVM and Operating System

---

- A multitasking operating system assigns CPU time (slices) to processes/threads via a kernel component called *scheduler*
- Preemption is defined as the ability of the scheduler to suspend a process/thread before the end of its execution (i.e., when its time slice is over)
- Small time-slices (5-20ms) provide the illusion of parallelism of different processes/threads (on multi-core machines it is a partial illusion)

# JVM and Operating System

---

- The JVM is a process and gets the CPU as assigned by the OS's scheduler
- Java is a specification with many different implementations\*
  - Some JVMs operate like a mini-OS and schedule their own threads
  - Most JVMs use the OS scheduler (a Java thread is actually mapped to a system thread)

\*[https://en.wikipedia.org/wiki/List\\_of\\_Java\\_virtual\\_machines](https://en.wikipedia.org/wiki/List_of_Java_virtual_machines)



# Why threads ?

---

- There are many reasons to use threads in your Java programs. If you use Android, Swing, JavaFX, Servlets, RMI, JavaBeans (EJB) you may already be using threads without realizing it.
- Main reasons for using threads:
  - make the UI more responsive
  - take advantage of multiprocessor systems
  - perform asynchronous operations or background processing



# Why threads ?

---

- Imagine a stock-broker application with three key capabilities:
  - Download stock prices
  - Store stock prices into a database
  - Short-term analysis (1 hour) for buy/sell signals



# Single-threaded scenario

---

- In a single-threaded runtime environment, actions execute one after another. An action happens only when the previous one is completed.
- If the download takes 10 minutes, should the user enjoy an unresponsive UI?
- If short-term analysis takes additional 10 minute the result may come too late. Prices could already have changed significantly!



# Multi-threaded scenario

---

- In a multi-threaded scenario
  - the **download can execute in background** (i.e. in another thread)
  - the **analysis, too, can execute in background** (i.e. in another thread) and eventually use multiple CPU cores for saving time
  - the user can enjoy a responsive GUI while awaiting for eventual notifications (buy/sell signals)



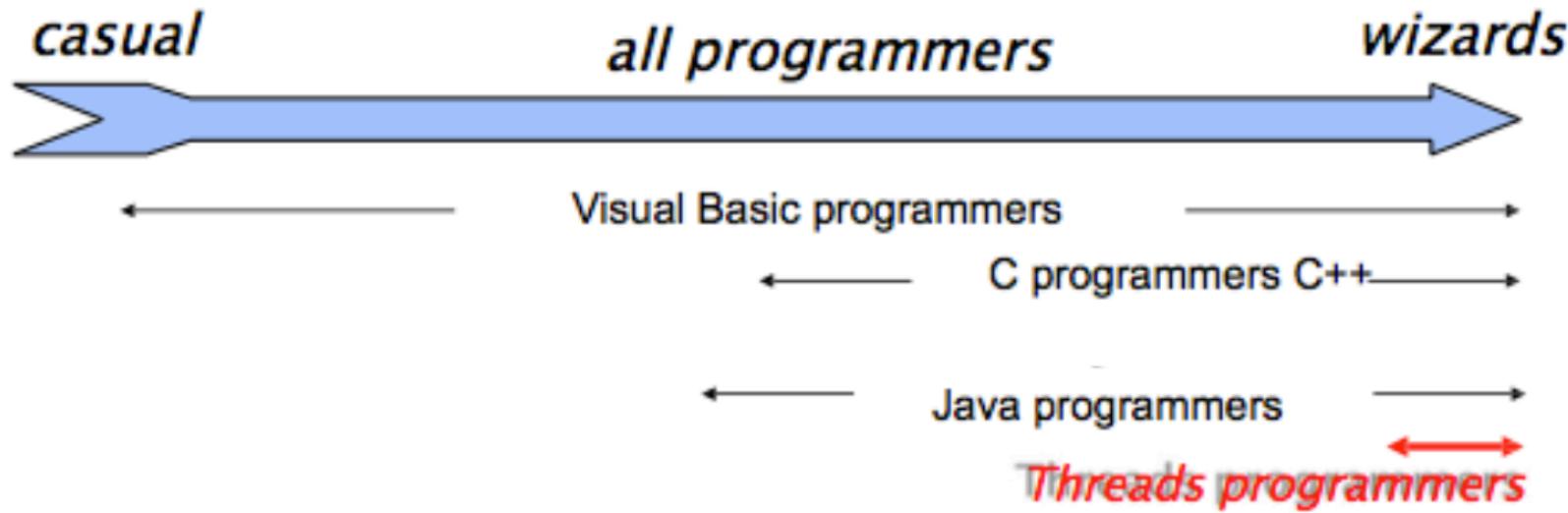
# The good

---

- Enable parallelism
- Lighter than processes for both
  - Creation(i.e., fork())
  - Communication (i.e., r/w pipes ...)

