# **CSE113: Parallel Programming** Feb. 2, 2022

- Topics:
  - Input/output queues
  - Producer consumer queues
    - Synchronous
    - Circular buffer



#### Announcements

- HW1 grades might be delayed until Monday
  - Let us know ASAP if there are issues
- Homework 2 is due today
  - Sanya has office hours
  - We will keep an eye on Piazza and try to ask questions asked before 5 pm
- Homwork 3 will be released today by midnight
  - Due in 2 weeks

#### Announcements

- Midterm is released on Monday
  - asynchronous, 1 week (no time limit)
  - Open note, open internet (to a reasonable extent: no googling exact questions or asking questions on forums)
  - do not discuss with classmates AT ALL while the test is active
  - No late tests will be accepted.
- Prioritize midterm next week!

## Homework clarifications

- Conditional variables
  - They are **not** allowed in your solution, but they are interesting
  - https://en.cppreference.com/w/cpp/thread/condition\_variable
- Part 2: reader/writer
  - You cannot significantly slow down readers in isolation
- Part 3: keeping the structure:
  - you can re-arrange functions, just no changing the high-level implementation

## Homework clarifications

• You can share results, but not code

## Today's Quiz

• Due Monday by class. Please do it!

What is the relationship between linearizable (L) and sequentially consistent (SC)?

○ Objects can be one or the other, but not both

○ Objects that are L are also SC, but not the other way around

○ Objects that are SC are also L, but not the other way around

○ SC and L are the different definitions for the same concept

Nonblocking states that:

 $\bigcirc$  threads do not share memory

 $\bigcirc$  threads will execute in a fair way

○ delays in one thread will not cause delays in other threads

 $\bigcirc$  no RMWs are used

Lock-free data structures are technically undefined because they contain data conflicts

⊖ True

 $\bigcirc$  False

Write a few sentences about the benefit of starting out implementations with specialized datastructions (e.g. input/output queues) rather than data structures that allow more general access patterns?

## Review



each command gets a linearization point.

You can place the point any where between its innovation and response!



each command gets a linearization point.

You can place the point any where between its innovation and response!

Project the linearization points to a global timeline



each command gets a linearization point.

You can place the point any where between its innovation and response!

Project the linearization points to a global timeline



slider game!



each command gets a linearization point.

You can place the point any where between its innovation and response!

Project the linearization points to a global timeline









How about a stack?













# Input/Output Queues

# Input/Output Queues

- Queue in which multiple threads read (deq), or write (enq), but not both.
- Why would we want a thing?
- Computation done in phases:
  - First phase prepares the queue (by writing into it)
  - All threads join
  - Second phase reads values from the queue.







does it matter which order threads add their data?



does it matter which order threads add their data?



does it matter which order threads add their data? No! Because there are no deqs!



## Now enqueue

#### enq

• Now we only do deqs



#### enq

• Now we only do deqs



• Now we only do deqs


• Now we only do deqs



• Now we only do deqs



• Now we only do deqs



```
class InputOutputQueue {
 private:
   atomic_int front;
   atomic int end;
   int list[SIZE];
 public:
   InputOutputQueue() {
       front = end = 0;
    void eng(int x) {
        int reserved index = atomic fetch add(&end, 1);
       list[reserved index] = x;
    void deq() {
       int reserved index = atomic fetch add(&front, 1);
      return list[reserved index];
     int size() {
       return end.load() - front.load();
```

```
class InputOutputQueue {
 private:
   atomic int front;
   atomic int end;
   int list[SIZE];
 public:
   InputOutputQueue() {
       front = end = 0;
    void enq(int x) {
        int reserved index = atomic fetch add(&end, 1);
       list[reserved index] = x;
    void deq() {
       int reserved index = atomic fetch add(&front, 1);
       return list[reserved index];
     int size() {
       return end.load() - front.load();
```

Does list need to be atomic?

How to make sure the queue has an element in it before you dequeue?

## What can go wrong if we deq and enq?

<u>Thread 0:</u> enq(1);

Thread 1: deq();

```
void enq(int x) {
    int reserved_index = atomic_fetch_add(&end, 1);
    list[reserved_index] = x;
```

void deq() {
 int reserved\_index = atomic\_fetch\_add(&front, 1);
 return list[reserved\_index];



# What can go wrong if we deq and enq?

Thread 0: enq(1);

<u>Thread 1:</u> deq();

```
void enq(int x) {
    int reserved_index = atomic_fetch_add(&end, 1);
    list[reserved_index] = x;
```

```
void deq() {
    int reserved_index = atomic_fetch_add(&front, 1);
    return list[reserved_index];
```



# What can go wrong if we deq and enq?

Thread 0: enq(1);

<u>Thread 1:</u> deq();

```
void enq(int x) {
    int reserved_index = atomic_fetch_add(&end, 1);
    list[reserved_index] = x;
```

```
void deq() {
    int reserved_index = atomic_fetch_add(&front, 1);
    return list[reserved_index];
```



# Blocking?

• Does the input/output queue block?















