EC 440 – Introduction to Operating Systems

Orran Krieger (BU)
Larry Woodman (Red Hat)

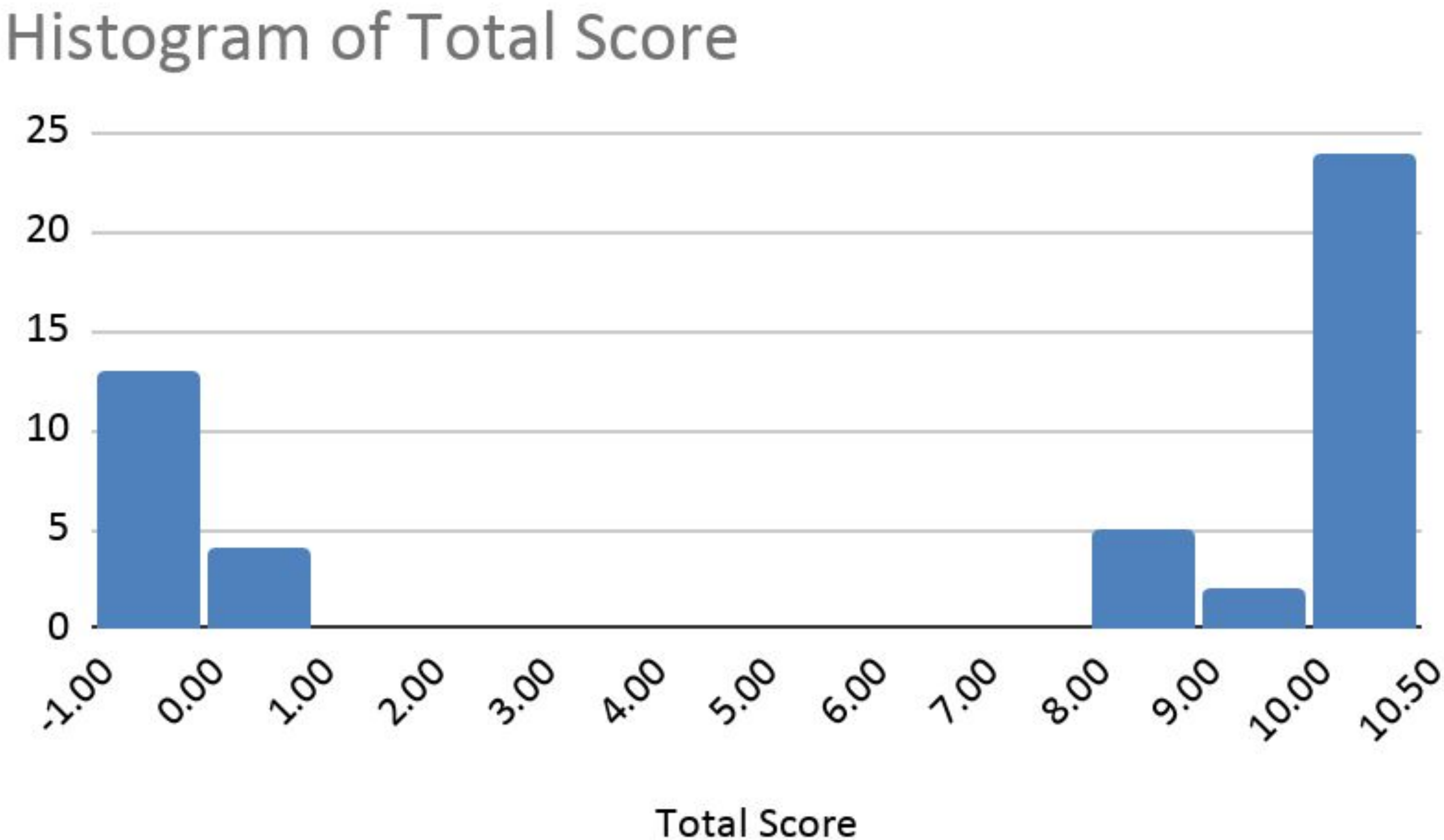
credit Christopher Kruegel, UCSB & Manuel Egele.

Administration and Review

Homework

- Shell submission date was before start of class:
  - Oral exams (~10 minutes) start today
- Threading assignment starts today
  - Due March 8. Start it this week! You will encounter new challenges.
  - Next lecture includes discussion.
- Kernel hacking workshop assigned today
  - Expect 3-6 hours total.
  - No *due date*, but it is a prerequisite to start HW5. Try to spend 1 hour/week.

Status HW2 as of 4:15 today



Retrospective on HW1

- Thank you for those of you helping in Piazza!
- How many of you were up all night?
- For people that have not submitted, why?
  - Post privately to us if you have a covid related problem.
- We are adding more extensive tests for HW2 to try to get more partial points

Observations

- Use proper makefile
- Write tests - test driven development
- Use your debugger
- Break things down into functions, and data structures, use enums, no magic constants... comment...
- Use style guild
- Use tools, e.g., valgrind/static analyzers

These do not slow you down!

Overview so far

- Overview of what OSes are, kernel, key services, ….
  - System calls, i.e., kernel API: processes, signals, memory, & files
- Gone into details on compute:
  - Processes, Programs, Threads
  - Scheduling
- This week we will talk about how synchronization happens: Modern Operating Systems (MOC) 2.3 & 2.5

Inter-process Communication and Synchronization

Inter-process Communication and Synchronization

- Processes/threads may need to exchange information
- Processes/threads should access resources in the right sequence
- Need to coordinate the activities of multiple threads
- Need to introduce the notion of *synchronization operations*
  - These operations allow threads to control the timing of their events relative to events in other threads

Asynchrony and Race Conditions

- Threads need to deal with asynchrony
- Asynchronous events occur arbitrarily during thread execution:
  - An interrupt causes control (CPU) being taken away from the current thread to the interrupt handler
  - A timer interrupt causes one thread to be suspended and another one to be resumed
  - Two threads running on different CPUs read and write the same memory
- Threads must be designed so that they can deal with such asynchrony
- (If not, the code must be protected from asynchrony)

Race Conditions

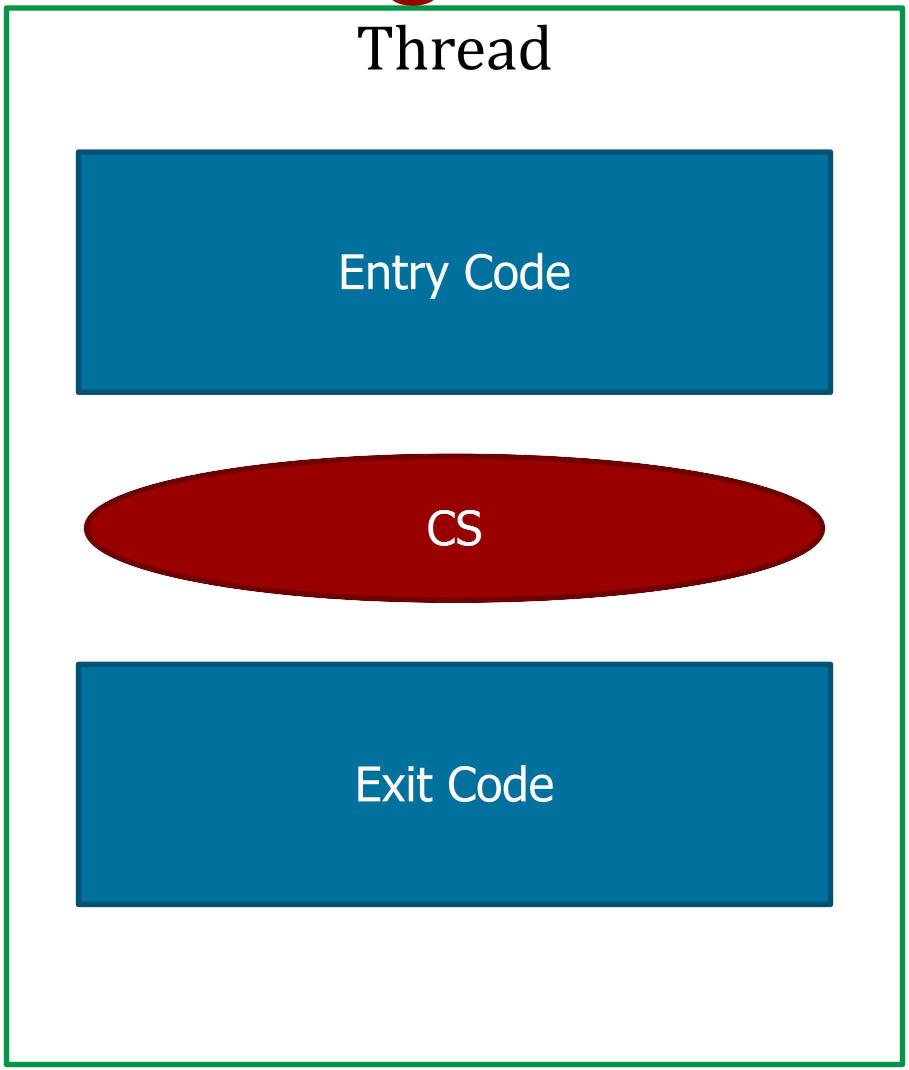
- Two threads, A and B, need to insert objects into a list, so that it can be processed by a third thread, C
- Both A and B
  - Check which is the first available slot in the list
  - Insert the object in the slot
- Everything seems to run fine until...
  - Thread A finds an available slot but gets suspended by the scheduler
  - Thread B finds the same slot and inserts its object
  - Thread B is suspended
  - Thread A is resumed and inserts the object in the same slot
- What did just happen?

B's object is lost!

Critical Regions and Mutual Exclusion

- The part of the program where shared memory is accessed is called a *critical region* (or *critical section*)
- Critical regions should be accessed in *mutual exclusion*
- Solution: Synchronization
  1. No two processes may be simultaneously inside the same critical region
  2. No process running outside the critical region should block another process
  3. No process should wait forever to enter its critical region
  4. No assumptions can be made about speed/number of CPUs

Entering and Exiting Critical Regions



Mutual Exclusion With Busy Waiting

First solution: Disable interrupts when in critical region

- What if the process “forgets” to re-enable interrupts?
- What if there are multiple CPUs?

Mutual Exclusion With Busy Waiting

First solution: Disable interrupts when in critical region

- What if the process "forgets" to re-enable interrupts?
- What if there are multiple CPUs?

Second solution: a lock variable

- Test if lock is 0
- If not, loop on check until 0
- When lock is 0, set it to 1 and start critical region
- Set it back to 0 when finished
- ... do you see any problem?

Mutual Exclusion With Busy Waiting

First solution: Disable interrupts when in critical region

- What if the process “forgets” to re-enable interrupts?
- What if there are multiple CPUs?

Second solution: a lock variable

- Test if lock is 0
- If not, loop on check until 0
- When lock is 0, set it to 1 and start critical region
- Set it back to 0 when finished
- ... do you see any problem?

