Shared Memory Programming with Pthreads

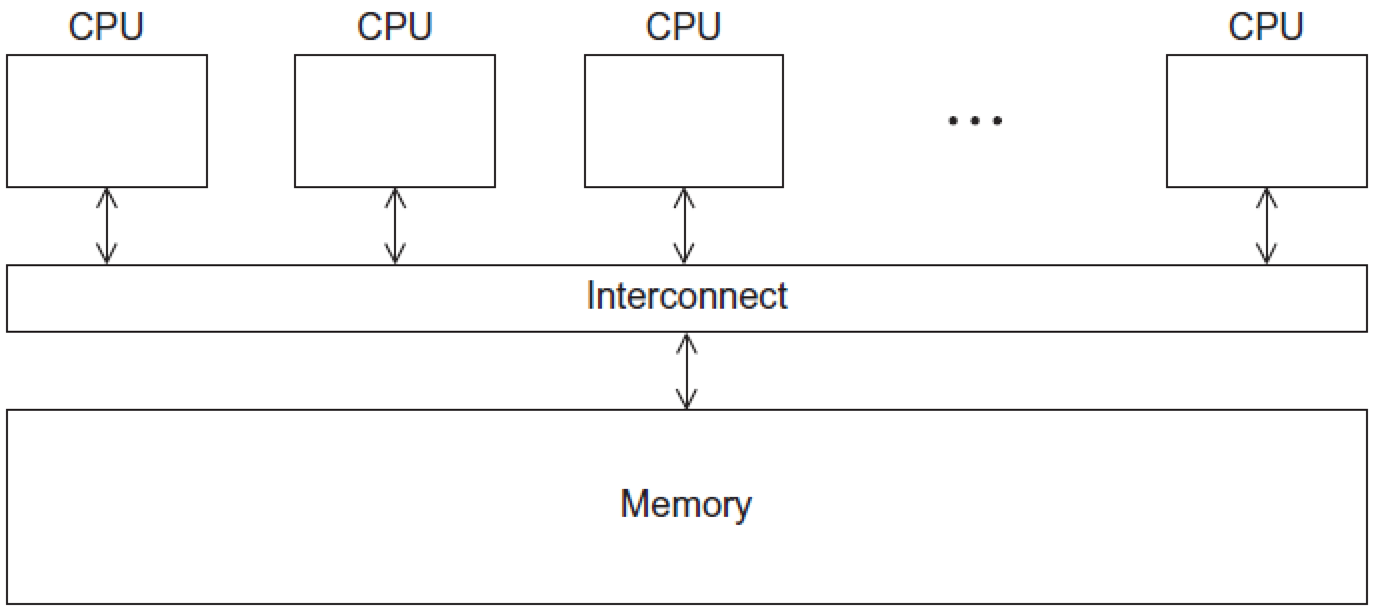
Pacheco. Chapter 4

T. Yang. UCSB CS140. Spring 2014

Outline

- **Shared memory programming: Overview**
- **POSIX pthreads**
- **Critical section & thread synchronization.**
  - Mutexes.
  - Producer-consumer synchronization and semaphores.
  - Barriers and condition variables.
  - Read-write locks.
- **Thread safety.**

Shared Memory Architecture



Processes and Threads

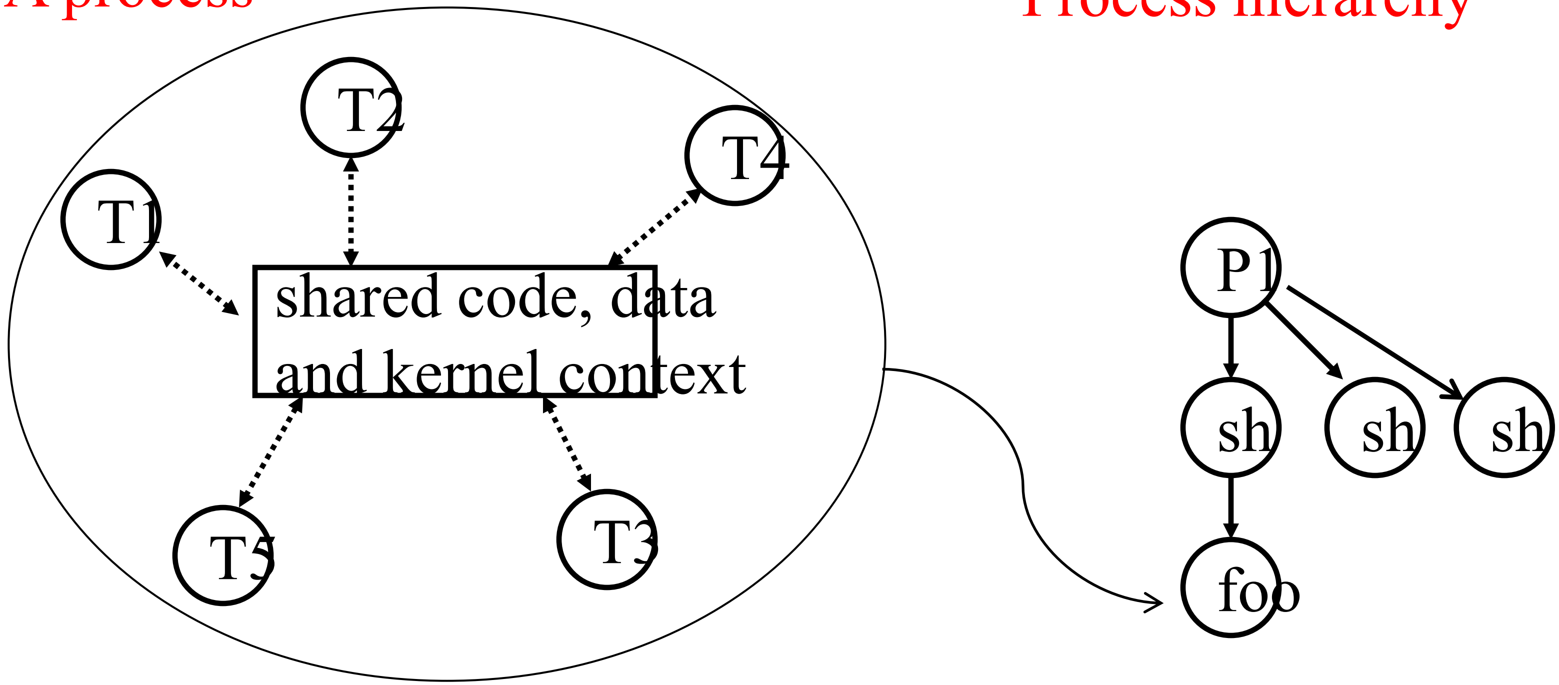
- **A process is an instance of a running (or suspended) program.**
- **Threads are analogous to a "light-weight" process.**
- **In a shared memory program a single process may have multiple threads of control.**

Logical View of Threads

- **Threads are created within a process**

A process

Process hierarchy

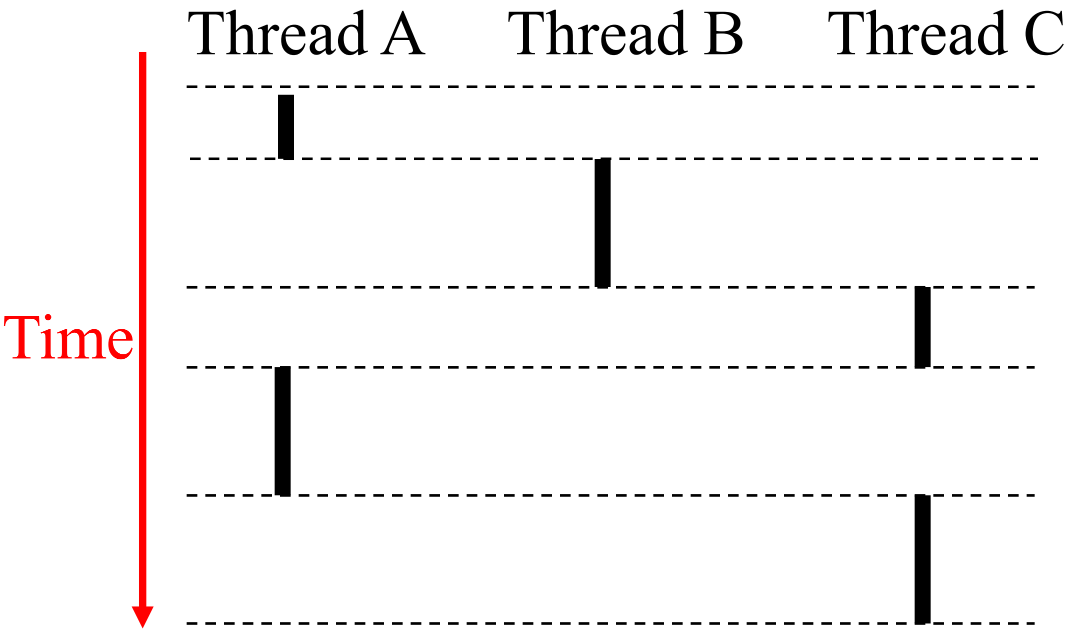


Concurrent Thread Execution

- **Two threads run concurrently if their logical flows overlap in time**
- **Otherwise, they are sequential (we'll see that processes have a similar rule)**
- **Examples:**
  - Concurrent:

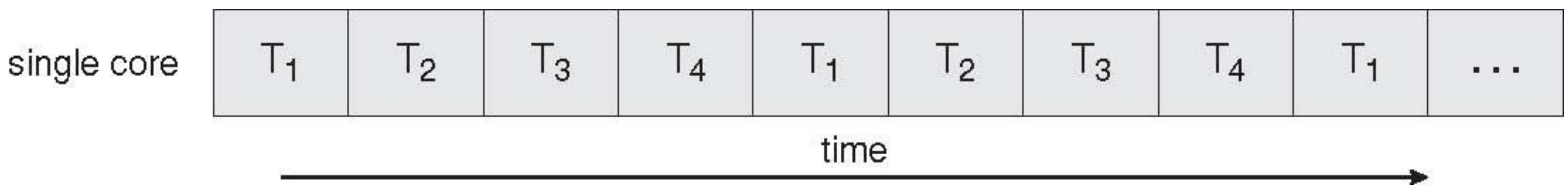
    A & B, A&C
  - Sequential:

    B & C

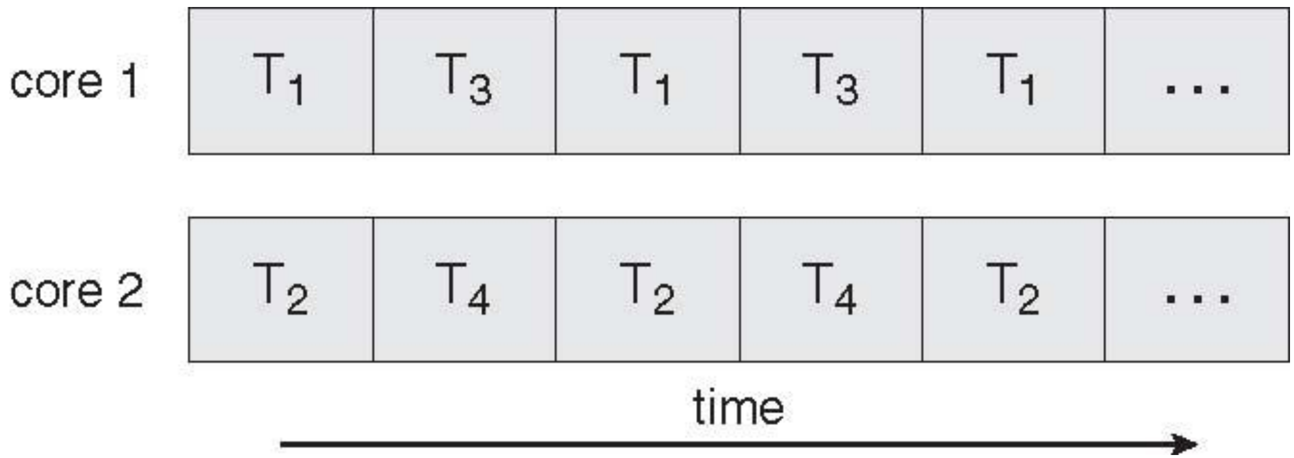


Execution Flow on one-core or multi-core systems

Concurrent execution on a single core system

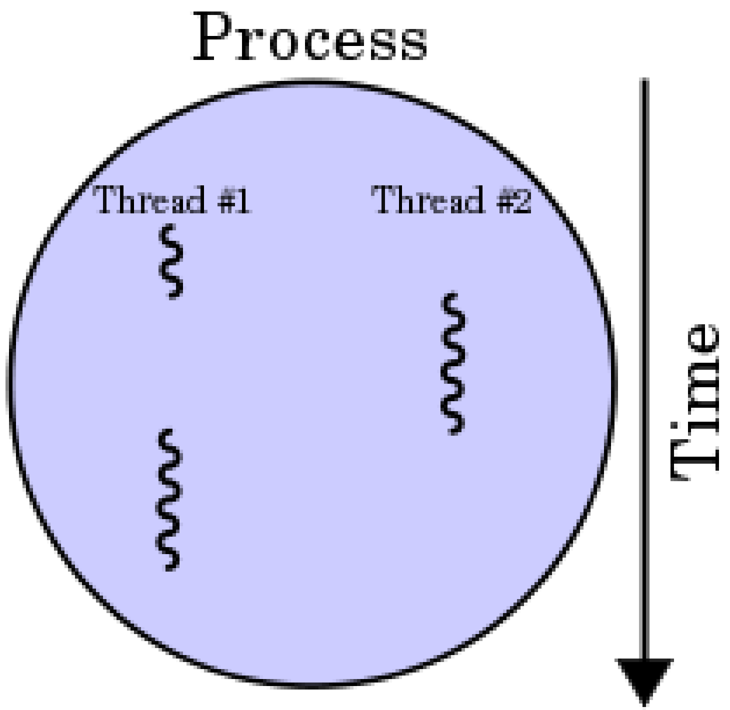


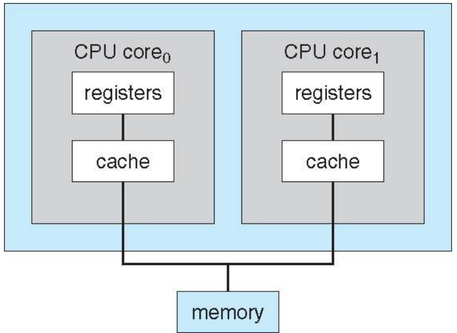
Parallel execution on a multi-core system



Benefits of multi-threading

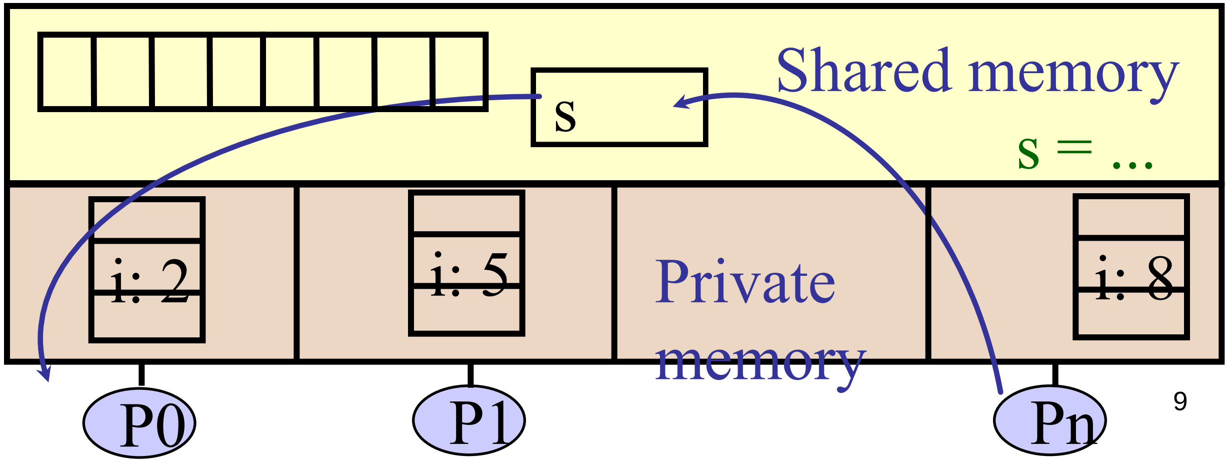
- **Responsiveness**
- **Resource Sharing**
  - Shared memory
- **Economy**
- **Scalability**
  - Explore multi-core CPUs





Thread Programming with Shared Memory

- **Program is a collection of threads of control.**
  - Can be created dynamically
- **Each thread has a set of private variables, e.g., local stack variables**
- **Also a set of shared variables, e.g., static variables, shared common blocks, or global heap.**
  - Threads communicate implicitly by writing and reading shared variables.
  - Threads coordinate by synchronizing on shared variables



Shared Memory Programming

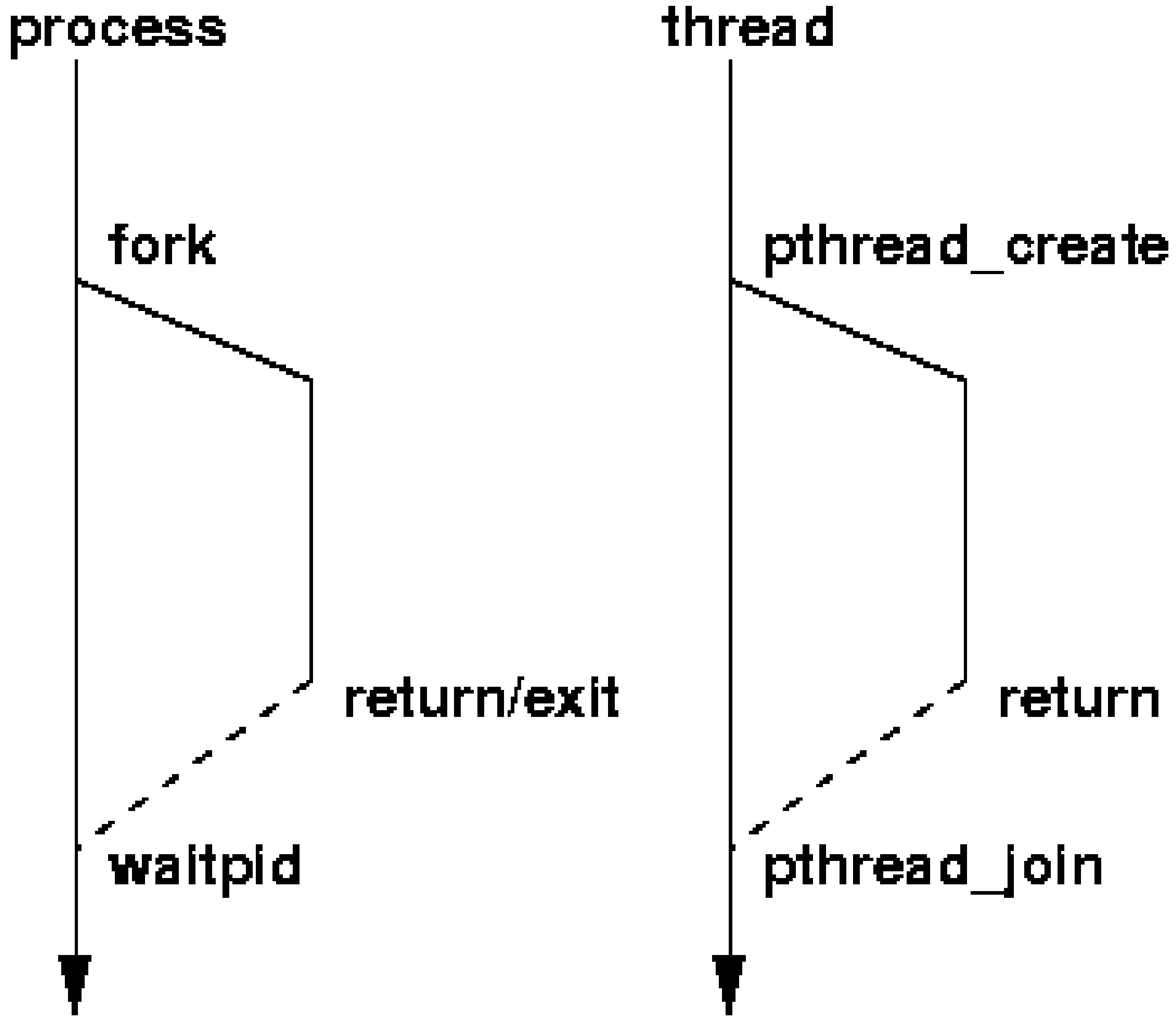
Several Thread Libraries/systems

- **Pthreads is the POSIX Standard**
  - Relatively low level
  - Portable but possibly slow; relatively heavyweight
- **OpenMP standard for application level programming**
  - Support for scientific programming on shared memory
  - http://www.openMP.org
- **TBB: Thread Building Blocks**
  - Intel
- **CILK: Language of the C "ilk"**
  - Lightweight threads embedded into C
- **Java threads**
  - Built on top of POSIX threads

Overview of POSIX Threads

- **POSIX: *Portable Operating System Interface for UNIX***
  - Interface to Operating System utilities
- **PThreads: The POSIX threading interface**
  - System calls to create and synchronize threads
  - In CSIL, compile a c program with gcc -lpthread
- **PThreads contain support for**
  - Creating parallelism and synchronization
  - No explicit support for communication, because shared memory is implicit; a pointer to shared data is passed to a thread

Creation of Unix processes vs. Pthreads



C function for starting a thread

pthread.h

One object for each thread.

pthread_t

```
int pthread_create (
        pthread_t*  thread_p /* out */ ,
        const pthread_attr_t*  attr_p /* in */ ,
        void*  (*start_routine ) ( void ) /* in */ ,
        void*  arg_p /* in */ ) ;
```

pthread_t objects

- Opaque
- The actual data that they store is system-specific.
- Their data members aren't directly accessible to user code.
- However, the Pthreads standard guarantees that a pthread_t object does store enough information to uniquely identify the thread with which it's associated.

A closer look (1)

```
int pthread_create (
        pthread_t*  thread_p /* out */ ,
        const pthread_attr_t*  attr_p /* in */ ,
        void*  (*start_routine ) ( void ) /* in */ ,
        void*  arg_p /* in */ ) ;
```

We won't be using, so we just pass NULL.

Allocate <u>before</u> calling.

A closer look (2)

```
int pthread_create (
        pthread_t*  thread_p /* out */ ,
        const pthread_attr_t*  attr_p /* in */ ,
        void*  (*start_routine ) ( void ) /* in */ ,
        void*  arg_p /* in */ ) ;
```

Pointer to the argument that should be passed to the function *start_routine*.

The function that the thread is to run.