Both threads need to execute this instruction. What happens if one is delayed?



Both threads need to execute this instruction. What happens if one is delayed? *It doesn't matter! The other thread can still keep going!* 



Both threads need to execute this instruction. What happens if one is delayed? *It doesn't matter! The other thread can still keep going!* 

```
class InputOutputQueue {
 private:
   atomic int front;
   atomic int end;
   int list[SIZE];
 public:
   InputOutputQueue() {
       front = end = 0;
    void enq(int x) {
        int reserved index = atomic fetch add(&end, 1);
       list[reserved index] = x;
     int deq() {
       int reserved index = atomic fetch add(&front, 1);
      return list[reserved index];
     int size() {
       return end.load() - front.load();
```

Could we implement a blocking version of this queue?

#### **class** InputOutputQueue { private: int front; int end; int list[SIZE]; mutex m; public: void enq(int x) { m.lock(); list[end] = x;end++; m.unlock(); } int deq() { m.lock(); int tmp = list[front]; front++; m.unlock(); return tmp;

Could we implement a blocking version of this queue?

Just add a mutex!

What are the pros and cons?









}





This implementation is blocking!

## On to new material!

## Schedule

- Producer Consumer queues
  - Synchronous
  - Circular buffer

# Producer Consumer Queues

- 1 enq, 1 deq
  - enq'er cannot deq
  - deq'er cannot enq
- Example: printf:
  - your program equeues values to print
  - the terminal process dequeues values and prints them

- First implementation:
  - Synchronous
  - Slow
  - Good for debugging

- First implementation:
  - Synchronous
  - Slow
  - Good for debugging
- enq does not return until value is deq'ed

Producer Thread
enq(7);



Consumer Thread
deq();



wait

Consumer Thread
deq();





Producer Thread
enq(7);



Consumer Thread
deq();

both can continue
Producer Thread
sleep();
enq(7);



Consumer Thread
deq();







They both can continue

Producer Thread
enq(7);



Consumer Thread
deq();













now the consumer can read from the box!











How to fix?



How to fix?





























### Schedule

- Producer Consumer Queues
  - Synchronous
  - Circular buffer

• Asynchronous:



• Asynchronous:



no waiting for producer (while there is room)

• Asynchronous:



no waiting for producer (while there is room)

• Asynchronous:



no waiting for producer (while there is room)
• Asynchronous:



no waiting for producer (while there is room)

• Asynchronous:



no waiting for producer (while there is room)

• Asynchronous:



no waiting for producer (while there is room)

returns 7

• Asynchronous:



no waiting for producer (while there is room)

returns 7

• Asynchronous:



no waiting for producer (while there is room)

returns 7

• Asynchronous:



no waiting for producer (while there is room)



• Asynchronous:



no waiting for producer (while there is room)

returns 7

• Asynchronous:



no waiting for producer (while there is room)

returns 8

• Asynchronous:



no waiting for producer (while there is room)

returns 8

• Asynchronous:



no waiting for producer (while there is room)

returns 8

• Asynchronous:



returns 9

• Asynchronous:



• Asynchronous:



• Asynchronous:



• Asynchronous:



blocks when there is nothing in the queue

• How do we implement it?

• Start with a fixed size array



• Start with a fixed size array



We will use what is called a *circular buffer method* 

• Start with a fixed size array



conceptually it is a circle

• Start with a fixed size array



• Start with a fixed size array



indexes will circulate in order and wrap around

conceptually it is a circle

• Start with a fixed size array

we will assume modular arithmetic:

if x = (SIZE - 1) then x + 1 == 0;



indexes will circulate in order and wrap around

conceptually it is a circle

• Start with a fixed size array

Two variables to keep track of where to deq and enq:

head and tail



• Start with a fixed size array

Two variables to keep track of where to deq and enq:

head and tail:

enq to the head, deq from the tail



indexes will circulate in order and wrap around

conceptually it is a circle

• Start with a fixed size array

Two variables to keep track of where to deq and enq:

head and tail

Empty queue is when head == tail



indexes will circulate in order and wrap around

conceptually it is a circle

• Start with a fixed size array

Two variables to keep track of where to deq and enq:

head and tail

Empty queue is when head == tail

Full queue is when head == tail?

conceptually it is a circle



but then

empty?