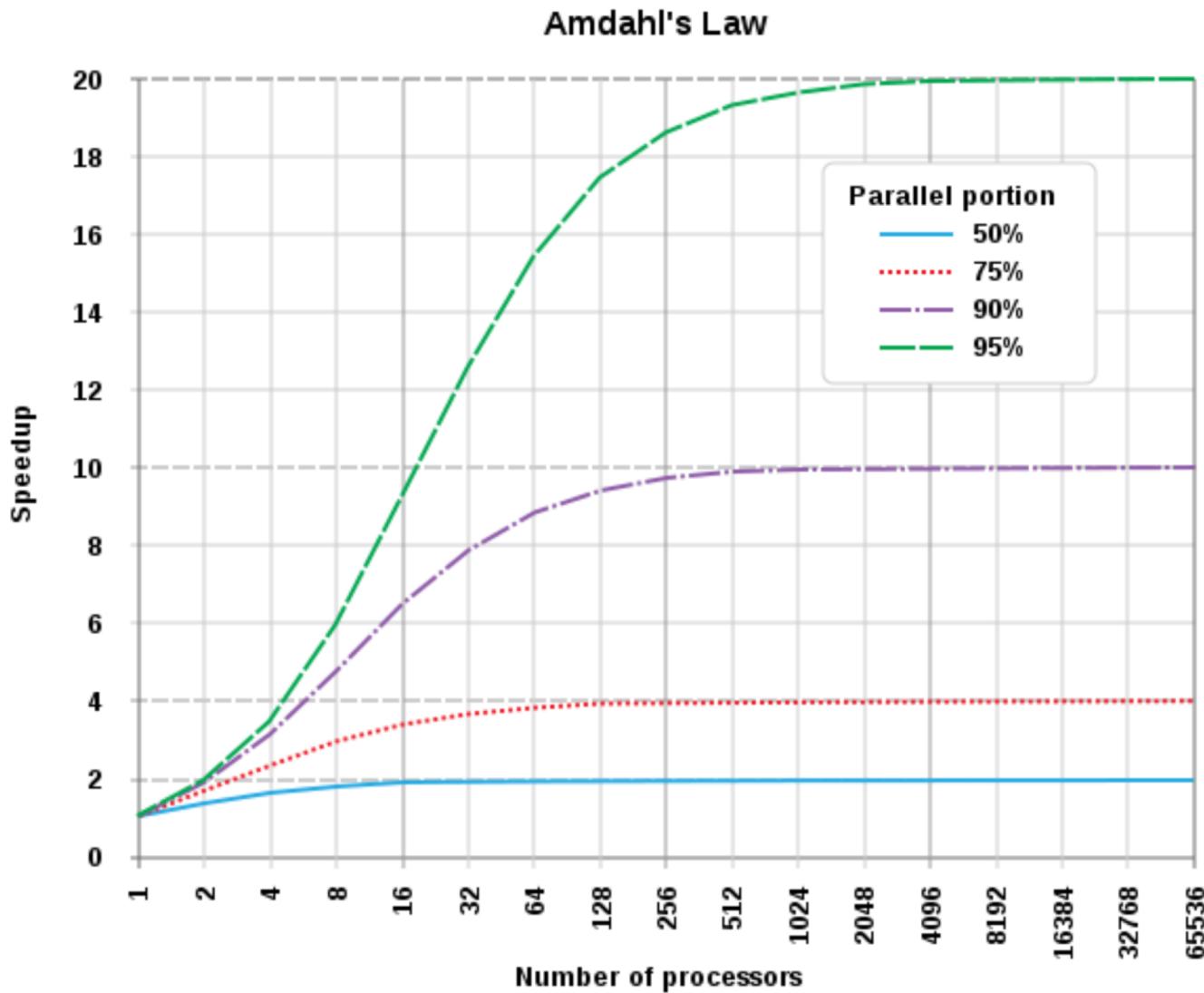


# The bad



- Hard for most programmers
- Even for experts, development is often painful
- **Threads break abstraction:** can't design modules independently.