Function started by pthread_create

- Prototype:
  void* thread_function (void* args_p) ;

- Void* can be cast to any pointer type in C.

- So args_p can point to a list containing one or more values needed by thread_function.
- Similarly, the return value of thread_function can point to a list of one or more values.

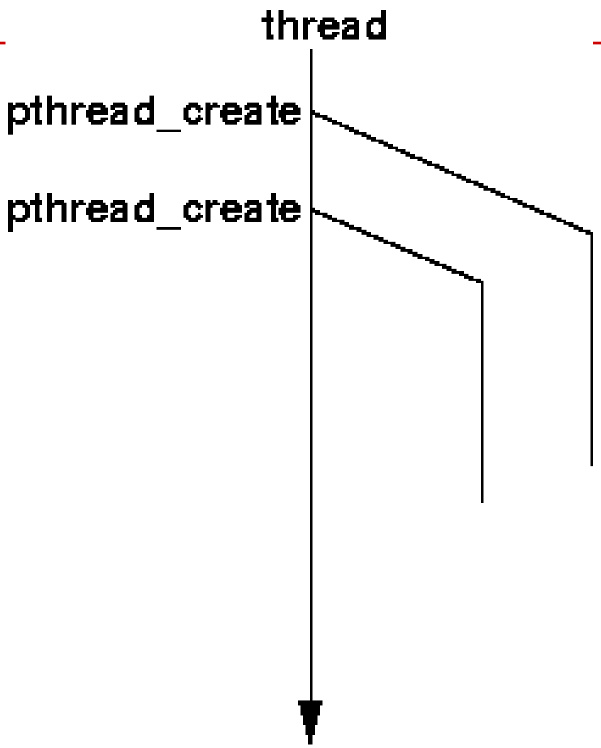
Wait for Completion of Threads

```
pthread_join(pthread_t *thread, void
   **result);
```

- Wait for specified thread to finish. Place exit value into *result.

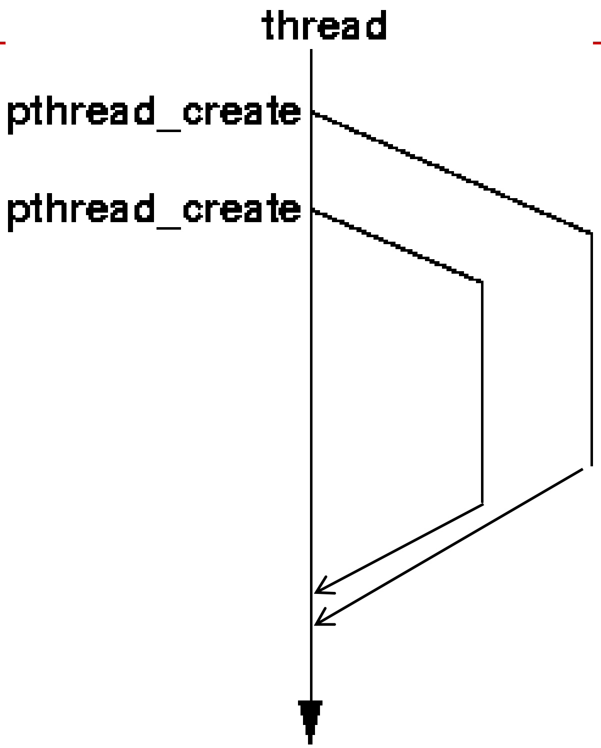
- **We call the function pthread_join once for each thread.**
- **A single call to pthread_join will wait for the thread associated with the pthread_t object to complete.**

Example of Pthreads




```
#include <pthread.h>
#include <stdio.h>
void *PrintHello(void * id){
   printf(“Thread%d: Hello World!\n", id);
}


void main (){
   pthread_t thread0, thread1;
   pthread_create(&thread0, NULL, PrintHello, (void *) 0);
   pthread_create(&thread1, NULL, PrintHello, (void *) 1);
}
```

Example of Pthreads with join




```
#include <pthread.h>
#include <stdio.h>
void *PrintHello(void * id){
   printf("Thread%d: Hello World!\n", id);
}

void main (){
   pthread_t thread0, thread1;
   pthread_create(&thread0, NULL, PrintHello, (void *) 0);
   pthread_create(&thread1, NULL, PrintHello, (void *) 1);
   pthread_join(thread0, NULL);
   pthread_join(thread1, NULL);
}
```

Some More Pthread Functions

- **`pthread_yield();`**
  - Informs the scheduler that the thread is willing to yield
- **`pthread_exit(void *value);`**
  - Exit thread and pass value to joining thread (if exists)

Others:

- **`pthread_t me; me = pthread_self();`**
  - Allows a pthread to obtain its own identifier pthread_t thread;
- **Synchronizing access to shared variables**
  - `pthread_mutex_init, pthread_mutex_[un]lock`
  - `pthread_cond_init, pthread_cond_[timed]wait`

Textbook Hello World example

declares the various Pthreads functions, constants, types, etc.

```
#include <stdio.h>
#include <stdlib.h>
#include <pthread.h>

/* Global variable:  accessible to all threads */
int thread_count;

void *Hello(void* rank);  /* Thread function */

int main(int argc, char* argv[]) {
   long        thread;  /* Use long in case of a 64-bit system */
   pthread_t* thread_handles;

   /* Get number of threads from command line */
   thread_count = strtol(argv[1], NULL, 10);

   thread_handles = malloc (thread_count*sizeof(pthread_t));
```

Hello World! (2)

```
for (thread = 0; thread < thread_count; thread++)
   pthread_create(&thread_handles[thread], NULL,
       Hello, (void*) thread);

printf("Hello from the main thread\n");

for (thread = 0; thread < thread_count; thread++)
   pthread_join(thread_handles[thread], NULL);

free(thread_handles);
return 0;
} /* main */
```

Hello World! (3)

```
void *Hello(void* rank) {
   long my_rank = (long) rank;  /* Use long in case of 64-bit system */

   printf("Hello from thread %ld of %d\n", my_rank, thread_count);

   return NULL;
}  /* Hello */
```

Compiling a Pthread program

gcc −g −Wall −o pth_hello pth_hello . c −lpthread

link in the Pthreads library

Running a Pthreads program

. /   pth_hello   <number of threads>

. / pth_hello 1

Hello from the main thread
Hello from thread 0 of 1

. / pth_hello 4

Hello from the main thread
Hello from thread 0 of 4
Hello from thread 1 of 4
Hello from thread 2 of 4
Hello from thread 3 of 4

Issues in Threads vs. Processes

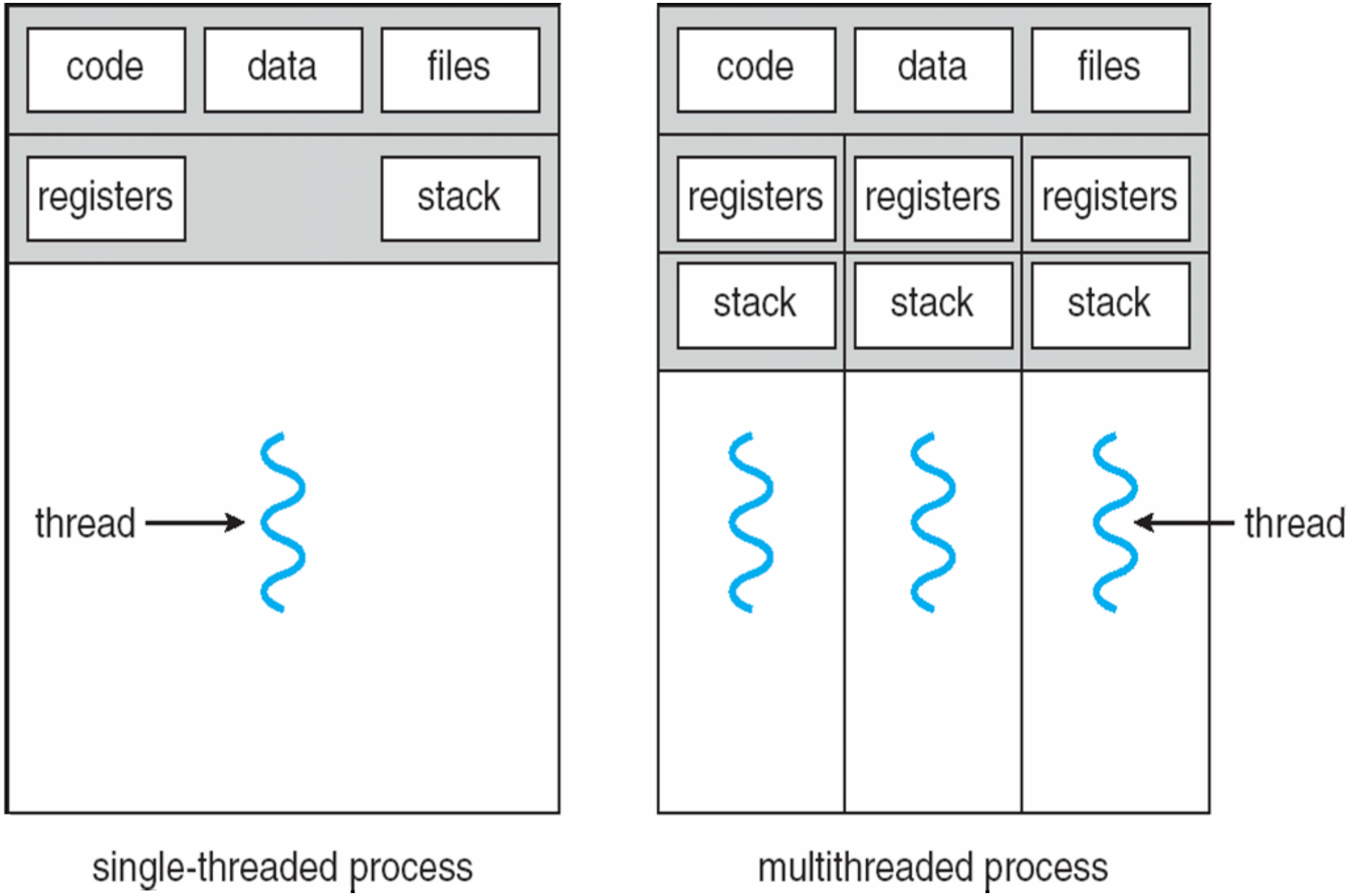
- **Shared variables as global variables exist in threads**
  - Can introduce subtle and confusing bugs!
  - Limit use of global variables to situations in which they’re really needed.

- **Starting threads**
  - Processes in MPI are usually started by a script.
  - In Pthreads the threads are started by the program executable.

