CSE 230
Concurrency: STM

The Grand Challenge
How to properly use multi-cores?
Need new programming models!

Parallelism vs Concurrency

- A parallel program exploits real parallel computing resources to run faster while computing the same answer.
  - Expectation of genuinely simultaneous execution
  - Deterministic

- A concurrent program models independent agents that can communicate and synchronize.
  - Meaningful on a machine with one processor
  - Non-deterministic

Concurrent Programming

Essential For Multicore Performance
Concurrent Programming

State-of-the-art is 30 years old!
Locks and condition variables
Java: synchronized, wait, notify

Locks etc. Fundamentally Flawed
“Building a sky-scraper out of matchsticks”

Even Worse! Locks Don’t Compose

What’s Wrong With Locks?

Races
Forgotten locks lead to inconsistent views

Deadlock
Locks acquired in “wrong” order

Lost Wakeups
Forgotten notify to condition variables

Diabolical Error recovery
Restore invariants, release locks in exception handlers

Even Worse! Locks Don’t Compose

class Account{
    float balance;

    synchronized void deposit(float amt) {
        balance += amt;
    }

    synchronized void withdraw(float amt) {
        if (balance < amt)
            throw new OutOfMoneyError();
        balance -= amt;
    }
}

A Correct bank Account class
Write code to transfer funds between accounts

1st Attempt transfer = withdraw then deposit

class Account{
    float balance;
    synchronized void deposit(float amt) {
        balance += amt;
    }
    synchronized void withdraw(float amt) {
        if (balance < amt)
            throw new OutOfMoneyError();
        balance -= amt;
    }
    void transfer(Account other, float amt) {
        other.withdraw(amt);
        this.deposit(amt);
    }
}
1st Attempt  transfer = withdraw then deposit

class Account{
  float balance;
  synchronized void deposit(float amt) {
    balance += amt;
  }
  synchronized void withdraw(float amt) {
    if(balance < amt)
      throw new OutOfMoneyError();
    balance -= amt;
  }
  void transfer(Account other, float amt) {
    other.withdraw(amt);
    this.deposit(amt);
  }
}

Race Condition Wrong sum of balances

Even Worse! Locks Don’t Compose

2st Attempt: synchronized transfer

class Account{
  float balance;
  synchronized void deposit(float amt) {
    balance += amt;
  }
  synchronized void withdraw(float amt) {
    if(balance < amt)
      throw new OutOfMoneyError();
    balance -= amt;
  }
  synchronized void transfer(Account other, float amt) {
    other.withdraw(amt);
    this.deposit(amt);
  }
}

Deadlocks with Concurrent reverse transfer

Even Worse! Locks Don’t Compose

Locks are absurdly hard to get right

Scalable double-ended queue: one lock per cell

No interference
If ends “far” apart

But watch out!
If queue is 0, 1, or 2 elements long
<table>
<thead>
<tr>
<th>Coding Style</th>
<th>Difficulty of queue implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential code</td>
<td>Undergraduate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coding Style</th>
<th>Difficulty of queue implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential code</td>
<td>Undergraduate</td>
</tr>
<tr>
<td>Locks &amp; Conditions</td>
<td>Major publishable result*</td>
</tr>
</tbody>
</table>

*Simple, fast, and practical non-blocking and blocking concurrent queue algorithms*

**What we have**

Locks and Conditions: Hard to use & Don’t compose

- Hardware (with Locks and condition variables)

**What we want**

Libraries Build Layered Concurrency Abstractions

- Concurrency primitives
- Hardware
Idea: Replace locks with atomic blocks

Atomic Blocks/STM: Easy to use & Do compose

Hardware

Atomic Blocks:
atomic, retry, orElse

Library

Library

Library

Library

Library

Library

Coding Style

Difficulty of queue implementation

Sequential code  Undergraduate

Locks & Conditions  Major publishable result*

Atomic blocks(STM)  Undergraduate

*Simple, fast, and practical non-blocking and blocking concurrent queue algorithms

Atomic Memory Transactions

atomic {...sequential code...}

Wrap atomic around sequential code
All-or-nothing semantics: atomic commit

Atomic Memory Transactions

atomic {...sequential code...}

Atomic Block Executes in Isolation
No Data Race Conditions!

cf “ACID” database transactions

cf “ACID” database transactions

Locks are absurdly hard to get right

Locks are absurdly hard to get right
Atomic Memory Transactions

There Are No Locks
Hence, no deadlocks!

atomic {...sequential code...}

cf “ACID” database transactions

How it Works

Optimistic Concurrency
Execute code without any locks.