# The important!



# Creating a Process

---

- In Java it is not possible to explicitly call the syscall fork() as in C. Syscalls fork() and exec() can be jointly via the Process Class
- Methods of Process Class also allow developers to acquire standard input, output, error, and exit value

```
/* Process p = Runtime.getRuntime().exec("/bin/ls -al /"); */  
Process p = (new ProcessBuilder("/bin/ls", "-al", "/")).start();
```

```
BufferedReader in = new BufferedReader(new  
    InputStreamReader(p.getInputStream()));  
while ((line = in.readLine()) != null) {  
    System.out.println(line);  
}  
in.close();  
System.out.println(p.exitValue());
```



# Creating a Thread

---

- Threads can be created by
  - extending the `Thread` class and overriding its `run()` method
  - passing to the `Thread` class constructor an object implementing the `java.lang.Runnable` Interface
- It is legal to create many threads using the same `Runnable` object as the target



# Creating a Thread

---

## Extending Thread

```
Class T extends Thread {  
    public void run() {  
        //code here  
    }  
}  
T t = new T();  
t.start();
```

## Implementing Runnable interface

```
Class R implements Runnable {  
    public void run() {  
        //code here  
    }  
}  
Thread t = new Thread(new R());  
t.start();
```



# Creating a Thread

---

```
class Counter implements Runnable {  
    public void run() {  
        for(int i=0; i<10; i++)  
            System.out.println(Thread.currentThread().getName() + i);  
    }  
}  
  
public class Runner{  
    public static void main(String[] args) {  
        Thread t1 = new Thread(new Counter(), "T_A");  
        Thread t2 = new Thread(new Counter(), "T_B");  
        Thread t3 = new Thread(new Counter(), "T_C");  
        t1.start(); t2.start(); t3.start();  
    }  
}
```



# Starting a Thread

---

- When a Thread object is created, **it does not start executing until its start() method is invoked.**
- When a Thread object exists but hasn't been started, **it is in the *new state* and is not considered alive.**
- Method start() can be called on a Thread object only once. If start() is called more than once on same object, it will throw a **RuntimeException**



# Starting a Thread

---

- It is not guaranteed that threads will start running in the same order in which their start() methods have been called.
- It is not guaranteed that a thread keeps executing until it's done (it is not guaranteed that its loop completes before another thread begins)
- **Nothing is guaranteed except:**
  - Each thread will start and will run to completion after acquiring the CPU a finite number of times

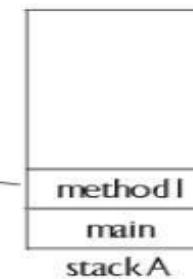


# Starting a Thread

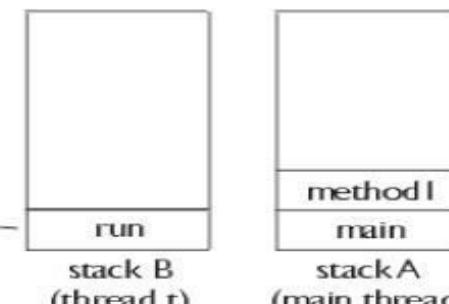
```
public static void main(String [] args) {  
    // running  
    // some code  
    // in main()  
    method1();  
    // running  
    // more code  
}  
  
void method1() {  
    Runnable r = new MyRunnable();  
    Thread t = new Thread(r);  
    t.start();  
    // do more stuff  
}
```



1) main() begins



2) main() invokes method1()



stackA  
(main thread)

3) method1() starts a new thread

# Thread priority

---

- By default, a thread gets the priority of the thread creating it.
- Priority values are defined between 1 and 10

Thread.MIN\_PRIORITY (== 1)

Thread.NORM\_PRIORITY (== 5)

Thread.MAX\_PRIORITY (== 10)