Difference between Single and Multithreaded Processes

Shared memory access for code/data

Separate control flow -> separate stack/registers



Matrix-Vector Multiplication with Pthreads

Textbook P.159-162

Sequential code

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} * \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1*1+2*2+3*3 \\ 4*1+5*2+6*3 \\ 7*1+8*2+9*3 \end{pmatrix} = \begin{pmatrix} 14 \\ 32 \\ 50 \end{pmatrix}$$

```
/* For each row of A */
for (i = 0; i < m; i++) {
    y[i] = 0.0;
    /* For each element of the row and each element of x */
    for (j = 0; j < n; j++)
        y[i] += A[i][j]* x[j];
}
```

$$\begin{array}{|c|c|c|c|} \hline a_{00} & a_{01} & \cdots & a_{0,n-1} \\ \hline a_{10} & a_{11} & \cdots & a_{1,n-1} \\ \hline \vdots & \vdots & & \vdots \\ \hline a_{i0} & a_{i1} & \cdots & a_{i,n-1} \\ \hline \vdots & \vdots & & \vdots \\ \hline a_{m-1,0} & a_{m-1,1} & \cdots & a_{m-1,n-1} \\ \hline \end{array} \quad \begin{array}{|c|} \hline x_0 \\ x_1 \\ \vdots \\ x_{n-1} \\ \hline \end{array} = \begin{array}{|c|} \hline y_0 \\ \hline y_1 \\ \hline \vdots \\ \hline y_i = a_{i0}x_0 + a_{i1}x_1 + \cdots a_{i,n-1}x_{n-1} \\ \hline \vdots \\ \hline y_{m-1} \\ \hline \end{array}$$

Partitioning for Matrix-Vector Multiplication

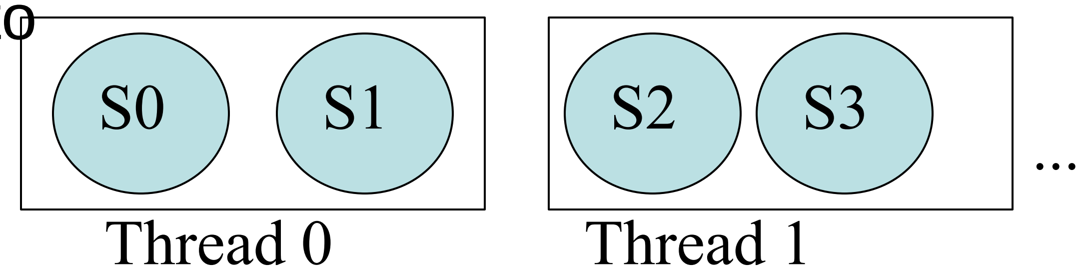
- Task partitioning

For (i=0; i<m; i=i+1)

```
Task Si for Row i
y[i]=0;
For (j=0; j<n; j=j+1)
      y[i]=y[i] +a[i][j]*x[j]
```

Task graph

Mapping to threads



Using 3 Pthreads for 6 Rows: 2 row per thread

Thread	Components of y	
0	y[0], y[1]	S0, S1
1	y[2], y[3]	S2, S3
2	y[4], y[5]	S4,S5

Code for S0

```
y[0] = 0.0;
for (j = 0; j < n; j++)
   y[0] += A[0][j]* x[j];
```

Code for Si

```
y[i] = 0.0;
for (j = 0; j < n; j++)
   y[i] += A[i][j]*x[j];
```

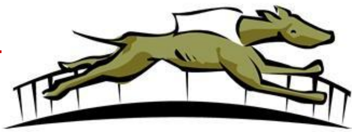
Pthread code for thread with ID rank

```
void *Pth_mat_vect(void* rank) {
   long my_rank = (long) rank;
   int i, j;
   int local_m = m/thread_count;
   int my_first_row = my_rank*local_m;
   int my_last_row = (my_rank+1)*local_m - 1;

   for (i = my_first_row; i <= my_last_row; i++) {
      y[i] = 0.0;
      for (j = 0; j < n; j++)
          y[i] += A[i][j]*x[j];
   }

   return NULL;
}  /* Pth_mat_vect */
```

Task Si

CRITICAL SECTIONS

Data Race Example



```
static int s = 0;
```

Thread 0

```
for i = 0, n/2-1
    s = s + f(A[i])
```

Thread 1

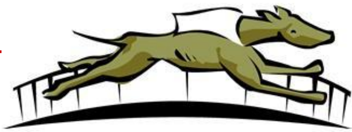
```
for i = n/2, n-1
    s = s + f(A[i])
```

- Also called critical section problem.
- A race condition or data race occurs when:
  - two processors (or two threads) access the same variable, and at least one does a write.
  - The accesses are concurrent (not synchronized) so they could happen simultaneously

Synchronization Solutions

1. **Busy waiting**
2. **Mutex (lock)**
3. **Semaphore**
4. **Conditional Variables**
5. **Barriers**

Example of Busy Waiting



```
static int s = 0;
static int flag=0
```

Thread 0

```
int temp, my_rank
for i = 0, n/2-1
    temp0=f(A[i])
    while flag!=my_rank;
    s = s + temp0
    flag= (flag+1) %2
```

Thread 1

```
int temp, my_rank
for i = n/2, n-1
    temp=f(A[i])
    while flag!=my_rank;
    s = s + temp
    flag= (flag+1) %2
```

- A thread repeatedly tests a condition, but, effectively, does no useful work until the condition has the appropriate value.
- Weakness: Waste CPU resource. Sometime not safe with compiler optimization.

Application Pthread Code: Estimating π

$$\pi = 4\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots + (-1)^n \frac{1}{2n+1} + \cdots\right)$$

```
double factor = 1.0;
double sum = 0.0;
for (i = 0; i < n; i++, factor = -factor) {
   sum += factor/(2*i+1);
}
pi = 4.0*sum;
```

Mapping for a multi-core machine

- Two thread distribution

Divide computation to 2 threads or more using block mapping. For example, n=20

Thread 0:	Thread 1:
Iterations 0, 1, 2, .., 9	Iterations 10, 11, 12, .., 19

- No of threads = thread_count
- No of iterations per thread   my_n= n/ thread_count
  - Assume it is an integer?
- Load assigned to my thread:
  - First iteration: my_n * my_rank
  - Last iteration: First iteration + my_n -1

A thread function for computing π

```
void* Thread_sum(void* rank) {
   long my_rank = (long) rank;
   double factor;
   long long i;
   long long my_n = n/thread_count;
   long long my_first_i = my_n*my_rank;
   long long my_last_i = my_first_i + my_n;

   if (my_first_i % 2 == 0) /* my_first_i is even */
      factor = 1.0;
   else /* my_first_i is odd */
      factor = -1.0;

   for (i = my_first_i; i < my_last_i; i++, factor = -factor) {
      sum += factor/(2*i+1);
   }

   return NULL;
}  /* Thread_sum */
```

Unprotected critical section.

Running results with 1 thread and 2 threads

	n			
	10^5	10^6	10^7	10^8
π	3.14159	3.141593	3.1415927	3.14159265
1 Thread	3.14158	3.141592	3.1415926	3.14159264
2 Threads	3.14158	3.141480	3.1413692	3.14164686

As n becomes larger,

- The one thread result becomes more accurate, gaining more correct digits
- The two-thread result is getting worse or strange

Possible race condition

Time	Thread 0	Thread 1
1	Started by main thread	
2	Call `Compute()`	Started by main thread
3	Assign `y = 1`	Call `Compute()`
4	Put `x=0` and `y=1` into registers	Assign `y = 2`
5	Add 0 and 1	Put `x=0` and `y=2` into registers
6	Store 1 in memory location `x`	Add 0 and 2
7		Store 2 in memory location `x`

Busy-Waiting

- **A thread repeatedly tests a condition, but, effectively, does no useful work until the condition has the appropriate value.**

- **Beware of optimizing compilers, though!**

```
y = Compute(my_rank);
while (flag != my_rank);
x = x + y;
flag++;
```

flag initialized to 0 by main thread

Pthreads global sum with busy-waiting

```
void* Thread_sum(void* rank) {
   long my_rank = (long) rank;
   double factor;
   long long i;
   long long my_n = n/thread_count;
   long long my_first_i = my_n*my_rank;
   long long my_last_i = my_first_i + my_n;

   if (my_first_i % 2 == 0)
      factor = 1.0;
   else
      factor = -1.0;

   for (i = my_first_i; i <
      while (flag != my_ran   );
      sum += factor/(2*i+1);
      flag = (flag+1) % thread_count;
   }

   return NULL;
}  /* Thread_sum */
```

sum is a shared global variable. Can we transform code and minimize thread interaction on this variable?

Global sum with local sum variable/busy waiting (1)

```
void* Thread_sum(void* rank) {
   long my_rank = (long) rank;
   double factor, my_sum = 0.0;
   long long i;
   long long my_n = n/thread_count;
   long long my_first_i = my_n*my_rank;
   long long my_last_i = my_first_i + my_n;

   if (my_first_i % 2 == 0)
      factor = 1.0;
   else
      factor = -1.0;
```

Global sum with local sum variable/busy waiting

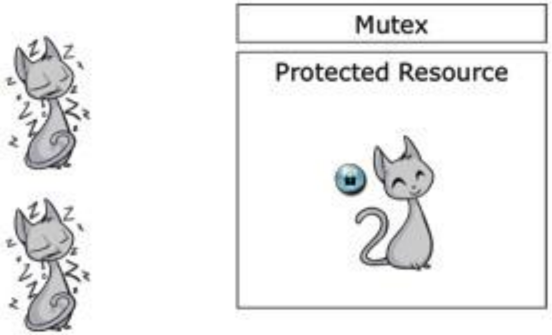
```
   for (i = my_first_i;                                    )
      my_sum += factor/(2*i

   while (flag != my_rank);
   sum += my_sum;
   flag = (flag+1) % thread_count;

   return NULL;
}  /* Thread_sum */
```

my_sum is a local variable, not shared.

Still have to contribute my_sum at the end to the global sum variable.

Mutexes (Locks)

- Code structure
  - Acquire mutex lock
  - Critical section
  - Unlock/Release mutex
- Mutex (mutual exclusion) is a special type of variable used to restrict access to a critical section to a single thread at a time.
- guarantee that one thread “excludes” all other threads while it executes the critical section**.**
- When A thread waits on a mutex/lock,
- CPU resource can be used by others.



Mutexes in Pthreads

- A special type for mutexes: pthread_mutex_t.

```
int pthread_mutex_init(
      pthread_mutex_t*            mutex_p   /* out */
      const pthread_mutexattr_t*  attr_p    /* in  */);
```

- To gain access to a critical section, call

```
int pthread_mutex_lock(pthread_mutex_t* mutex_p  /* in/out */);
```

- To release

```
int pthread_mutex_unlock(pthread_mutex_t* mutex_p  /* in/out */);
```

- When finishing use of a mutex, call

```
int pthread_mutex_destroy(pthread_mutex_t* mutex_p  /* in/out */);
```

Global sum function that uses a mutex (1)

```
void* Thread_sum(void* rank) {
   long my_rank = (long) rank;
   double factor;
   long long i;
   long long my_n = n/thread_count;
   long long my_first_i = my_n*my_rank;
   long long my_last_i = my_first_i + my_n;
   double my_sum = 0.0;

   if (my_first_i % 2 == 0)
      factor = 1.0;
   else
      factor = -1.0;
```

Global sum function that uses a mutex (2)

```
    for (i = my_first_i; i < my_last_i; i++, factor = -factor) {
        my_sum += factor/(2*i+1);
    }
    pthread_mutex_lock(&mutex);
    sum += my_sum;
    pthread_mutex_unlock(&mutex);

    return NULL;
}   /* Thread_sum */
```

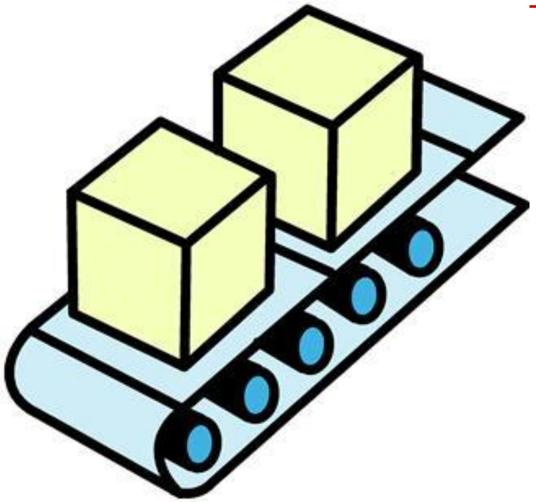
Threads	Busy-Wait	Mutex
1	2.90	2.90
2	1.45	1.45
4	0.73	0.73
8	0.38	0.38
16	0.50	0.38
32	0.80	0.40
64	3.56	0.38

$$\frac{T_{\text{serial}}}{T_{\text{parallel}}} \approx \texttt{thread\_count}$$

Run-times (in seconds) of π programs using n = 108 terms on a system with two four-core processors.