Third solution: strict alternation

Taking Turns...

`turn` Initially set to 0

Thread 0

```
while (turn != 0) { }
```

CS

```
turn=1;
```

Thread 1

```
while (turn != 1) { }
```

CS

```
turn=0;
```

Taking Turns...

What if thread 0 is much faster than thread 1?

Thread 0 may be waiting for its turn even if thread 1 is outside the critical region

We said:

- No process running outside the critical region should block another process

Need for something better:
Peterson's algorithm

Peterson's Algorithm

Thread 0

```
interested_0 = TRUE;
turn = 0;
while (interested_1 == TRUE
        && turn == 0) { };
```

CS

```
interested_0 = FALSE;
```

Thread 1

```
interested_1 = TRUE;
turn = 1;
while (interested_0 == TRUE
          && turn == 1) { };
```

CS

```
interested_1 = FALSE;
```

Test And Set Lock Instruction

If the hardware (that is, the CPU) provides an *atomic* way of testing and setting a lock, life is easier … Why?

Test And Set Lock Instruction

If the hardware (that is, the CPU) provides an *atomic* way of testing and setting a lock, life is easier … Why?

- `TSL RX, LOCK`
  - Reads contents of address LOCK into RX
  - Stores a nonzero value into location LOCK
- Now back to lock variables

```
enter: TSL RX, LOCK
       CMP RX, #0
       JNE enter
       RET
leave: MOV LOCK, #0
       RET
```

Test And Set Lock Instruction

No TSL insn on Intel, use `CMPXCHG` instead

Implicitly uses the accumulator register AL/AX/EAX/RAX for the comparison

Does all this as an atomic instruction

- i.e., either the entire instruction executes, or none of it

CMPXCHG—Compare and Exchange

Opcode/ Instruction	Op/ En	64-Bit Mode	Compat/ Leg Mode	Description
0F B0/*r* CMPXCHG *r/m8, r8*	MR	Valid	Valid*	Compare AL with *r/m8*. If equal, ZF is set and *r8* is loaded into *r/m8*. Else, clear ZF and load *r/m8* into AL.

Sleep and Wakeup

- Busy waiting is a waste of CPU

Sleep and Wakeup

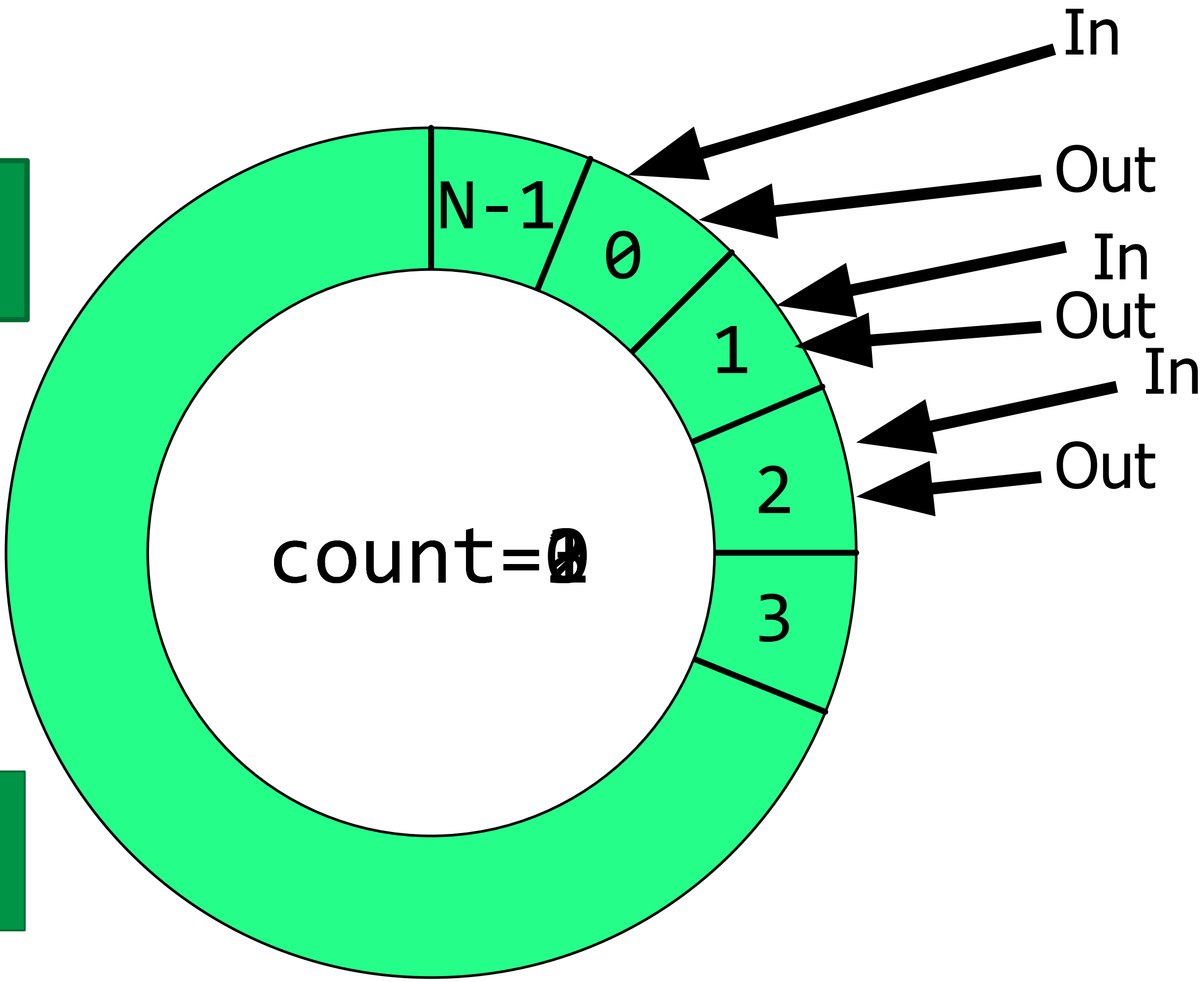
- Busy waiting is a waste of CPU
- Need to provide a mechanism so that a thread can suspend when a critical region cannot be entered
  - `Sleep()` blocks the thread
  - `Wakeup()` resumes a thread

Sleep and Wakeup

- Busy waiting is a waste of CPU
- Need to provide a mechanism so that a thread can suspend when a critical region cannot be entered
  - `Sleep()` blocks the thread
  - `Wakeup()` resumes a thread
- Classical problem:
  - Producer and Consumer communicating through a set of buffers
- Number of buffers (N) is limited
  - 0 buffers available ➔ consumer must wait
  - N buffers filled ➔ producer must wait

Producer/Consumer Problem

`count=0` →
Consumer must wait



`count=N` →
Producer must wait

Producer/Consumer

```
in = 0;
out = 0;
count = 0;
```

Producer:

```
while (1) {
     item = produce_item();
     if (count == N) sleep();
     buff[in]=item;
     in=(in+1) % N;
     count=count+1;
     if (count == 1)
         wakeup(consumer)
}
```

Consumer:

```
while (1) {
     if (count == 0) sleep();
     item = buff[out];
     out=(out+1) % N;
     count = count-1;
     if (count == N-1)
        wakeup(producer)
     consume_item(item);
}
```

Producer/Consumer

```
in = 0;
out = 0;
count = 0;
```

What is broken?

```
while (1
    item = produce_item();
    if (count == N) sleep();
    buff[in]=item;
    in=(in+1) % N;
    count=count+1;
    if (count == 1)
        wakeup(consumer)
}
```

```
    if (count == 0) sleep();
    item = buff[out];
    out=(out+1) % N;
    count = count-1;
    if (count == N-1)
       wakeup(producer)
    consume_item(item);
}
```

Missing the Wake Up Call

1. Buffer is empty
2. Consumer reads counter and gets 0
3. Before falling asleep, there is a context switch to the Producer thread
4. Producer inserts item and, since count==1, sends a wakeup
5. Consumer is not sleeping and wakeup signal gets lost
6. Control returns to Consumer that falls asleep (the check on count has been done before)
7. Producer continues until count reaches N and then falls asleep …

Game Over!

Producer/Consumer

1. Buffer is empty
2. Consumer reads counter and gets 0
3. Before falling asleep, there is a context switch to the Producer thread
4. Producer inserts item and, since count==1, sends a wakeup
5. Consumer is not sleeping and wakeup signal gets lost
6. Control returns to Consumer that falls asleep (the check on count has been done before)
7. Producer continues until count reaches N and then falls asleep …

Producer:

```
while (1) {
     item = produce_item();
     if (count == N) sleep();
     buff[in]=item;
     in=(in+1) % N;
     count=count+1;
     if (count == 1)
         wakeup(consumer)
}
```

Consumer:

```
while (1) {
     if (count == 0) sleep();
     item = buff[out];
     out=(out+1) % N;
     count = count-1;
     if (count == N-1)
         wakeup(producer)
     consume_item(item);
}
```

Semaphores

- Edward Dijkstra suggested to use an integer variable to count the number of wakeups issued (in 1962)
- New type, the **Semaphore**
  - Semaphore(count) creates and initializes to count
  - `P()` or `down()`
    - If the counter is greater than 0 then decrements the counter and returns
    - If counter = 0 the process suspends. When it wakes up decrements the counter and returns
  - `V()` or `up()`
    - Increments the counter
    - If there are any processes waiting on the semaphore one is woken up
    - Returns
  - `down()` and `up()` are ATOMIC operations