• Start with a fixed size array

Two variables to keep track of where to deq and enq:

head and tail

Empty queue is when head == tail

Full queue is when head == tail?

conceptually it is a circle



• Start with a fixed size array

Two variables to keep track of where to deq and enq:

head and tail

Empty queue is when head == tail

Full queue is when head + 1 == tail

conceptually it is a circle





```
class ProdConsQueue {
  private:
    atomic_int head;
    atomic_int tail;
    int buffer[SIZE];

  public:
    void enq(int x) {
        // store value at head
        // increment head
     }
}
```



```
class ProdConsQueue {
   private:
     atomic_int head;
     atomic_int tail;
     int buffer[SIZE];

   public:
     void enq(int x) {
        // store value at head
        // increment head
     }
}
```



```
class ProdConsQueue {
  private:
    atomic_int head;
    atomic_int tail;
    int buffer[SIZE];

  public:
    void enq(int x) {
        // store value at head
        // increment head
     }
}
```



class ProdConsQueue { private: atomic\_int head; atomic\_int tail; int buffer[SIZE]; public: void enq(int x) { // store value at head // increment head int deq() { // get value at tail // increment tail



class ProdConsQueue { private: atomic int head; atomic\_int tail; int buffer[SIZE]; public: void enq(int x) { // store value at head // increment head int deq() { // get value at tail // increment tail

This looks like the two threads don't even share head and tail! What is missing?





what happens if we try to dequeue here?



```
class ProdConsQueue {
 private:
    atomic int head;
    atomic_int tail;
    int buffer[SIZE];
 public:
   void enq(int x) {
      // store value at head
      // increment head
    int deq() {
      // wait while queue is empty
      // get value at tail
      // increment tail
```



```
class ProdConsQueue {
 private:
    atomic_int head;
    atomic_int tail;
    int buffer[SIZE];
 public:
   void enq(int x) {
      // store value at head
      // increment head
    int deq() {
      // wait while queue is empty
      // get value at tail
      // increment tail
```


```
class ProdConsQueue {
 private:
    atomic_int head;
    atomic_int tail;
    int buffer[SIZE];
 public:
   void enq(int x) {
      // store value at head
        increment head
    int deq() {
      // wait while queue is empty
      // get value at tail
      // increment tail
}
```



```
class ProdConsQueue {
 private:
    atomic int head;
    atomic_int tail;
    int buffer[SIZE];
 public:
   void enq(int x) {
      // store value at head
        increment head
    int deq() {
      // wait while queue is empty
         get value at tail
      // increment tail
```



```
class ProdConsQueue {
 private:
    atomic int head;
    atomic_int tail;
    int buffer[SIZE];
 public:
   void enq(int x) {
      // store value at head
      // increment head
    int deq() {
      // wait while queue is empty
      // get value at tail
      // increment tail
}
```



class ProdConsQueue { private: atomic int head; atomic\_int tail; int buffer[SIZE]; public: void enq(int x) { // store value at head // increment head int deq() { // wait while queue is empty // get value at tail // increment tail

similarly for enqueue

but why can't we enqueue?



```
class ProdConsQueue {
 private:
    atomic int head;
    atomic_int tail;
    int buffer[SIZE];
 public:
   void enq(int x) {
      // store value at head
      // increment head
    int deq() {
      // wait while queue is empty
      // get value at tail
      // increment tail
```

incrementing the head would make it empty!



```
class ProdConsQueue {
 private:
    atomic int head;
    atomic_int tail;
    int buffer[SIZE];
 public:
    void enq(int x) {
      // wait for their to be room
      // store value at head
      // increment head
    int deq() {
      // wait while queue is empty
      // get value at tail
      // increment tail
```

## Other questions:



```
class ProdConsQueue {
 private:
    atomic int head;
    atomic int tail;
    int buffer[SIZE];
 public:
   void enq(int x) {
      // wait for their to be room
      // store value at head
      // increment head
    int deq() {
      // wait while queue is empty
      // get value at tail
      // increment tail
```



Other questions:

Do these need to be atomic RMWs?

<b>class</b> ProdConsQueue {
private:
<pre>atomic_int head;</pre>
<pre>atomic_int tail;</pre>
<pre>int buffer[SIZE];</pre>
public:
<pre>void enq(int x) {</pre>
// wait for their to be room
// store value at head
// increment head
}
<b>int</b> deq() {
// wait while queue is empty
// get value at tail
// increment tail
}
}

## Next week

- Work stealing and generalized concurrent objects
- Get HW 2 turned in today!
- HW 3 is out today. You can get started on Part 1
- Prepare for midterm on Monday