Time	flag	Thread				
		0	1	2	3	4
0	0	crit sect	busy wait	susp	susp	susp
1	1	terminate	crit sect	susp	busy wait	susp
2	2	—	terminate	susp	busy wait	busy wait
⋮	⋮			⋮	⋮	⋮
?	2	—	—	crit sect	susp	busy wait

Possible sequence of events with busy-waiting and more threads than cores.



Producer-consumer Synchronization and Semaphores

Issues

- **Busy-waiting enforces the order threads access a critical section.**
- **Using mutexes, the order is left to chance and the system.**
- **There are applications where we need to control the order threads access the critical section.**

Why Semaphores?

Synchronization	Functionality/weakness
Busy waiting	Spinning for a condition. Waste resource. Not safe
Mutex lock	Support code with simple mutual exclusion
Semaphore	Handle more complex signal-based synchronization

- **Examples of complex synchronization**
  - Allow a resource to be shared among multiple threads.
    - Mutex: no more than 1 thread for one protected region.
  - Allow a thread waiting for a condition after a signal
    - E.g. Control the access order of threads entering the critical section.
    - For mutexes, the order is left to chance and the system.

Problems with a mutex solution in multiplying many matrices

product_mat= A*B*C

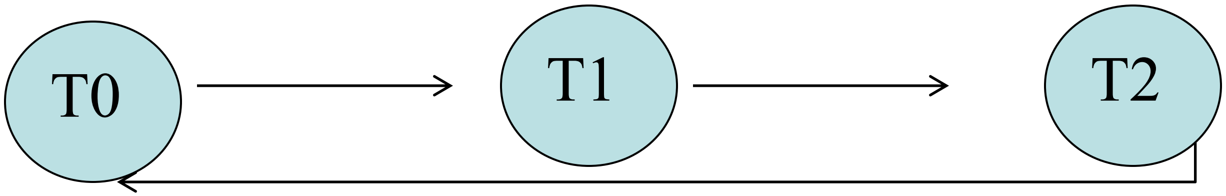
Out of order multiplication → product_mat= A*C*B

That is wrong

```
/* n and product_matrix are shared and initialized by the main thread *
/* product_matrix is initialized to be the                            *
void* Thread_work(void* rank) {
   long my_rank = (long) rank;
   matrix_t my_mat = Allocate_matrix(n);
   Generate_matrix(my_mat);
   pthread_mutex_lock(&mutex);
   Multiply_matrix(product_mat, my_mat);
   pthread_mutex_unlock(&mutex);
   Free_matrix(&my_mat);
   return NULL;
}  /* Thread_work */
```

The order of multiplication is not defined

Producer-Consumer Example



- Thread x produces a message for Thread x+1.
  - Last thread produces a message for thread 0.
- Each thread prints a message sent from its source.
- Will there be many null messages printed?
  - A consumer thread prints its source message before this message is produced.
  - How to avoid that?

First attempt at sending messages using pthreads

```
/* messages has type char**. It's allocated in main. */
/* Each entry is set to NULL in main.                */
void *Send_msg(void* rank) {
   long my_rank = (long) rank;
   long dest = (my_rank + 1) % thread_count;
   long source = (my_rank + thread_count - 1) % thread_count;
   char* my_msg = malloc(MSG_MAX*sizeof(char));

   sprintf(my_msg, "Hello to
   messages[dest] = my_msg;

   if (messages[my_rank] != NULL)
      printf("Thread %ld > %s\n", my_rank, messages[my_rank]);
   else
      printf("Thread %ld > No message from %ld\n", my_rank, source);

   return NULL;
}  /* Send_msg */
```

Produce a message for a destination thread

Consume a message

Semaphore: Generalization from mutex locks

- **Semaphore *S* – integer variable**
- **Can only be accessed /modified via two (atomic) operations**
  - wait (S) {   //also called P()

        while S <= 0 wait in a queue;

         S--;

      }
  - post(S) {   //also called V()

        S++;

        Wake up a thread that waits in the queue.

      }

'I think Lassie is trying to tell us something, ma.'

Syntax of the various semaphore functions

Semaphores are not part of Pthreads;
you need to add this.

```
#include <semaphore.h>

int sem_init(
      sem_t*      semaphore_p    /* out */,
      int         shared         /* in  */,
      unsigned    initial_val    /* in  */);


int sem_destroy(sem_t*   semaphore_p   /* in/out */);
int sem_post(sem_t*      semaphore_p   /* in/out */);
int sem_wait(sem_t*      semaphore_p   /* in/out */);
```

Message sending with semaphores

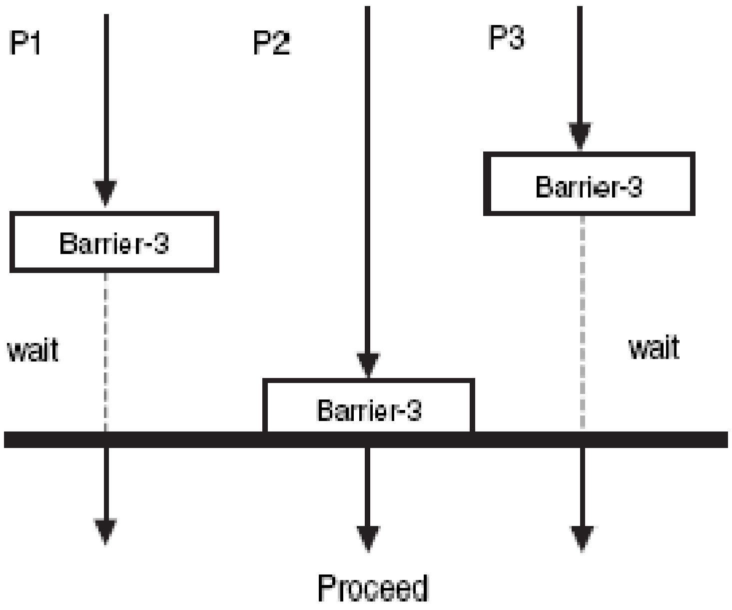
```
sprintf(my_msg, "Hello to %ld from %ld", dest, my_rank);
messages[dest] = my_msg;

sem_post(&semaphores[dest]);
                                /* signal the dest thread*/
sem_wait(&semaphores[my_rank]);
      /* Wait until the source message is created */

printf("Thread %ld > %s\n", my_rank,
  messages[my_rank]);
```

BARRIERS AND CONDITION VARIABLES

Barriers

- Synchronizing the threads to make sure that they all are at the same point in a program is called a barrier.

- No thread can cross the barrier until all the threads have reached it.



Application: Start timing of all threads at a fixed point.

```
/* Shared */
double elapsed_time;
. . .
/* Private */
double my_start, my_finish, my_elapsed;
. . .
Synchronize threads;
Store current time in my_start;
/* Execute timed code */
. . .
Store current time in my_finish;
my_elapsed = my_finish − my_start;

elapsed = Maximum of my_elapsed values;
```

Using barriers for debugging

```
point in program we want to reach;
barrier;
if (my_rank == 0) {
    printf("All threads reached this point\n");
    fflush(stdout);
}
```

Implement a barrier with busy-waiting and a mutex

- A shared counter as # of threads waiting in this point.

```
/* Shared and initialized by the main thread */
int counter;  /* Initialize to 0 */
int thread_count;
pthread_mutex_t barrier_mutex;
. . .

void* Thread_work(. . .) {
   . . .
   /* Barrier */
   pthread_mutex_lock(&barrier_mutex);
   counter++;
   pthread_mutex_unlock(&barrier_mutex);
   while (counter < thread_count);
   . . .
}
```

Need one counter variable for each instance of the barrier, otherwise problems are likely to occur.

Implementing a barrier with semaphores

```
/* Shared variables */
int counter;            /* Initialize to 0 */
sem_t count_sem;        /* In
sem_t barrier_sem;      /* Initialize to 0 */
. . .
void* Thread_work(...) {
   . . .
   /* Barrier */
   sem_wait(&count_sem);
   if (counter == thread_count-1) {
      counter = 0;
      sem_post(&count_sem);
      for (j = 0; j < thread_count-1; j++)
         sem_post(&barrier_sem);
   } else {
      counter++;
      sem_post(&count_sem);
      sem_wait(&barrier_sem);
   }
   . . .
```

Protect counter

Wait all threads to come

Condition Variables

- Why?
- More programming primitives to simplify code for synchronization of threads

Synchronization	Functionality
Busy waiting	Spinning for a condition. Waste resource. Not safe
Mutex lock	Support code with simple mutual exclusion
Semaphore	Signal-based synchronization. Allow sharing (not wait unless semaphore=0)
Barrier	Rendezvous-based synchronization
Condition variables	More complex synchronization: Let threads wait until a user-defined condition becomes true

Synchronization Primitive: Condition Variables

- Used together with a lock
- One can specify more general waiting condition compared to semaphores.
- A thread is blocked when condition is no true:
  - placed in a waiting queue, yielding CPU resource to somebody else.
  - Wake up until receiving a signal

Pthread synchronization: Condition variables

int status;   pthread_condition_t cond;

const pthread_condattr_t attr;

pthread_mutex mutex;

status = pthread_cond_init(&cond,&attr);

status = pthread_cond_destroy(&cond);

status = pthread_cond_wait(&cond,&mutex);

-wait in a queue until somebody wakes up. Then the mutex is reacquired.

status = pthread_cond_signal(&cond);

- wake up one waiting thread.

status = pthread_cond_broadcast(&cond);