Semaphores and Mutual Exclusion

`mutex` Semaphore with count = 1, initial value 1

Thread 0

```
mutex.down();
```

CS

```
mutex.up();
```

Thread 1

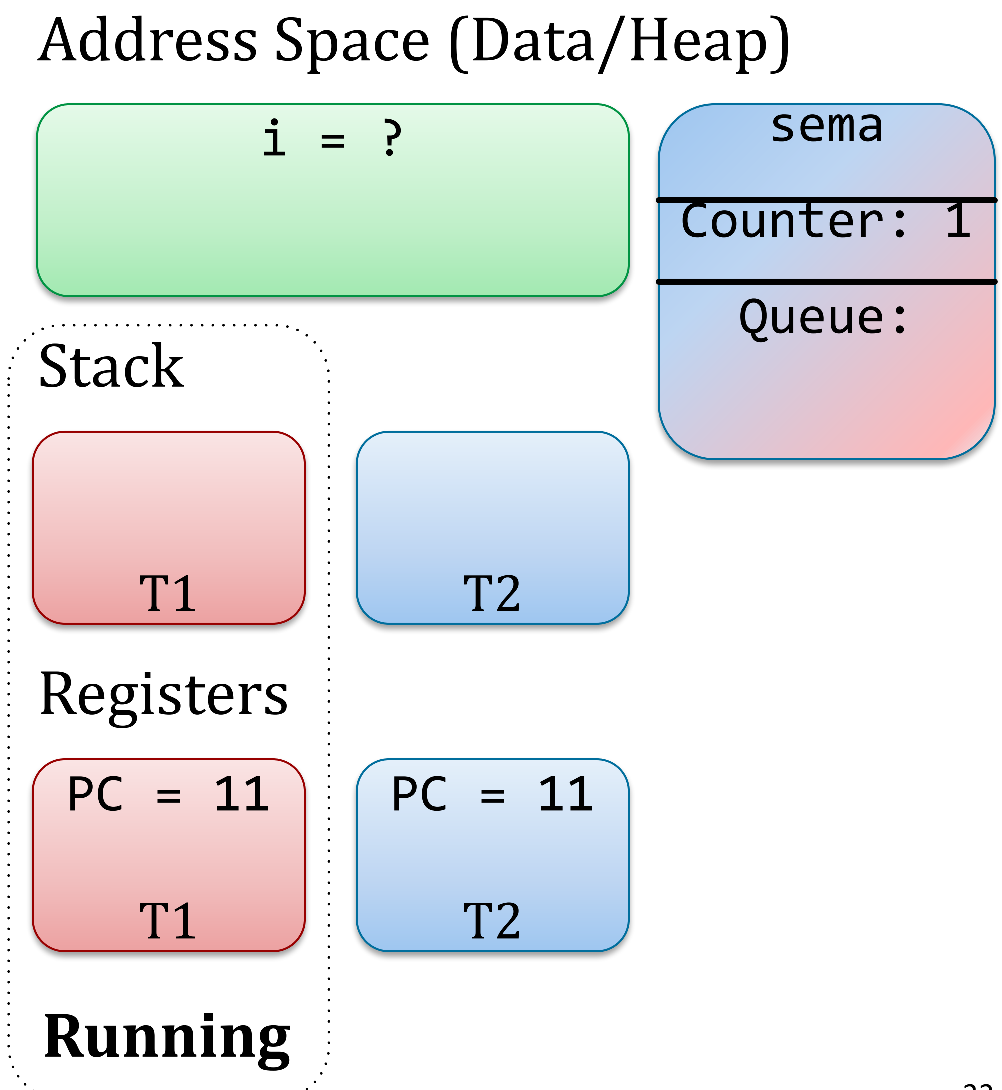
```
mutex.down();
```

CS

```
mutex.up();
```

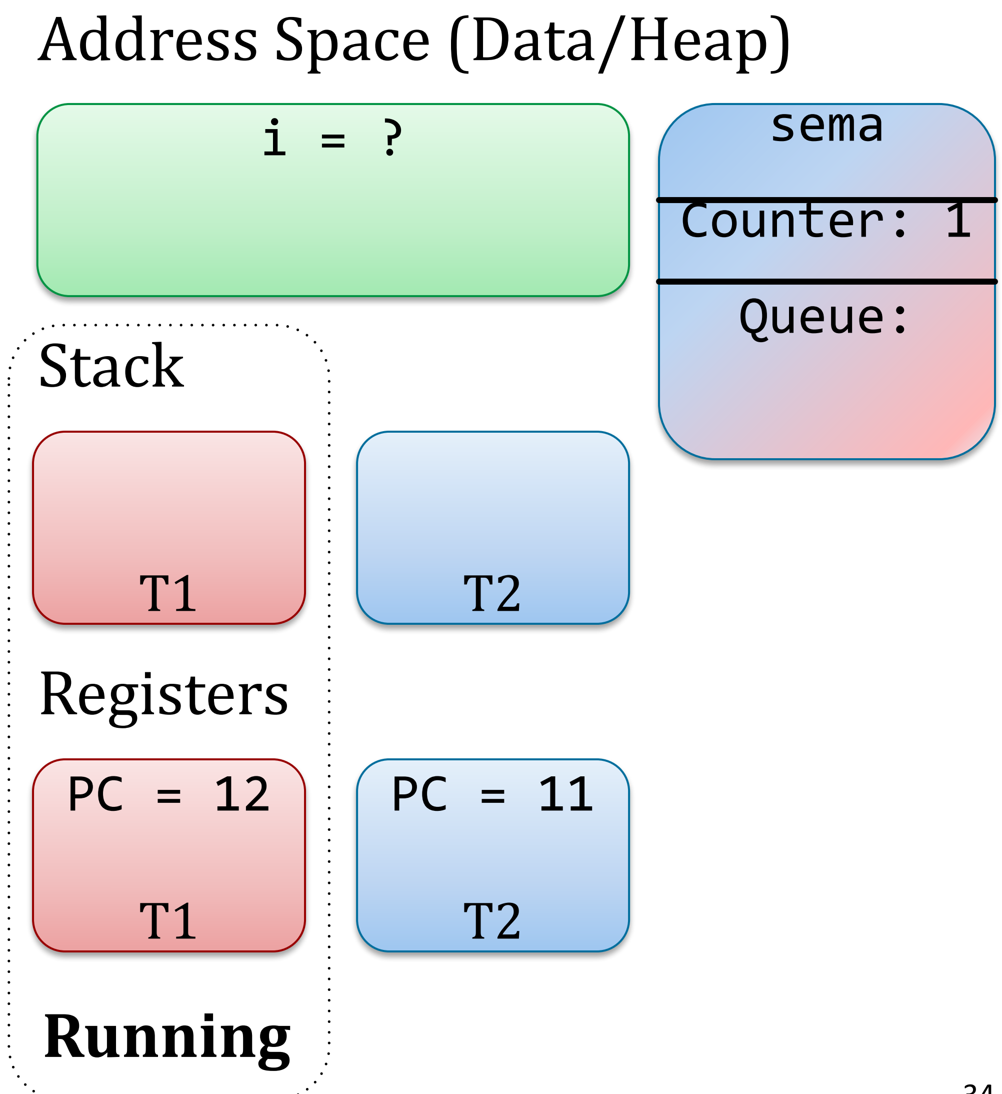
Threads – Revisited

```
 1: int i;
 2: Semaphore sema;
 3:
 4: f()
 5: {
 6:     printf("i is %d\n", i);
 7: }
 8:
 9: int main(int argc, char **argv)
10: {
11:    ... (do stuff here) ...
12:    P(sema);
13:    i = get_input();
14:    f();
15:    V(sema);
16:    return 0;
17: }
```



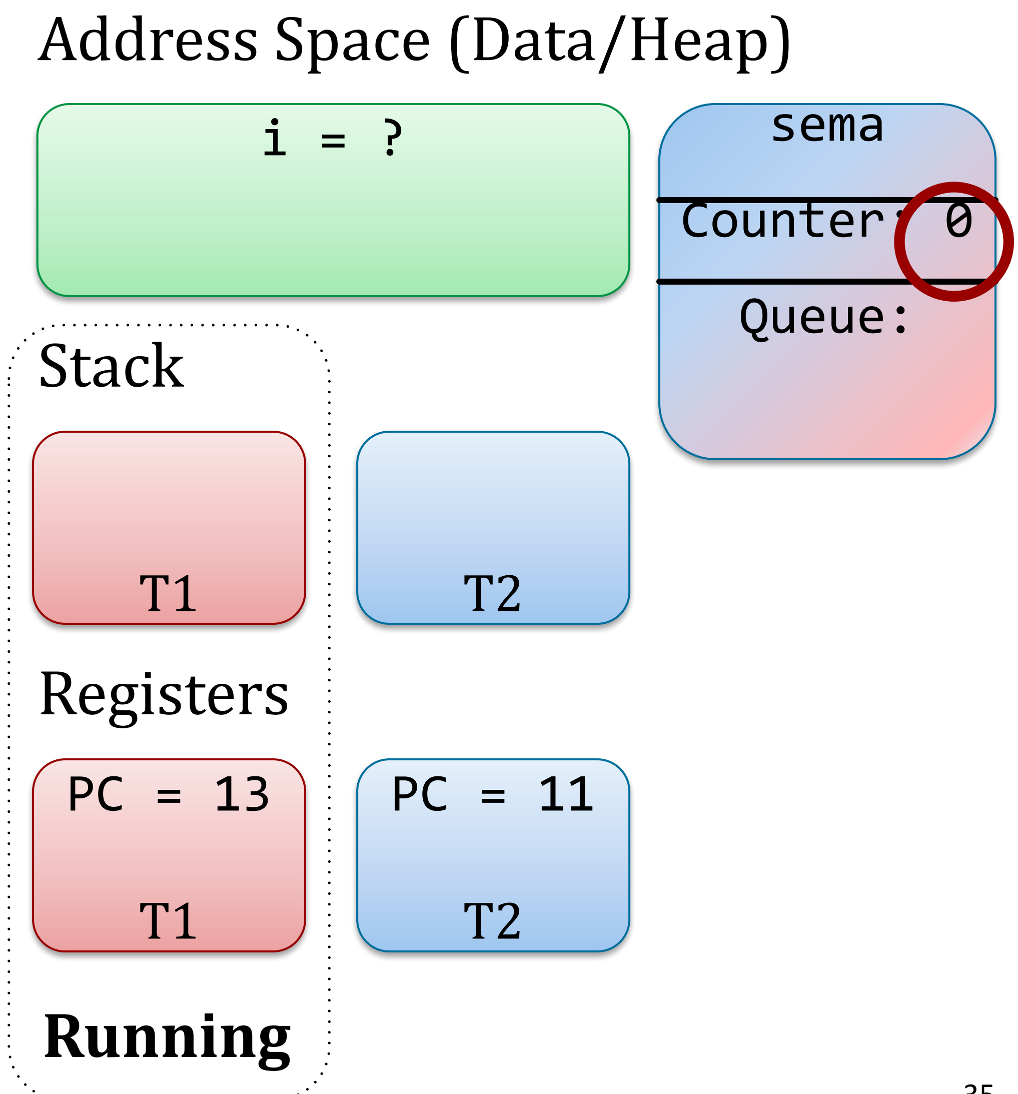
Threads – Revisited

```
 1: int i;
 2: Semaphore sema;
 3:
 4: f()
 5: {
 6:     printf("i is %d\n", i);
 7: }
 8:
 9: int main(int argc, char **argv)
10: {
11:   … (do stuff here) …
12:   P(sema);
13:   i = get_input();
14:   f();
15:   V(sema);
16:   return 0;
17: }
```



