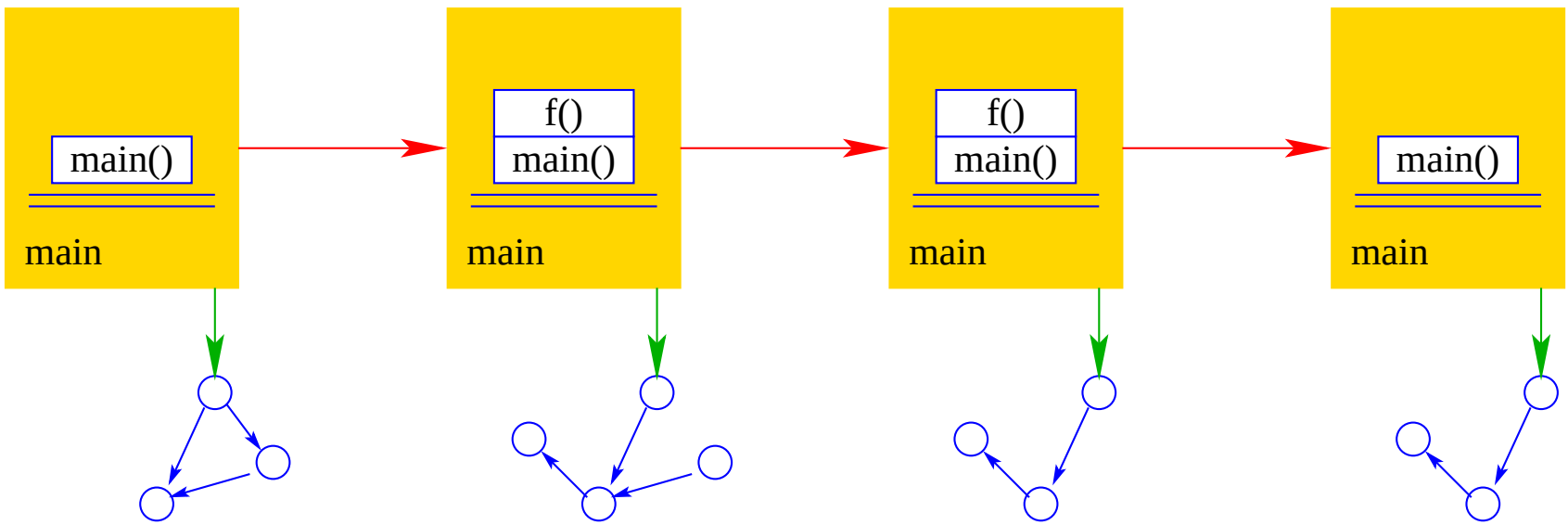
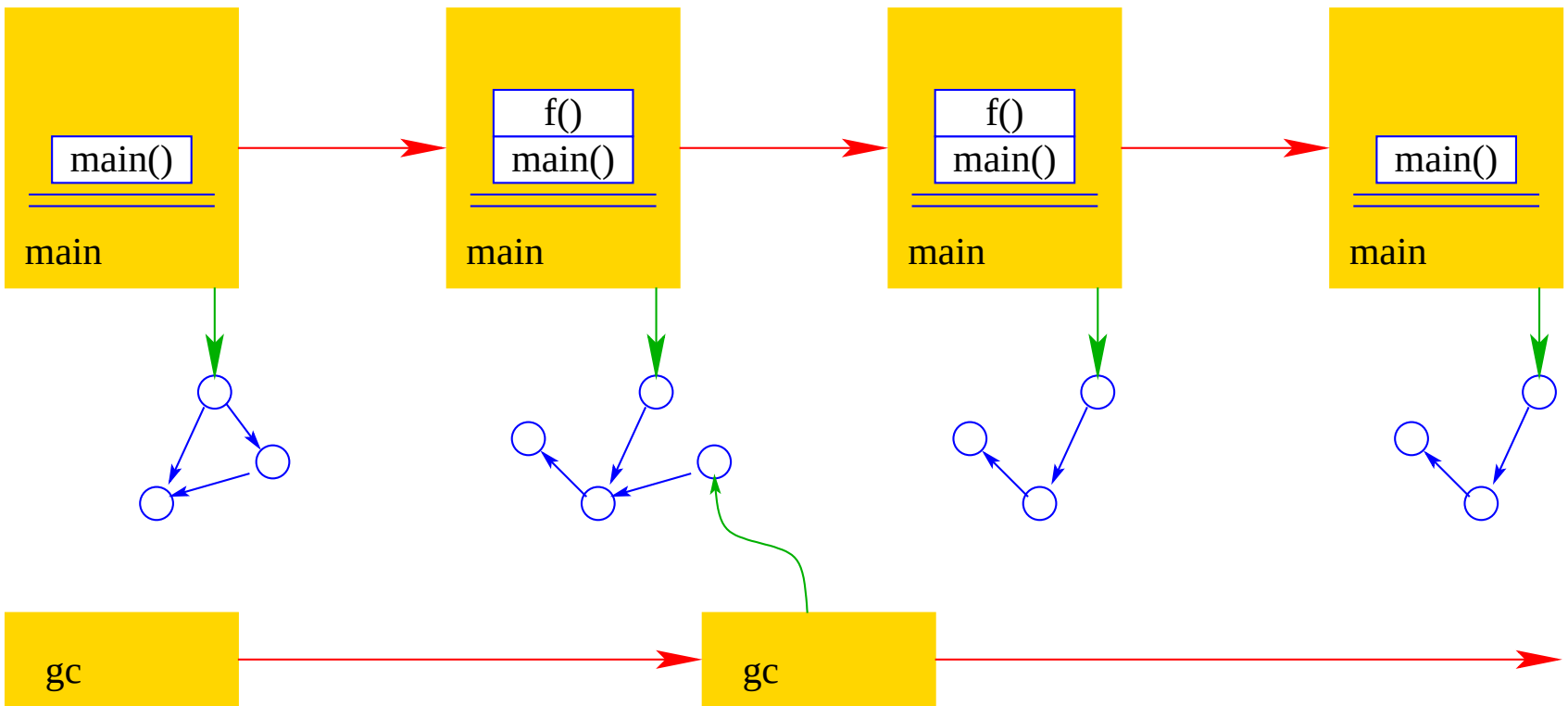
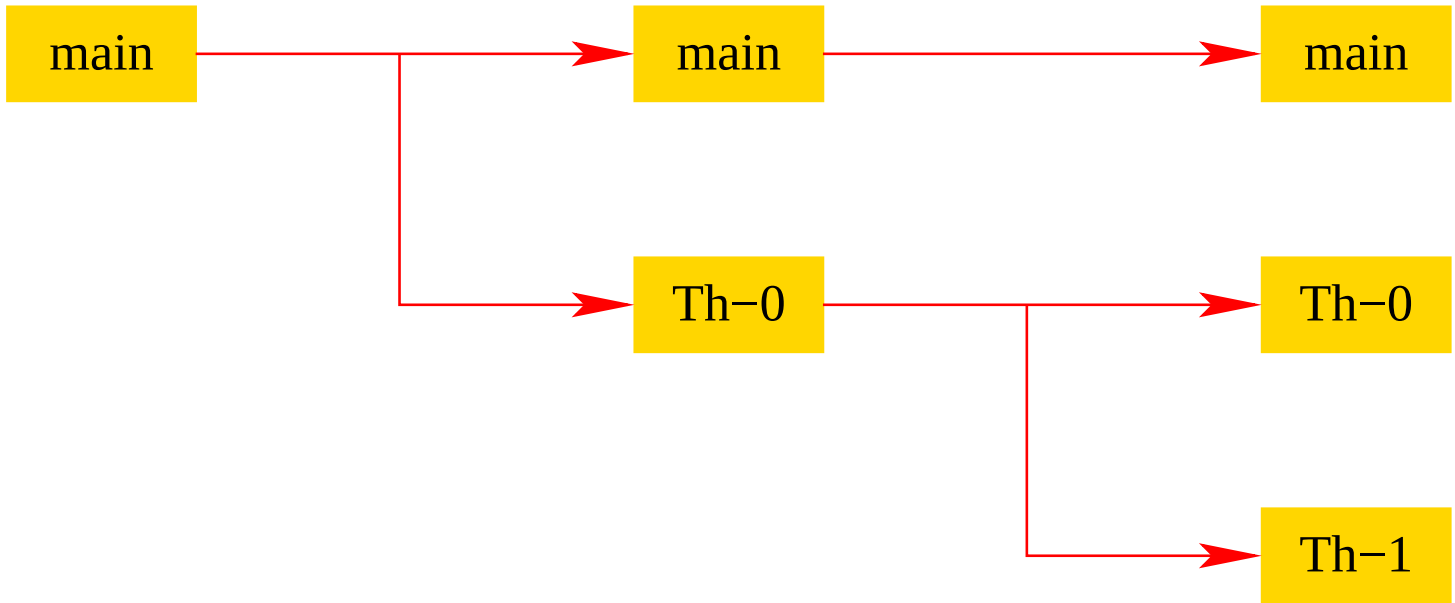
18 Threads

- Die Ausführung eines Java-Programms besteht in Wahrheit nicht aus einem, sondern mehreren parallel laufenden Threads.
- Ein Thread ist ein sequentieller Ausführungs-Strang.
- Der Aufruf eines Programms startet einen Thread `main`, der die Methode `main()` des Programms ausführt.
- Ein weiterer Thread, den das Laufzeitsystem parallel startet, ist die Garbage Collection.
- Die Garbage Collection soll mittlerweile nicht mehr erreichbare Objekte beseitigen und den von ihnen belegten Speicherplatz der weiteren Programm-Ausführung zur Verfügung stellen.


main()
main
f()
main()
main
f()
main()
main
main()
main


main()
main
f()
main()
main
f()
main()
main
main()
main
gc
gc

- Mehrere Threads sind auch nützlich, um
  - ... mehrere Eingabe-Quellen zu überwachen (z.B. Mouse-Klicks und Tastatur-Eingaben) ↑Graphik;
  - ... während der Blockierung einer Aufgabe etwas anderes Sinnvolles erledigen zu können;
  - ... die Rechenkraft mehrerer Prozessoren auszunutzen.
- Neue Threads können deshalb vom Programm selbst erzeugt und gestartet werden.
- Dazu stellt Java die Klasse `Thread` und das Interface `Runnable` bereit.


main
main
main
Th−0
Th−0
Th−1

Beispiel:

```
public class MyThread extends Thread {
   public void hello(String s) {
      System.out.println(s);
   }
   public void run() {
      hello("I'm running ...");
   }  // end of run()
   public static void main(String[] args) {
      MyThread t = new MyThread();
      t.start();
      System.out.println("Thread has been started ...");
   }  // end of main()
}     // end of class MyThread
```

- Neue Threads werden für Objekte aus (Unter-) Klassen der Klasse `Thread` angelegt.
- Jede Unterklasse von `Thread` sollte die Objekt-Methode `public void run();` implementieren.
- Ist t ein `Thread`-Objekt, dann bewirkt der Aufruf `t.start();` das folgende:

  1. ein neuer Thread wird initialisiert;
  2. die (parallele) Ausführung der Objekt-Methode `run()` für t wird angestoßen;
  3. die eigene Programm-Ausführung wird hinter dem Aufruf fortgesetzt.

Beispiel:

```
public class MyRunnable implements Runnable {
   public void hello(String s) {
      System.out.println(s);
   }
   public void run() {
      hello("I'm running ...");
   }  // end of run()
   public static void main(String[] args) {
      Thread t = new Thread(new MyRunnable());
      t.start();
      System.out.println("Thread has been started ...");
   }  // end of main()
}     // end of class MyRunnable
```

- Auch das Interface `Runnable` verlangt die Implementierung einer Objekt-Methode `public void run();`
- `public Thread(Runnable` obj`);` legt für ein `Runnable`-Objekt obj ein `Thread`-Objekt an.
- Ist `t` das `Thread`-Objekt für das `Runnable` obj, dann bewirkt der Aufruf `t.start();` das folgende:
  1. ein neuer Thread wird initialisiert;
  2. die (parallele) Ausführung der Objekt-Methode `run()` für obj wird angestoßen;
  3. die eigene Programm-Ausführung wird hinter dem Aufruf fortgesetzt.

Mögliche Ausführungen:

```
Thread has been started ...
I'm running ...
```

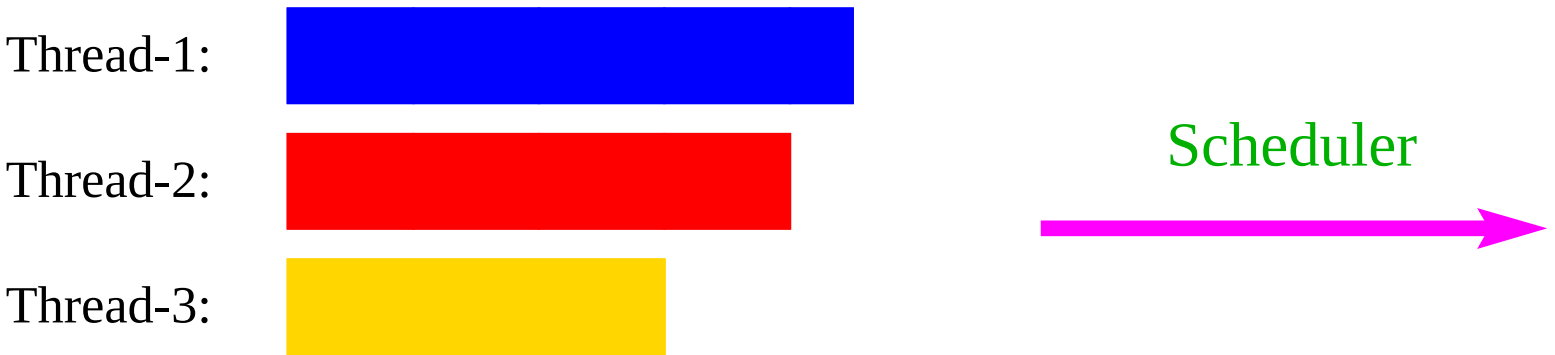
... oder:

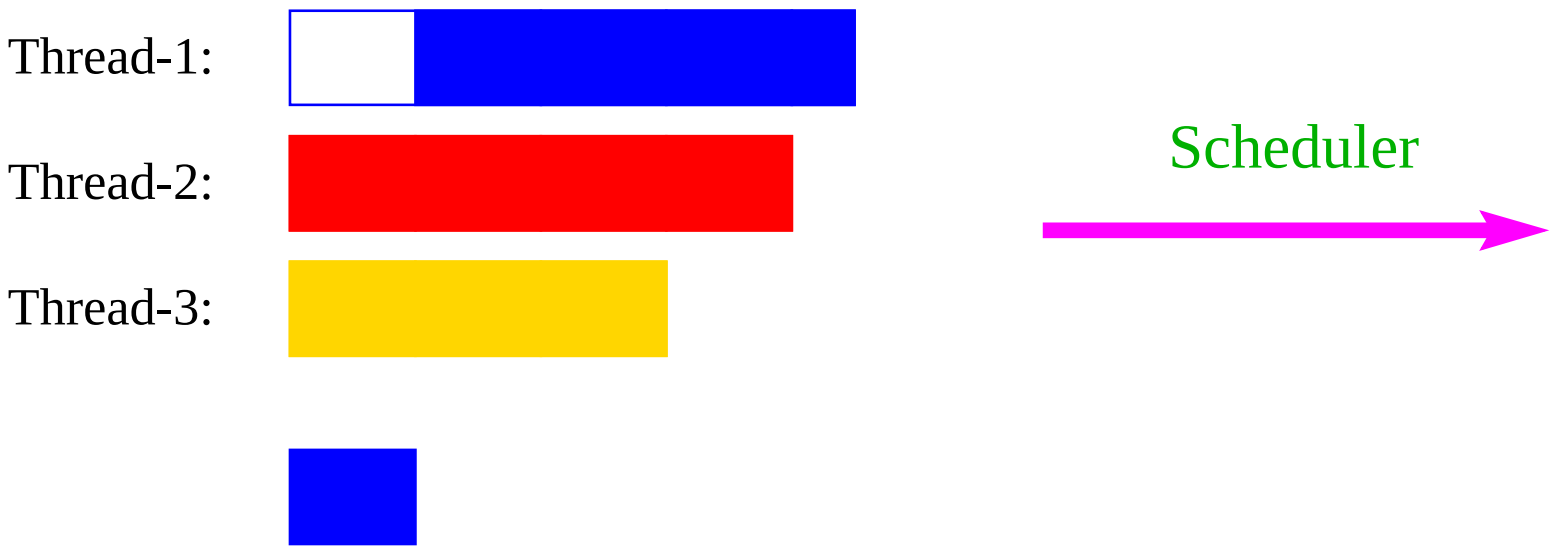
```
I'm running ...
Thread has been started ...
```

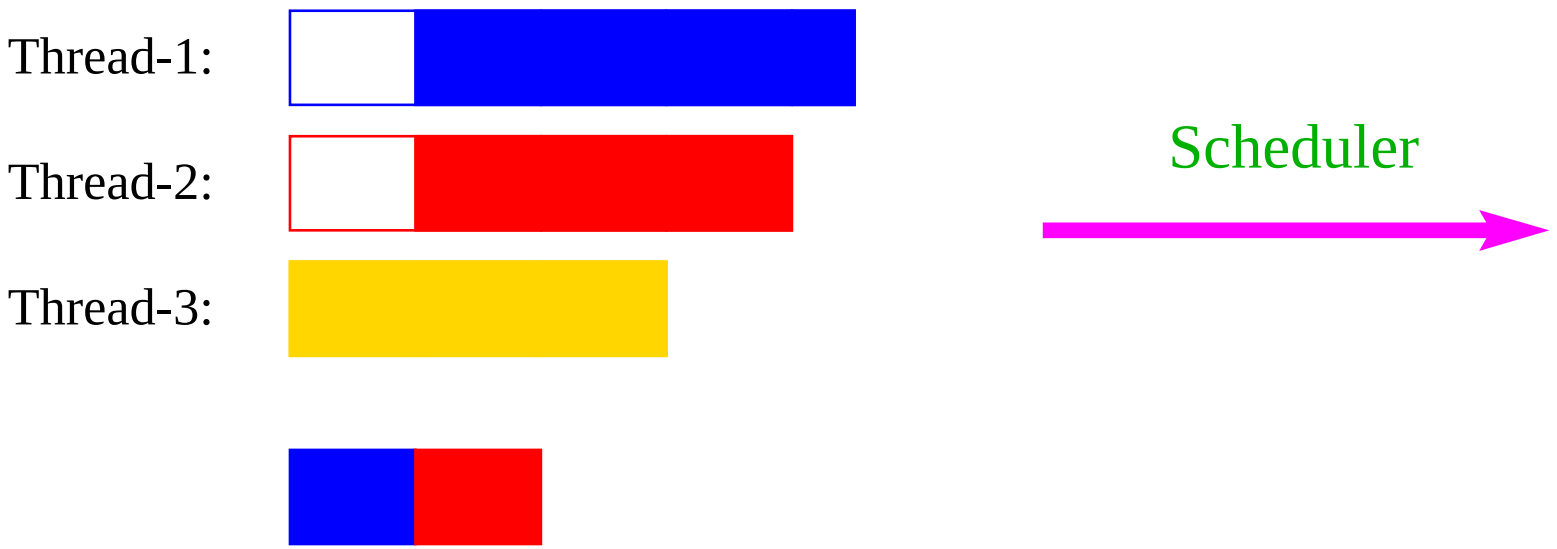
- Ein Thread kann nur eine Operation ausführen, wenn ihm ein Prozessor (CPU) zur Ausführung zugeteilt worden ist.
- Im Allgemeinen gibt es mehr Threads als CPUs.
- Der Scheduler verwaltet die verfügbaren CPUs und teilt sie den Threads zu.
- Bei verschiedenen Programm-Läufen kann diese Zuteilung verschieden aussehen!!!
- Es gibt verschiedene Politiken, nach denen sich Scheduler richten können ↑Betriebssysteme.

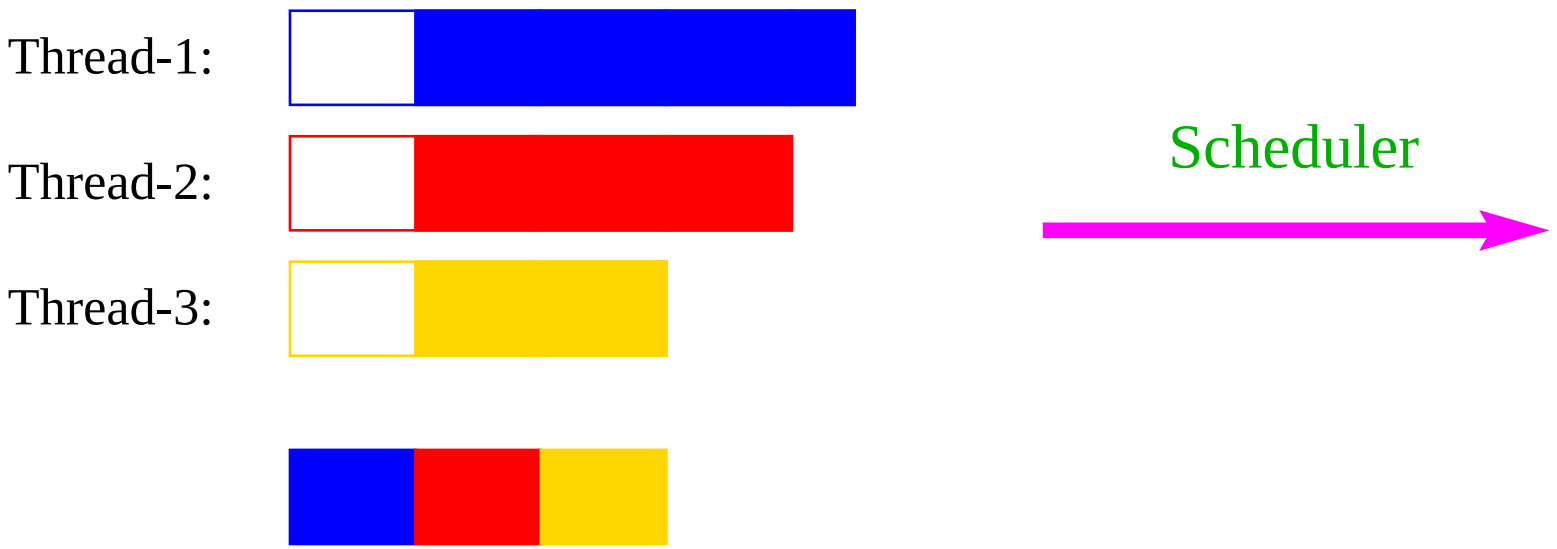
1. Zeitscheiben-Verfahren:

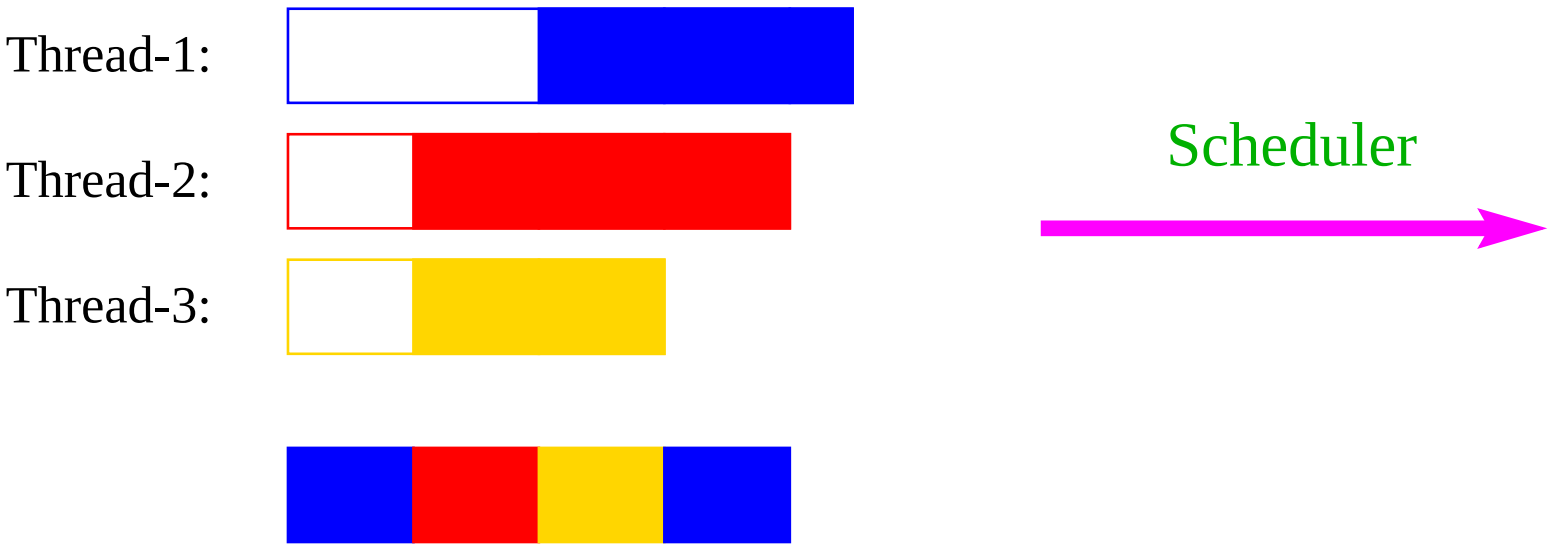
- Ein Thread erhält eine CPU nur für eine bestimmte Zeitspanne (Time Slice), in der er rechnen darf.
- Danach wird er unterbrochen. Dann darf ein anderer.