- Priority can be set

```
Thread t = new Thread(new Runnable());  
t.setPriority(Thread.MAX_PRIORITY);  
t.start();
```



# JVM scheduling policy

---

- A thread always runs with a priority number
- The scheduler in most JVMs uses **time-sliced, preemptive, priority-based** scheduling
- JVM specification does not require a VM to implement a time-slicing scheduler !
  - some JVM may use a scheduler that lets one thread stay running until it completes its run() method



# Checking JVM scheduler

---

```
public class CheckPreemption implements Runnable {  
    @Override  
    public void run() {  
        while (true) {  
            System.out.println(Thread.currentThread().getName());  
        }  
    }  
  
    public static void main(String argv[]) {  
        CheckPreemption c = new CheckPreemption();  
        new Thread(c, "To be").start();  
        new Thread(c, "Not to be").start();  
    }  
}
```

- If the scheduler is non-preemptive the first thread chosen runs forever
- If the scheduler is preemptive both threads randomly alternate



# Terminating a Thread

---

- If the parent thread terminates, all of its child threads terminate as well.
- Child threads share resources with the parent thread, including variables. When the parent thread terminates, the child threads will not be able to access to those resources that the parent thread owns.
- Thus, if the parent thread terminates earlier than its own child threads, synchronization mechanisms are required.



# Terminating a Thread

---

```
/* start children threads */
producer.start();
consumer.start();

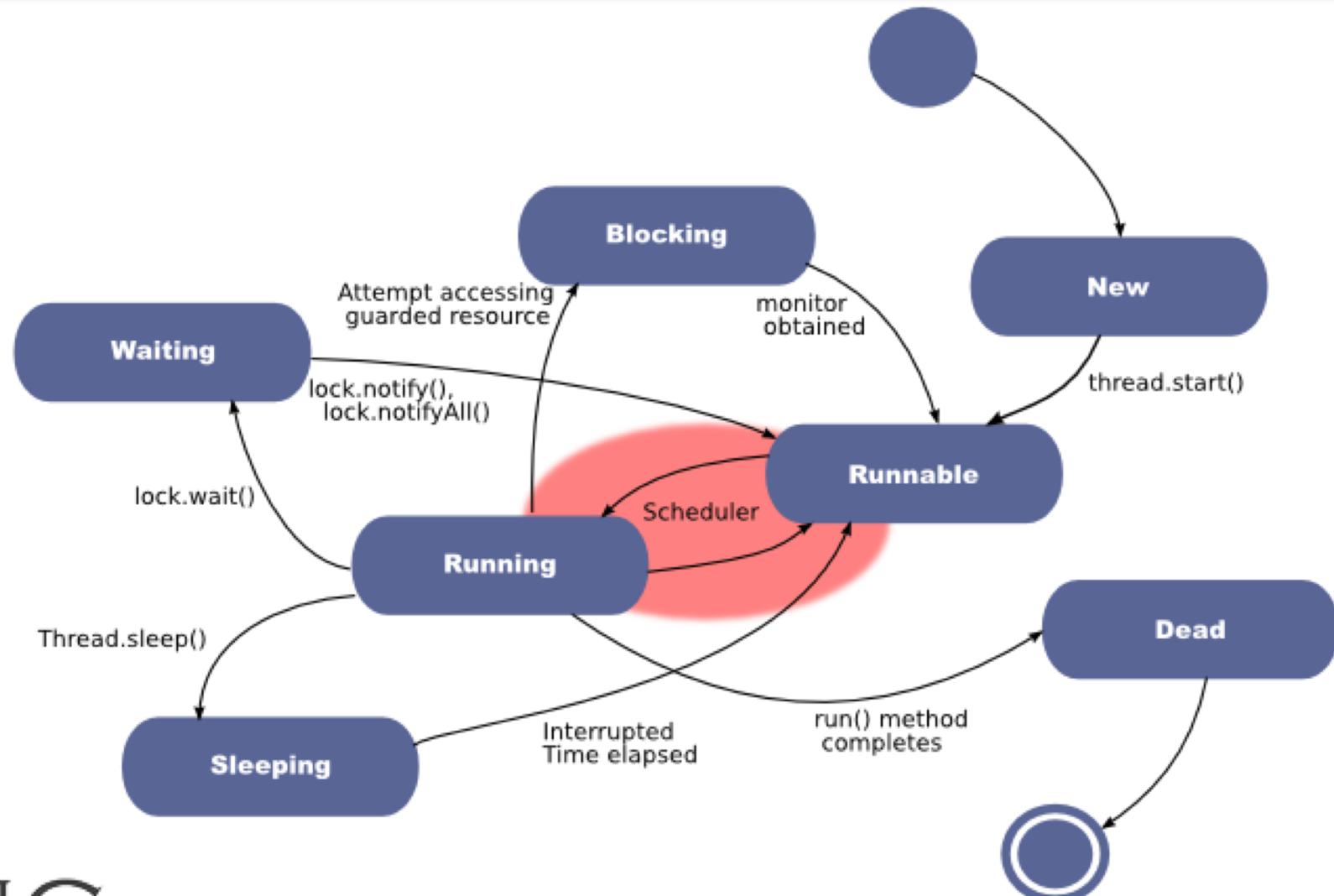
/* wait 1/10 of second */
Thread.sleep(100);

/* gracefully shut down children threads */
producer.running = false;
consumer.running = false;

/* wait for children before exit */
producer.join();
consumer.join();
```