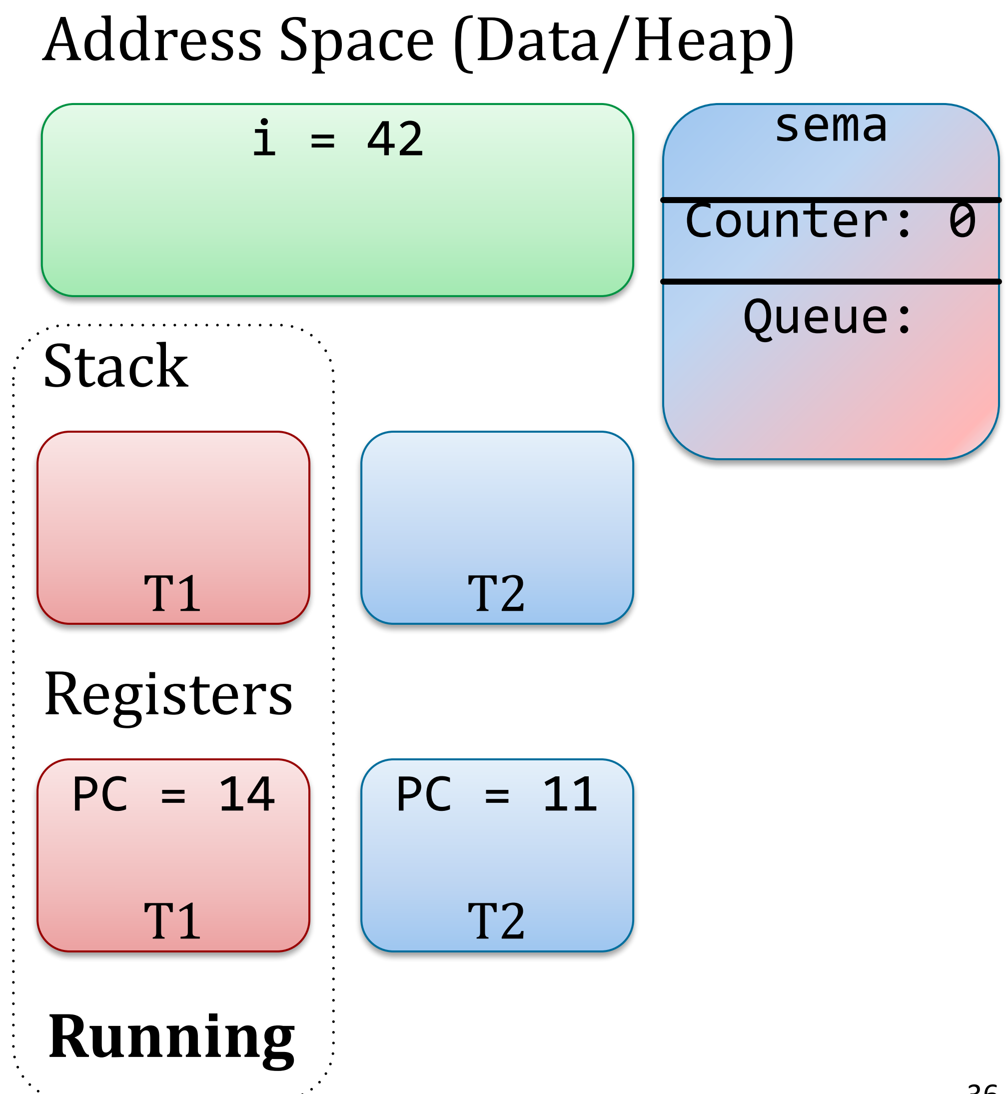
Threads – Revisited

```
 1: int i;
 2: Semaphore sema;
 3:
 4: f()
 5: {
 6:     printf("i is %d\n", i);
 7: }
 8:
 9: int main(int argc, char **argv)
10: {
11:   … (do stuff here) …
12:   P(sema);
13:   i = get_input();
14:   f();
15:   V(sema);
16:   return 0;
17: }
```



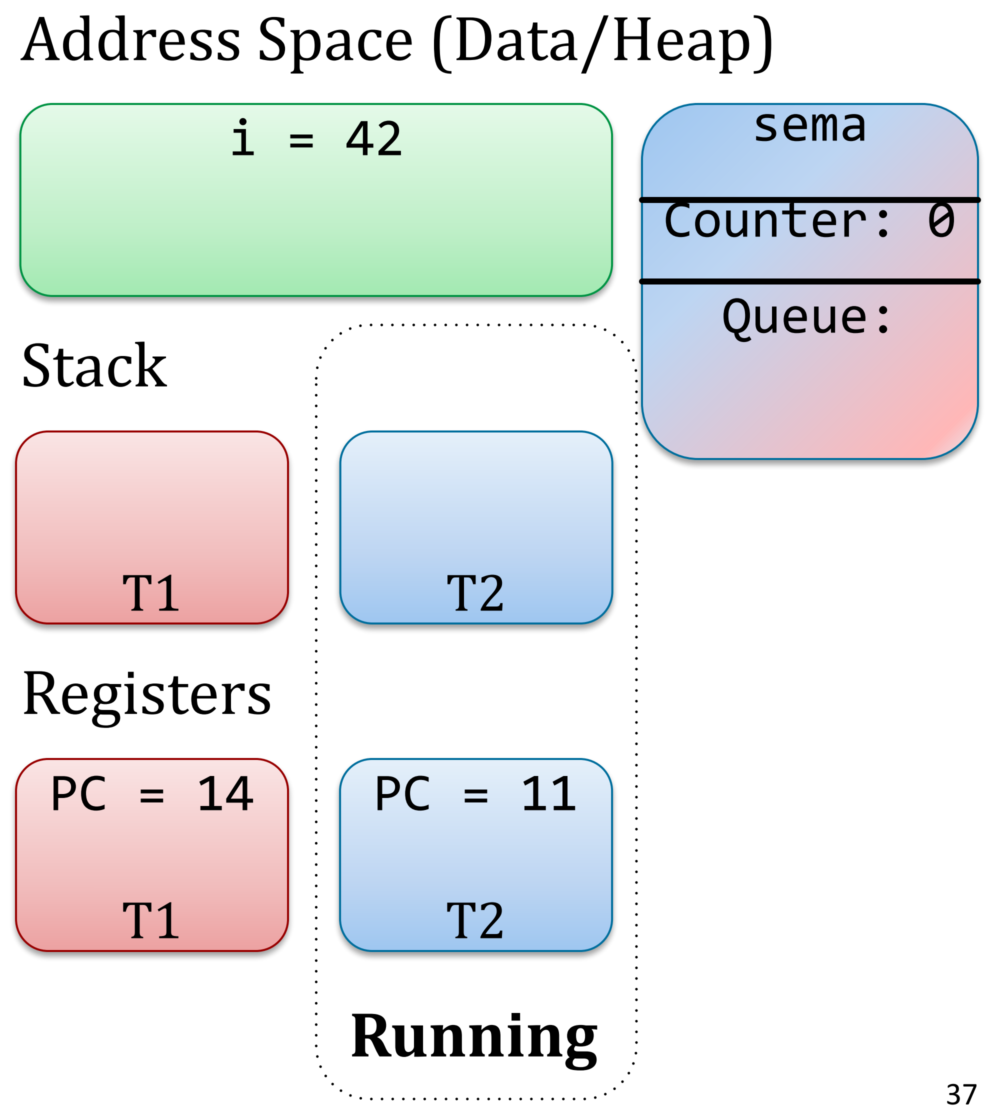
Threads – Revisited

```
 1: int i;
 2: Semaphore sema;
 3:
 4: f()
 5: {
 6:     printf("i is %d\n", i);
 7: }
 8:
 9: int main(int argc, char **argv)
10: {
11:    … (do stuff here) …
12:    P(sema);
13:    i = get_input();
14:    f();
15:    V(sema);
16:    return 0;
17: }
```



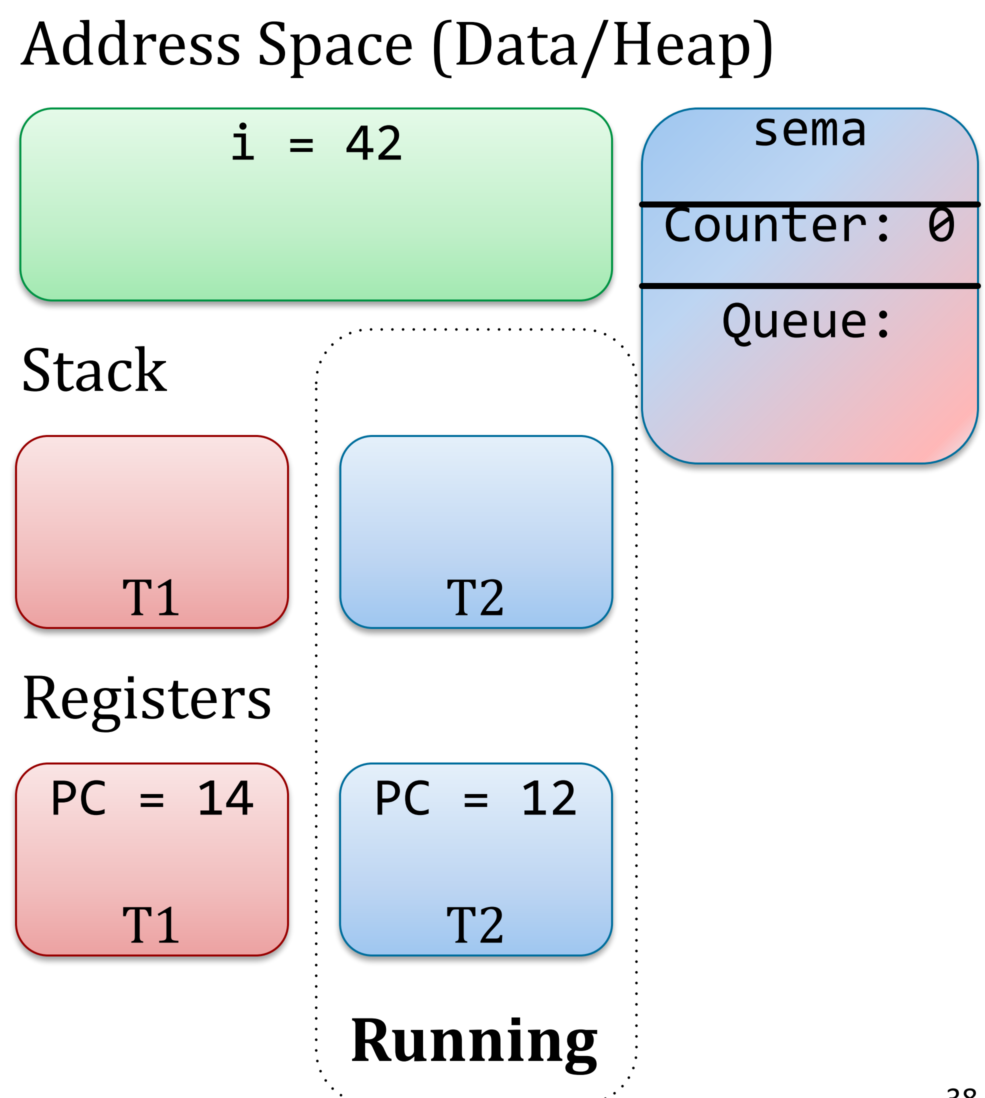
Threads – Revisited

```
 1: int i;
 2: Semaphore sema;
 3:
 4: f()
 5: {
 6:     printf("i is %d\n", i);
 7: }
 8:
 9: int main(int argc, char **argv)
10: {
11:   ... (do stuff here) ...
12:   P(sema);
13:   i = get_input();
14:   f();
15:   V(sema);
16:   return 0;
17: }
```