Record reads/writes in thread-local transaction
Writes go to the log only, not to memory.

At the end, transaction validates the log
If valid, atomically commit changes to memory
If invalid, re-run from start, discarding changes

Why it Doesn’t Work...

Logging Memory Effects is Expensive
Huge slowdown on memory read/write

Cannot “Re-Run”, Arbitrary Effects
How to “retract” email?
How to “un-launch” missile?

STM in Haskell
<table>
<thead>
<tr>
<th>Haskell Fits the STM Shoe</th>
<th>Issue: Logging Memory Is Expensive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haskellers brutally trained from birth to use memory/IO effects sparingly!</td>
<td>Haskell already partitions world into Immutable values (zillions and zillions) Mutable locations (very few)</td>
</tr>
<tr>
<td></td>
<td>Solution: Only log mutable locations!</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Issue: Logging Memory Is Expensive</th>
<th>Issue: Undoing Arbitrary IO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haskell already paid the bill!</td>
<td>Types control where IO effects happen</td>
</tr>
<tr>
<td>Reading and Writing locations are Expensive function calls</td>
<td>Easy to keep them out of transactions</td>
</tr>
<tr>
<td>Logging Overhead</td>
<td>Monads Ideal For Building Transactions</td>
</tr>
<tr>
<td>Lower than in imperative languages</td>
<td>Implicitly (invisibly) passing logs</td>
</tr>
</tbody>
</table>
Tracking Effects with Types

```haskell
main = do { putStr (reverse "yes"); putStr "no" }
```

(reverse "yes") :: String  -- No effects
(putStr "no") :: IO ()   -- Effects okay

Main program is a computation with effects

```haskell
main :: IO ()
```

Mutable State via the IO Monad

```haskell
newRef :: a -> IO (IORef a)
readRef :: IORef a -> IO a
writeRef :: IORef a -> a -> IO ()
```

Reads and Writes are 100% Explicit

(r+6) is rejected as r :: IORef Int
Concurrency in Haskell

**forkIO** function spawns a thread
Takes an IO action as argument

\[
\text{forkIO} :: \text{IO } a \rightarrow \text{IO } \text{ThreadId}
\]

Atomic Blocks in Haskell

\[
\text{atomically} :: \text{IO } a \rightarrow \text{IO } a
\]

**atomically** `act`
Executes `act` atomically

\[
\text{main} = \text{do } r \leftarrow \text{newRef } 0
\text{forkIO } \$ \text{atomically } \text{incR } r
\text{incR } r
\text{print } s
\]

Data Race

Atomic Blocks in Haskell

\[
\text{atomically} :: \text{IO } a \rightarrow \text{IO } a
\]

\[
\text{main} = \text{do } r \leftarrow \text{newRef } 0
\text{forkIO } \$ \text{atomically } \text{incR } r
\text{atomically } \text{incR } r
\text{atomic} \text{ Ensures No Data Races!}
\]
Atomic Blocks in Haskell

What if we use \texttt{incR} outside block?
Yikes! Races in code inside & outside!

\begin{verbatim}
main = do r <- newRef 0
  forkIO $ incR r
  atomically $ incR r
\end{verbatim}

Data Race

A Better Type for Atomic

\begin{verbatim}
atomic :: STM a -> IO a
newTVar :: a -> STM (TVar a)
readTVar :: TVar a -> STM a
writeTVar :: TVar a -> a -> STM ()
\end{verbatim}

Types ensure \texttt{Tvar} only touched in \texttt{STM} action

Type System Guarantees

You cannot forget \texttt{atomically}
Only way to execute \texttt{STM} action

\begin{verbatim}
incT :: TVar Int -> STM ()
incT r = do v <- readTVar r
          writeTVar r (v+1)
main = do r <- atomically $ newVar 0 TVar
          forkIO $ atomically $ incT r
          atomically $ incT r
          ...
\end{verbatim}

Outside Atomic Block
Can’t fiddle with TVars

Inside Atomic Block
Can’t do IO, Can’t manipulate imperative variables

\texttt{atomic} $\quad$ if \texttt{x<y} then \texttt{launchMissiles}
Type System Guarantees

Note: atomically is a function not a special syntactic construct ...and, so, best of all...

(Unlike Locks) STM Actions Compose!