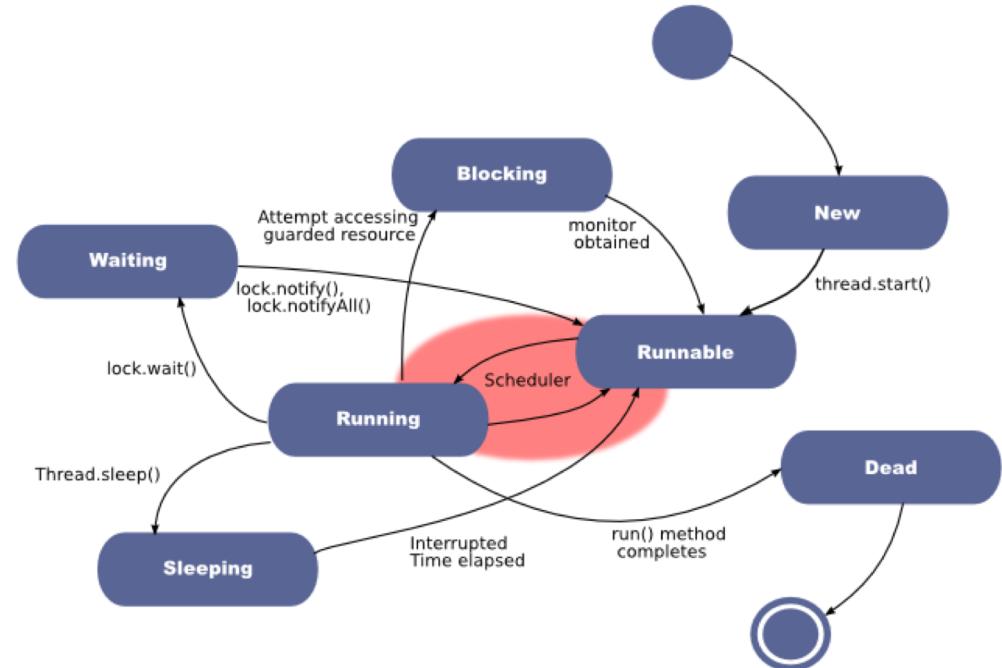


# Thread states



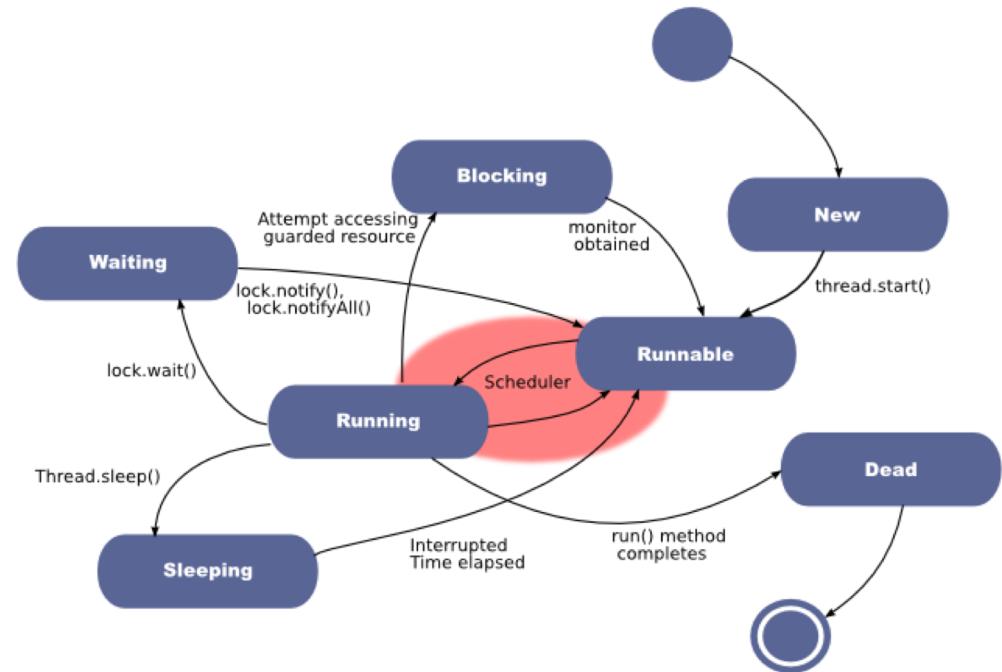
# Thread state: Running

- The thread has been selected (from the runnable pool) to be **the currently executing thread**.
- Transitions:
  - Thread.yield() (**RUNNABLE**)
  - Thread.sleep() (**SLEEPING**)
  - Thread.join() (**WAITING**)
  - Can acquire a resource but there is no work to do (**WAITING**)