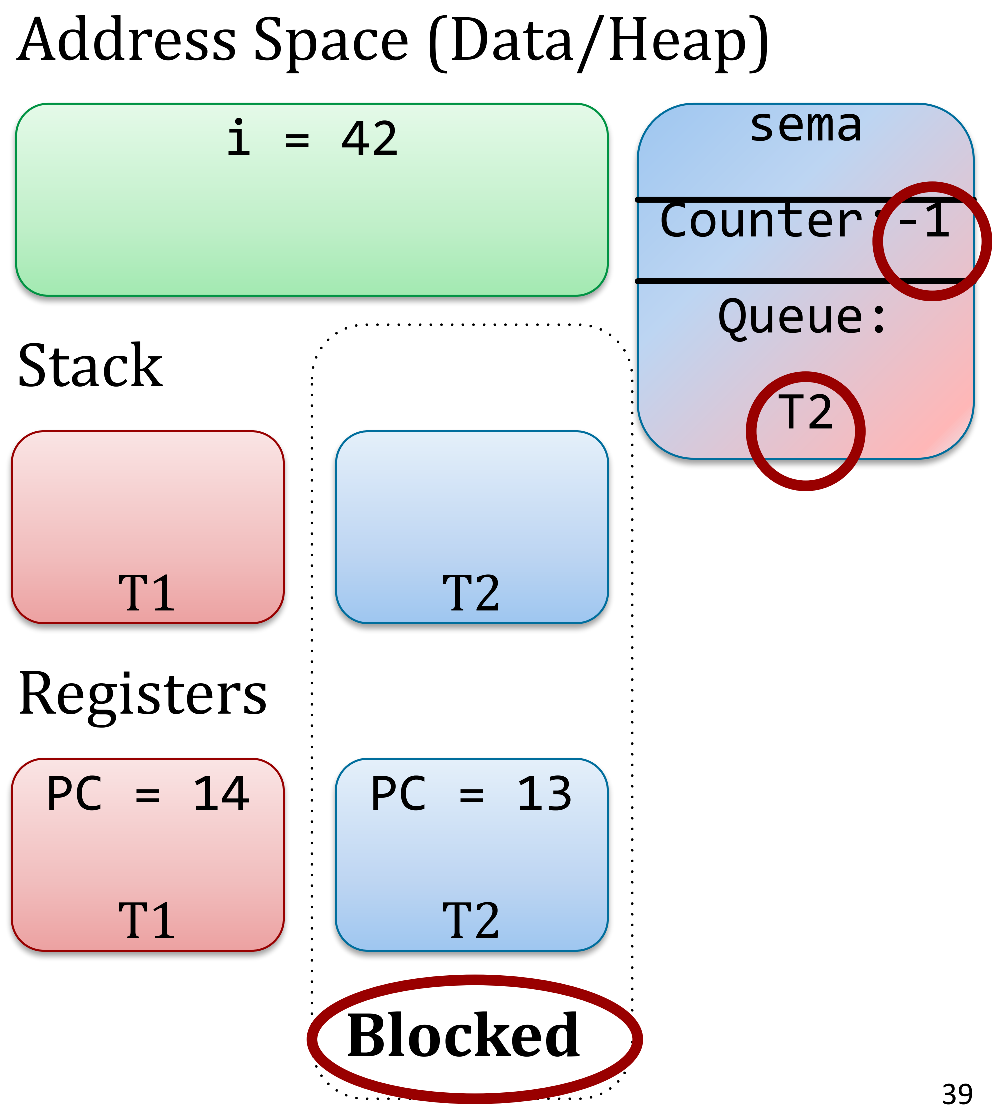
Threads – Revisited

```
 1: int i;
 2: Semaphore sema;
 3:
 4: f()
 5: {
 6:    printf("i is %d\n", i);
 7: }
 8:
 9: int main(int argc, char **argv)
10: {
11:   ... (do stuff here) ...
12:   P(sema);
13:   i = get_input();
14:   f();
15:   V(sema);
16:   return 0;
17: }
```



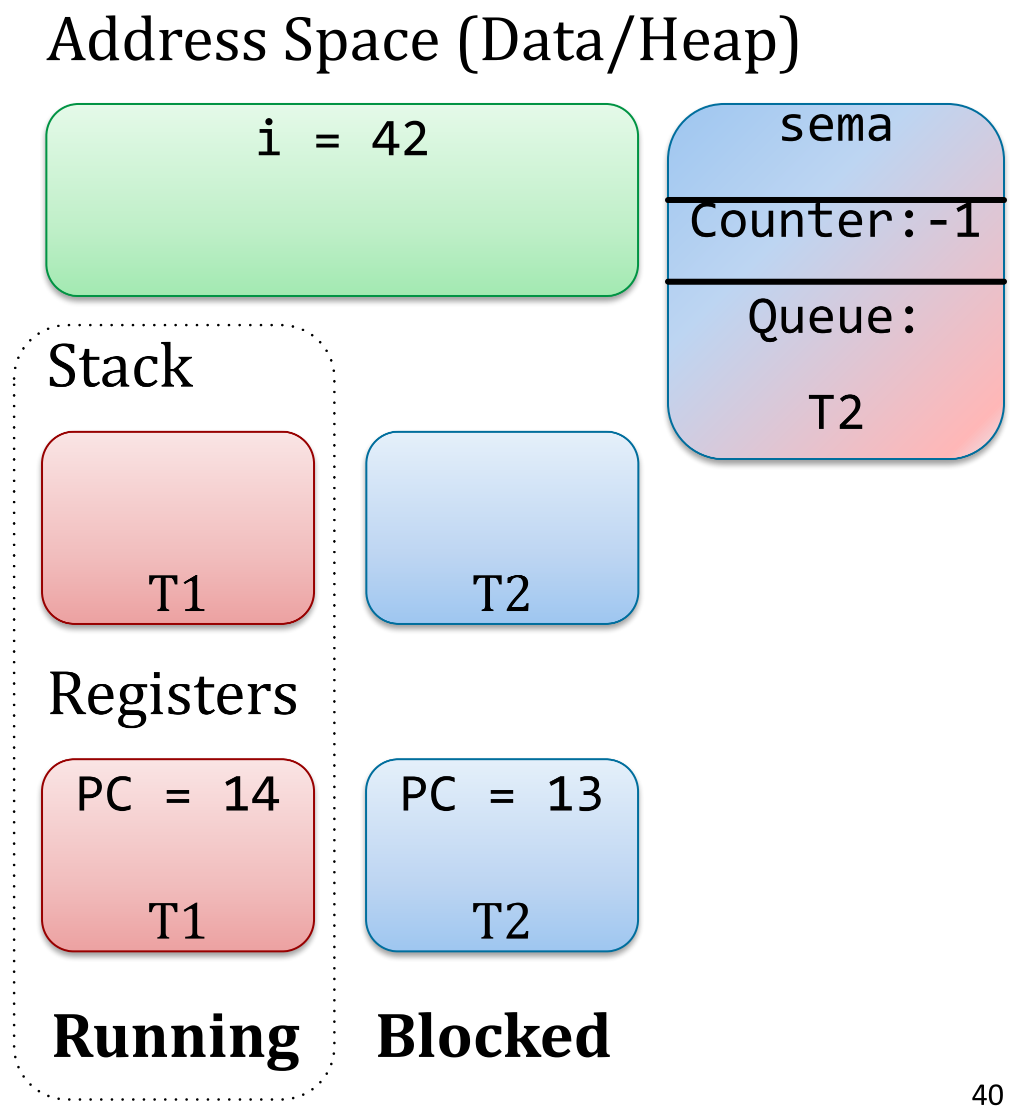
Threads – Revisited

```
 1: int i;
 2: Semaphore sema;
 3:
 4: f()
 5: {
 6:     printf("i is %d\n", i);
 7: }
 8:
 9: int main(int argc, char **argv)
10: {
11:    … (do stuff here) …
12:    P(sema);
13:    i = get_input();
14:    f();
15:    V(sema);
16:    return 0;
17: }
```



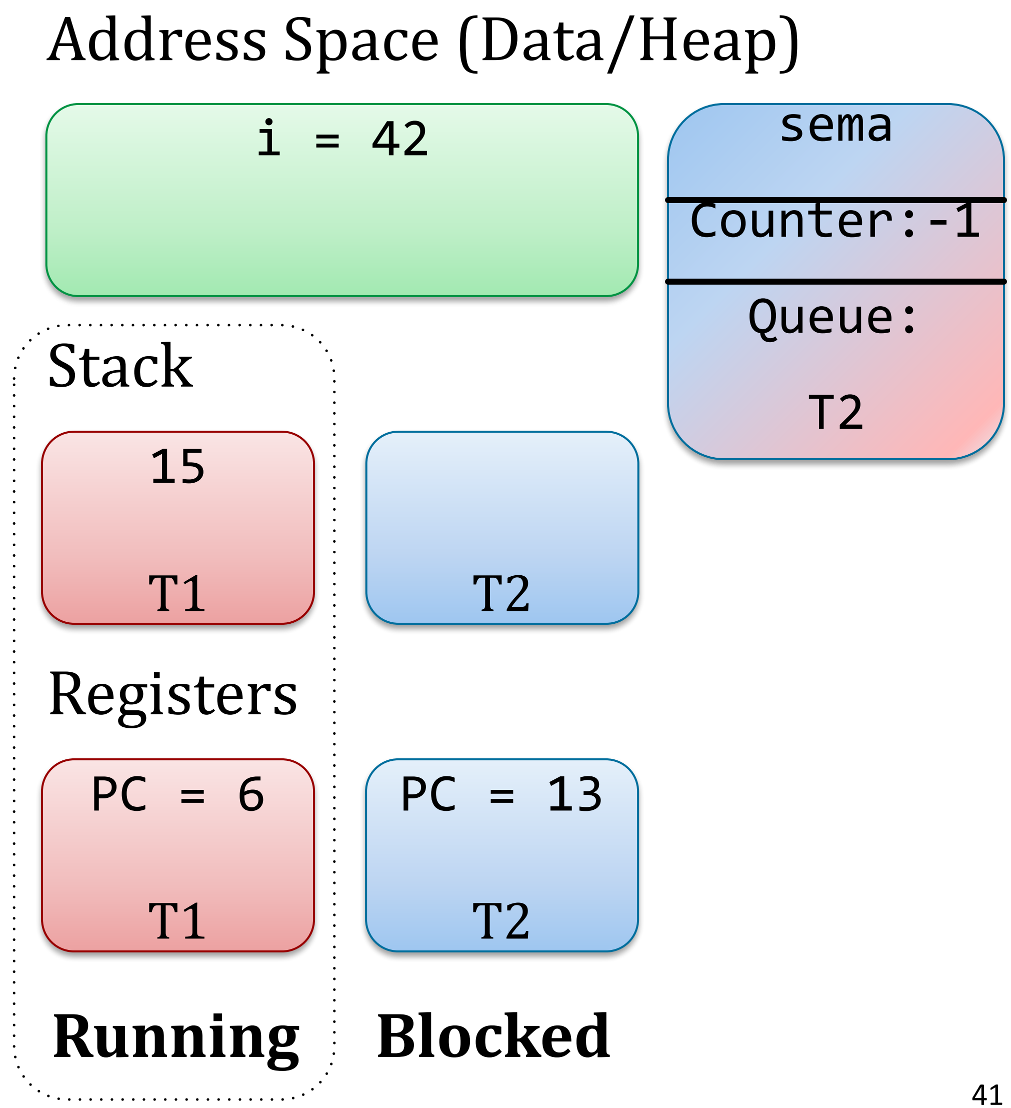
Threads – Revisited

```
 1: int i;
 2: Semaphore sema;
 3:
 4: f()
 5: {
 6:     printf("i is %d\n", i);
 7: }
 8:
 9: int main(int argc, char **argv)
10: {
11:    … (do stuff here) …
12:    P(sema);
13:    i = get_input();
14:    f();
15:    V(sema);
16:    return 0;
17: }
```