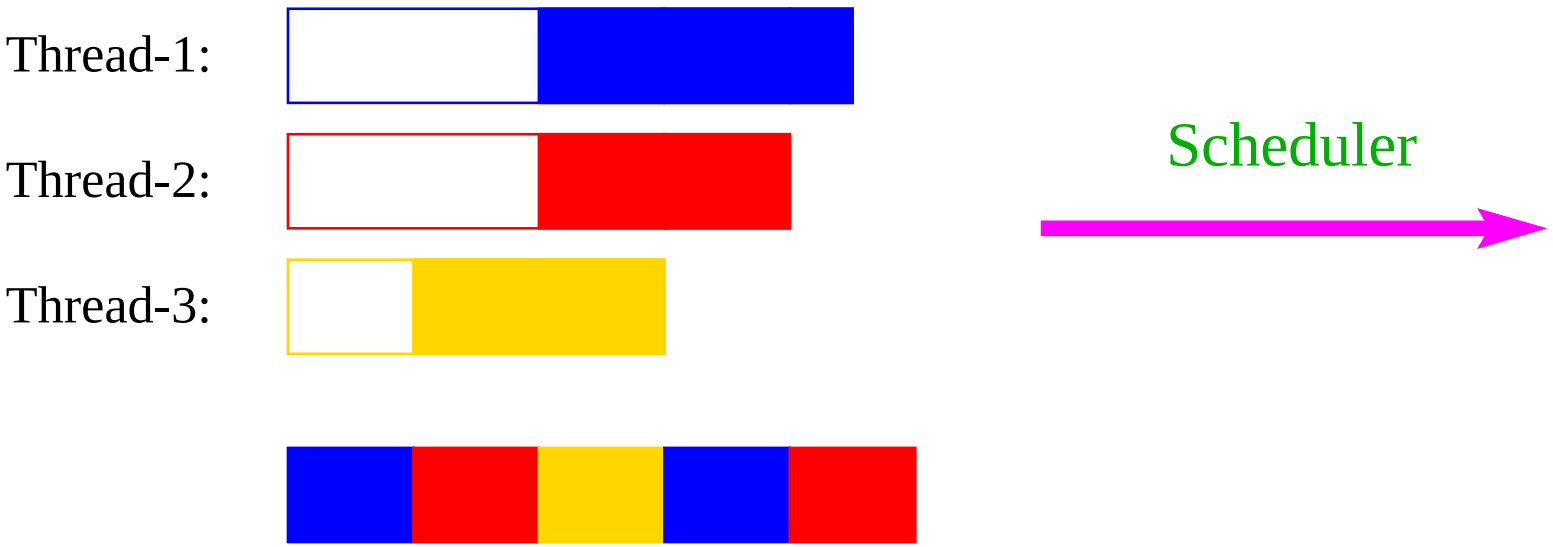

Thread-1:
Thread-2:
Thread-3:
Scheduler

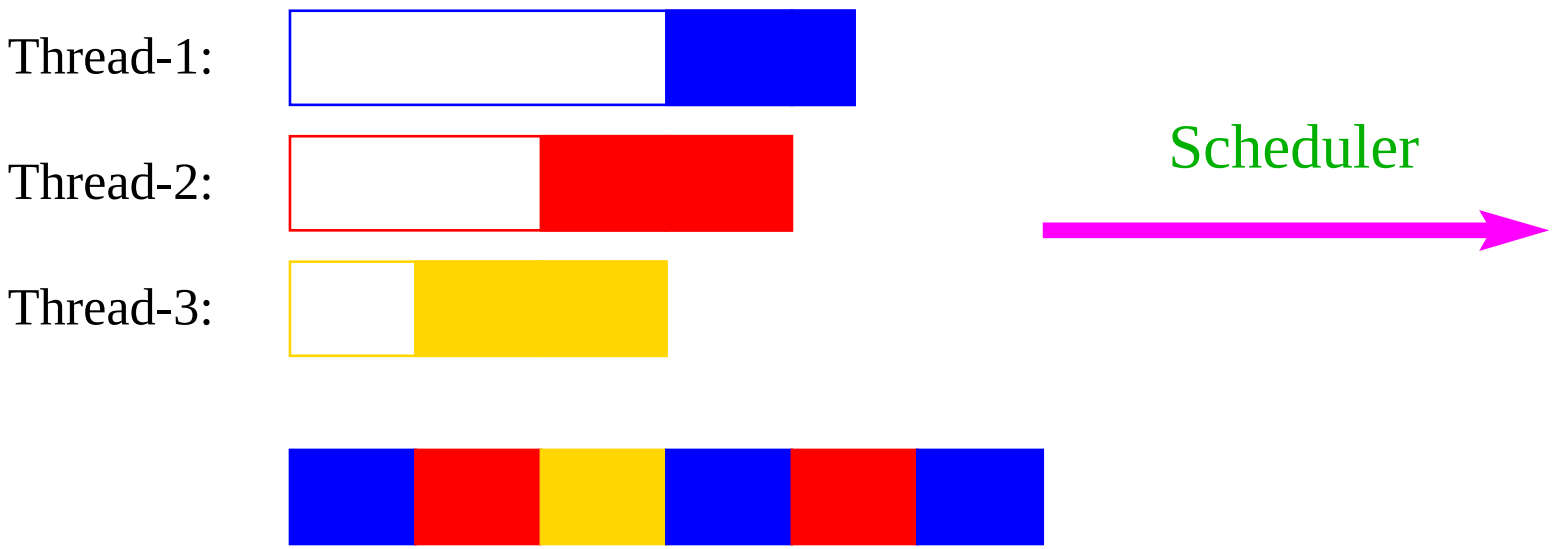

Thread-1:
Thread-2:
Thread-3:
Scheduler

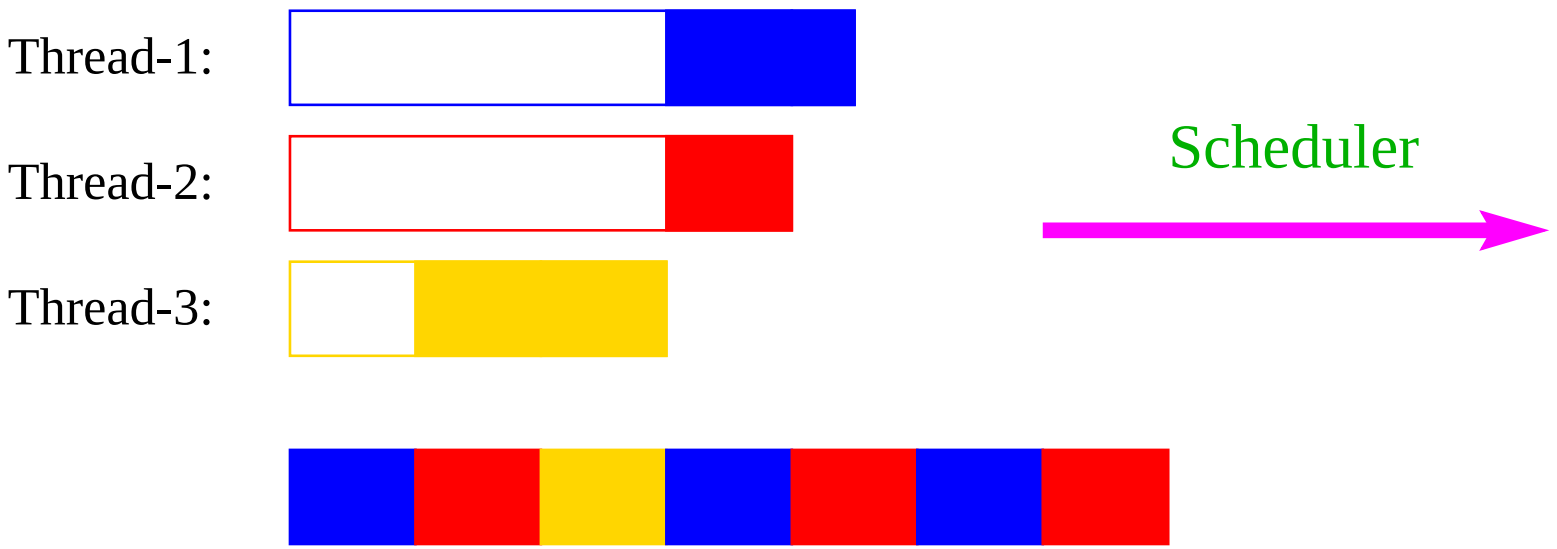

Thread-1:
Thread-2:
Thread-3:
Scheduler

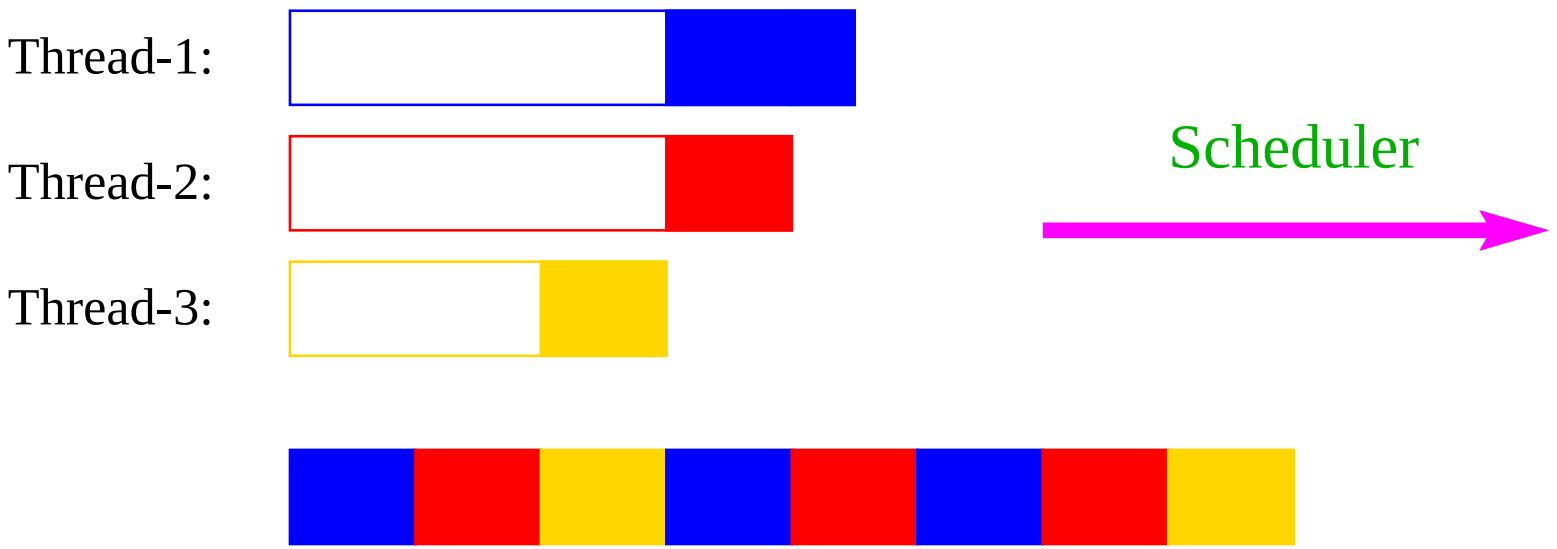

Thread-1:
Thread-2:
Thread-3:
Scheduler

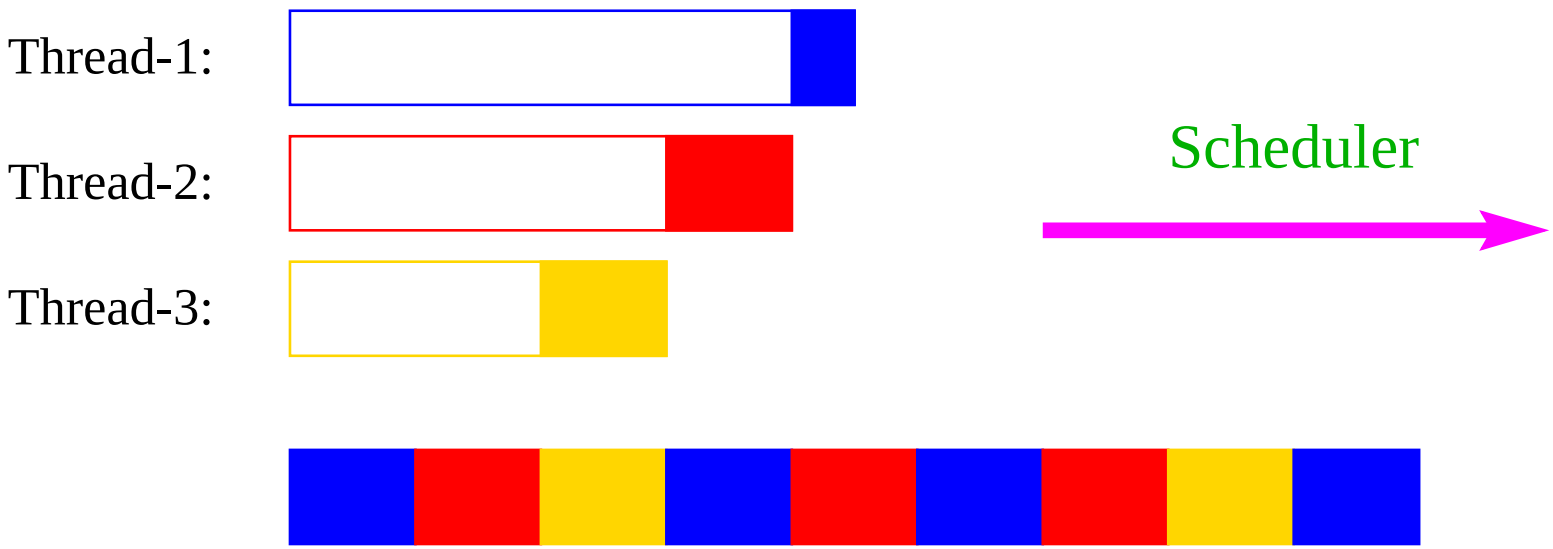

Thread-1:
Thread-2:
Thread-3:
Scheduler

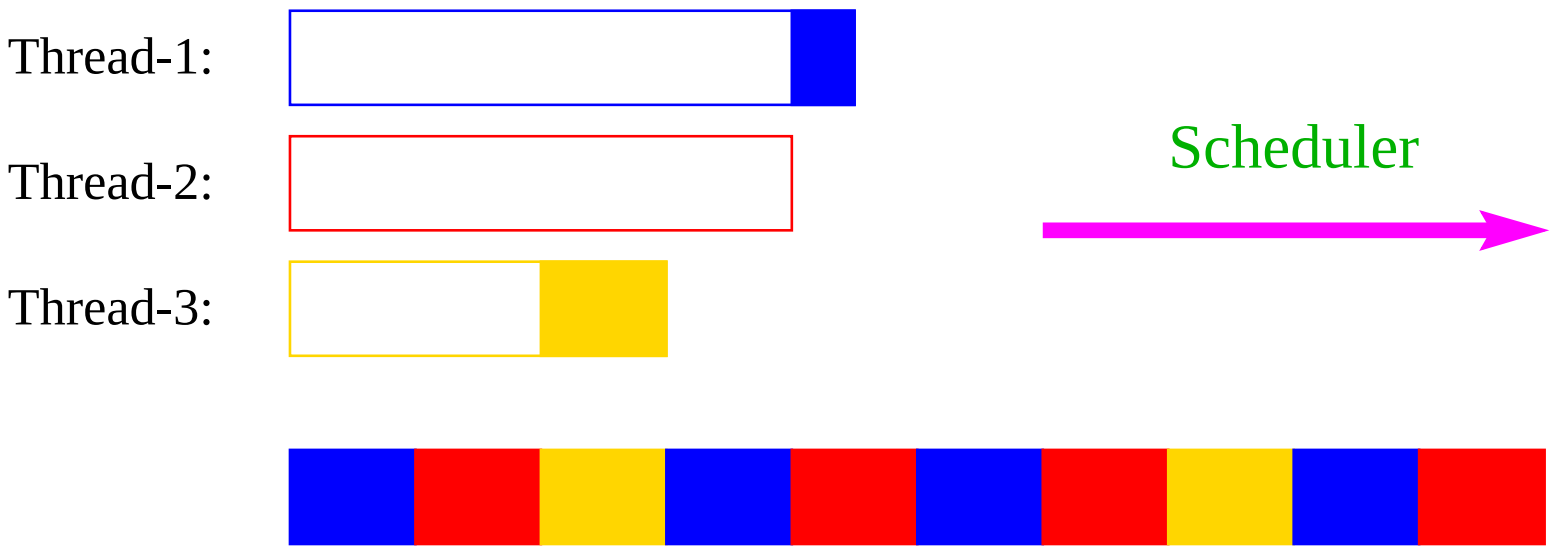

Thread-1:
Thread-2:
Thread-3:
Scheduler

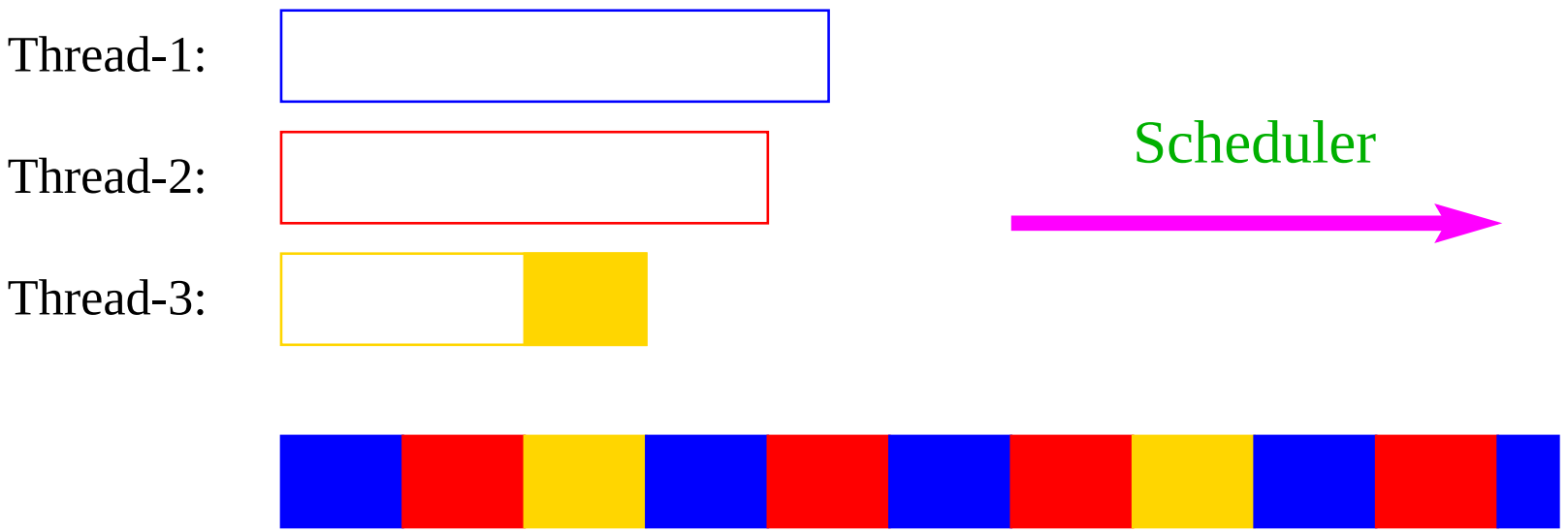

Thread-1:
Thread-2:
Thread-3:
Scheduler

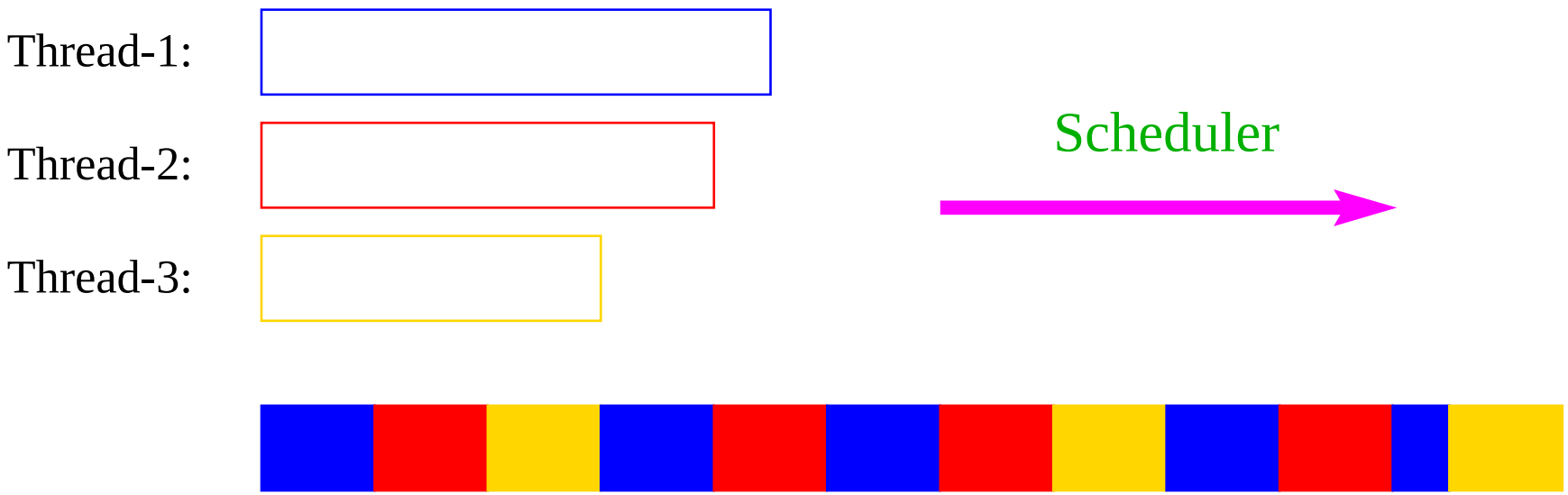

Thread-1:
Thread-2:
Thread-3:
Scheduler


Thread-1:
Thread-2:
Thread-3:
Scheduler


Thread-1:
Thread-2:
Thread-3:
Scheduler

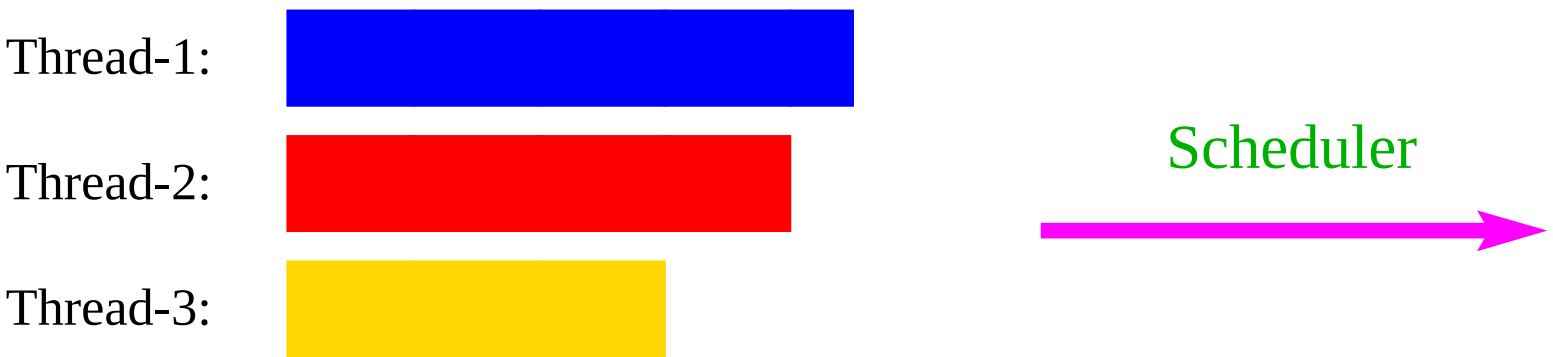

Thread-1:
Thread-2:
Thread-3:
Scheduler


Thread-1:
Thread-2:
Thread-3:
Scheduler


Thread-1:
Thread-2:
Thread-3:
Scheduler

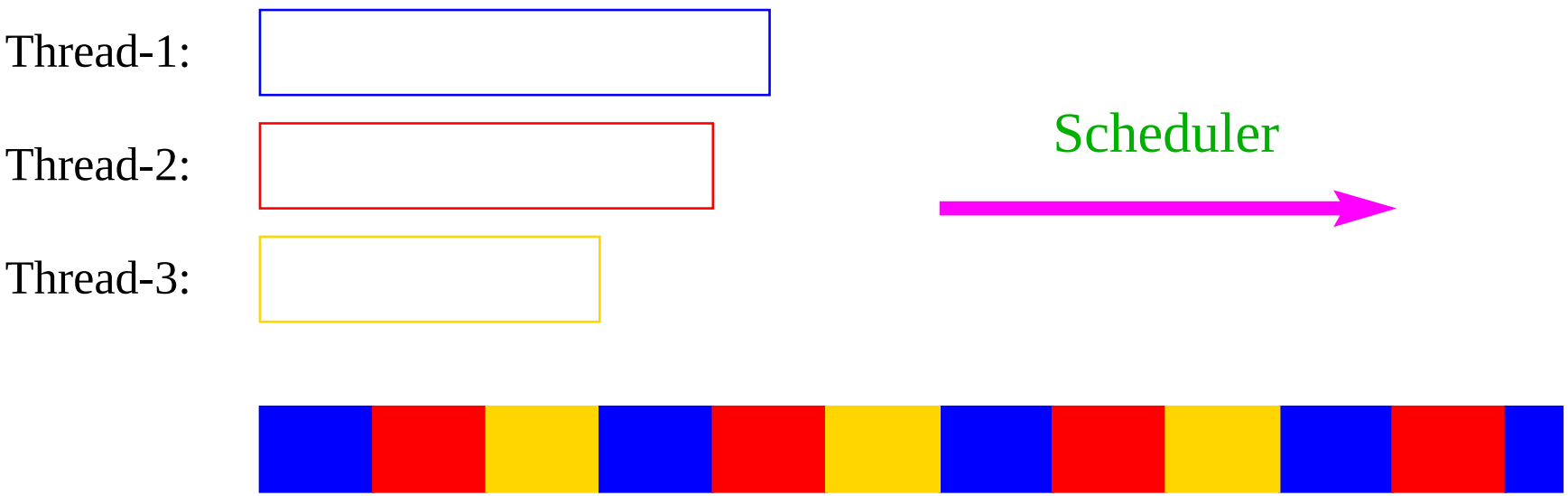
Achtung:

Eine andere Programm-Ausführung mag dagegen liefern:



Achtung:

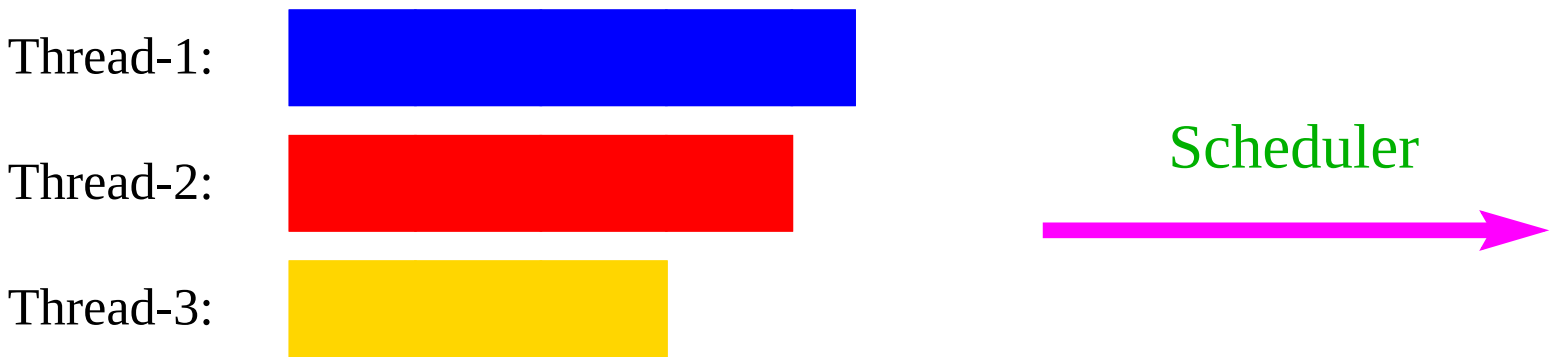
Eine andere Programm-Ausführung mag dagegen liefern:

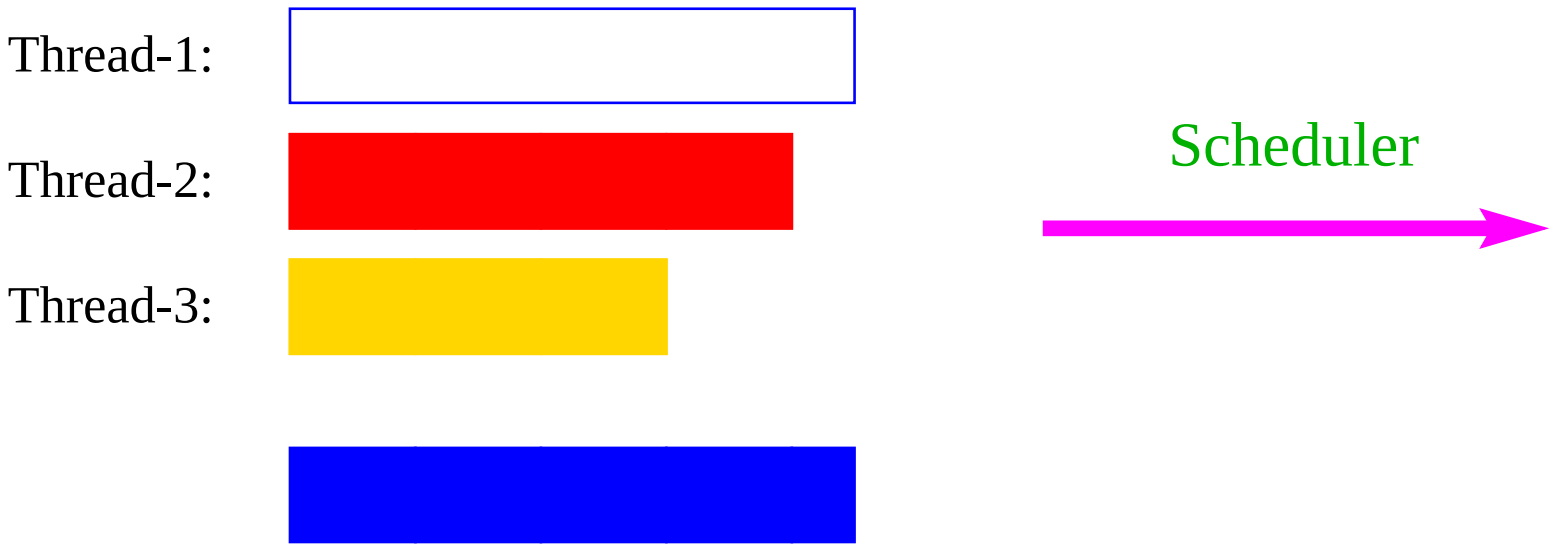


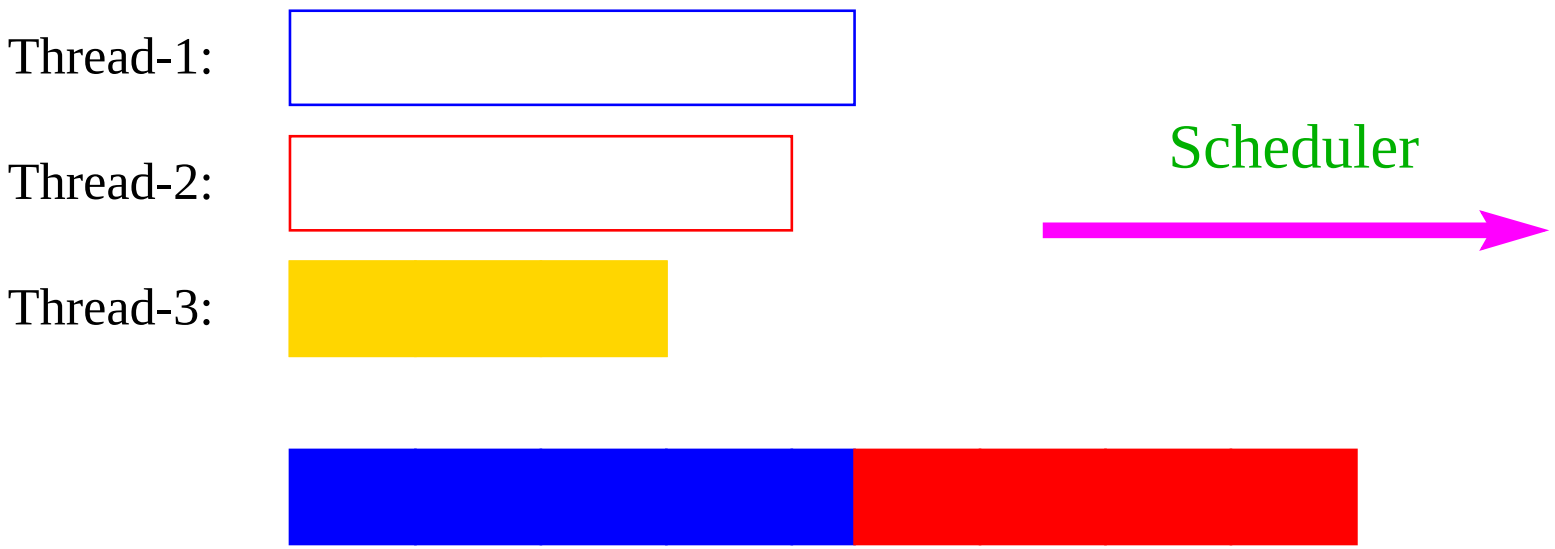
- Ein Zeitscheiben-Scheduler versucht, jeden Thread fair zu behandeln, d.h. ab und zu Rechenzeit zuzuordnen – egal, welche Threads sonst noch Rechenzeit beanspruchen.
- Kein Thread hat jedoch Anspruch auf einen bestimmten Time-Slice.
- Für den Programmierer sieht es so aus, als ob sämtliche Threads "echt" parallel ausgeführt werden, d.h. jeder über eine eigene CPU verfügt.

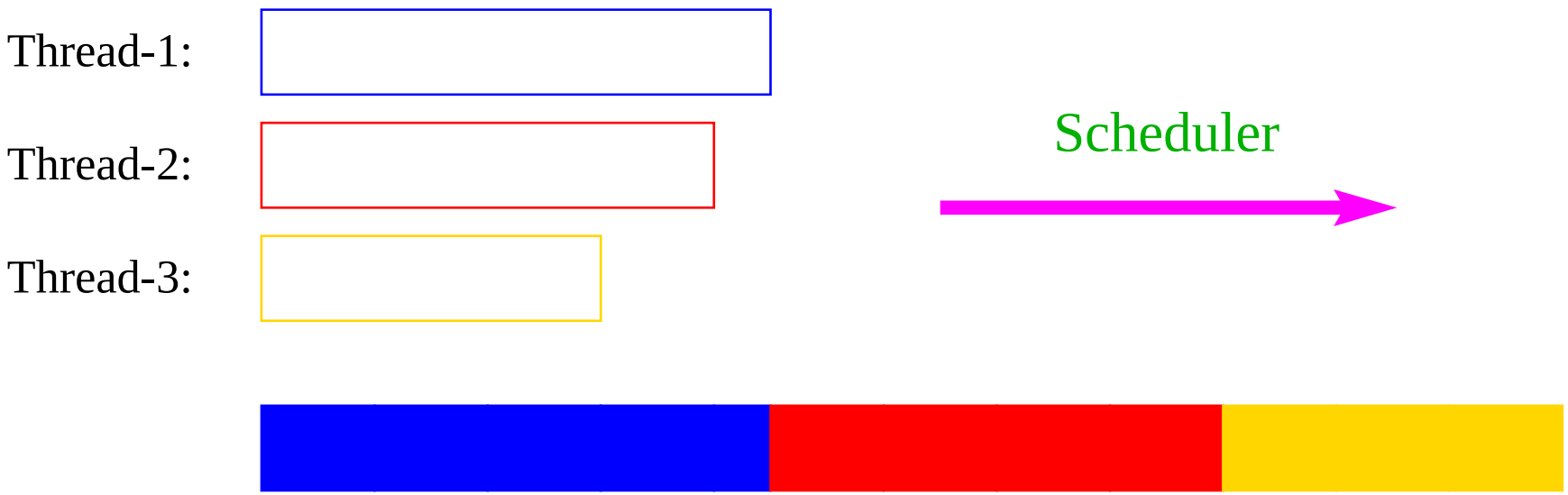
2. Naives Verfahren:

- Erhält ein Thread eine CPU, darf er laufen, so lange er will ...
- Gibt er die CPU wieder frei, darf ein anderer Thread arbeiten ...


Thread-1:
Thread-2:
Thread-3:
Scheduler


Thread-1:
Thread-2:
Thread-3:
Scheduler


Thread-1:
Thread-2:
Thread-3:
Scheduler


Thread-1:
Thread-2:
Thread-3:
Scheduler

Beispiel:

```
public class Start extends Thread {
public void run() {
   System.out.println("I'm running ...");
   while(true) ;
}
public static void main(String[] args) {
   (new Start()).start();
   (new Start()).start();
   (new Start()).start();
   System.out.println("main is running ...");
   while(true) ;
}
}  // end of class Start
```

... liefert als Ausgabe (bei naivem Scheduling und einer CPU) :

```
main is running ...
```

... liefert als Ausgabe (bei naivem Scheduling und einer CPU) :

```
main is running ...
```

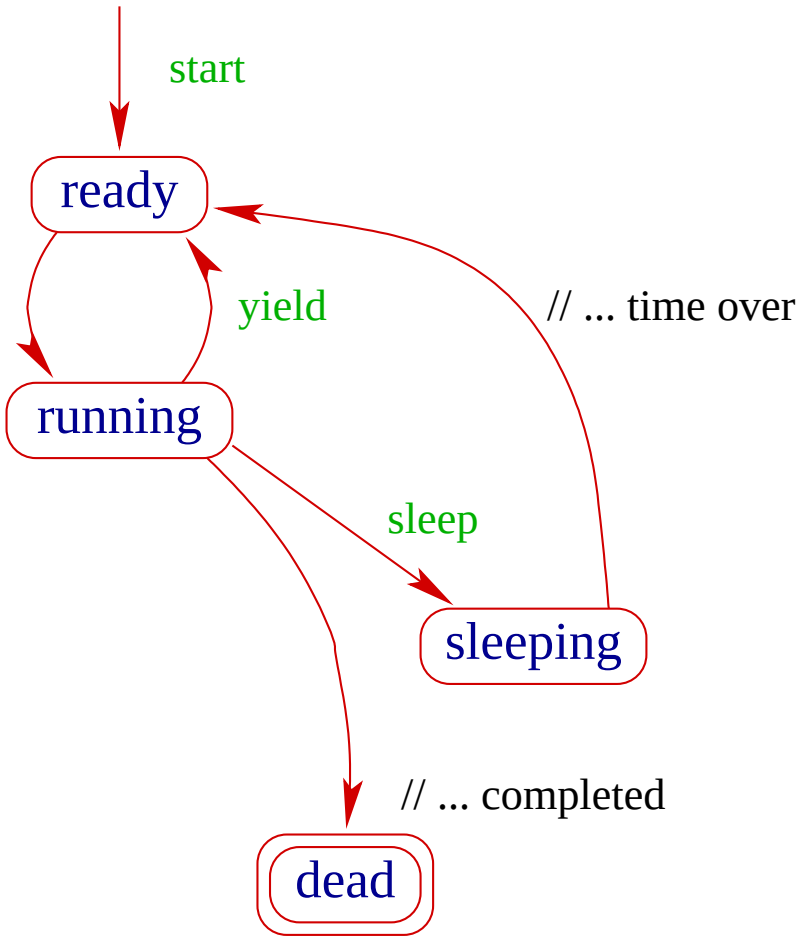
- Weil `main` nie fertig wird, erhalten die anderen Threads keine Chance, sie verhungern.
- Faires Scheduling mit einem Zeitscheiben-Verfahren würde z.B. liefern:

```
I'm running ...
main is running ...
I'm running ...
I'm running ...
```

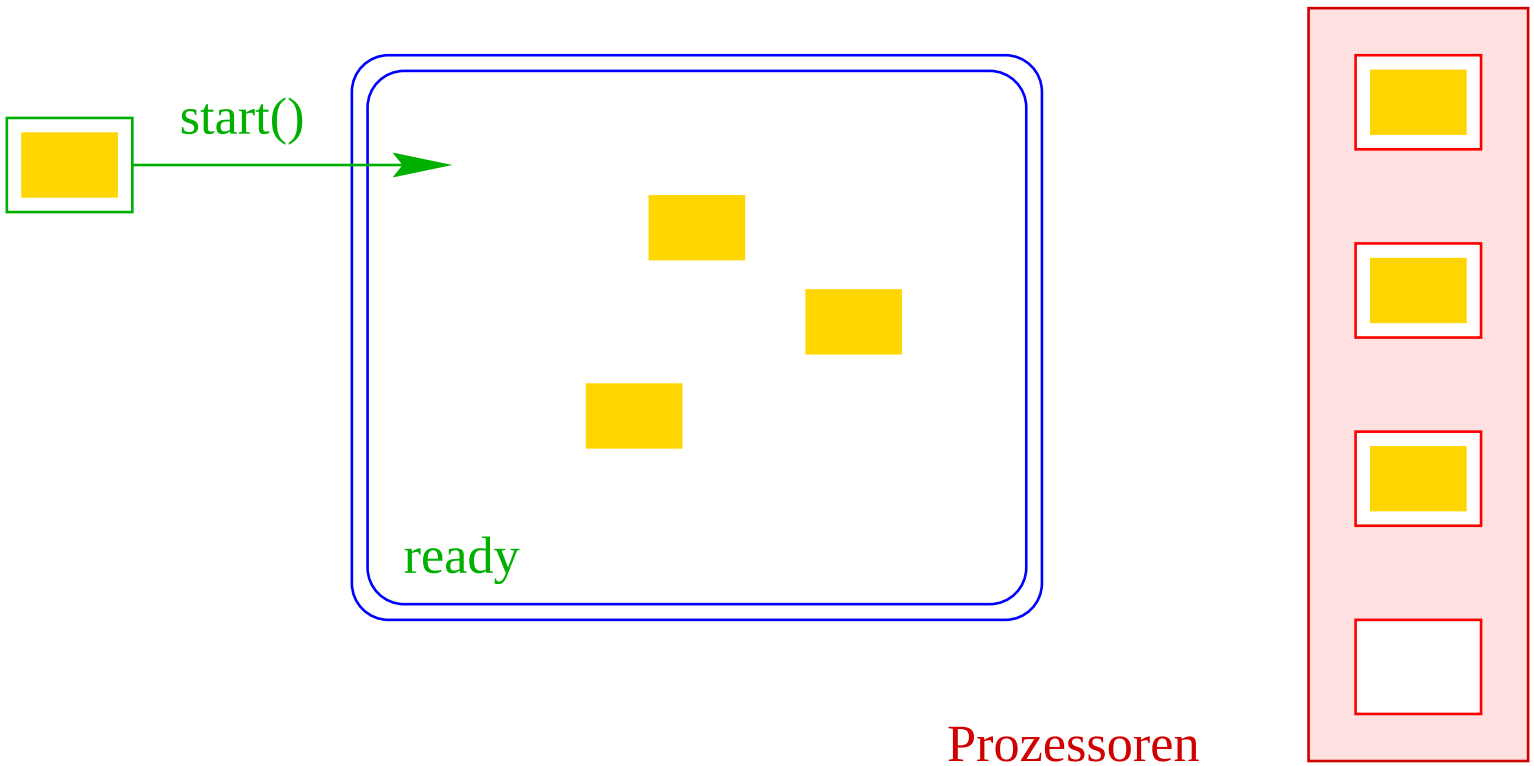
- Java legt nicht fest, wie intelligent der Scheduler ist.
- Die aktuelle Implementierung unterstützt faires Scheduling.
- Programme sollten aber für jeden Scheduler das gleiche Verhalten zeigen. Das heißt:
- ... Threads, die aktuell nichts sinnvolles zu tun haben, z.B. weil sie auf Verstreichen der Zeit oder besseres Wetter warten, sollten stets ihre CPU anderen Threads zur Verfügung stellen.
- ... Selbst wenn Threads etwas Vernünftiges tun, sollten sie ab und zu andere Threads laufen lassen.