  - Awaiting for a resource or I/O completion (**BLOCKING**)
  - End of run() method (**DEAD**)



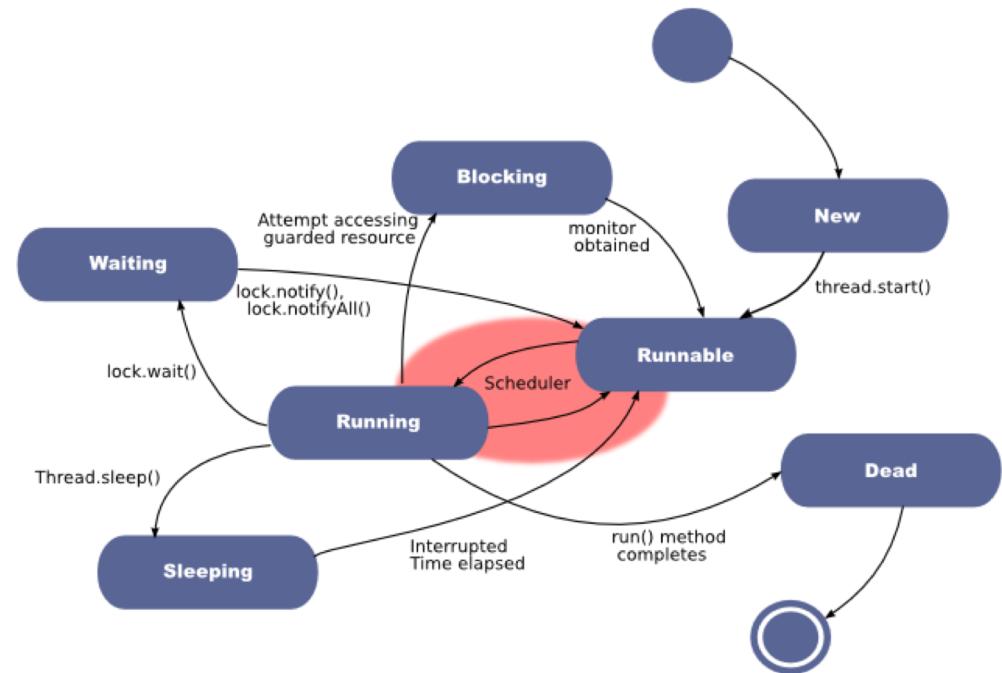
# Thread state: Runnable

- A thread which is eligible to run, but the scheduler has not selected it to be the running thread
- A thread first enters the runnable state when the start() method is invoked
- A thread can also return to the runnable state after either the running, blocked, waiting, or sleeping state