Threads – Revisited

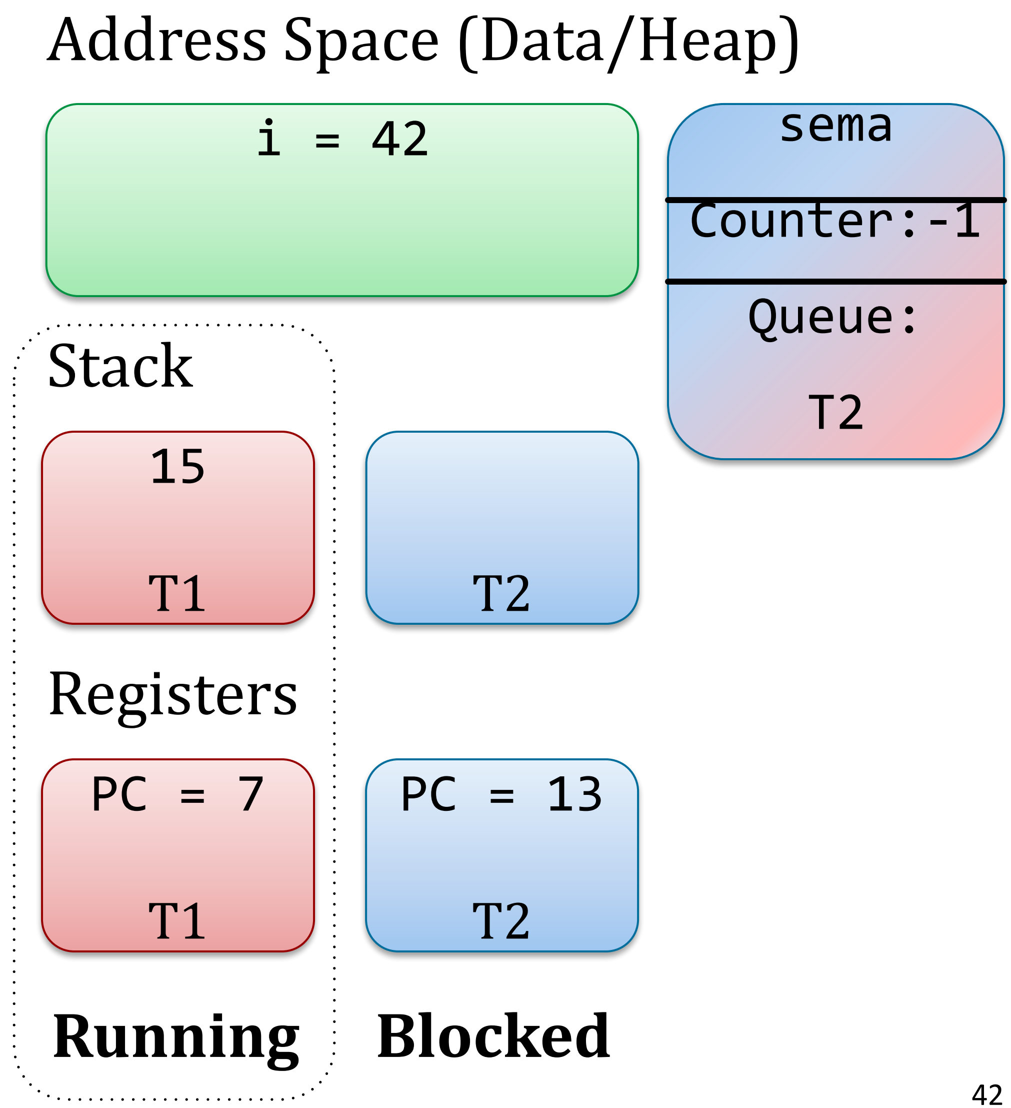
```
 1: int i;
 2: Semaphore sema;
 3:
 4: f()
 5: {
 6:     printf("i is %d\n", i);
 7: }
 8:
 9: int main(int argc, char **argv)
10: {
11:    … (do stuff here) …
12:    P(sema);
13:    i = get_input();
14:    f();
15:    V(sema);
16:    return 0;
17: }
```



Threads – Revisited

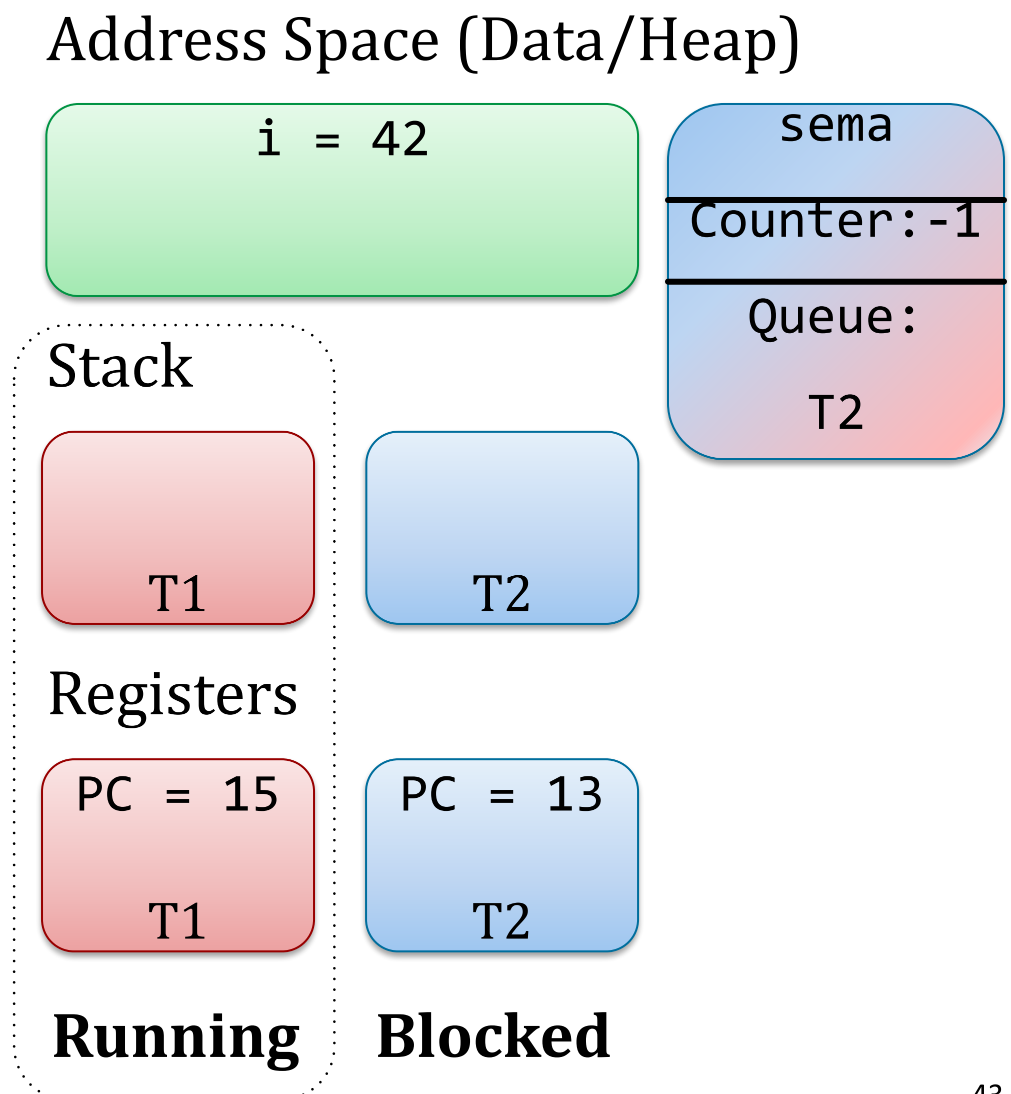
```
 1: int i;
 2: Semaphore sema;
 3:
 4: f()
 5: {
 6:    printf("i is %d\n", i);
 7: }
 8:
 9: int main(int argc, char **argv)
10: {
11:   ... (do stuff here) ...
12:   P(sema);
13:   i = get_input();
14:   f();
15:   V(sema);
16:   return 0;
17: }
```