- wake up all waiting threads in that condition

How to Use Condition Variables: Typical Flow

- Thread 1: //try to get into critical section and wait for the condition

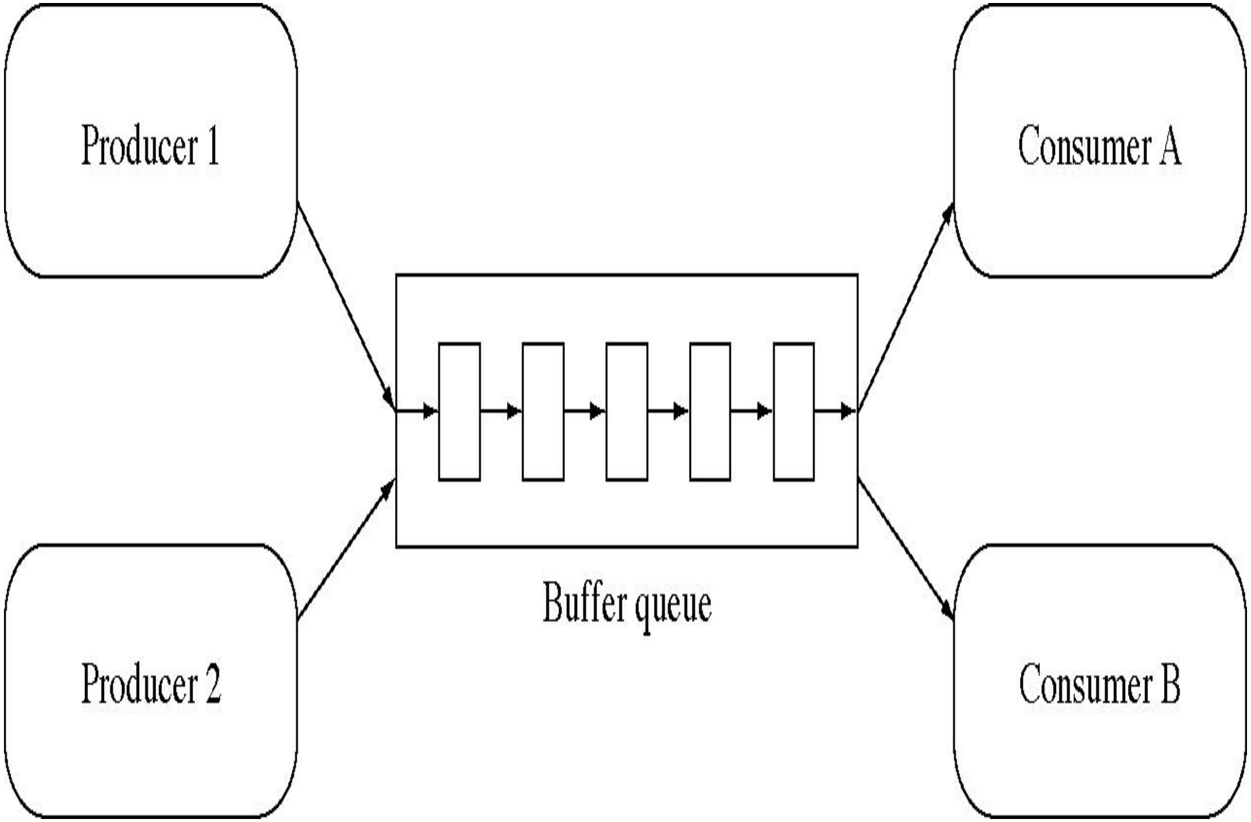
Mutex_lock(mutex);
While (condition is not satisfied)
Cond_Wait(mutex, cond);
Critical Section;
Mutex_unlock(mutex)

- Thread 2: // Try to create the condition.

Mutex_lock(mutex);
When condition can satisfy, Signal(cond);
Mutex_unlock(mutex);

Condition variables for in producer-consumer problem with unbounded buffer

Producer deposits data in a buffer for others to consume



Condition Variables for consumer-producer problem with unbounded buffer

- int avail=0; // # of data items available for consumption
- Pthread mutex m and condition cond;
- Consumer thread:

```
multex_lock(&m)
 while (avail <=0) Cond_Wait(&cond, &m);
  Consume next item;  avail = avail-1;
 mutex_unlock(&mutex)
```

- *Producer thread:*

```
mutex_lock(&m);
Produce next item;  availl = avail+1;
Cond_signal(&cond); //notify an item is available
mutex_unlock(&m);
```

When to use condition broadcast?

- When waking up one thread to run is not sufficient.
- Example: concurrent malloc()/free() for allocation and deallocation of objects with non-uniform sizes.

Running trace of malloc()/free()

- Initially 10 bytes are free.
- m() stands for malloc(). f() for free()

Thread 1:	Thread 2:	Thread 3:
m(10) – succ	m(5) – wait	m(5) – wait
f(10) –broadcast		
	Resume m(5)-succ	
		Resume m(5)-succ
m(7) – wait		
		m(3) –wait
	f(5) –broadcast	
Resume m(7)-wait		Resume m(3)-succ

Time

Implementing a barrier with condition variables

Text book p.180

```
/* Shared */
int counter = 0;
pthread_mutex_t mutex;
pthread_cond_t cond_var;
. . .
void* Thread_work(. . .) {
    . . .
    /* Barrier */
    pthread_mutex_lock(&mutex);
    counter++;
    if (counter == thread_count) {
        counter = 0;
        pthread_cond_broadcast(&cond_var);
    } else {
        while (pthread_cond_wait(&cond_var, &mutex) != 0);
    }
    pthread_mutex_unlock(&mutex);
    . . .
}
```

READ-WRITE LOCKS

Synchronization Example for Readers-Writers Problem

- **A data set is shared among a number of concurrent threads.**
  - Readers – only read the data set; they do **not** perform any updates
  - Writers – can both read and write
- **Requirement:**
  - allow multiple readers to read at the same time.
  - Only one writer can access the shared data at the same time.
- **Reader/writer access permission table:**

	Reader	Writer
Reader	OK	No
Writer	NO	No

Readers-Writers (First try with 1 mutex lock)

- **writer**

```
do {
        mutex_lock(w);
        //    writing is performed
        mutex_unlock(w);
} while (TRUE);
```

- **Reader**

```
do {
        mutex_lock(w);
        //    reading is performed
        mutex_unlock(w);
} while (TRUE);
```

	Reader	Writer
Reader	?	?
Writer	?	?

Readers-Writers (First try with 1 mutex lock)

- **writer**

```
do {
        mutex_lock(w);
        //    writing is performed
        mutex_unlock(w);
} while (TRUE);
```

- **Reader**

```
do {
        mutex_lock(w);
        //    reading is performed
        mutex_unlock(w);
} while (TRUE);
```

	Reader	Writer
Reader	no	no
Writer	no	no

2nd try using a lock + readcount

- **writer**

```
do {
        mutex_lock(w);// Use writer mutex lock
        //    writing is performed
        mutex_unlock(w);
} while (TRUE);
```

- **Reader**

```
do {
        readcount++; // add a reader counter.
        if(readcount==1) mutex_lock(w);
        //    reading is performed
        readcount--;
         if(readcount==0)  mutex_unlock(w);
} while (TRUE);
```

Readers-Writers Problem with semaphone

- **Shared Data**
  - Data set
  - Lock mutex (to protect readcount)
  - Semaphore wrt initialized to 1 (to synchronize between readers/writers)
  - Integer readcount initialized to 0

Readers-Writers Problem

- **A writer**

```
do {
    sem_wait(wrt) ;  //semaphore wrt

    // writing is performed

    sem_post(wrt) ;  //
} while (TRUE);
```

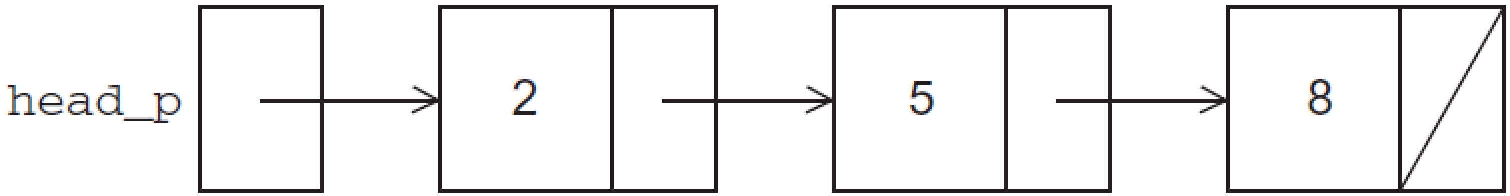
Readers-Writers Problem (Cont.)

- **Reader**

```
do {
            mutex_lock(mutex);
            readcount ++ ;
            if (readcount == 1)
                    sem_wait(wrt);  //check if anybody is writing
            mutex_unlock(mutex)

             // reading is performed

             mutex_lock(mutex);
             readcount  - - ;
             if (readcount  == 0)
                    sem_post(wrt) ;  //writing is allowed now
             nlock(mutex) ;
      } while (TRUE);
```

Application case: Sharing a sorted linked list of integers

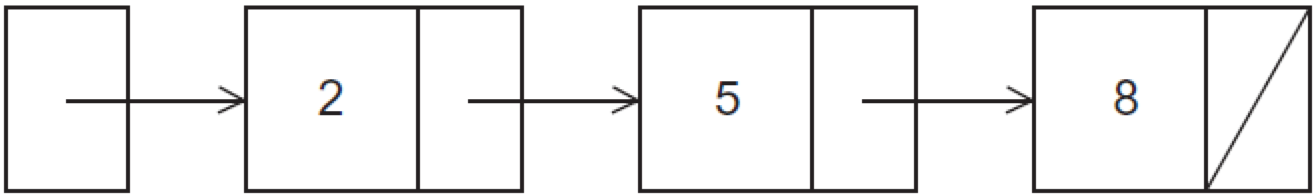
- **Demonstrate controlling of access to a large, shared data structure**
- **Operations supported**
  - Member, Insert, and Delete.



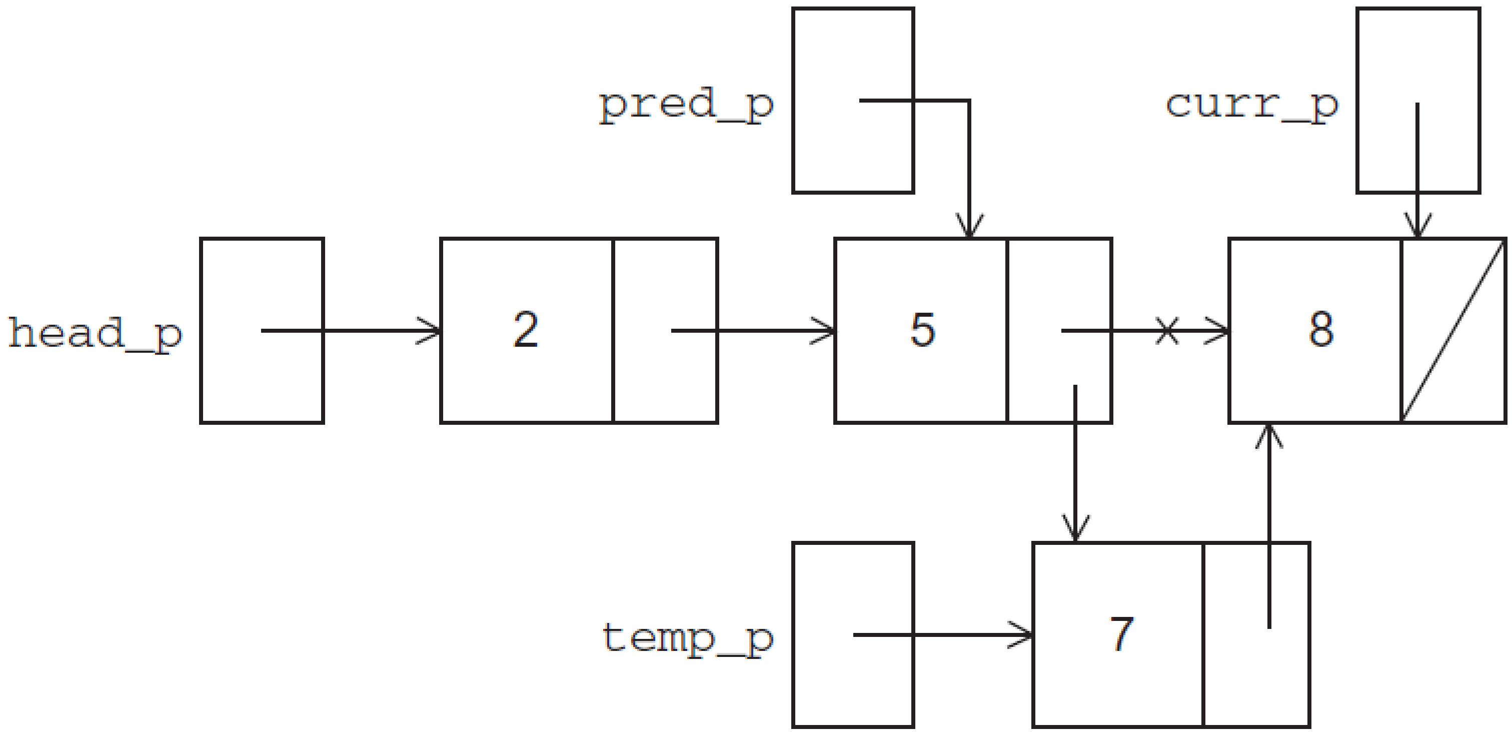

```
struct list_node_s {
    int data;
    struct list_node_s* next;
}
```