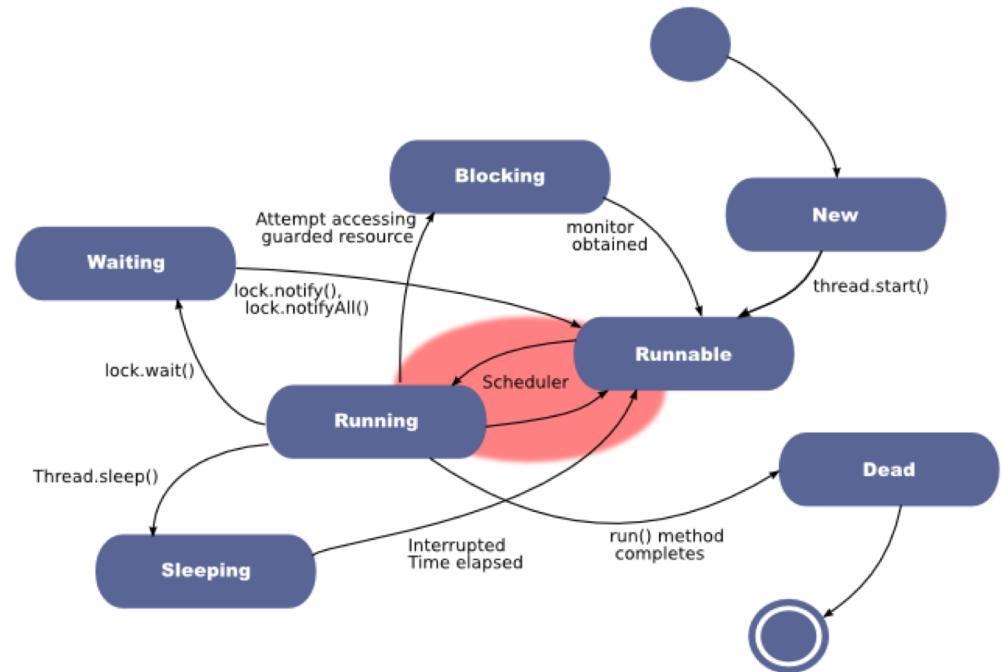
# Thread state: Waiting

- A thread that can acquire a resource but there is no work to do
- A thread awaiting for its children (`Thread.join()`)
- The thread calls `object.wait()` and waits for another thread to call `object.notify()` or `object.notifyAll()`



# Thread state: Blocking

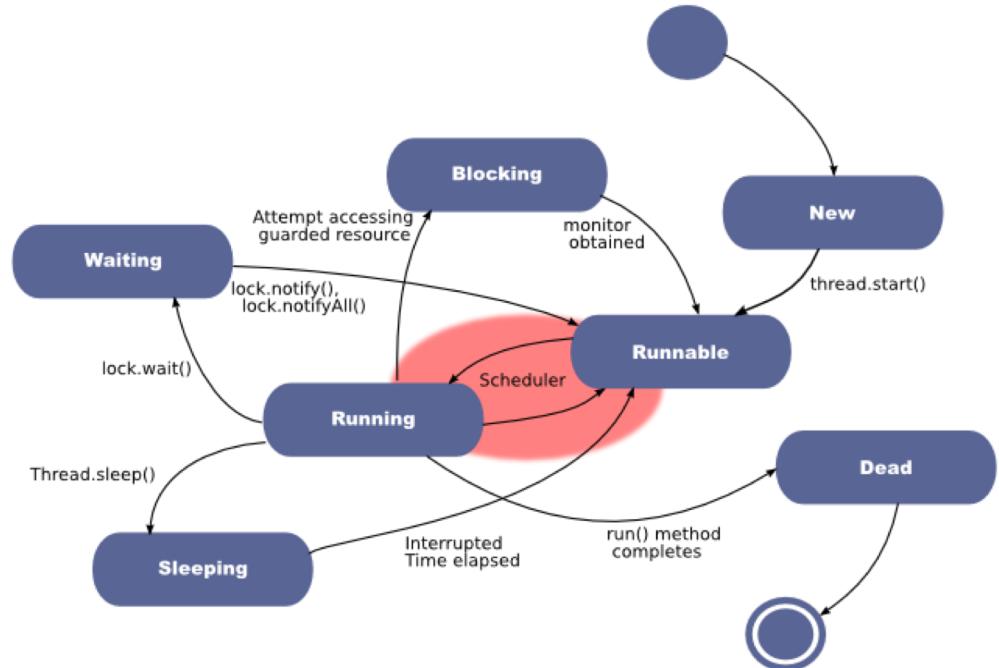
- A thread waiting for a resource
- For example, awaiting for the completion of I/O operations or for an object's lock