i is 42



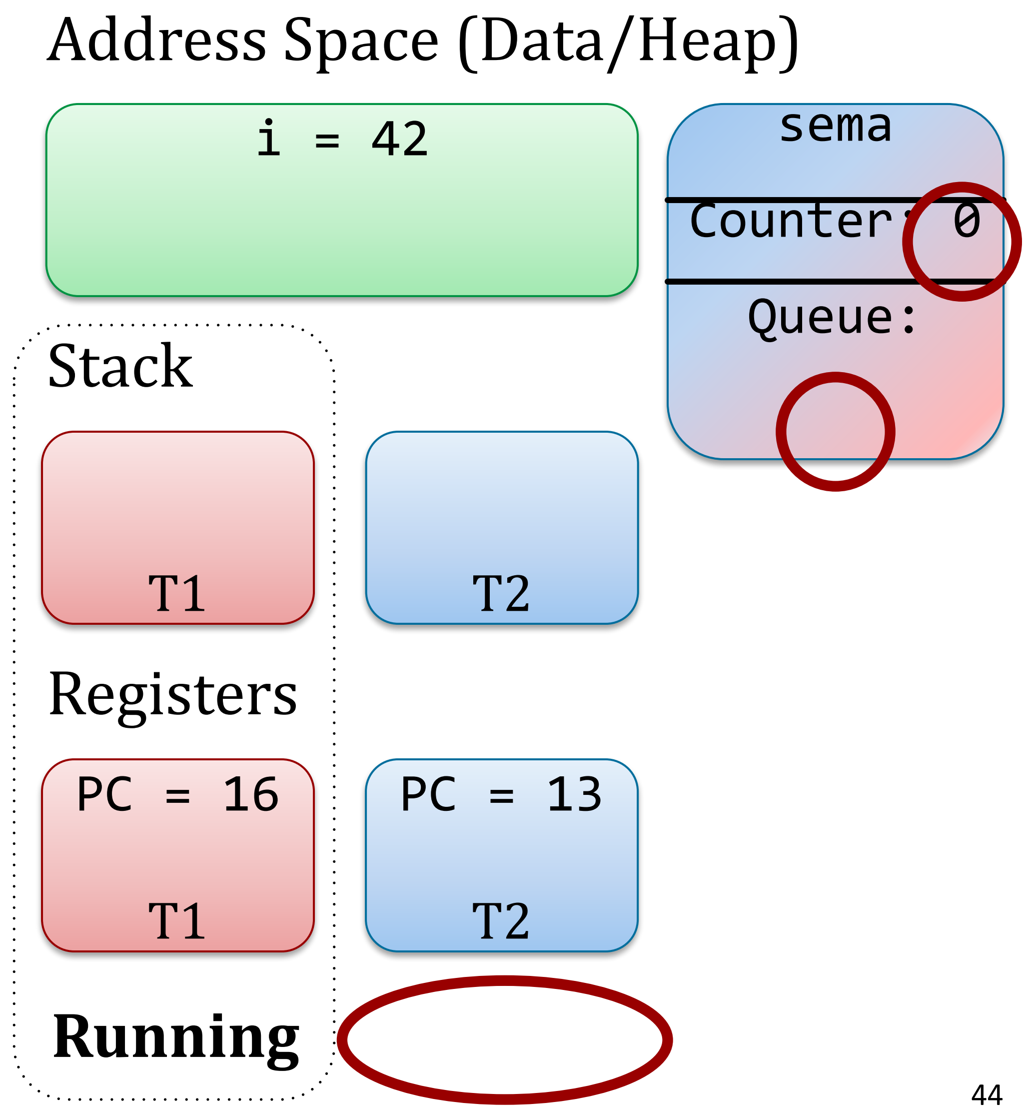
Threads – Revisited

```
1: int i;
2: Semaphore sema;
3:
4: f()
5: {
6:     printf("i is %d\n", i);
7: }
8:
9: int main(int argc, char **argv)
10: {
11:    ... (do stuff here) ...
12:    P(sema);
13:    i = get_input();
14:    f();
15:    V(sema);
16:    return 0;
17: }
```



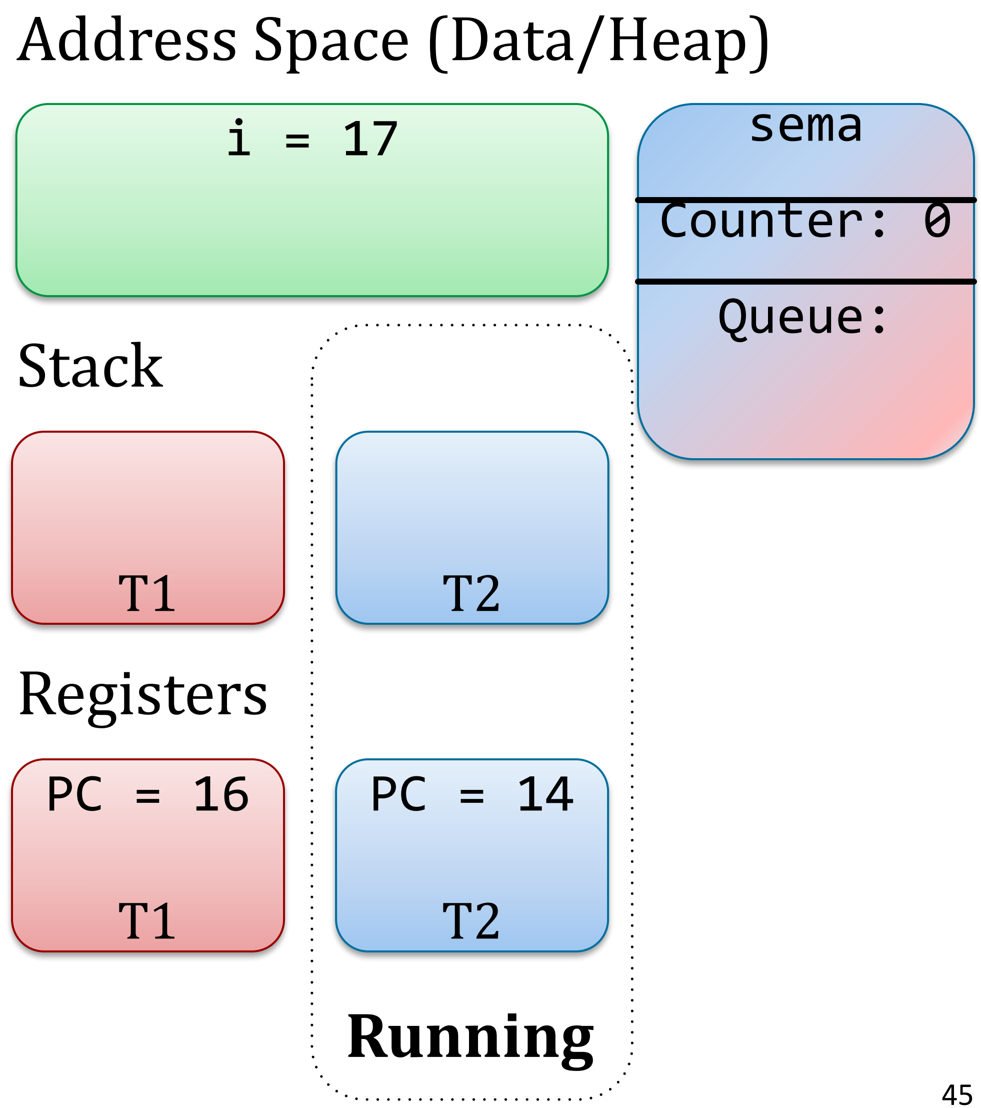
Threads – Revisited

```
 1: int i;
 2: Semaphore sema;
 3:
 4: f()
 5: {
 6:     printf("i is %d\n", i);
 7: }
 8:
 9: int main(int argc, char **argv)
10: {
11:    … (do stuff here) …
12:    P(sema);
13:    i = get_input();
14:    f();
15:    V(sema);
16:    return 0;
17: }
```



Threads – Revisited

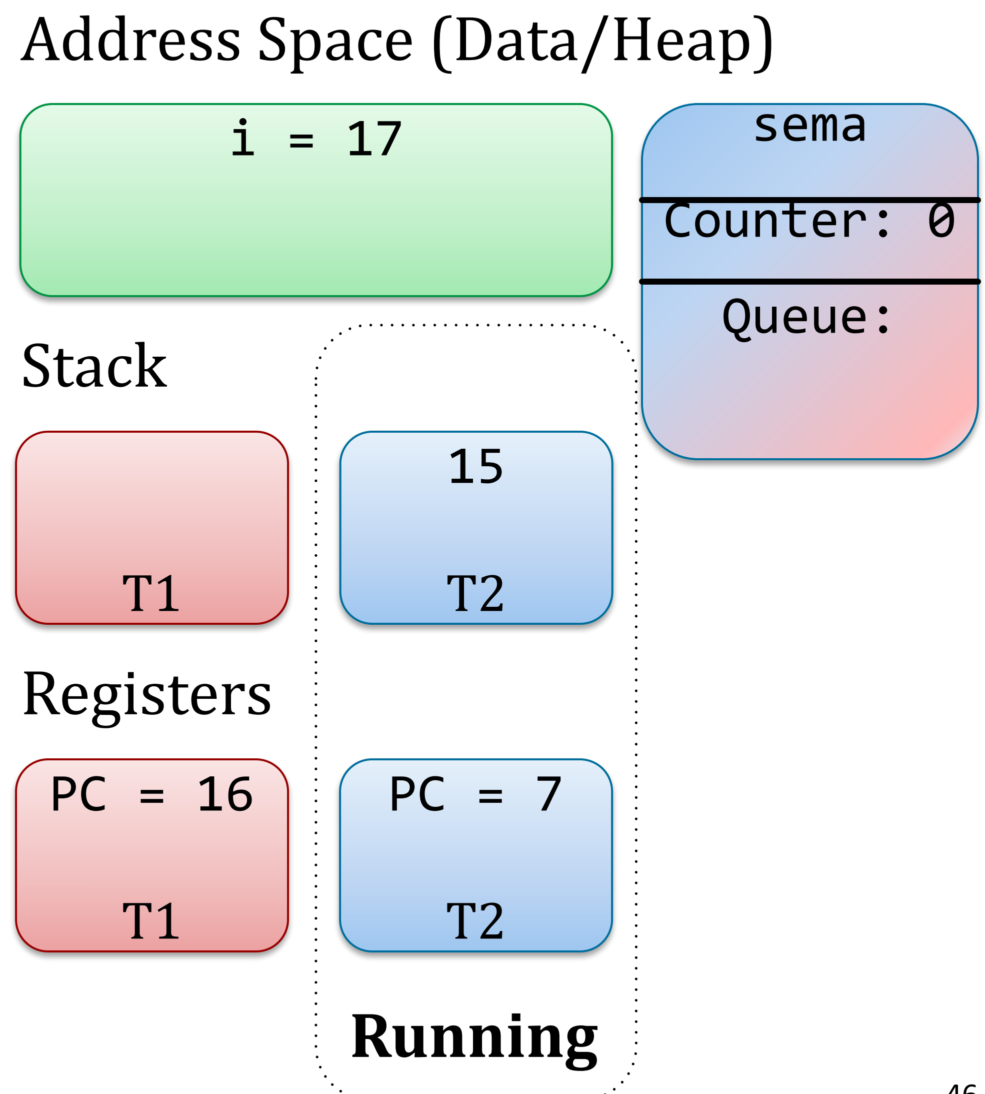
```
 1: int i;
 2: Semaphore sema;
 3:
 4: f()
 5: {
 6:     printf("i is %d\n", i);
 7: }
 8:
 9: int main(int argc, char **argv)
10: {
11:    … (do stuff here) …
12:    P(sema);
13:    i = get_input();
14:    f();
15:    V(sema);
16:    return 0;
17: }
```