Membership operation for a linked list

```
int  Member(int value, struct list_node_s* head_p) {
   struct list_node_s* curr_p = head_p;

   while (curr_p != NULL && curr_p->data < value)
      curr_p = curr_p->next;

   if (curr_p == NULL || curr_p->data > value) {
      return 0;
   } else {
      return 1;
   }
}  /* Member */
```

head_p


Insert operation: Inserting a new node

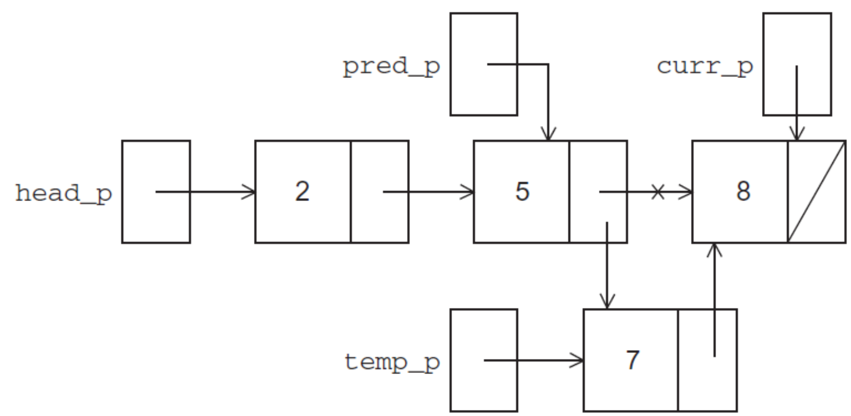


Inserting a new node into a list

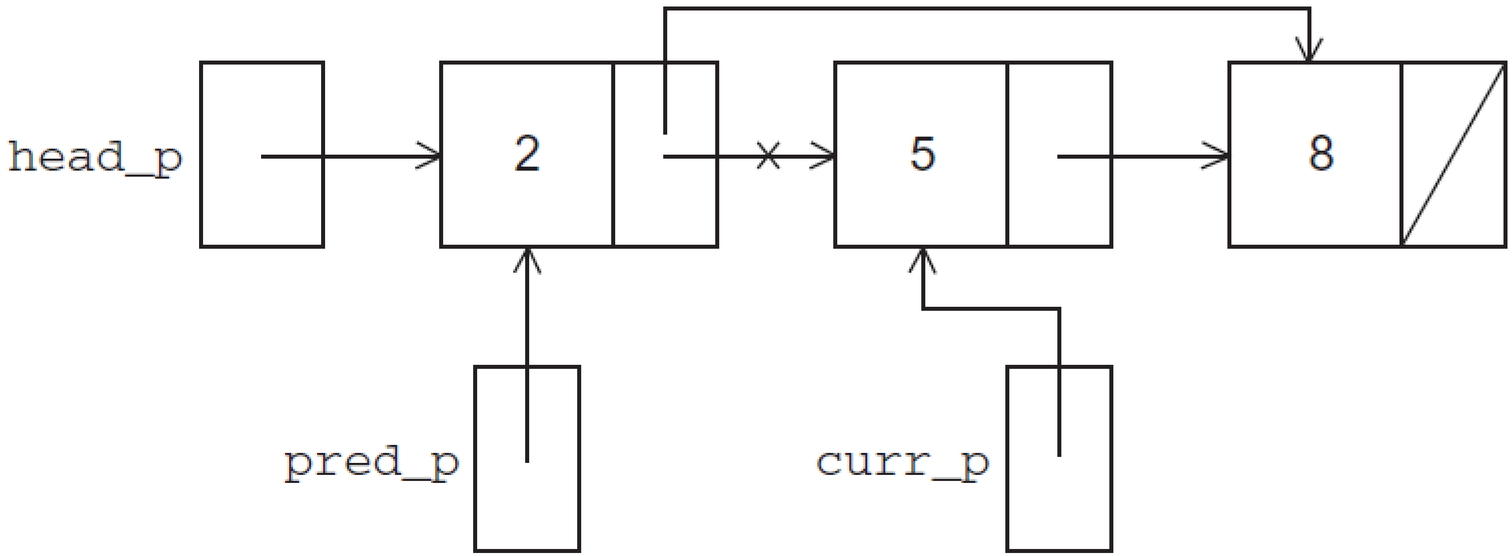
```
int Insert(int value, struct list_node_s** head_pp) {
   struct list_node_s* curr_p = *head_pp;
   struct list_node_s* pred_p = NULL;
   struct list_node_s* temp_p;

   while (curr_p != NULL && curr_p->data <
      pred_p = curr_p;
      curr_p = curr_p->next;
   }

   if (curr_p == NULL || curr_p->data > value) {
      temp_p = malloc(sizeof(struct list_nod
      temp_p->data = value;
      temp_p->next = curr_p;
      if (pred_p == NULL) /* New first node */
         *head_pp = temp_p;
      else
         pred_p->next = temp_p;
      return 1;
   } else { /* Value already in list
      return 0;
   }
}  /* Insert */
```

Find the right position in the sorted list

Insert to this position



Delete operation: remove a node from a linked list

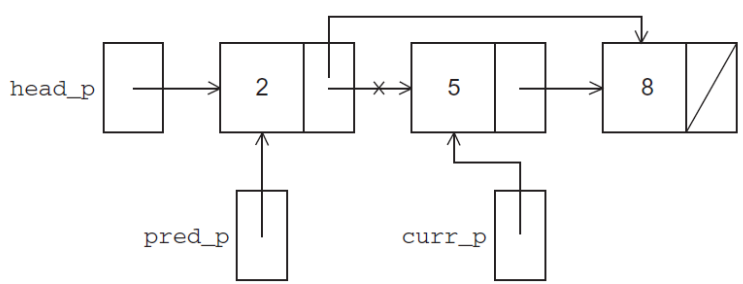


Deleting a node from a linked list

```
int Delete(int value, struct list_node_s** head_pp) {
   struct list_node_s* curr_p = *head_pp;
   struct list_node_s* pred_p = NULL;

   while (curr_p != NULL && curr_p->data
      pred_p = curr_p;
      curr_p = curr_p->next;
   }

   if (curr_p != NULL && curr_p->data == value) {
      if (pred_p == NULL) { /* Deleting first node in list */
         *head_pp = curr_p->next;
         free(curr_p);
      } else {
         pred_p->next = curr_p->next;
         free(curr_p);
      }
      return 1;
   } else { /* Value isn't in list */
      return 0;
   }
}  /* Delete */
```

Find a node with the given value

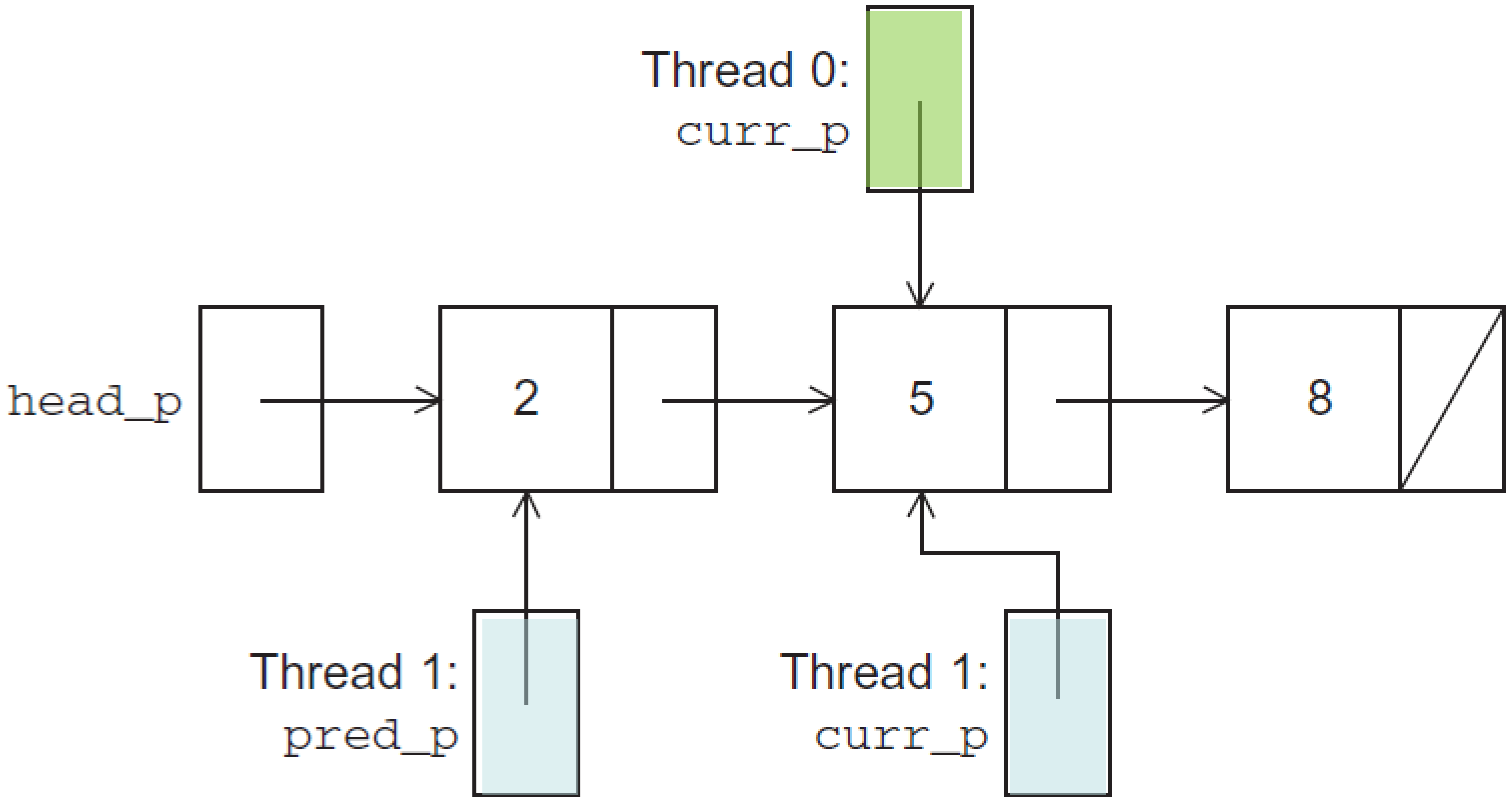
Remove this node



A Multi-Threaded Linked List

- Allow a sorted linked list to be accessed by multiple threads
- In order to share access to the list, define head_p to be a global variable.
  - This will simplify the function headers for Member, Insert, and Delete,
  - since we won't need to pass in either head_p or a pointer to head_p: we'll only need to pass in the value of interest.