# Thread state: Sleeping

- A thread which is sleeping after an explicit call to Thread.sleep()
- Back to Runnable state when the thread wakes up because its sleep time has expired.

```
try {  
    Thread.sleep(1000); //  
one second  
} catch  
(InterruptedException ex) {  
}
```



# Leaving the running state (explicitly)

---

- There are 3 ways for a thread to do it:
  - `sleep()`: the currently running thread stops executing for at least the specified sleep duration
  - `yield()`: the currently running thread moves back to runnable to give room to other threads
  - `join()`: the currently running thread stop executing until the thread it joins completes



# sleep()

---

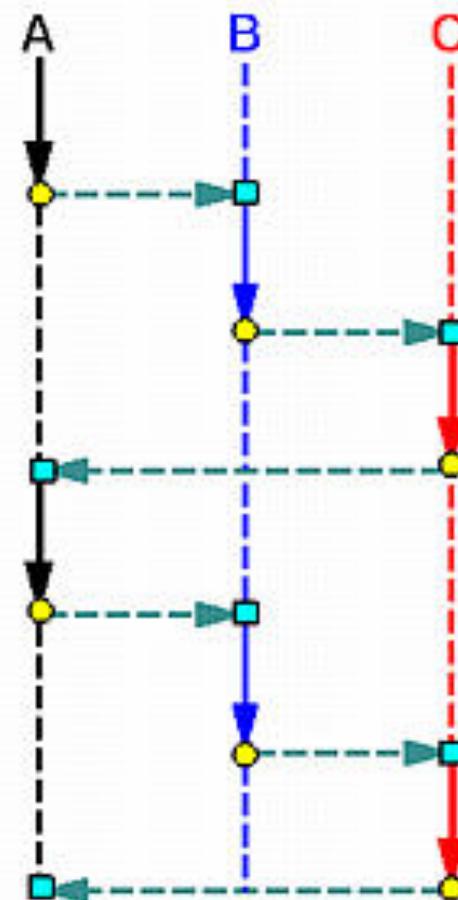
```
try {  
    // Sleep for 1 second  
    Thread.sleep(1000);  
} catch (InterruptedException ex) {  
    //  
}
```



# yield()

- The method `yield()` make the currently running thread back to Runnable state
  - It allows other threads to get their turn
  - However, it might have no effect at all. In fact, there's no guarantee the yielding thread won't be scheduled again for execution.

● **thread yield**  
● **execution resume**



# yield()

---

- Code is less dependent from the scheduler type, because threads release CPU when needed.
- Frequently used when computation is not possible (no work to do) in a specific time slice.



# join()

---

- A thread can execute a thread join to wait until the other thread terminates. In general, thread join is for a parent to join with one of its child threads. Thread join has the following activities, assuming that a parent thread P wants to join with one of its child threads C
  - When P executes Thread.join() in order to join with C, which is still running, P is suspended until C terminates. Once C terminates, P resumes
  - When P executes Thread.join() and C has already terminated, P continues as if no such thread join has ever executed

# join()

---

The join() method lets one thread "join onto the end" of another thread.

```
Thread t = new Thread();
t.start();
t.join();
```

Caller move to Waiting state and it will be Runnable when thread t is dead. A timeout can be set to wait for a thread's end

```
t.join(5000);
// wait t for 5 seconds: if t is not finished
// then current thread is Runnable again
```



# A word of advice

---

- Some methods may look like they tell another thread to block, but they don't.
- If `t` is a thread object reference, you can write something like: `t.sleep()` or `t.yield()`
- However, `they are static methods of Thread`:
  - they don't affect the instance `t` !!!
  - instead they affect the thread which is currently in execution
- Using an instance variable to access a static method is error-prone!