Threads – Revisited

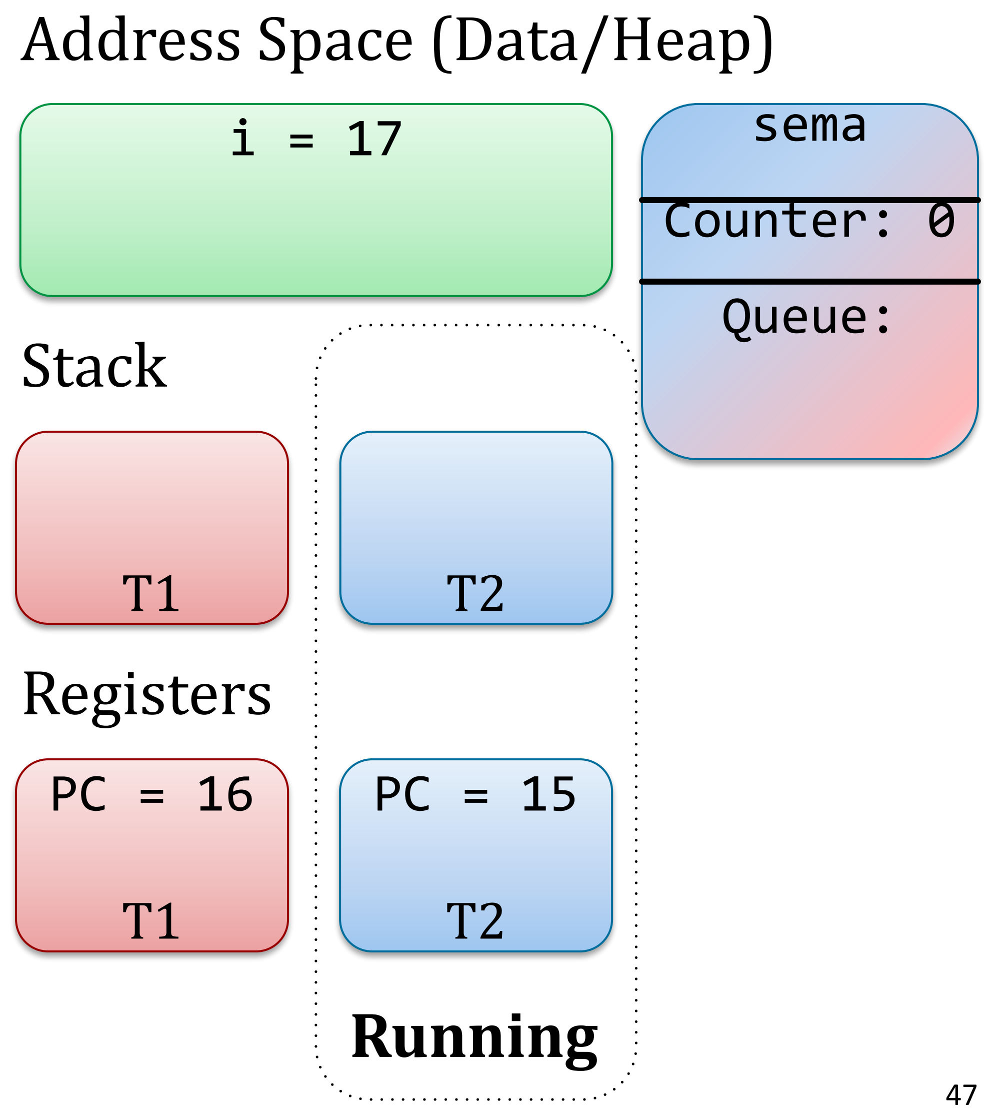
```
 1: int i;
 2: Semaphore sema;
 3:
 4: f()
 5: {
 6:    printf("i is %d\n", i);
 7: }
 8:
 9: int main(int argc, char **argv)
10: {
11:    ... (do stuff here) ...
12:    P(sema);
13:    i = get_input();
14:    f();
15:    V(sema);
16:    return 0;
17: }
```

i is 17



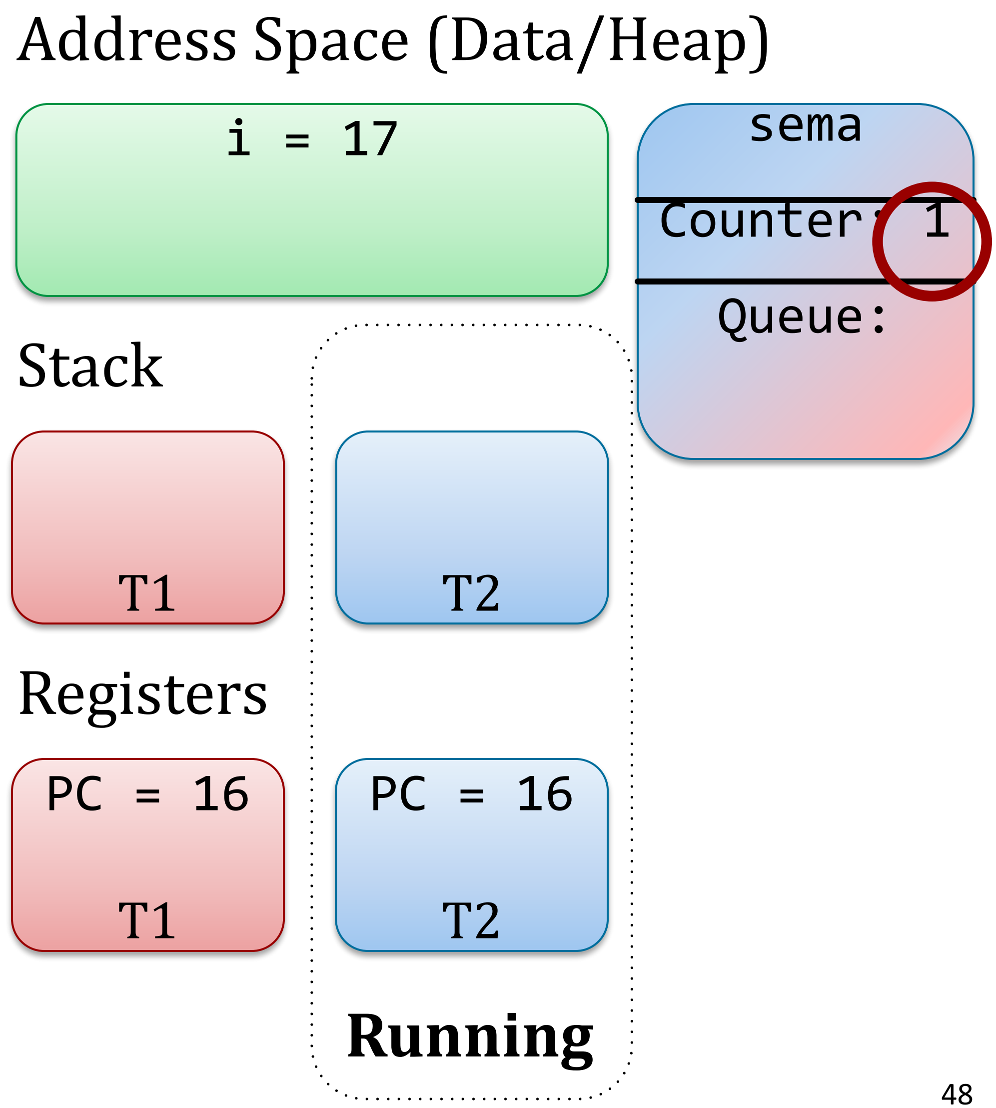
Threads – Revisited

```
 1: int i;
 2: Semaphore sema;
 3:
 4: f()
 5: {
 6:     printf("i is %d\n", i);
 7: }
 8:
 9: int main(int argc, char **argv)
10: {
11:    … (do stuff here) …
12:    P(sema);
13:    i = get_input();
14:    f();
15:    V(sema);
16:    return 0;
17: }
```



Threads – Revisited

```
 1: int i;
 2: Semaphore sema;
 3:
 4: f()
 5: {
 6:     printf("i is %d\n", i);
 7: }
 8:
 9: int main(int argc, char **argv)
10: {
11:    … (do stuff here) …
12:    P(sema);
13:    i = get_input();
14:    f();
15:    V(sema);
16:    return 0;
17: }
```



Producer/Consumer with Semaphores

Three semaphores:

1. `full`: counts the number of slots that are full
2. `empty`: keeps track of the empty slots
3. `mutex`: makes sure producer and consumer do not access the buffers at the same time

Initially:

- `full` = 0
- `empty` = N
- `mutex` = 1

Producer/Consumer with Semaphores

Producer

```
item = produce_item()
```

```
empty.down();
mutex.down();
```

```
insert_item(item);
```

```
mutex.up();
full.up();
```

Consumer

```
full.down();
mutex.down();
```

```
item = remove_item();
```

```
mutex.up();
empty.up();
```

```
consume_item(item);
```

Producer/Consumer with a Mistake …

Producer

```
item = produce_item()

mutex.down();
empty.down();

insert_item(item);

mutex.up();
full.up();
```

Consumer

```
full.down();
mutex.down();

item = remove_item();

mutex.up();
empty.up();

consume_item(item);
```

Monitors

- A monitor is collection of procedures, variables, and data structures grouped together in a special module
- Only one thread can be active in a monitor at any instant
- Mutual exclusion is enforced by the compiler and therefore it is less prone to errors
- Monitors introduce the concept of condition variables

```
monitor example
    integer i;
    condition c;

    procedure producer( );
    .
    .
    .
    end;

    procedure consumer( );
    .
    .
    .
    end;
end monitor;
```

Condition Variables

- Condition variables support two operations
  - Wait
  - Signal
- wait(condition): the calling thread blocks and allows another thread to enter the monitor
- signal(condition): the calling thread wakes up a thread blocked on the condition variable
  - If more than one thread is waiting, only one is selected by the scheduler
  - The signal operation must be the last statement executed, so that the caller immediately exits the monitor
- Condition variables do not keep track of signals as semaphores do

Producer/Consumer with Monitors

```
monitor ProducerConsumer
    condition full, empty;
    integer count;
    procedure insert(item: integer);
    begin
        if count = N then wait(full);
        insert_item(item);
        count := count + 1;
        if count = 1 then signal(empty)
    end;
    function remove: integer;
    begin
        if count = 0 then wait(empty);
        remove = remove_item;
        count := count − 1;
        if count = N − 1 then signal(full)
    end;
    count := 0;
end monitor;
```

```
procedure producer;
begin
    while true do
    begin
        item = produce_item;
        ProducerConsumer.insert(item)
    end
end;
procedure consumer;
begin
    while true do
    begin
        item = ProducerConsumer.remove;
        consume_item(item)
    end
end;
```

More on monitors

- Important to understand; history of OS
- Synchronization is one of the huge challenges: TUNIS demonstrated monitors hugely simplify
- Java implements in user level
- Irrelevant in modern OSes
  - Language level so not used in today's OSes
  - This is about locking code; that was fine with one core, but what about 16 cores or 32
  - Atomic operations now implemented in cache, enormously less expensive

The Readers and Writers Problem

The Readers and Writers Problem

- Multiple threads can read from a database at the same time
- If one thread is writing data into the db, no process should be reading or writing at the same time
- First reader gets a hold of a lock on the db
- Subsequent readers just increment the reader counter (critical section with a mutex)
- When they are finished they decrement the counter (critical section with a mutex)
- Last reader does an up() on the database lock letting the writer access the db

Reader/Writer Solution

```
reader() {
  mutex.down();
  readerCount++;
  if (readerCount==1) db.down();
  mutex.up();

  read_db();

  mutex.down();
  readerCount--;
  if (readerCount==0) db.up();
  mutex.up();
  use_db_data();
}
```

```
writer() {
  prepare_db_data();
  db.down();
  write_db_data();
  db.up();
}
```

Writer may starve if readers are too “active”