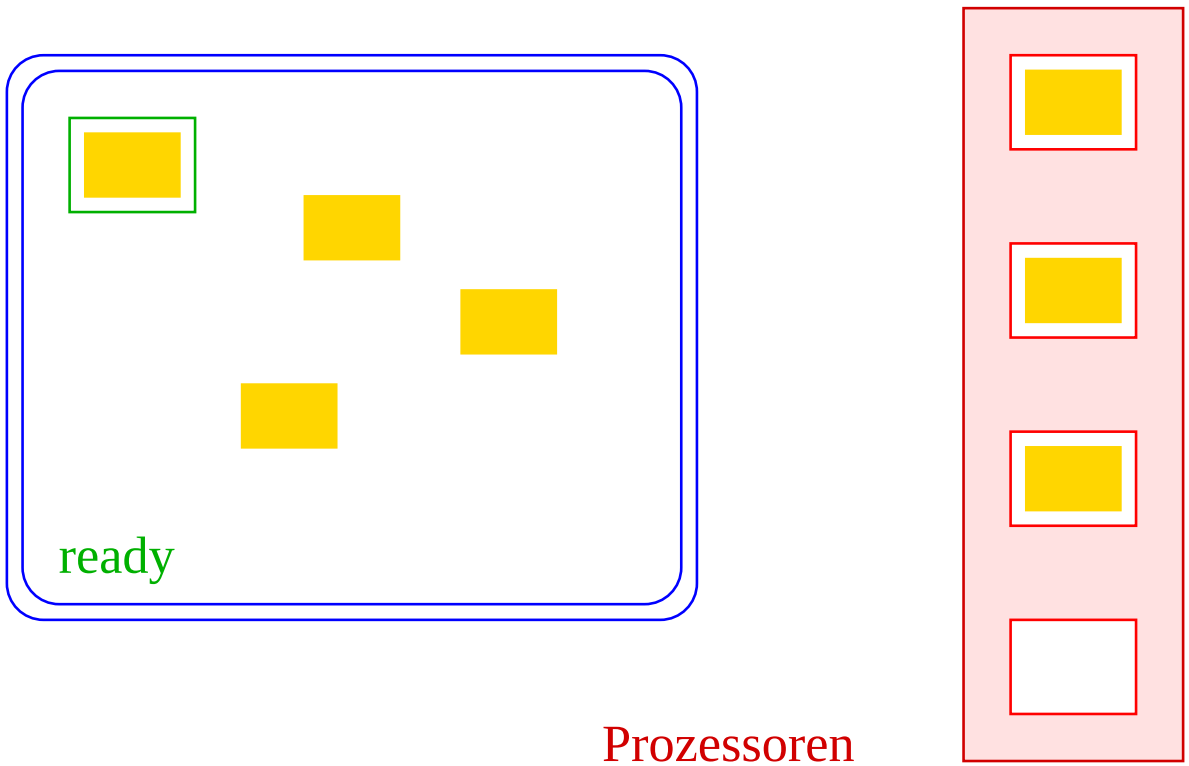
  (Achtung: Wechsel des Threads ist teuer!!!)
- Dazu verfügt jeder Thread über einen Zustand, der bei der Vergabe von Rechenzeit beücksichtigt wird.

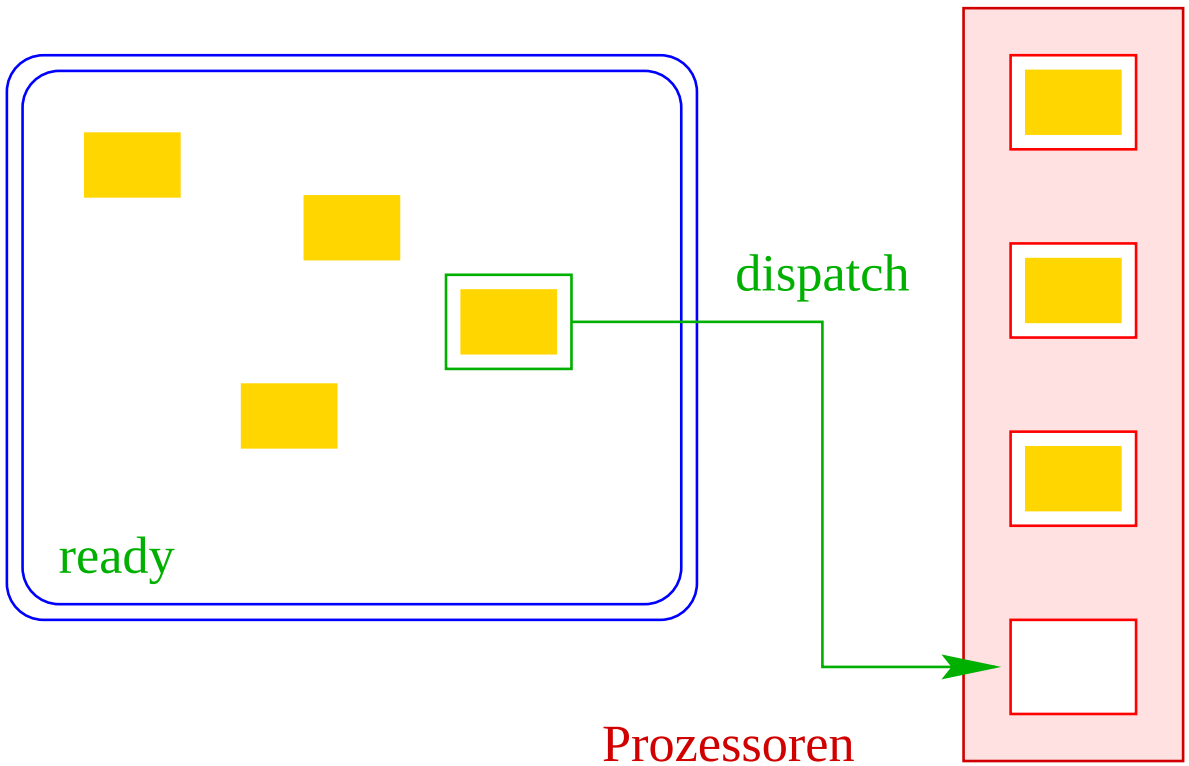
Einige Thread-Zustände:

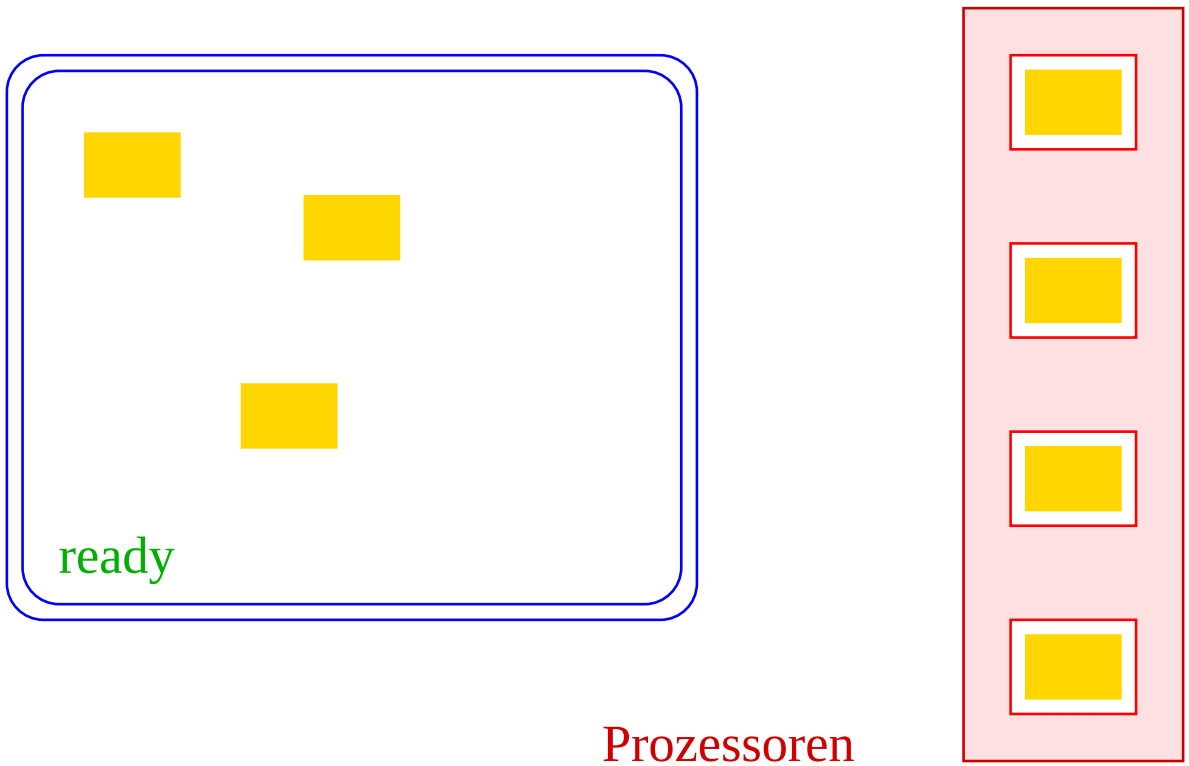


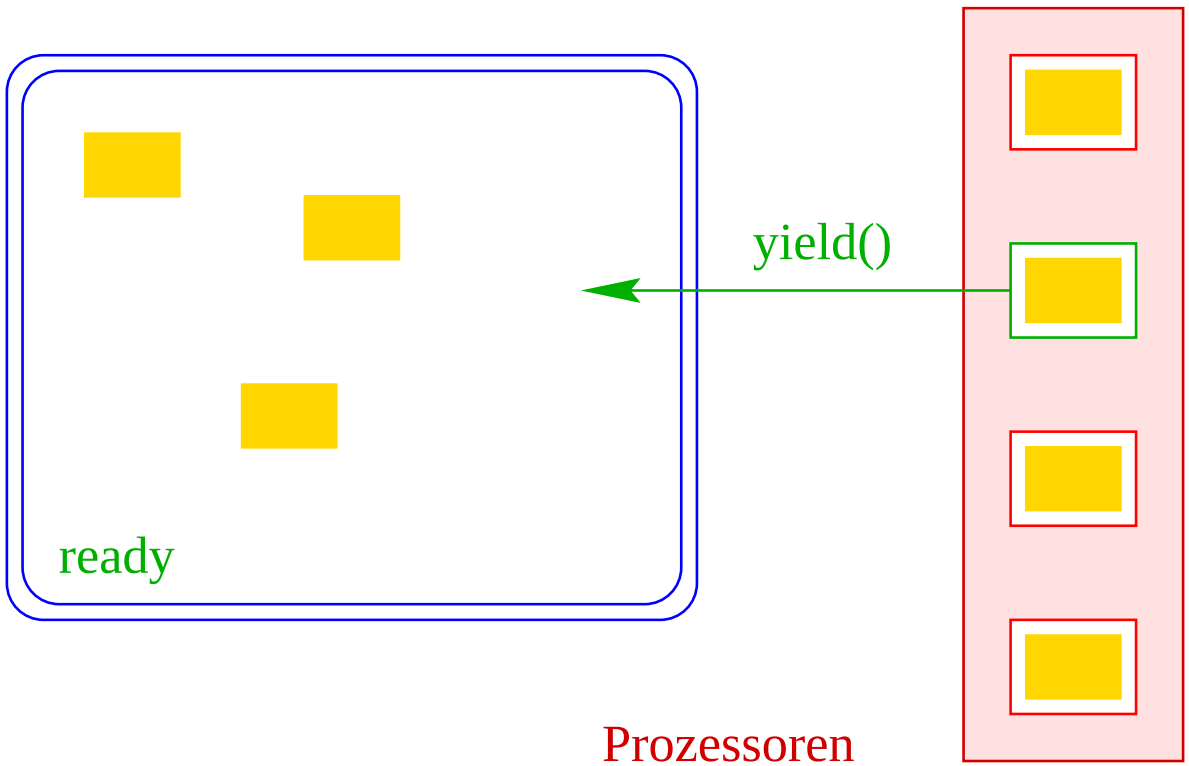
- `public void start();` legt einen neuen Thread an, setzt den Zustand auf ready und übergibt damit den Thread dem Scheduler zur Ausführung.
- Der Scheduler ordnet den Threads, die im Zustand ready sind, Prozessoren zu ("dispatching"). Aktuell laufende Threads haben den Zustand running.
- `public static void yield();` setzt den aktuellen Zustand zurück auf ready und unterbricht damit die aktuelle Programm-Ausführung. Andere ausführbare Threads erhalten die Gelegenheit zur Ausführung.
- `public static void sleep(int msec) throws InterruptedException;` legt den aktuellen Thread für `msec` Millisekunden schlafen, indem der Thread in den Zustand sleeping wechselt.