Simultaneous access by two threads



Solution #1

- **An obvious solution is to simply lock the list any time that a thread attempts to access it.**
- **A call to each of the three functions can be protected by a mutex.**

```
Pthread_mutex_lock(&list_mutex);
Member(value);
Pthread_mutex_unlock(&list_mutex);
```

In place of calling Member(value).

Issues

- We’re serializing access to the list.
- If the vast majority of our operations are calls to Member, we’ll fail to exploit this opportunity for parallelism.
- On the other hand, if most of our operations are calls to Insert and Delete,
  - This may be the bes
    - since serializati
      performance im
    - Easy to implem

Mem			
Inser			
Delete	?	?	?

List-level	Member	Insert	Delete
Member	no	no	no
Insert	no	no	no
Delete	no	no	no

Solution #2

- **Instead of locking the entire list, lock individual nodes.**
  - A “finer-grained” approach: One mutex lock per node

```
struct list_node_s {
   int data;
   struct list_node_s* next;
   pthread_mutex_t mutex;
}
```

Node-level	Member	Insert	Delete
Member	no	no	no
Insert	no	no	no
Delete	no	no	no

Implementation of Member with one mutex per list node (1)

```
int  Member(int value) {
   struct list_node_s* temp_p;

   pthread_mutex_lock(&head_p_mutex);
   temp_p = head_p;
   while (temp_p != NULL && temp_p->data < value) {
      if (temp_p->next != NULL)
         pthread_mutex_lock(&(temp_p->next->mutex));
      if (temp_p == head_p)
         pthread_mutex_unlock(&head_p_mutex);
      pthread_mutex_unlock(&(temp_p->mutex));
      temp_p = temp_p->next;
   }
```

Implementation of Member with one mutex per list node (2)

```
    if (temp_p == NULL || temp_p->data > value) {
       if (temp_p == head_p)
          pthread_mutex_unlock(&head_p_mutex);
       if (temp_p != NULL)
          pthread_mutex_unlock(&(temp_p->mutex));
       return 0;
    } else {
       if (temp_p == head_p)
          pthread_mutex_unlock(&head_p_mutex);
       pthread_mutex_unlock(&(temp_p->mutex));
       return 1;
    }
}  /* Member */
```

Issues

- Much more complex than the original Member function.
- Much slower,
  - each time a node is accessed, a mutex must be locked and unlocked.
- Significant space cost
  - Adding a mutex field to each node

Motivation for using Pthreads Read-Write Locks

- Neither of our multi-threaded linked lists exploits the potential for simultaneous access to any node by threads that are executing Member.
- The first solution only allows one thread to access the entire list at any instant.
- The second only allows one thread to access any given node at any instant.

Pthreads Read-Write Locks

- A read-write lock is somewhat like a mutex except that it provides two lock functions.
  - The first lock function locks the read-write lock for reading, while the second locks it for writing.
- Example for

a linked list

	Member	Insert	Delete
Member	?	?	?
Insert	?	?	?
Delete	?	?	?

```
pthread_rwlock_rdlock(&rwlock);
Member(value);
pthread_rwlock_unlock(&rwlock);
. . .
pthread_rwlock_wrlock(&rwlock);
Insert(value);
pthread_rwlock_unlock(&rwlock);
. . .
pthread_rwlock_wrlock(&rwlock);
Delete(value);
pthread_rwlock_unlock(&rwlock);
```

Pthreads Read-Write Locks

- Multiple threads can simultaneously obtain the lock by calling the read-lock function, while only one thread can obtain the lock by calling the write-lock function.
- If any threads own the lock for reading, any threads that want to obtain the lock for writing will block in the call to the write-lock function.
- If any thread owns the loc
threads that want to obtai
or writing will block in thei
functions.

List-level	Member	Insert	Delete
Member	yes	no	no
Insert	no	no	no
Delete	no	no	no

A performance comparison of 3 implementations for a linked list

Total time in second for executing 100,000 operations.

99.9% Member

0.05% Insert

0.05% Delete

Implementation	Number of Threads 1	2	4	8
Read-Write Locks	0.213	0.123	0.098	0.115
One Mutex for Entire List	0.211	0.450	0.385	0.457
One Mutex per Node	1.680	5.700	3.450	2.700

Linked List Performance: Comparison

Total time in seconds for executing 100,000 operations

80% Member

10% Insert

10% Delete

Implementation	Number of Threads 1	2	4	8
Read-Write Locks	2.48	4.97	4.69	4.71
One Mutex for Entire List	2.50	5.13	5.04	5.11
One Mutex per Node	12.00	29.60	17.00	12.00

Issues with Threads: False Sharing, Deadlocks, Thread-safety

Caches, Cache-Coherence, and False Sharing

- Recall that chip designers have added blocks of relatively fast memory to processors called cache memory.
- The use of cache memory can have a huge impact on shared-memory.
- A write-miss occurs when a core tries to update a variable that's not in cache, and it has to access main memory.

Block-based pthreads matrix-vector multiplication

```
void *Pth_mat_vect(void* rank) {
   long my_rank = (long) rank;
   int i, j;
   int local_m = m/thread_count;
   int my_first_row = my_rank*local_m;
   int my_last_row = (my_rank+1)*local_m - 1;

   for (i = my_first_row; i <= my_last_row; i++) {
      y[i] = 0.0;
      for (j = 0; j < n; j++)
          y[i] += A[i][j]*x[j];
   }

   return NULL;
}  /* Pth_mat_vect */
```

Run-times and efficiencies of matrix-vector multiplication

Threads	Matrix Dimension					
	$8,000,000 \times 8$		8000×8000		$8 \times 8,000,000$	
	Time	Eff.	Time	Eff.	Time	Eff.
1	0.393	1.000	0.345	1.000	0.441	1.000
2	0.217	0.906	0.188	0.918	0.300	0.735
4	0.139	0.707	0.115	0.750	0.388	0.290

(times are in seconds)

Deadlock and Starvation

- **Deadlock** – two or more threads are waiting indefinitely for an event that can be only caused by one of these waiting threads
- **Starvation** – indefinite blocking (in a waiting queue forever).
  - Let S and Q be two mutex locks:

P_0	P_1
Lock(S);	Lock(Q);
Lock(Q);	Lock(S);
.	.
.	.
.	.
Unlock(Q);	Unlock(S);
Unlock(S);	Unlock(Q);

Deadlock Avoidance

- **Order the locks and always acquire the locks in that order.**
- **Eliminate circular waiting**
  - :

P_0	P_1
Lock(S);	Lock(S);
Lock(Q);	Lock(Q);
.	.
.	.
.	.
Unlock(Q);	Unlock(Q);
Unlock(S);	Unlock(S);

Thread-Safety

- A block of code is thread-safe if it can be simultaneously executed by multiple threads without causing problems.
- When you program your own functions, you know if they are safe to be called by multiple threads or not.
- You may forget to check if system library functions used are thread-safe.
  - Unsafe function: strtok()from C string.h library
  - Other example.
    - The random number generator random in stdlib.h.
    - The time conversion function localtime in time.h.

Example of using strtok()

- **“Tokenize” a English text file**
  - Tokens are contiguous sequences of characters separated by a white-space, a tab, or a newline.
  - Example: “Take UCSB CS140”

  → Three tokens: “Take”, “UCSB”, “CS140”
- **Divide the input file into lines of text and assign the lines to the threads in a round-robin fashion.**
  - Each thread tokenizes a line using strtok()
  - Line 1 → thread 0, Line 2 → thread 1, . . . , the tth goes to thread t, the t +1st goes to thread 0, etc.
  - Serialize access to input lines using semaphores

The strtok function

- **The first time it's called,**
  - the string argument is the text to be tokenized (Our line of input)
  - strtok caches a pointer to string
- **For subsequent calls, it returns successive tokens <u>taken from the cached copy</u>**
  - the first argument should be NULL.

```
char* strtok(
      char*         string      /* in/out */,
      const char*   separators  /* in     */);
```

Multi-threaded tokenizer (1)

```
void *Tokenize(void* rank) {
   long my_rank = (long) rank;
   int count;
   int next = (my_rank + 1) % thread_count;
   char *fg_rv;
   char my_line[MAX];
   char *my_string;

   sem_wait(&sems[my_rank]);
   fg_rv = fgets(my_line, MAX, stdin);
   sem_post(&sems[next]);
   while (fg_rv != NULL) {
      printf("Thread %ld > my line = %s", my_rank, my_line);
```

Multi-threaded tokenizer (2)

```
   count = 0;
   my_string = strtok(my_line, " \t\n");
   while ( my_string != NULL ) {
      count++;
      printf("Thread %ld > string %d = %s\n", my_rank, count,
            my_string);
      my_string = strtok(NULL, " \t\n");
   }

   sem_wait(&sems[my_rank]);
   fg_rv = fgets(my_line, MAX, stdin);
   sem_post(&sems[next]);
}

return NULL;
}  /* Tokenize */
```

First token

Next token

Running with one thread

Input file:

Pease porridge hot.

Pease porridge cold.

Pease porridge in the pot

Nine days old.

- **It correctly tokenizes the input stream with 1 thread**

Pease

porridge

hot

. . .

Running with two threads

```
Thread 0 > my line = Pease porridge hot.
Thread 0 > string 1 = Pease
Thread 0 > string 2 = porridge
Thread 0 > string 3 = hot.
Thread 1 > my line = Pease porridge cold.
Thread 0 > my line = Pease porridge in the pot
Thread 0 > string 1 = Pease
Thread 0 > string 2 = porridge
Thread 0 > string 3 = in
Thread 0 > string 4 = the
Thread 0 > string 5 = pot
Thread 1 > string 1 = Pease
Thread 1 > my line = Nine days old.
Thread 1 > string 1 = Nine
Thread 1 > string 2 = days
Thread 1 > string 3 = old.
```

Oops!

What happened?

- strtok caches the input line by declaring a variable to have static (persistent) storage class.
  - Unfortunately this cached string is shared, not private.
- Thus, thread 0's call to strtok with the third line of the input has apparently <u>overwritten</u> the contents of thread 1's call with the second line.
- So the strtok function is not thread-safe. If multiple threads call it simultaneously, the output may not be correct.

"re-entrant" (thread safe) functions

- **In some cases, the C standard specifies an alternate, thread-safe, version of a function.**

```
char* strtok_r(
      char*          string        /* in/out */,
      const char*    separators,   /* in     */
      char**         saveptr_p     /* in/out */);
```

Concluding Remarks (1)

- A thread in shared-memory programming is analogous to a process in distributed memory programming.
  - However, a thread is often lighter-weight than a full-fledged process.
- In Pthreads programs, all the threads have access to global variables, while local variables usually are private to the thread running the function.
- When multiple threads access a shared resource without controling, it may result in an error: we have a race condition.
  - A critical section is a block of code that updates a shared resource that can only be updated by one thread at a time

Concluding Remarks (2)

- **Busy-waiting can be used for critical sections with a flag variable and a while-loop**
  - It can be very wasteful of CPU cycles.
  - It can also be unreliable if compiler optimization is turned on.
- **A mutex arrange for mutually exclusive access to a critical section.**
- **A semaphore**
  - It is an unsigned int together with two operations: sem_wait and sem_post.
  - Semaphores are more powerful than mutexes since they can be initialized to any nonnegative value.

Concluding Remarks (3)

- A barrier is a point in a program at which the threads block until all of the threads have reached it.
- A read-write lock is used when it's safe for multiple threads to simultaneously read a data structure, but if a thread needs to modify or write to the data structure, then only that thread can access the data structure during the modification.
- Some C functions cache data between calls by declaring variables to be static, causing errors when multiple threads call the function.
  - This type of function is not thread-safe