Glue STM Actions Arbitrarily
Wrap with atomic to get an IO action
Types ensure STM action is atomic

STM Type Supports Exceptions

No need to restore invariants, or release locks!
In `atomically act` if `act` throws exception:
1. Transaction is aborted with no effect,
2. Exception is propagated to enclosing IO code*

Transaction Combinators

*Composable Memory Transactions
retry :: STM ()

“Abort current transaction & re-execute from start”

Implementation Avoids Busy Waiting
Uses logged reads to block till a read-var (eg. acc) changes

retry :: STM ()

withdraw :: TVar Int -> Int -> STM ()
withdraw acc n = do bal <- readTVar acc
                if bal < n then retry
                writeTVar acc (bal-n)

No Condition Variables!
Uses logged reads to block till a read-var (eg. acc) changes
Retrying thread is woken on write, so no forgotten notifies

retry :: STM ()

No Condition Variables!
No danger of forgetting to test conditions
On waking as transaction runs from the start.

withdraw :: TVar Int -> Int -> STM ()
withdraw acc n = do bal <- readTVar acc
                if bal < n then retry
                writeTVar acc (bal-n)
Why is retry Compositional?

Can appear anywhere in an STM Transaction
Nested arbitrarily deeply inside a call

atomic $ do withdraw a1 3
         withdraw a2 7

Waits until `a1>3` AND `a2>7`
Without changing/knowing `withdraw` code

Hoisting Guards Is Not Compositional

atomic (a1>3 && a2>7) {
  ...stuff...
}

Breaks abstraction of “…stuff…”
Need to know code to expose guards

#2 orElse: Choice

How to transfer 3$ from a1 or a2 to b?

Try this…
...and if it retries, try this
atomically $ do withdraw a1 3 `orElse` withdraw a2 3
          deposit b 3

...and and then do this

orElse :: STM a -> STM a -> STM a

Choice Is Composable Too!

transfer a1 a2 b = do withdraw a1 3 `orElse` withdraw a2 3
                    deposit b 3

atomically $ transfer a1 a2 b
              `orElse`
              transfer a3 a4 b

transfer calls orElse

But calls to it can be composed with orElse
Ensuring Correctness of Concurrent Accesses?

e.g. account should never go below 0

Assumed on Entry, Verified on Exit

Only Tested If Invariant’s TVar changes

### #3 always: Enforce Invariants

```haskell
always :: STM Bool -> STM ()
checkBal :: TVar Int -> STM Bool
checkBal v = do cts <- readTVar v
               return (v > 0)
newAccount :: STM (TVar Int)
newAccount = do v <- newTVar 0
               always $ checkBal v
               return v
```

An arbitrary boolean valued STM action

Every Transaction that touches acct will check invariant
If the check fails, the transaction restarts

Add a new invariant to a global pool
Conceptually, all invariants checked on all commits

Implementation Checks Relevant Invariants
That read TVars written by the transaction
## Recap: Composing Transactions

A transaction is a value of type STM a.
Transactions are first-class values.

Big Tx By Composing Little Tx
sequence, choice, block ...

To Execute, Seal The Transaction
atomically :: STM a -> IO a

## Complete Implementation in GHC6

Performance is similar to Shared-Var
Need more experience using STM in practice...

You can play with it*
Final will have some STM material 😊

* Beautiful Concurrency

## STM in Mainstream Languages

Proposals for adding STM to Java etc.

```java
class Account {
    float balance;
    void deposit(float amt) {
        atomic { balance += amt; }
    }
    void withdraw(float amt) {
        atomic {
            if(balance < amt) throw new OutOfMoneyError();
            balance -= amt;
        }
    }
    void transfer(Acct other, float amt) {
        atomic { // Can compose withdraw and deposit.
            other.withdraw(amt);
            this.deposit(amt);
        }
    }
}
```

Mainstream Types Don’t Control Effects

So Code Inside Tx Can Conflict with Code Outside!

**Weak Atomicity**
Outside code sees inconsistent memory
Avoid by placing all shared mem access in Tx

**Strong Atomicity**
Outside code guaranteed consistent memory view
Causes big performance hit
## A Monadic Skin

**In C/Java, IO is Everywhere**
No need for special type, all code is in “IO monad”

**Haskell Gives You A Choice**
When to be in IO monad vs when to be purely functional

**Haskell Can Be Imperative BUT C/Java Cannot Be Pure!**
Mainstream PLs lack a statically visible pure subset

The separation facilitates concurrent programming...

## Conclusions

STM raises abstraction for concurrent programming
Think high-level language vs assembly code
Whole classes of low-level errors are eliminated.

But not a silver bullet!
Can still write buggy programs
Concurrent code still harder than sequential code
Only for shared memory, not message passing

There is a performance hit
But it seems acceptable, and things can only get better...