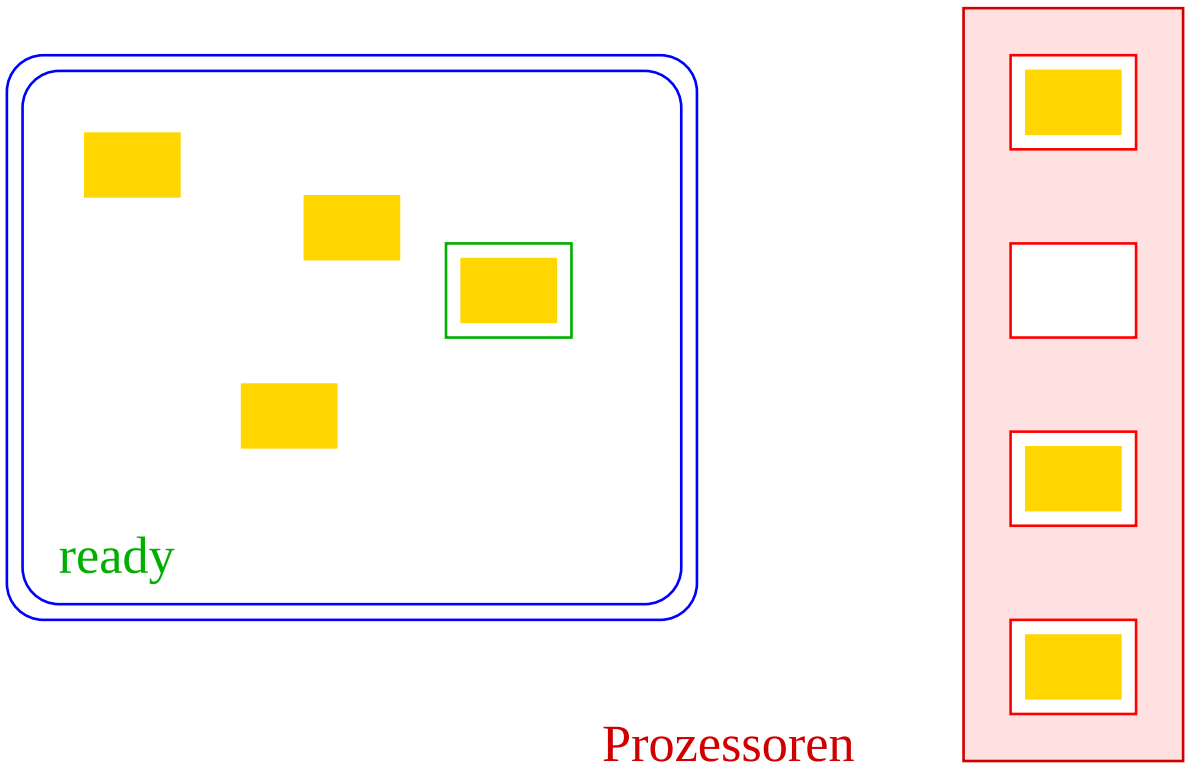

start()
ready
Prozessoren

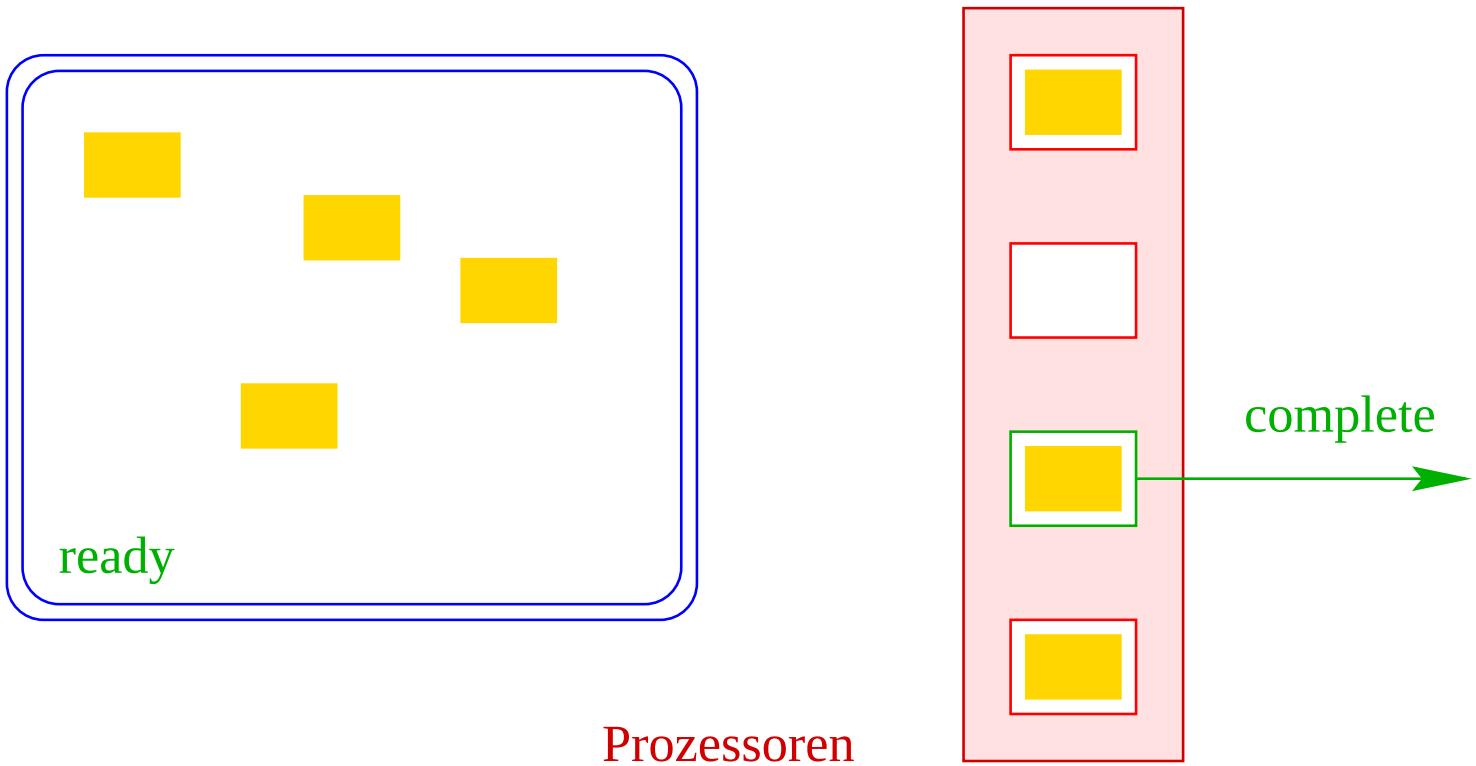

ready
Prozessoren

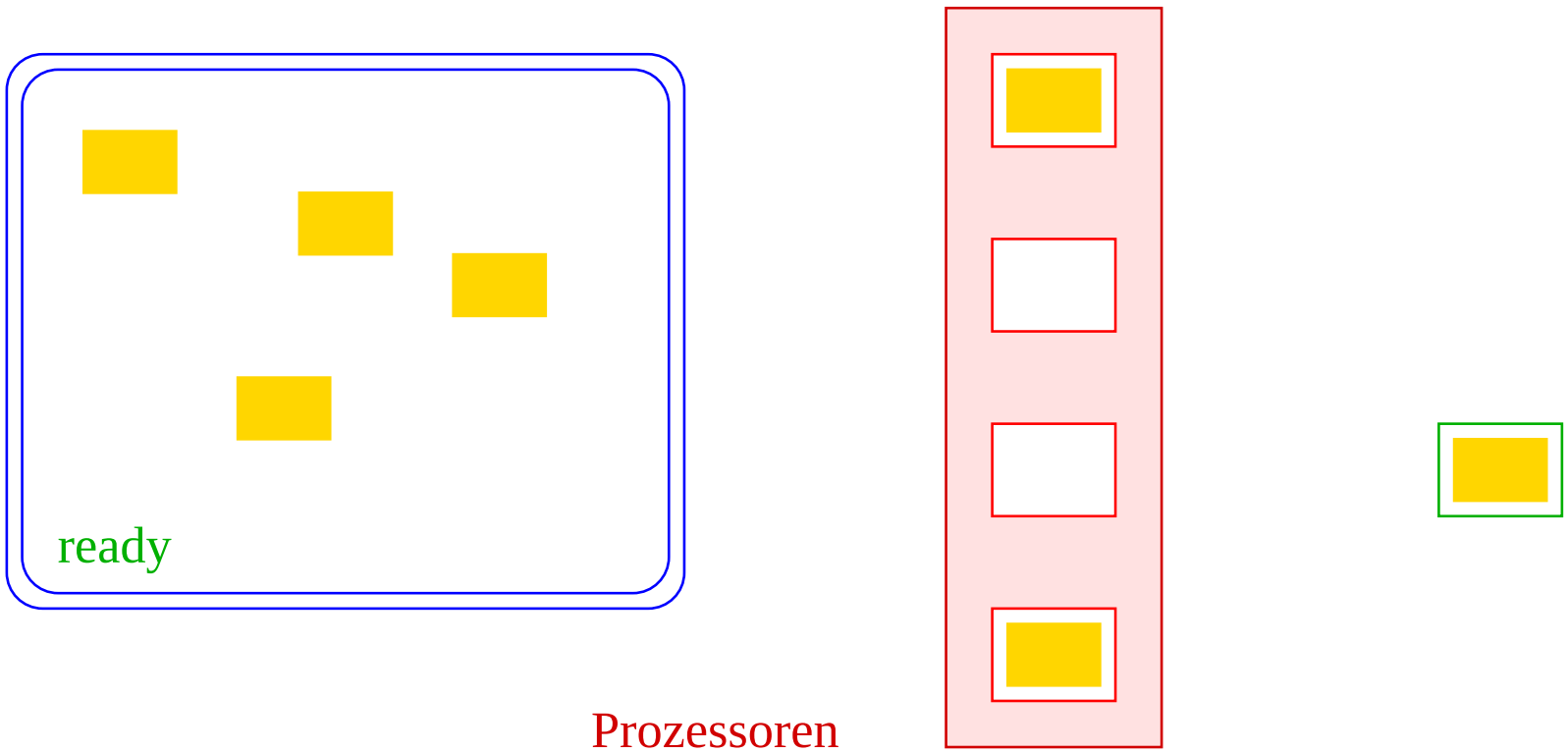

dispatch
ready
Prozessoren


ready
Prozessoren


yield()
ready
Prozessoren


ready
Prozessoren


ready
complete
Prozessoren


ready
Prozessoren

18.1 Futures

- Die Berechnung eines Zwischenergebnisses kann lange dauern.
- Während dieser Berechnung kann möglicherweise etwas anderes sinnvolles berechnet werden.

18.1 Futures

- Die Berechnung eines Zwischenergebnisses kann lange dauern.
- Während dieser Berechnung kann möglicherweise etwas anderes sinnvolles berechnet werden.

Idee:

- Berechne das Zwischenergebnisses in einem eigenen Thread.
- Greife auf den Wert erst zu, wenn sich der Thread beendet hat.

$\Longrightarrow$ Futures

Eine Future startet die Berechnung eines Werts, auf den später zugegriffen wird ...

Das generische Interface

```
public interface Callable<T> {
   T call throws Exception ();
}
```

aus `java.util.concurrency` beschreibt Klassen, für deren Objekte ein Wert vom Typ `T` berechnet werden kann.

Eine Future startet die Berechnung eines Werts, auf den später zugegriffen wird ...

Das generische Interface

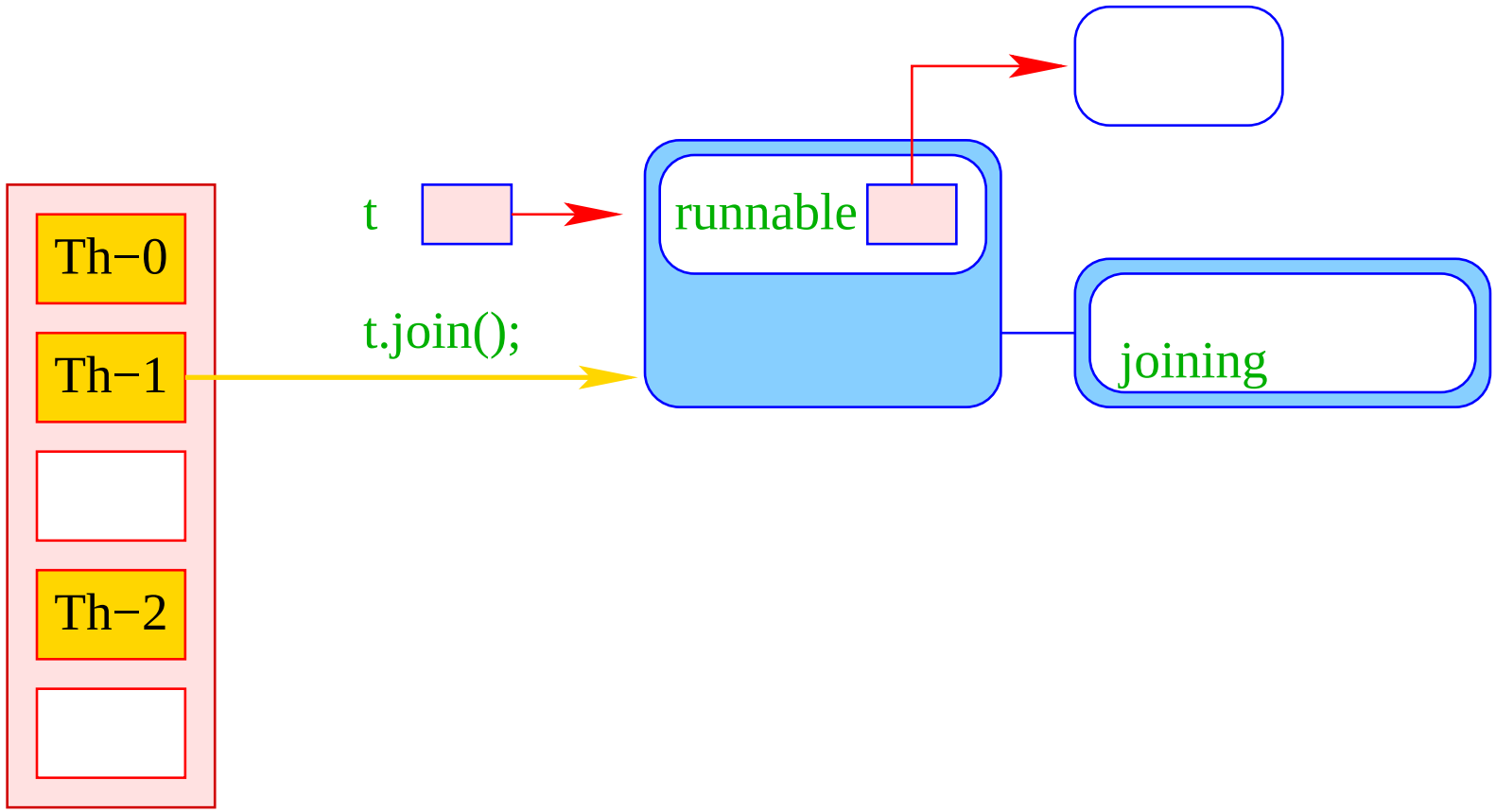
```
public interface Callable<T> {
   T call throws Exception ();
}
```

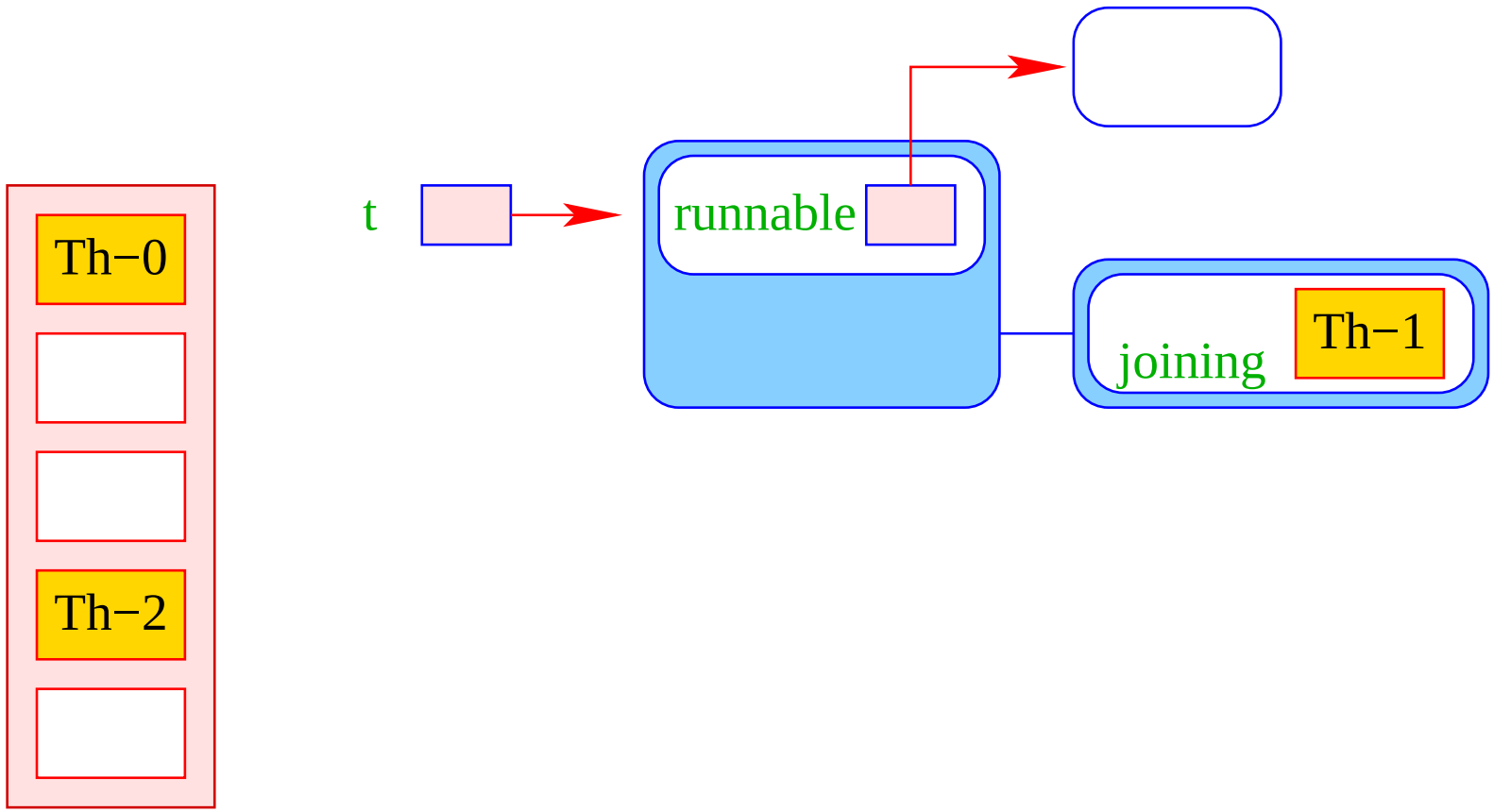
aus `java.util.concurrency` beschreibt Klassen, für deren Objekte ein Wert vom Typ `T` berechnet werden kann. Wir implementieren:

```
public class Future<T> implements Runnable {
   private T value = null;
   private Exception exc = null;
   private Callable<T> work;
   private Thread task;
...
```

```
...
public Future<T>(Callable<T> w) {
   work = w;
   task = new Thread (this);
   task.start();
   }
public void run() {
   try {value = work.call();}
   catch (Exception e) { exc = e;}
   }
public T get() throws Exception {
   task.join();
   if (value == null) throw exc;
   return value;
   }
}
```

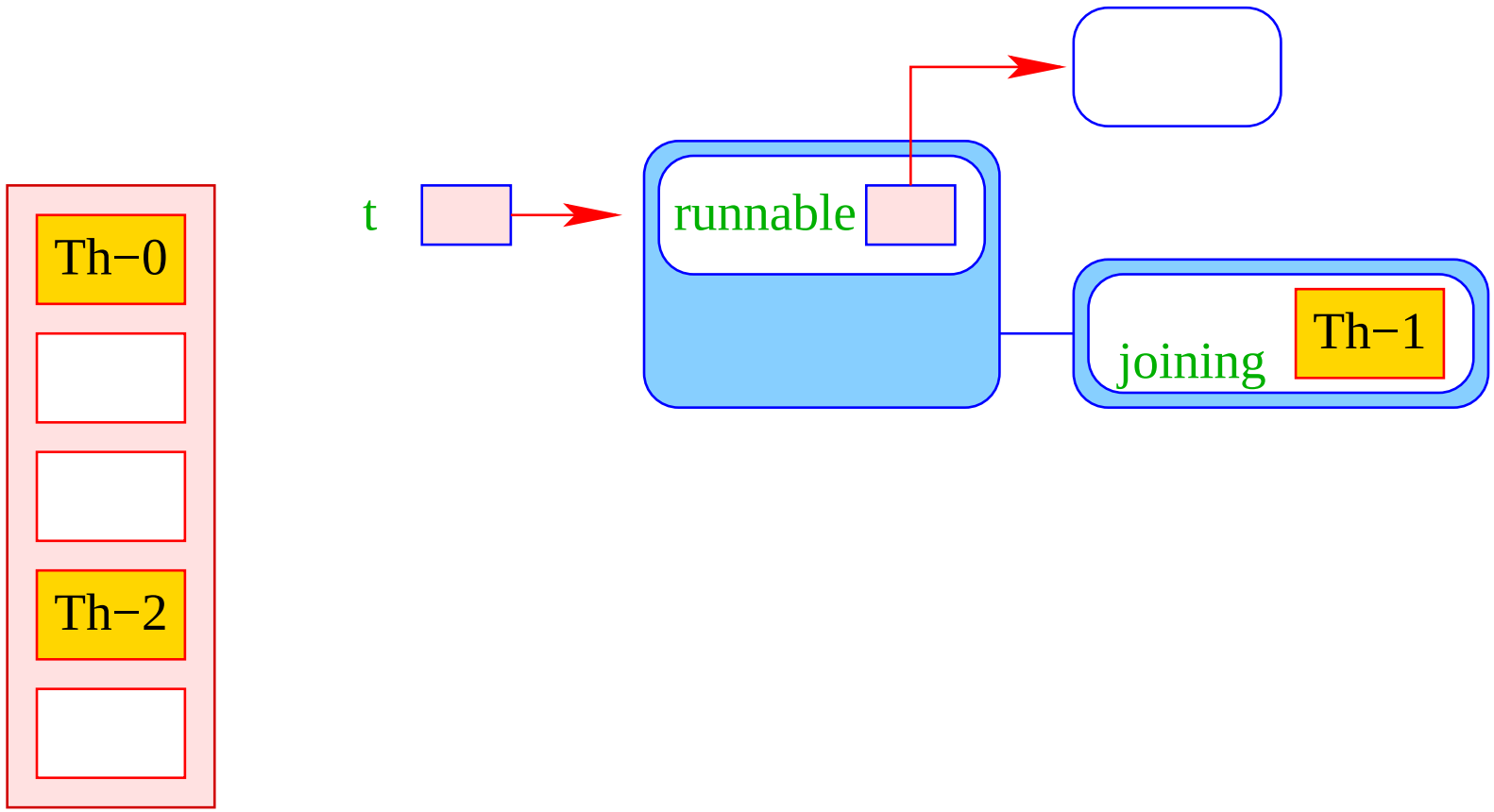
- Der Konstruktor erhält ein `Callable`-Objekt.
- Die Methode `run()` ruft für dieses Objekt die Methode `call()` auf und speichert deren Ergebnis in dem Attribut `value` — bzw. eine geworfene Exception in `exc` ab.
- Der Konstruktor legt ein Thread-Objekt für die Future an und startet diesen Thread, der dann `run()` ausführt.
- Die Methode `get()` wartet auf Beendigung des Threads. Dazu verwendet sie die Objekt-Methode `public final void join() throws InterruptedException` der Klasse `Thread` ...
- Dann liefert `get()` den berechneten Wert zurück — falls dieser nicht `null` ist. Andernfalls wird die Exception `exc` geworfen.

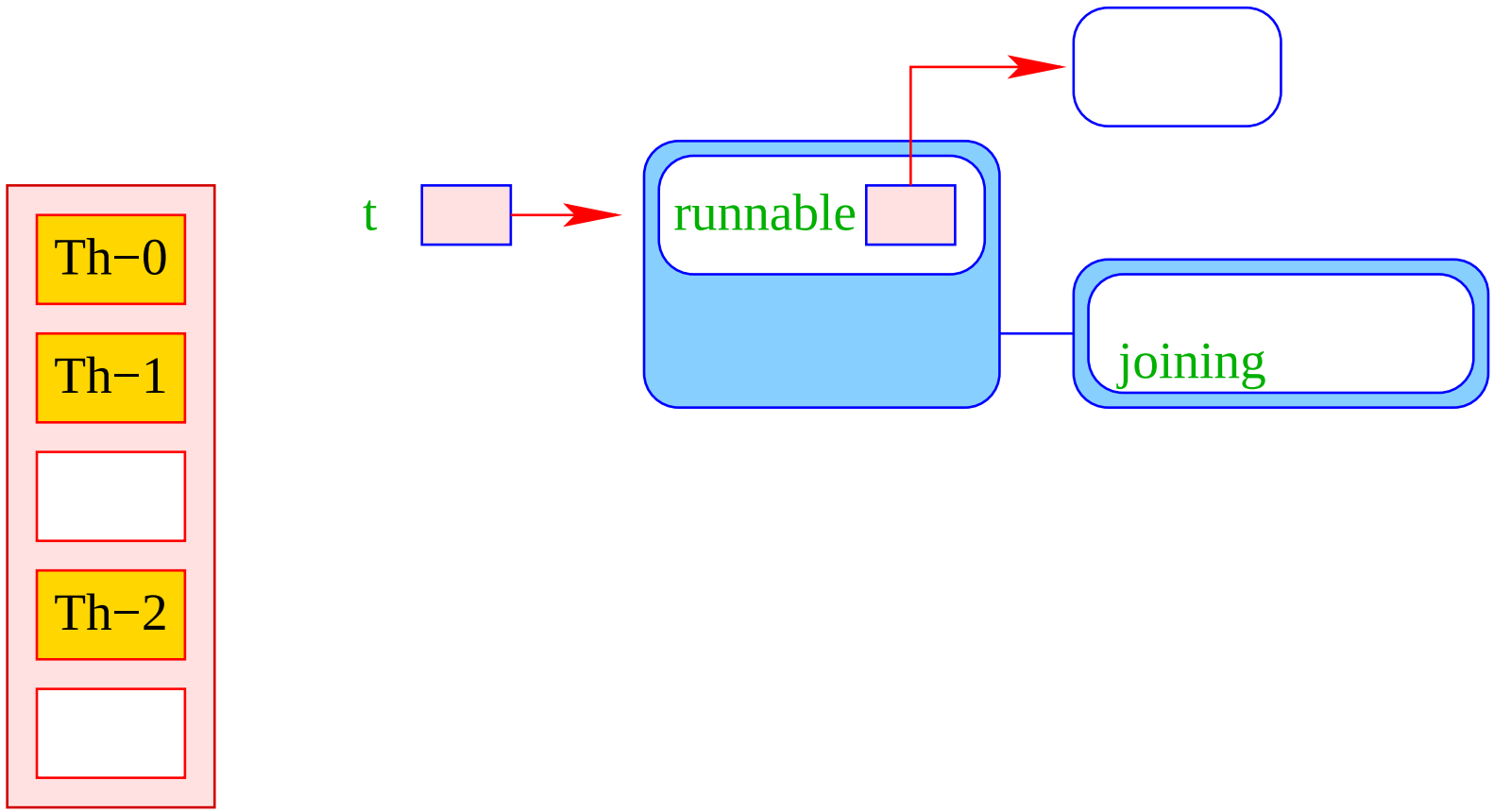

Th−0
Th−1
Th−2
t
t.join();
runnable
joining


Th−0
Th−2
t
runnable
joining
Th−1

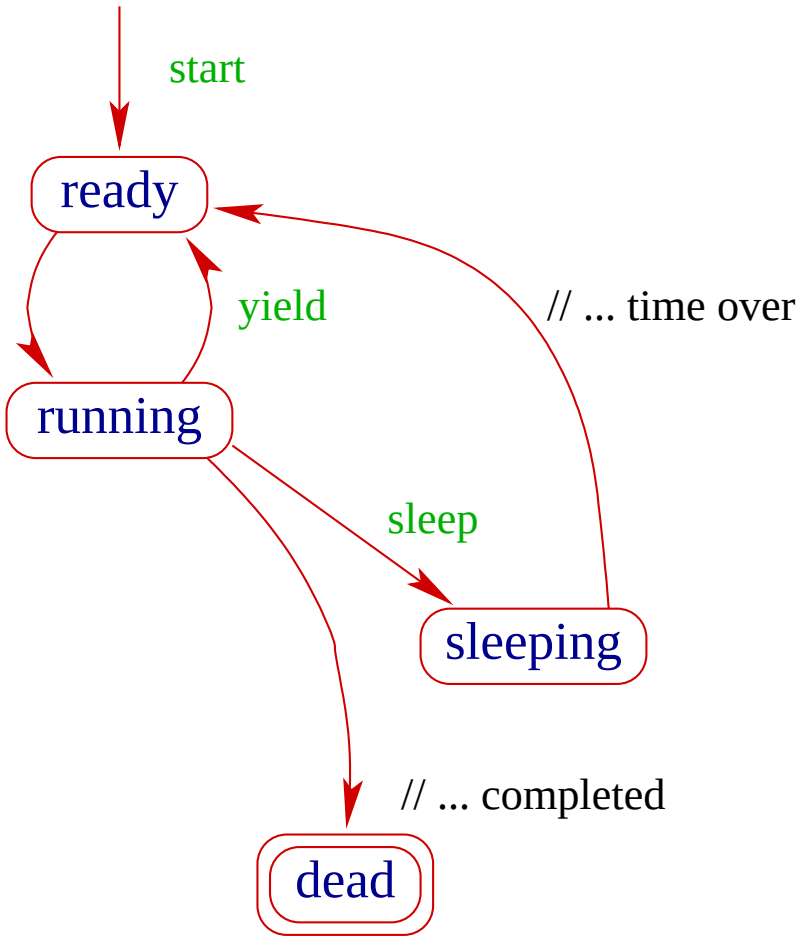
Beachte:

- Für jedes Threadobjekt t gibt es eine Schlange `ThreadQueue joiningThreads`.
- Threads, die auf Beendigung des Threads t warten, werden in diese Schlange eingefügt.
- Dabei gehen sie konzeptuell in einen Zustand joining über und werden aus der Menge der ausführbaren Threads entfernt.
- Beendet sich ein Thread, werden alle Threads, die auf ihn warteten, wieder aktiviert ...


Th−0
Th−2
t
runnable
joining
Th−1


Th−0
Th−1
Th−2
t
runnable
joining

Erweitertes Zustandsdiagramm:



Erweitertes Zustandsdiagramm: