

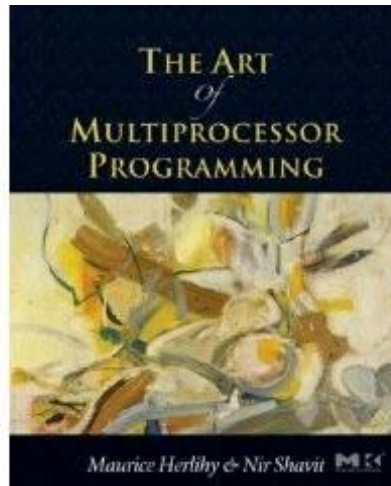
NON-BLOCKING DATA STRUCTURES AND TRANSACTIONAL MEMORY

Tim Harris, 27 November 2015

Lecture 8

- Problems with locks
- Atomic blocks and composition
- Hardware transactional memory
- Software transactional memory

Transactional Memory



Companion slides for
The Art of Multiprocessor Programming
by Maurice Herlihy & Nir Shavit

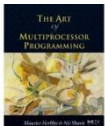
Our Vision for the Future

In this course, we covered

Best practices ...

New and clever ideas ...

And common-sense observations.



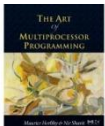
Our Vision for the Future

Nevertheless ...

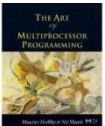
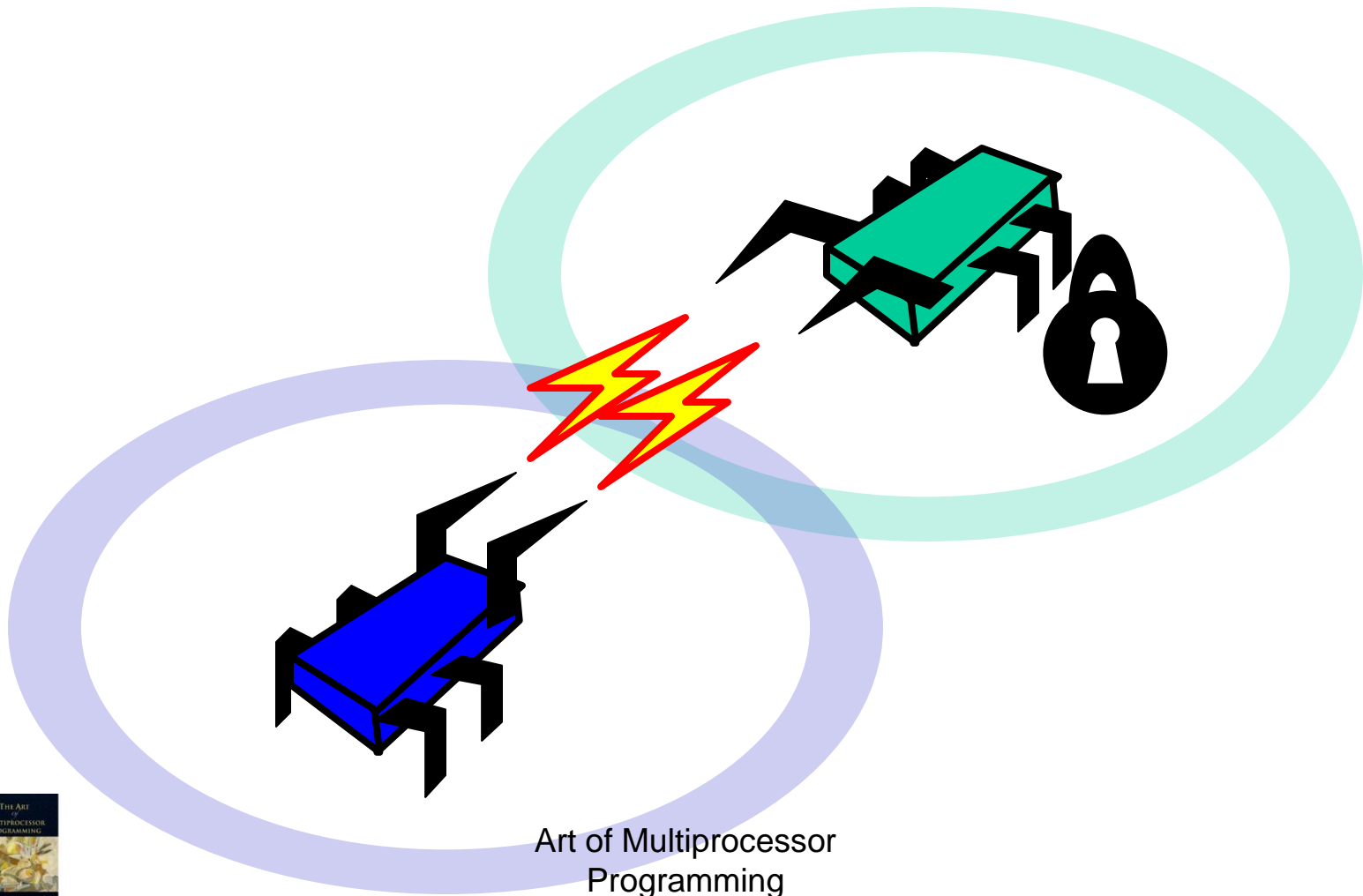
Concurrent programming is still too hard ...

Here we explore why this is

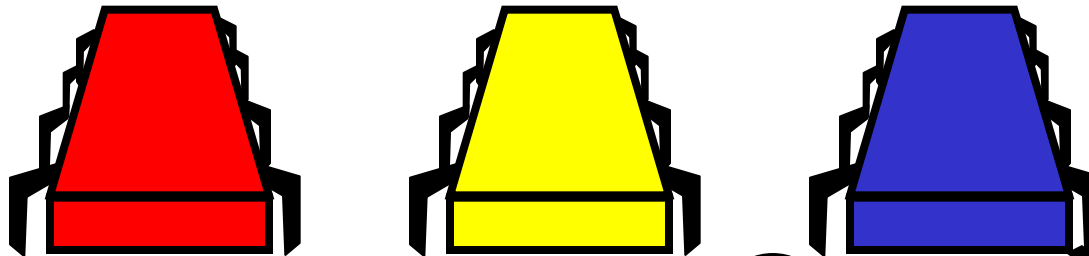
And what we can do about it.



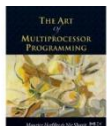
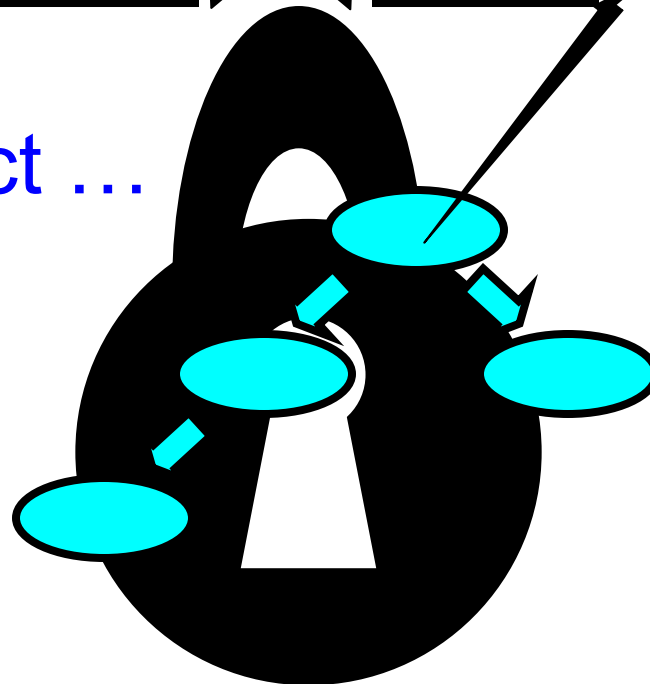
Locking



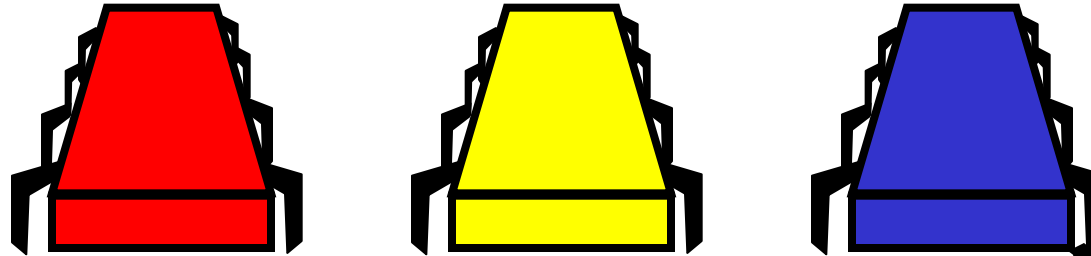
Coarse-Grained Locking



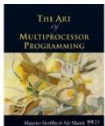
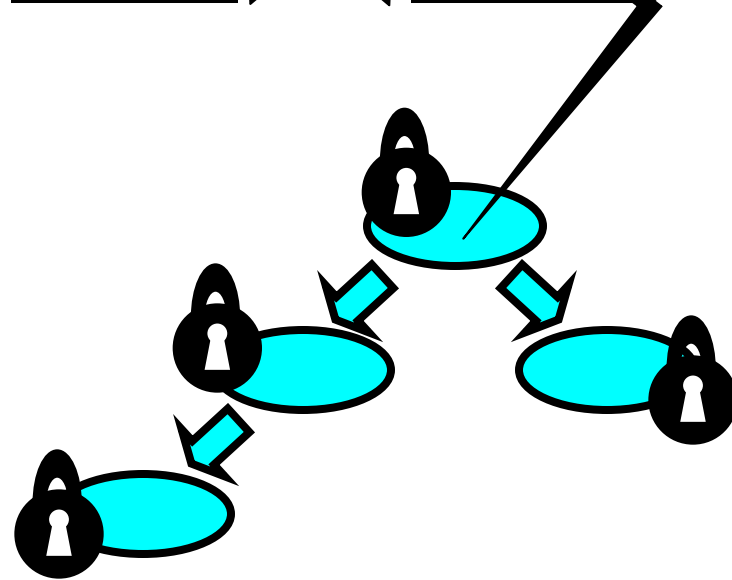
Easily made correct ...
But not scalable.



Fine-Grained Locking

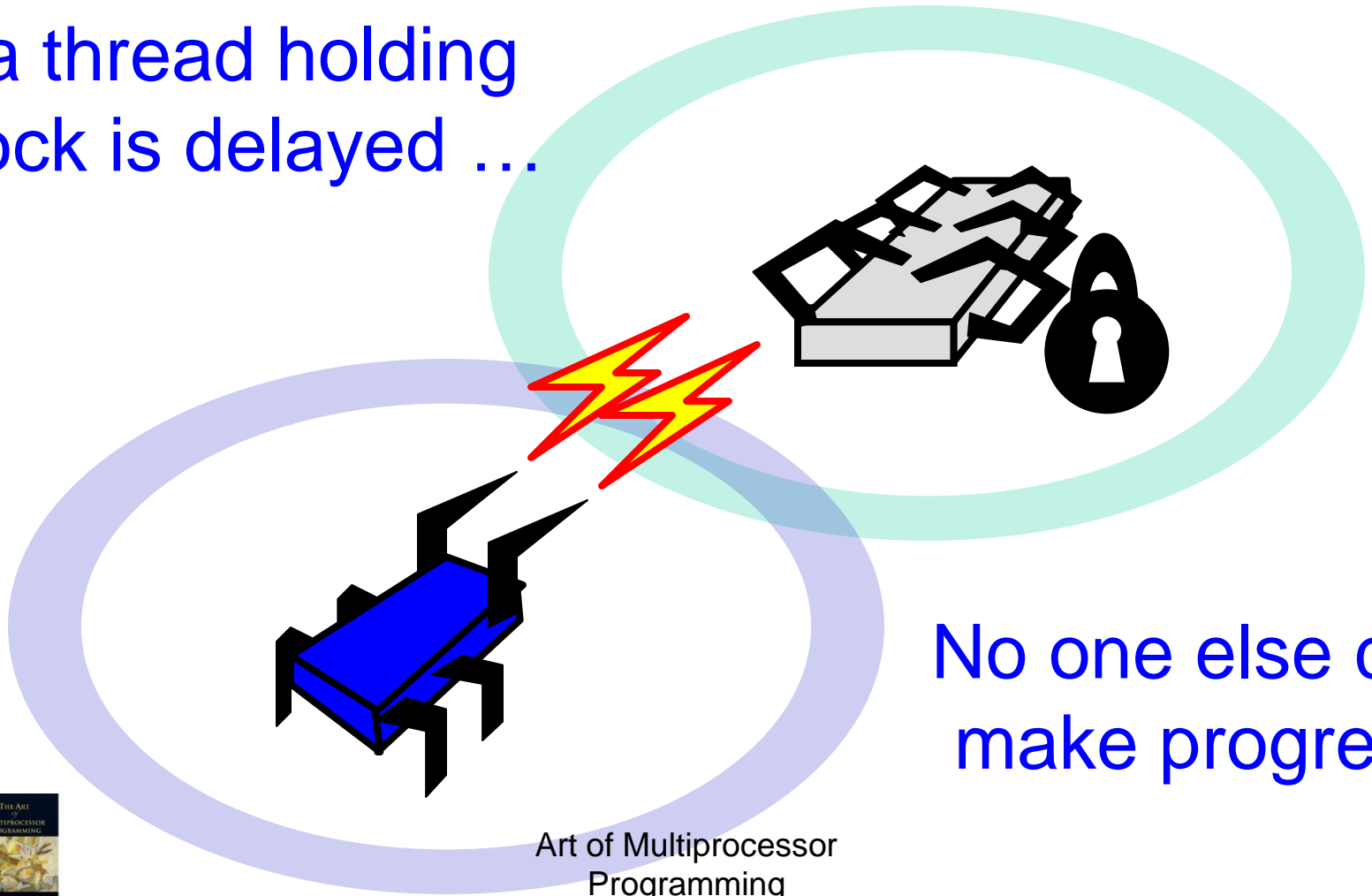


Can be tricky ...

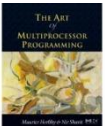


Locks are not Robust

If a thread holding a lock is delayed ...



No one else can make progress

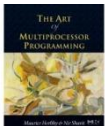


Locking Relies on Conventions

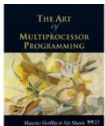
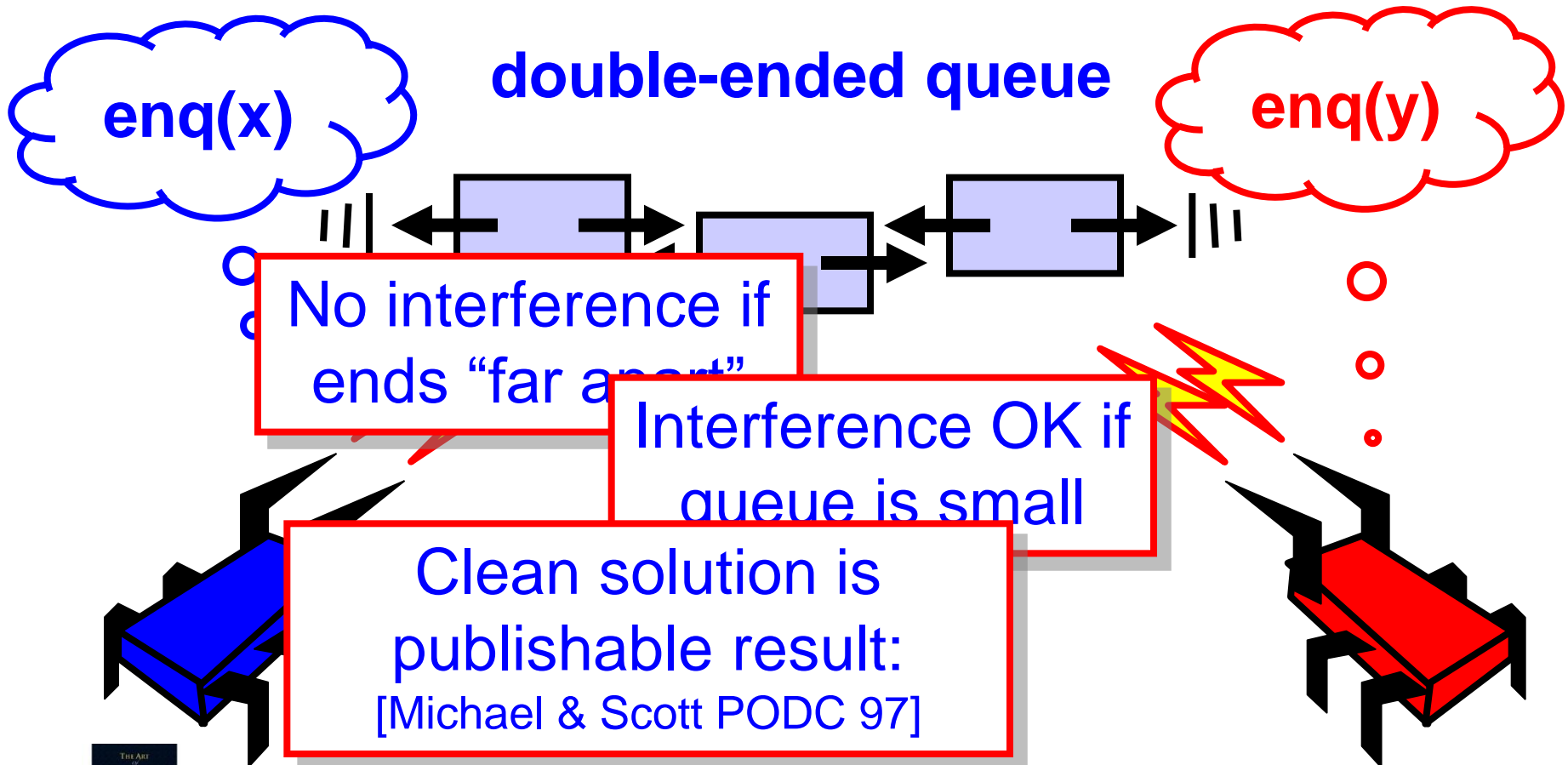
- Relation between
 - Locks and objects
 - Exists only in programmer's

Actual comment
from Linux Kernel
(hat tip: Bradley Kuszmaul)

```
/*  
 * When a locked buffer is visible to the I/O layer  
 * BH_Laundry is set. This means before unlocking  
 * we must clear BH_Laundry, mb() on alpha and then  
 * clear BH_Lock, so no reader can see BH_Laundry set  
 * on an unlocked buffer and then risk to deadlock.  
 */
```



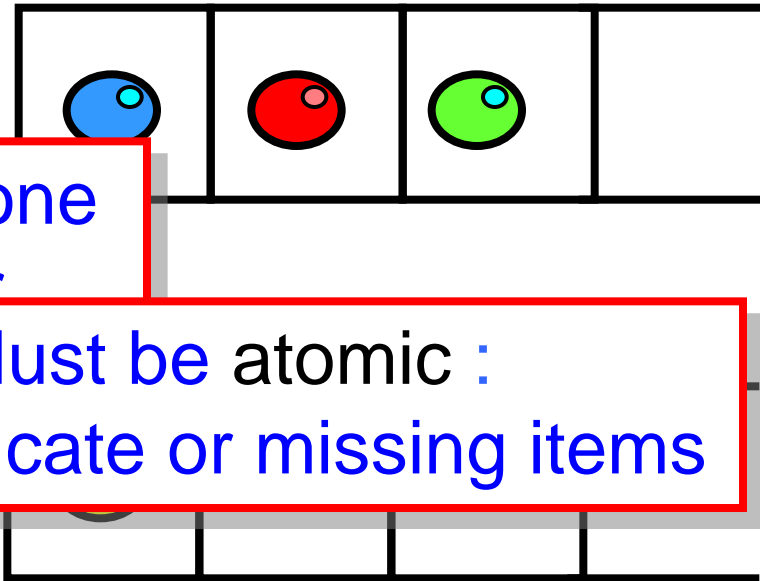
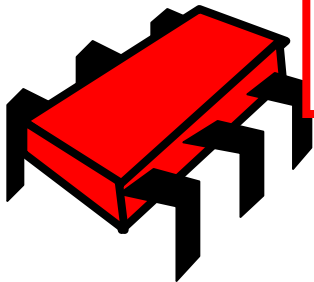
Simple Problems are hard



Locks Not Composable

Transfer item from one
queue to another

Must be atomic :
No duplicate or missing items



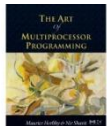
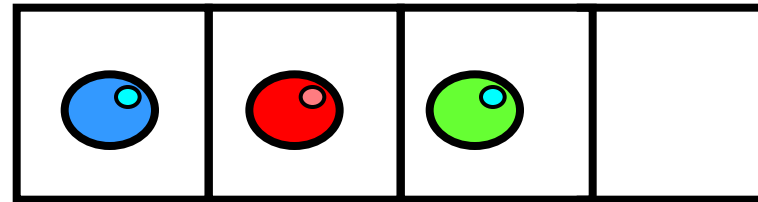
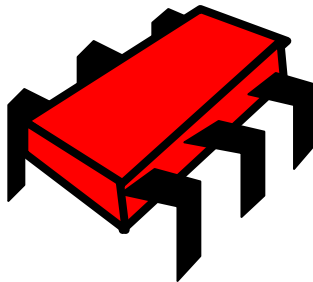
Locks Not Composable



Lock source

Unlock source
& target

Lock target



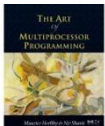
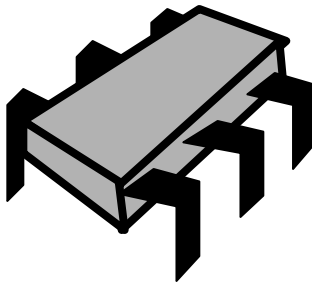
Locks Not Composable

Methods cannot provide
internal synchronization

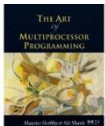
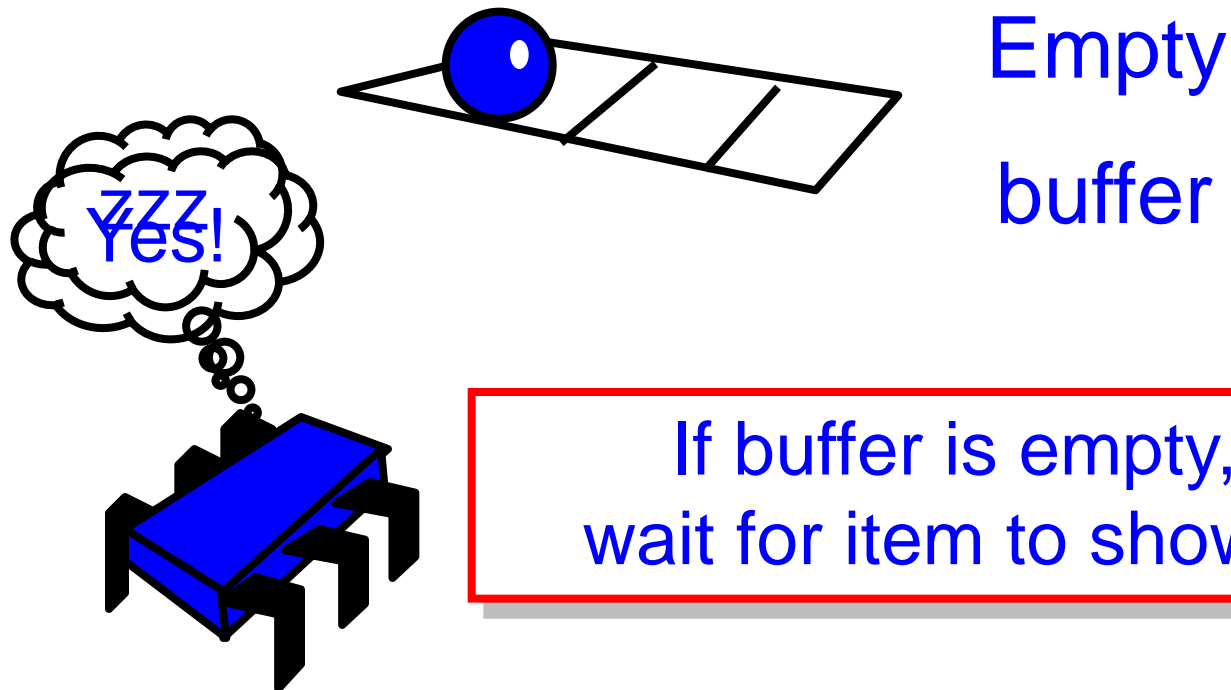
Objects must expose
locking protocols to clients

Clients must devise and
follow protocols

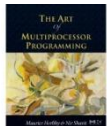
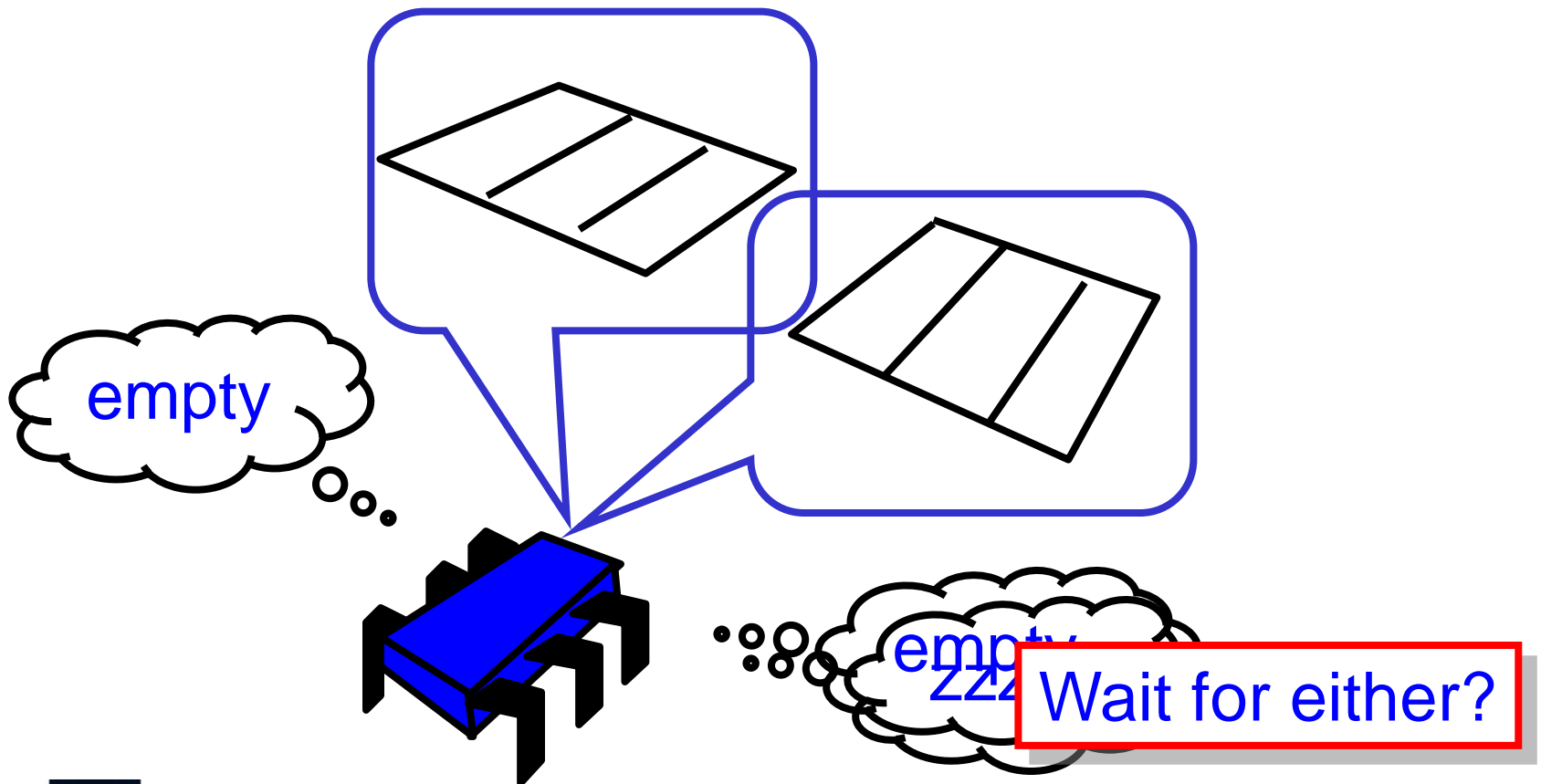
Abstraction broken!



Monitor Wait and Signal

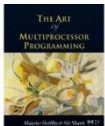


Wait and Signal do not Compose



The Transactional Manifesto

- Current practice inadequate
 - to meet the multicore challenge
- Research Agenda
 - Replace **locking** with a **transactional API**
 - **Design** languages or libraries
 - **Implement** efficient run-time systems



Transactions

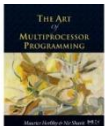
Block of code

Atomic: appears to happen instantaneously

Serializable: all appear to happen in one-at-a-time

Commit: takes effect (atomically)

Abort: has no effect (typically restarted)



Atomic Blocks

```
atomic {  
  x.remove(3);  
  y.add(3);  
}
```

```
atomic {  
  y = null;  
}
```



Atomic Blocks

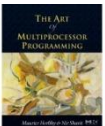
```
atomic {  
  x.remove(3);  
  y.add(3);  
}
```

y.add(3);

```
atomic {  
  y = null;  
}
```

y = null;

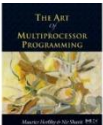
No data race



A Double-Ended Queue

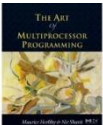
```
public void LeftEnq(item x) {  
    Qnode q = new Qnode(x);  
    q.left = left;  
    left.right = q;  
    left = q;  
}
```

Write sequential Code



A Double-Ended Queue

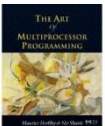
```
public void LeftEnq(item x)
atomic {
    Qnode q = new Qnode(x);
    q.left = left;
    left.right = q;
    left = q;
}
}
```



A Double-Ended Queue

```
public void LeftEnq(item x) {  
    atomic {  
        Qnode q = new Qnode(x);  
        q.left = left;  
        left.right = q;  
        left = q;  
    }  
}
```

Enclose in atomic block

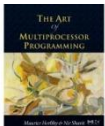
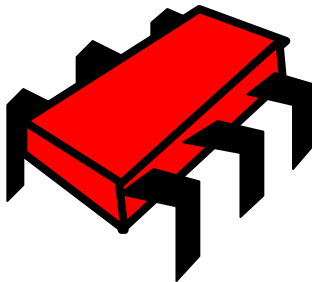
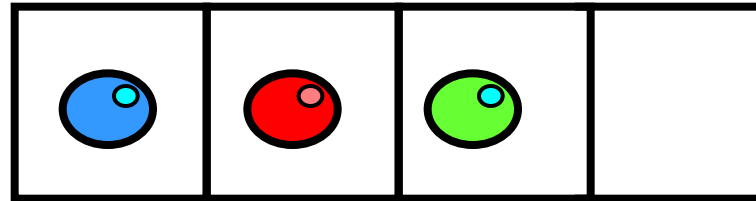


Warning

- Not always this simple
 - Conditional waits
 - Enhanced concurrency
 - Complex patterns
- But often it is...



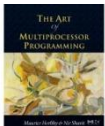
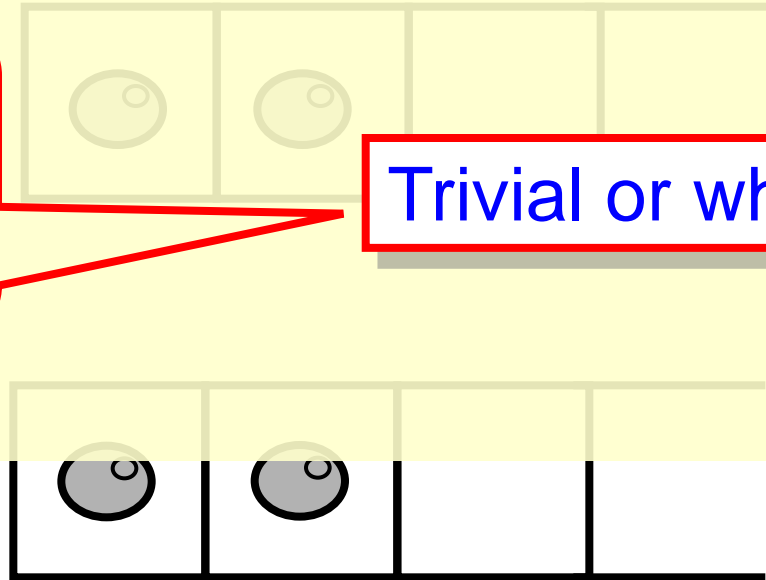
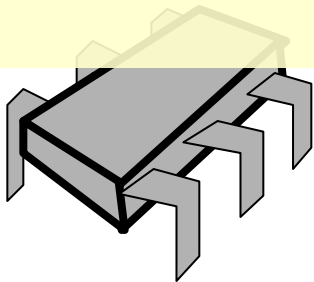
Composition?



Composition?

```
public void Transfer (Queue<T> q1, q2)
{
  atomic {
    T x = q1.deq();
    q2.enq(x);
  }
}
```

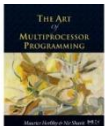
Trivial or what?



Conditional Waiting

```
public T LeftDeq() {  
    atomic {  
        if (left == null)  
            retry;  
        ...  
    }  
}
```

**Roll back transaction
and restart when
something changes**



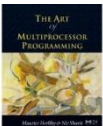
Composable Conditional Waiting

```
atomic {  
  x = q1.deq();  
} orElse {  
  x = q2.deq();  
}
```

Run 1st method. If it retries

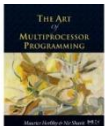
Run 2nd method. If it retries ...

Entire statement retries



Hardware Transactional Memory

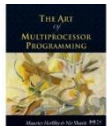
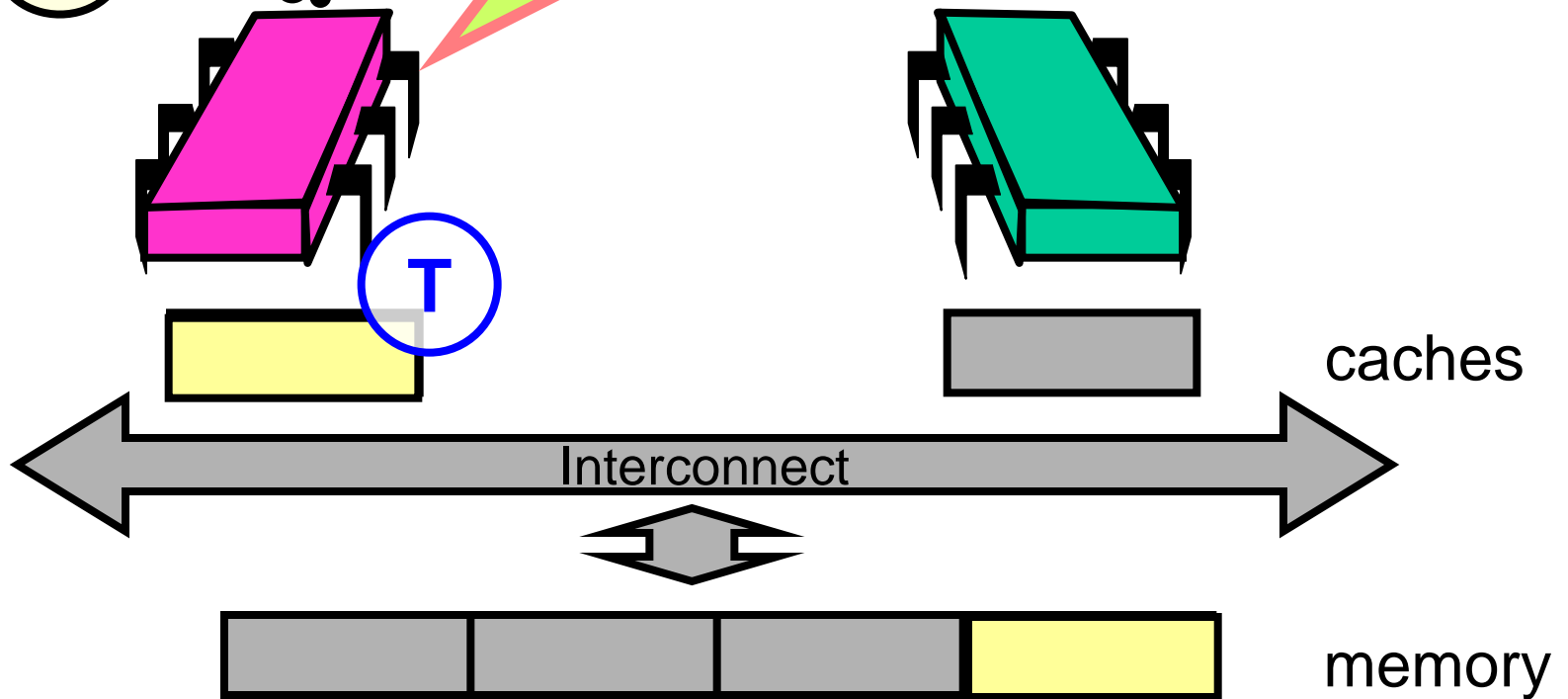
- Exploit Cache coherence
- Already almost does it
 - Invalidation
 - Consistency checking
- Speculative execution
 - Branch prediction = optimistic synch!



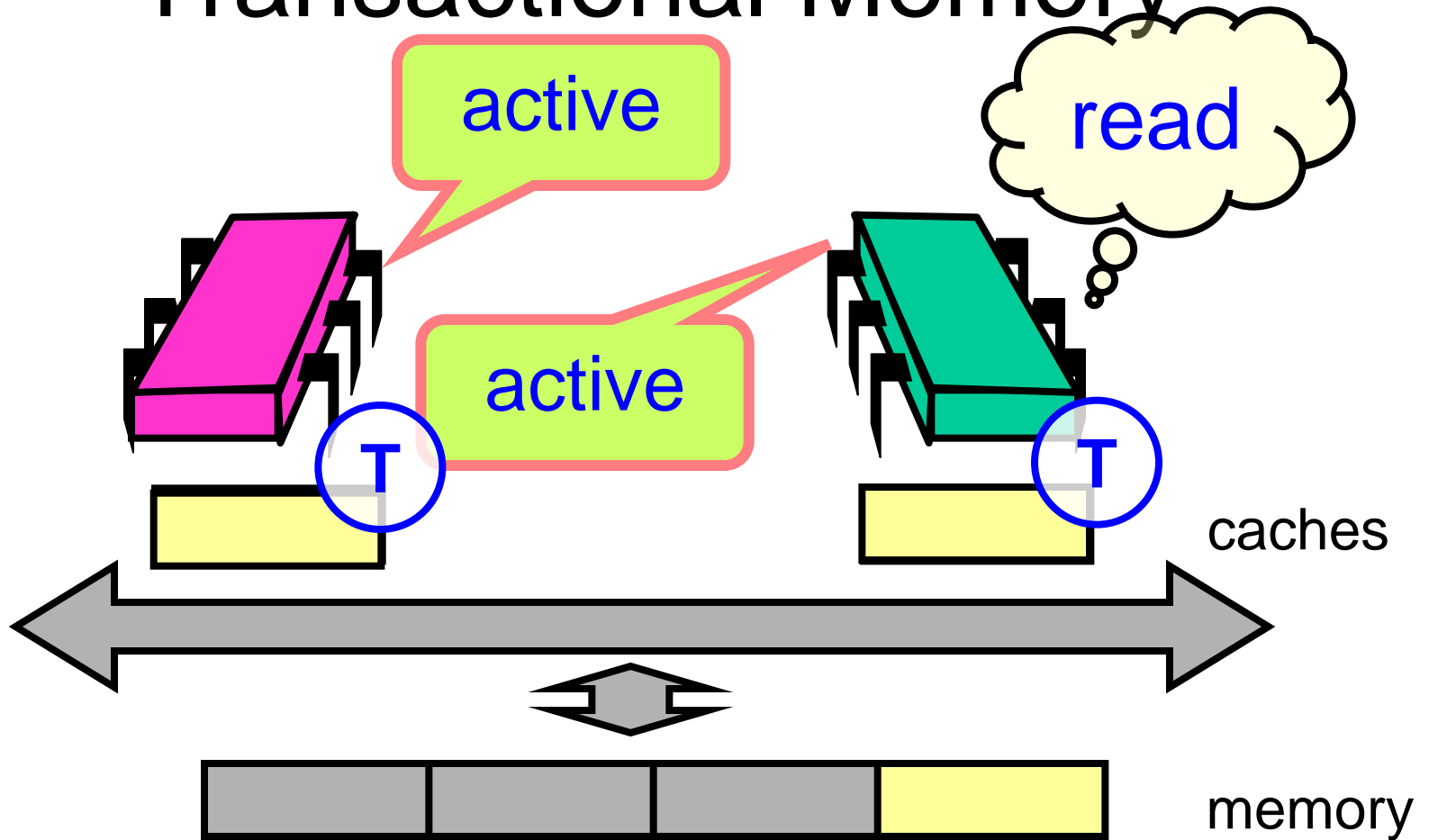
HW Transactional Memory

read

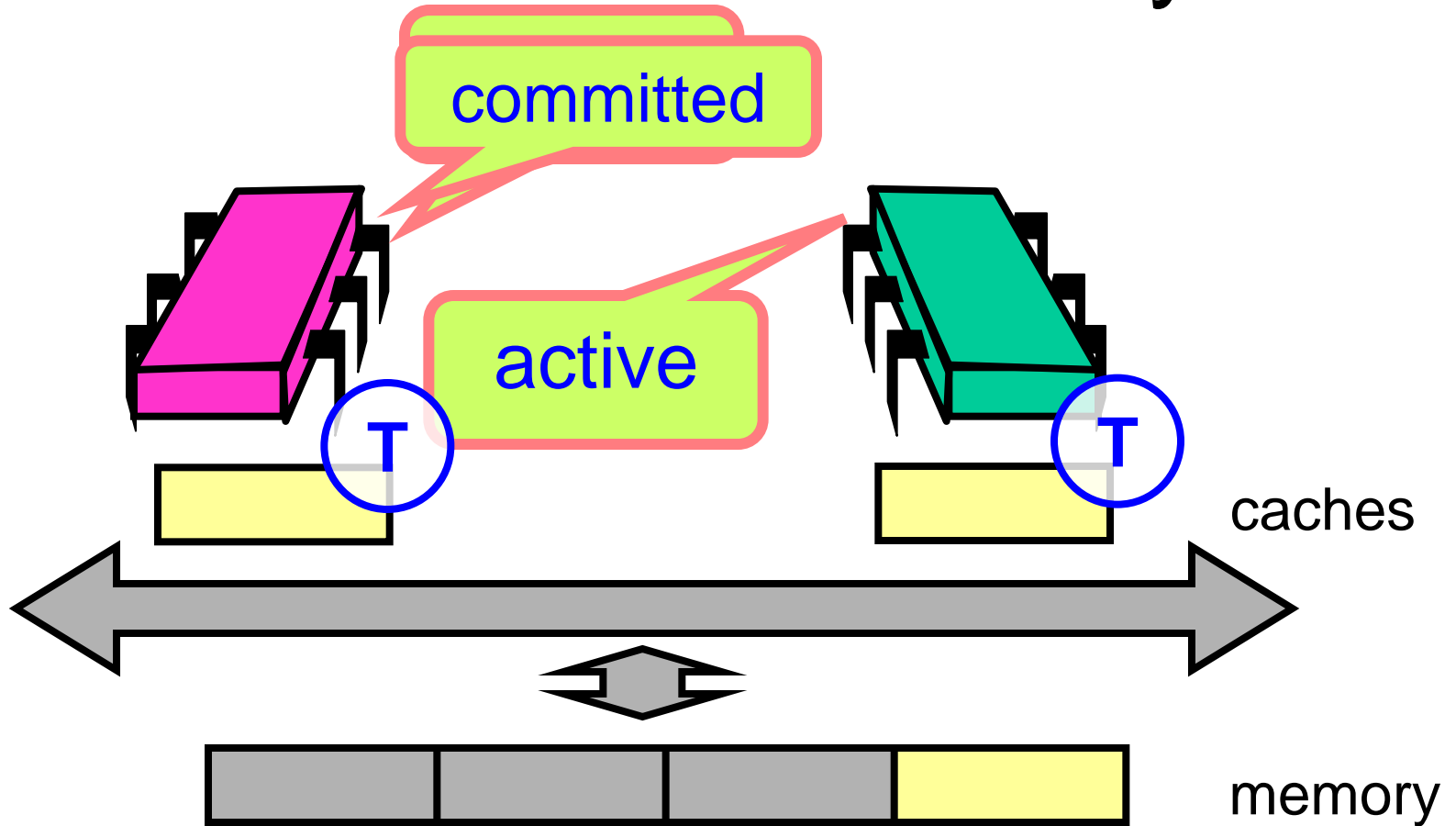
active



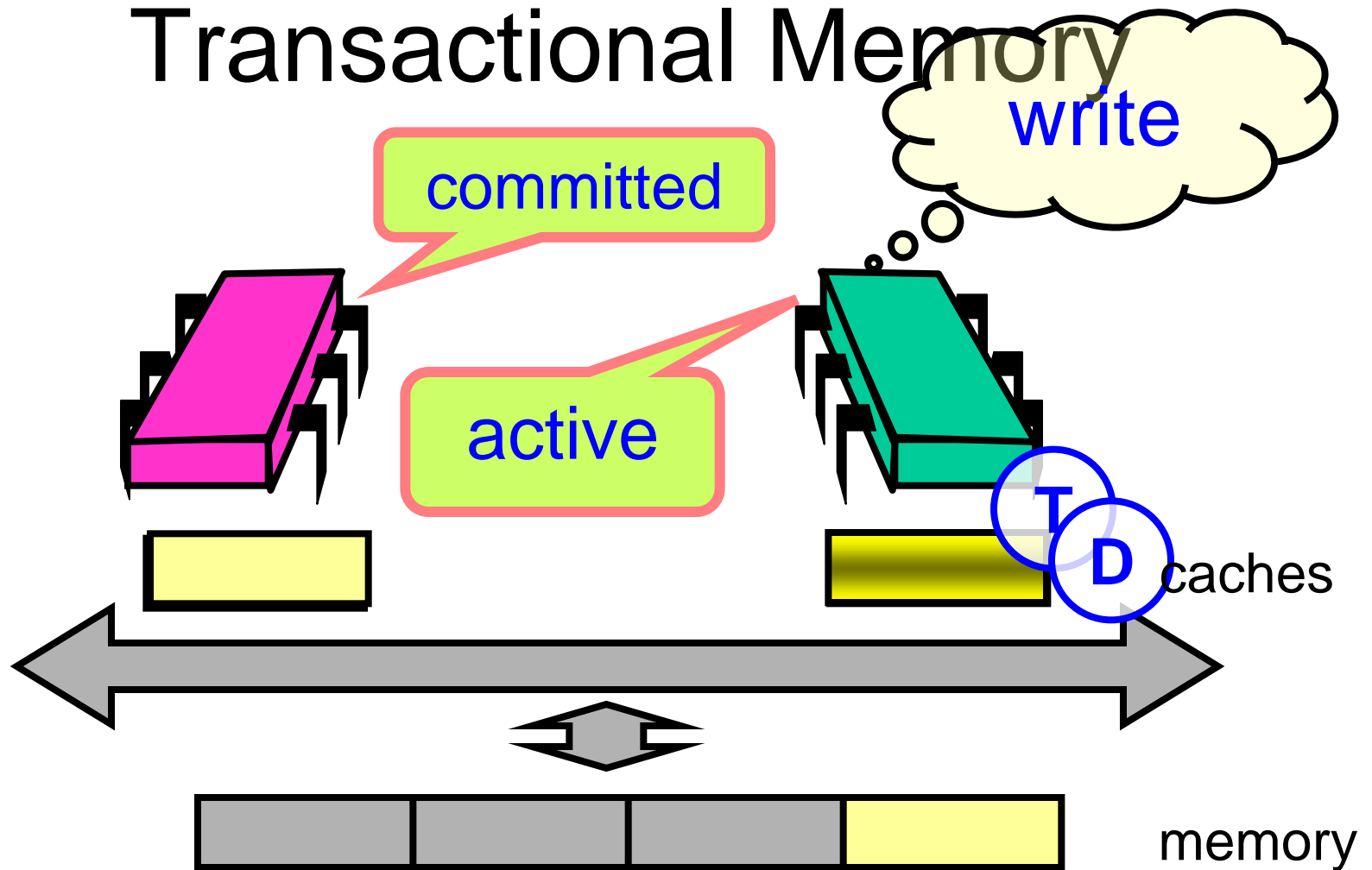
Transactional Memory

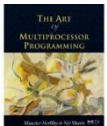
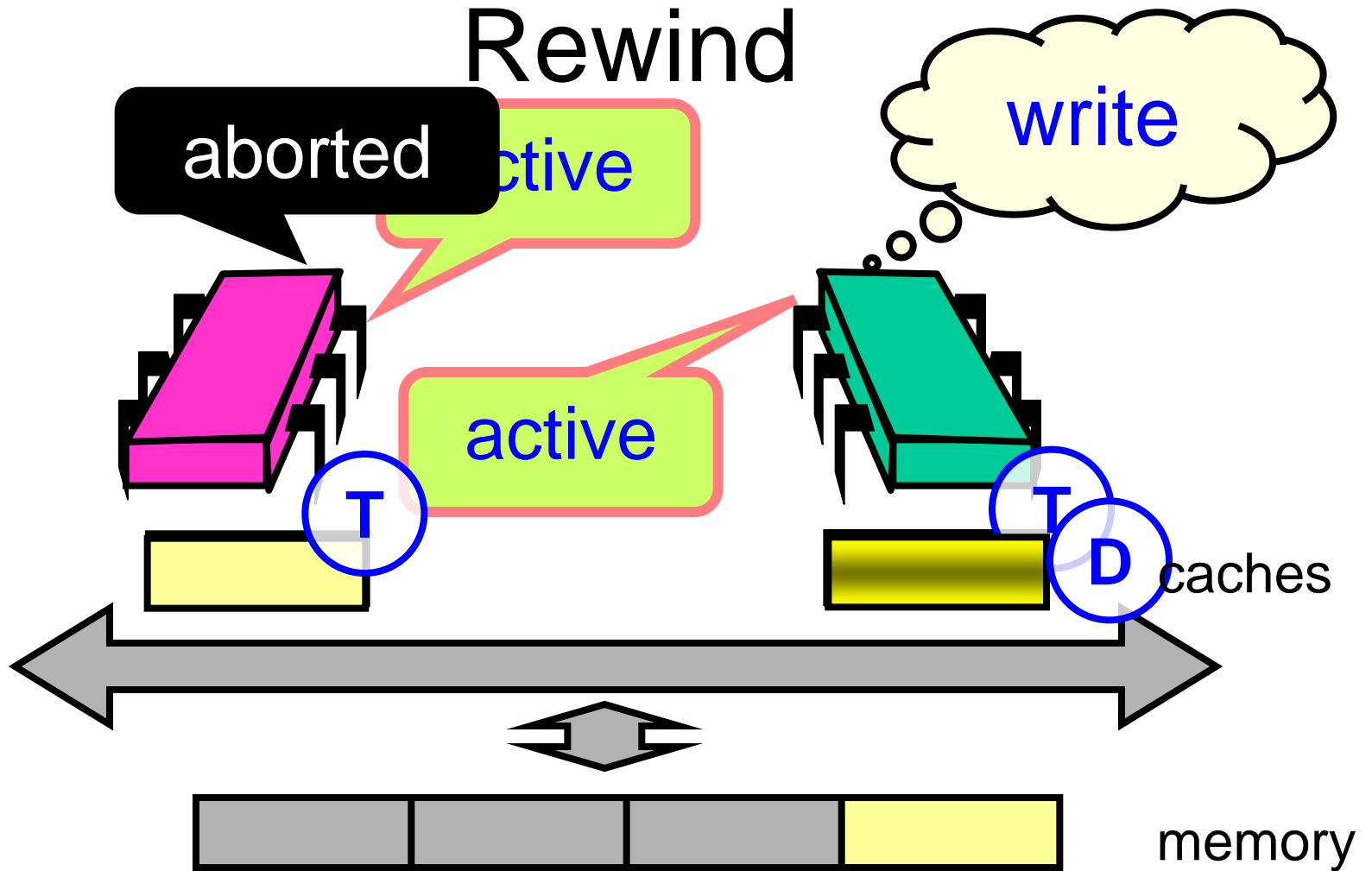


Transactional Memory



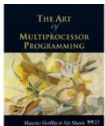
Transactional Memory





Transaction Commit

- At commit point
 - If no cache conflicts, we win.
- Mark transactional entries
 - Read-only: valid
 - Modified: dirty (eventually written back)
- That's all, folks!
 - Except for a few details ...



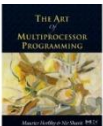
Not all  and



- Limits to
 - Transactional cache size
 - Scheduling quantum
- Transaction cannot commit if it is
 - Too big
 - Too slow
 - Actual limits platform-dependent

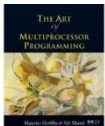
HTM Strengths & Weaknesses

- Ideal for lock-free data structures



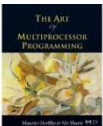
HTM Strengths & Weaknesses

- Ideal for lock-free data structures
- Practical proposals have limits on
 - Transaction size and length
 - Bounded HW resources
 - Guarantees vs best-effort



HTM Strengths & Weaknesses

- Ideal for lock-free data structures
- Practical proposals have limits on
 - Transaction size and length
 - Bounded HW resources
 - Guarantees vs best-effort
- On fail
 - Diagnostics essential
 - Try again in software?



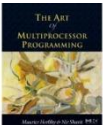
Composition

Locks don't compose, transactions do.

Composition necessary for Software Engineering.

But practical HTM doesn't *really* support composition!

Why we need STM



Note: this post-publication cited in the publication cited in the Appendix has been fixed, and an Appendix

Abstract

Writing concurrent programs is notoriously difficult. Increasing practical importance. A particular source of concern is that even correctly-implemented concurrency abstractions cannot be composed together to form larger abstractions. In this paper we present a new concurrency model, based on *transactional memory* (TM), which allows composition. All the usual benefits of transactional memory (e.g. freedom from deadlock), but in addition, it provides forms of blocking and choice

- We present a new concurrency model to the program

Composability Transactions
publication version 2006

Simon Peyton Jones
Cambridge
[simonpj@microsoft.com]

Tim Harris

{tharris, simonpj}@microsoft.com

2005

vent threads from providing a convincing story. However, we resolve these shortcomings with the following contributions:

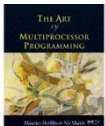
the ideas of transactional memory (Section 3). This is much more than a declarative language, and we show, STM can be expressed in a far stronger guarantee. More transactions together to

interfaces and it did not block. In this case, make the

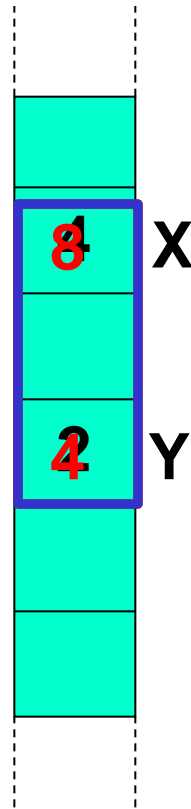
in the setting which appears more than a a. STM can be expressed in a far stronger guarantee. More transactions together to which appears more than a a. STM can be expressed in a far stronger guarantee. More transactions together to

Transactional Consistency

- Memory Transactions are collections of reads and writes executed atomically
- They should maintain *consistency*
 - *External*: with respect to the interleavings of other transactions (*linearizability*).
 - *Internal*: the transaction itself should operate on a consistent state.



External Consistency



Invariant $x = 2y$

Transaction A:

Write x

Write y

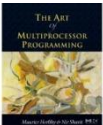
Transaction B:

Read x

Read y

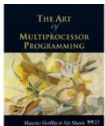
Compute $z = 1/(x-y) = 1/2$

Application
Memory



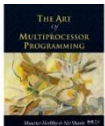
A Simple Lock-Based STM

- STMs come in different forms
 - Lock-based
 - Lock-free
- Here : a simple lock-based STM
- Lets start by Guaranteeing External Consistency

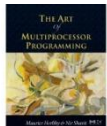
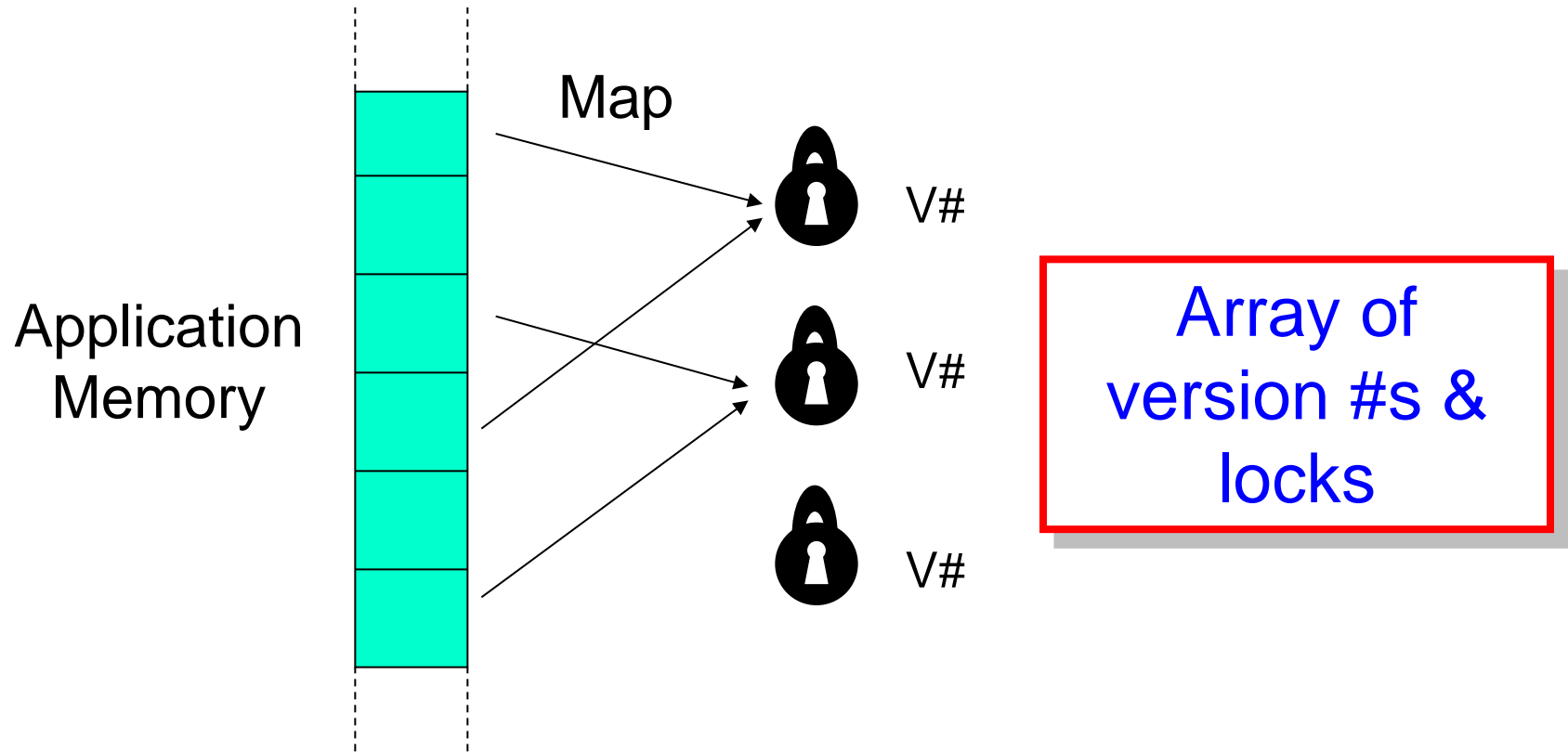


Synchronization

- Transaction keeps
 - **Read set**: locations & values read
 - **Write set**: locations & values to be written
- Deferred update
 - Changes installed at commit
- Lazy conflict detection
 - Conflicts detected at commit



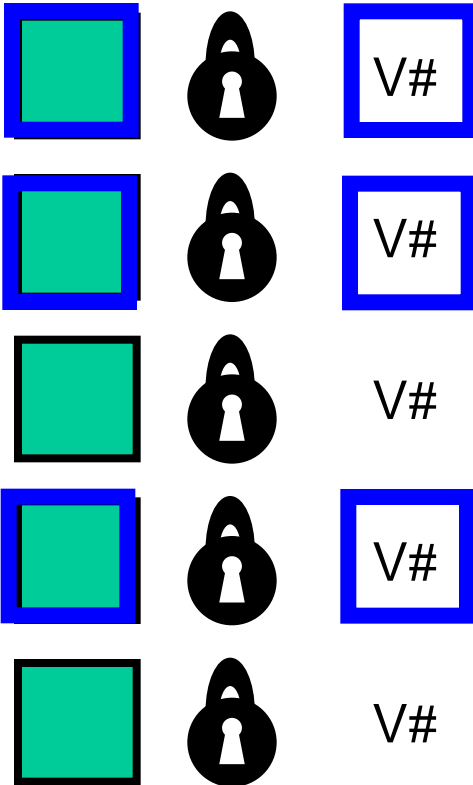
STM: Transactional Locking



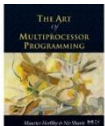
Reading an Object

Mem

Locks



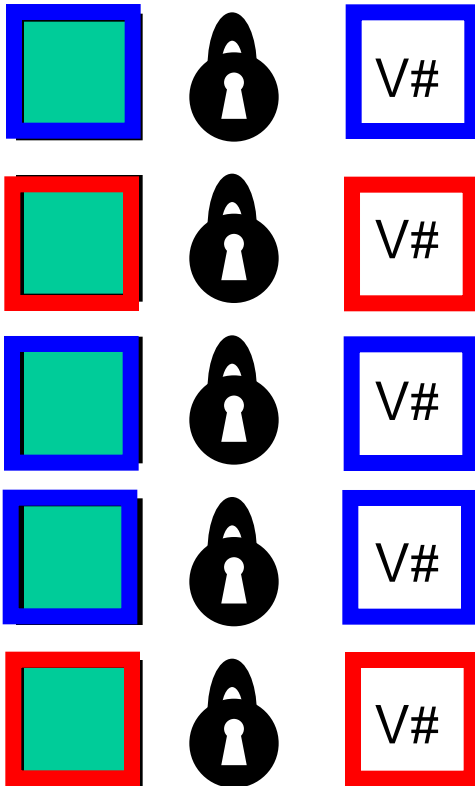
Add version numbers
& values to read set



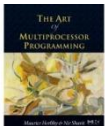
To Write an Object

Mem

Locks



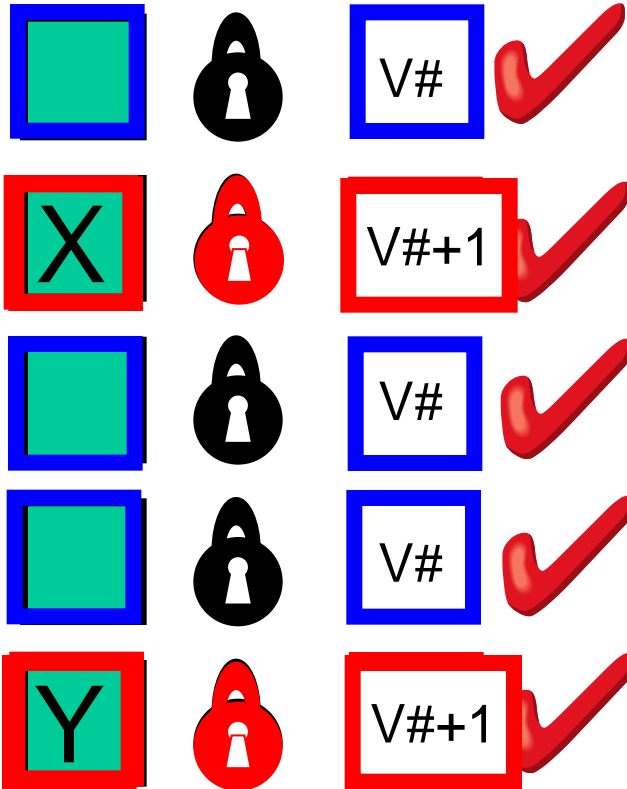
Add version numbers & new values to write set



To Commit

Mem

Locks



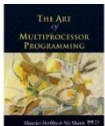
Acquire write locks

Check version numbers
unchanged

Install new values

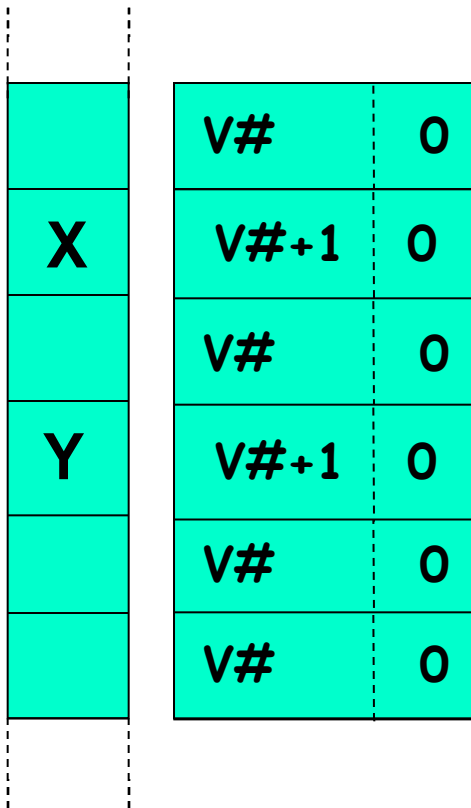
Increment version numbers

Unlock.



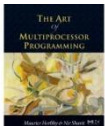
Encounter Order Locking (Undo Log)

Mem Locks

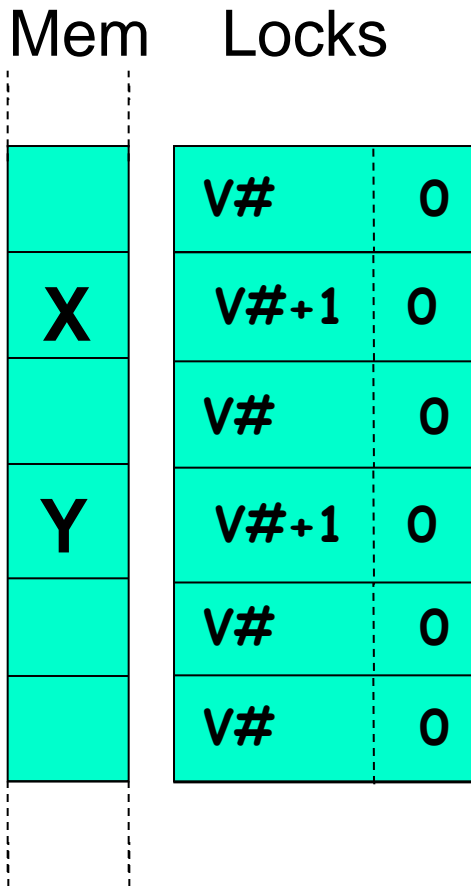


1. To Read: load lock + location
2. Check unlocked add to Read-Set
3. To Write: lock location, store value
4. Add old value to undo-set
5. Validate read-set v#'s unchanged
6. Release each lock with v#+1

Quick read of values freshly written by the reading transaction



Commit Time Locking (Write Buff)

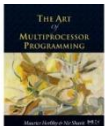
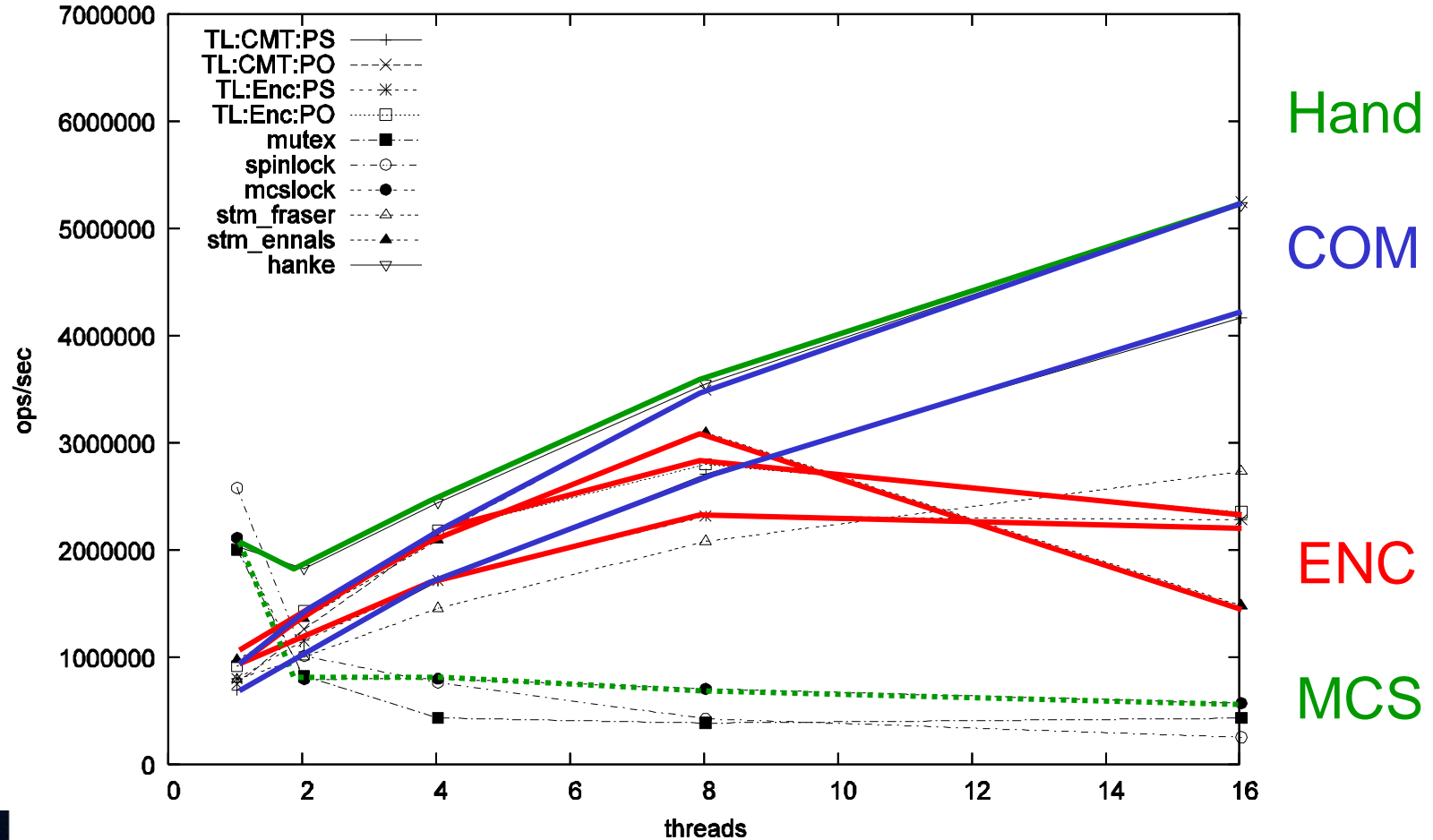


1. To Read: load lock + location
2. Location in write-set? (Bloom Filter)
3. Check unlocked add to Read-Set
4. To Write: add value to write set
5. **Acquire Locks**
6. Validate read/write v#'s unchanged
7. **Release each lock with v#+1**

Hold locks for very short duration

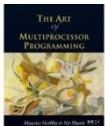
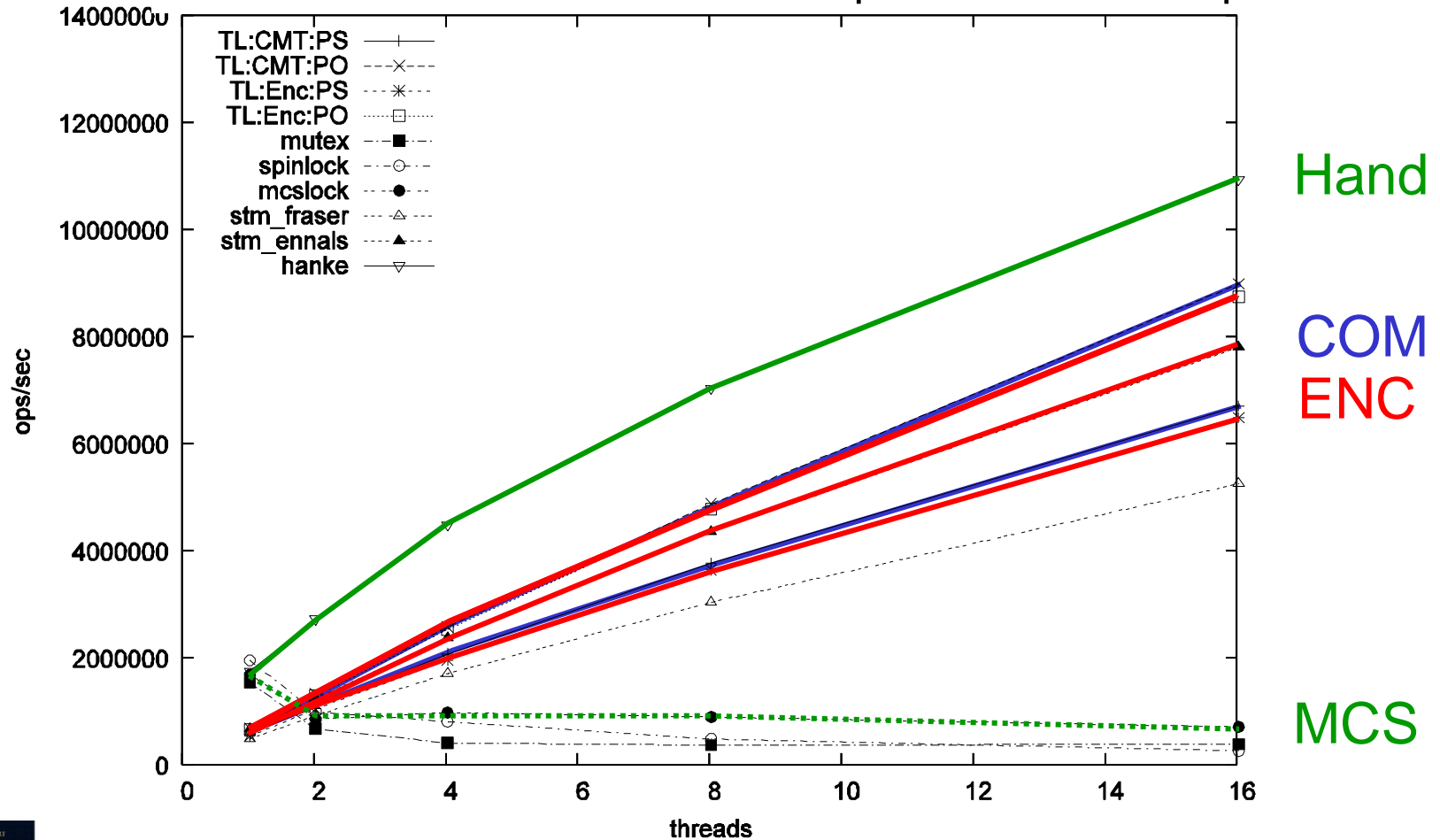
COM vs. ENC High Load

Red-Black Tree 20% Delete 20% Update 60% Lookup



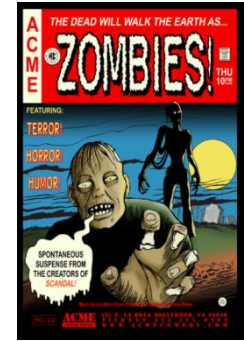
COM vs. ENC Low Load

Red-Black Tree 5% Delete 5% Update 90% Lookup

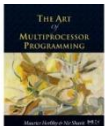




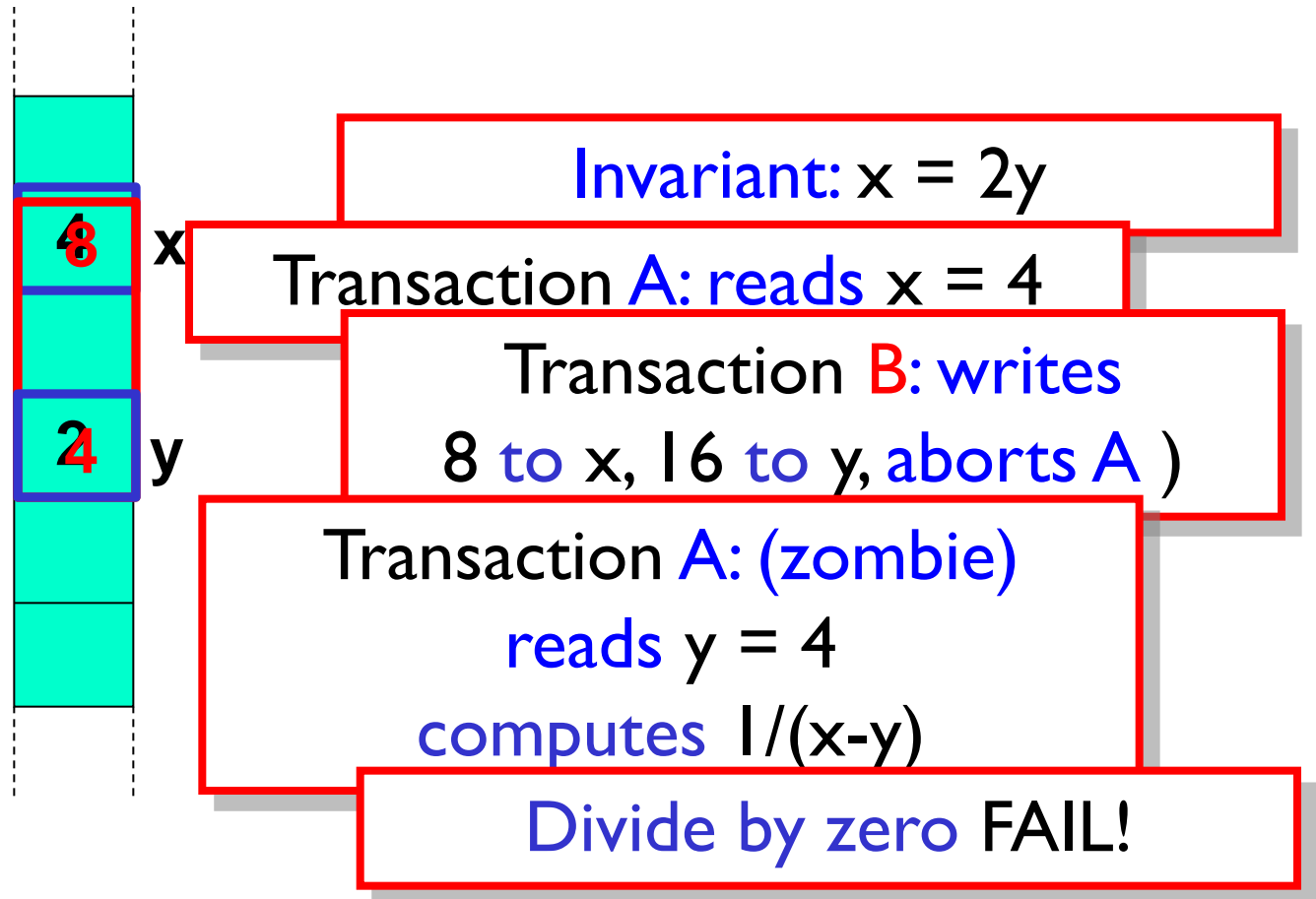
Problem: Internal Inconsistency



- A Zombie is an active transaction destined to abort.
- If Zombies see inconsistent states bad things can happen

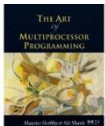


Internal Consistency



Solution: The Global Clock (The TL2 Algorithm)

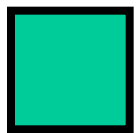
- Have one shared global clock
- Incremented by (small subset of) writing transactions
- Read by all transactions
- Used to validate that state worked on is always consistent



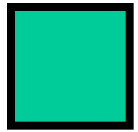
Read-Only Transactions

Mem

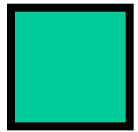
Locks



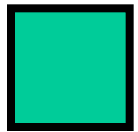
12



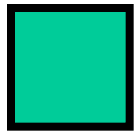
32



56

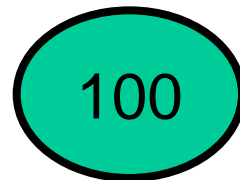


19

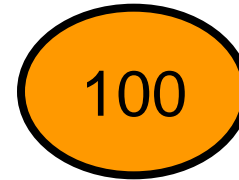


17

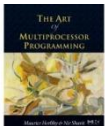
Copy version clock to local read version clock



Shared Version Clock



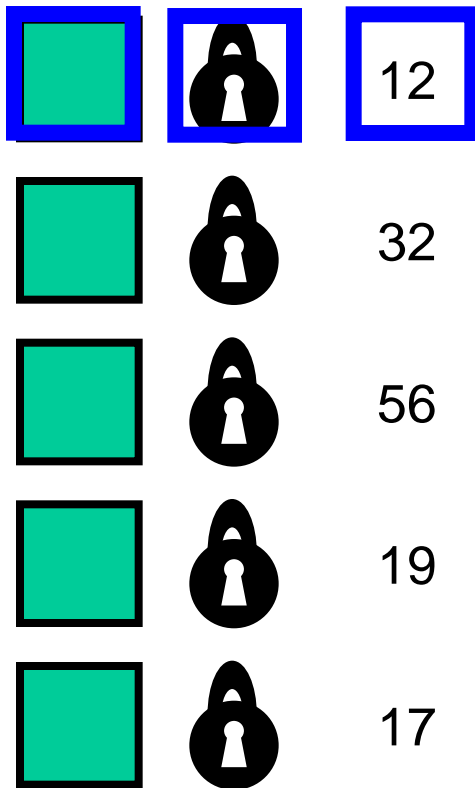
Private Read Version (RV)



Read-Only Transactions

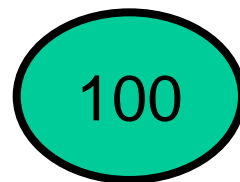
Mem

Locks

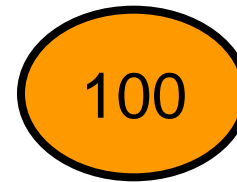


Copy version clock to local

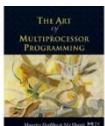
Read lock, version #, and memory



Shared Version Clock



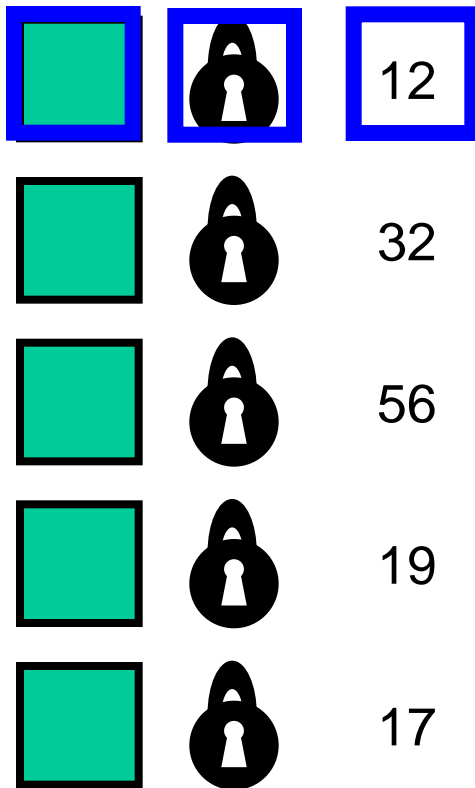
Private Read Version (RV)



Read-Only Transactions

Mem

Locks



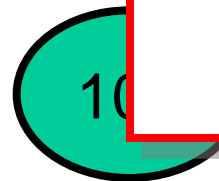
Copy version clock to local

Read lock, version #, and

memory address to local version #s

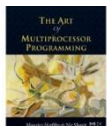
On Commit:

check unlocked &
version #s less than
local read clock



Shared Version
Clock

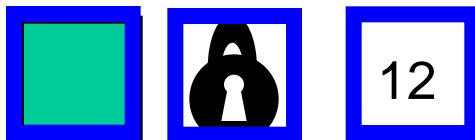
Private Read
Version (RV)



Read-Only Transactions

Mem

Locks



Copy version clock to local

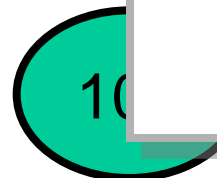


Read lock, version #, and

We have taken a snapshot without keeping an explicit read set!

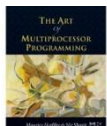


version #s less than local lock

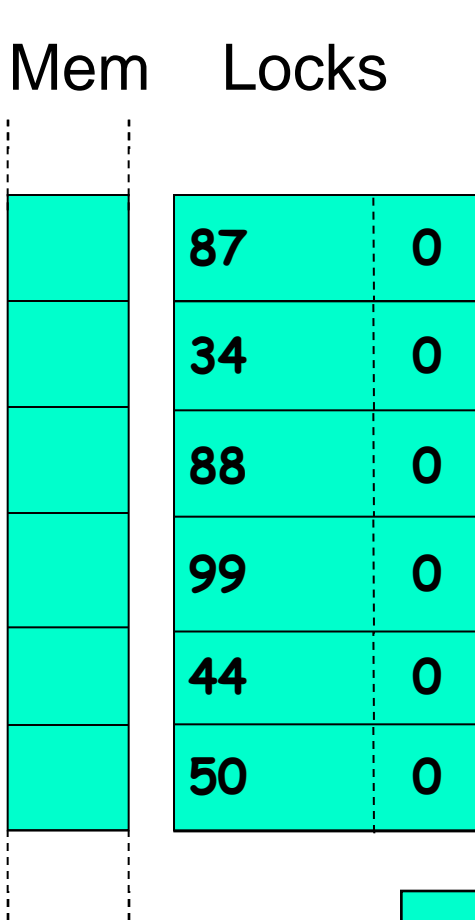


Shared Version Clock

Private Read Version (RV)



Example Execution: Read Only Trans



100

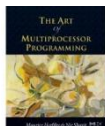
Shared Version Clock

1. $RV \leftarrow$ Shared Version Clock
2. On Read: read lock, read mem, read lock: check unlocked, unchanged, and $v\# \leq RV$
3. Commit.

Reads form a snapshot of memory.
No read set!

100

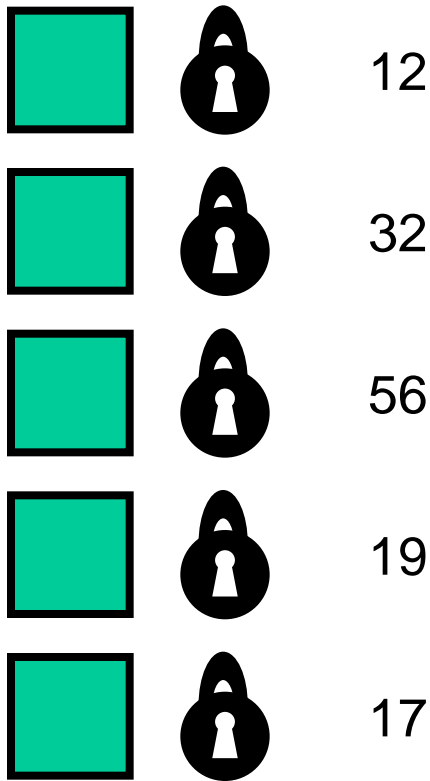
RV



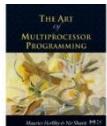
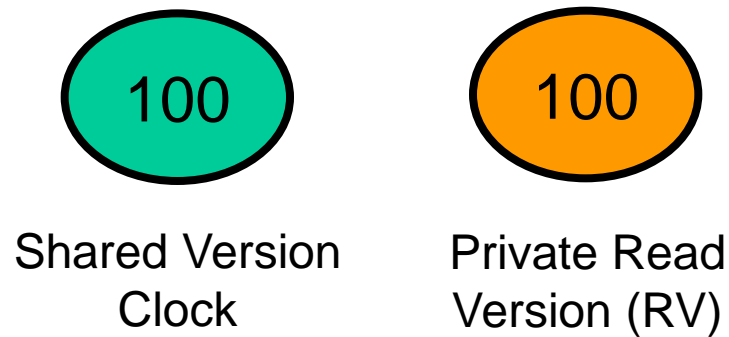
Ordinary (Writing) Transactions

Mem

Locks



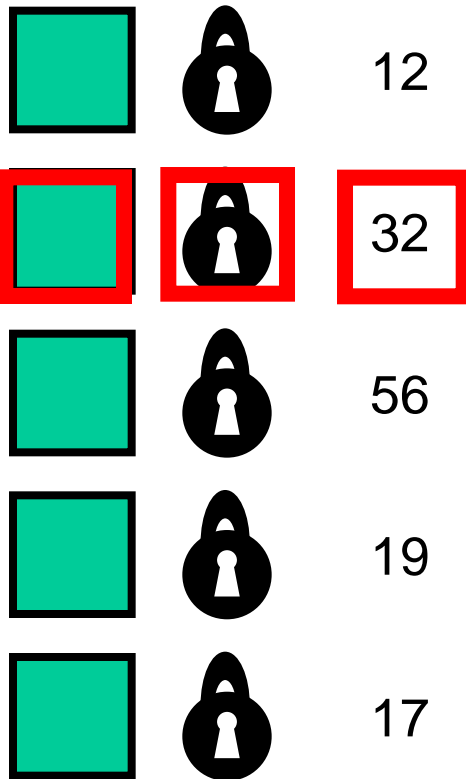
Copy version clock to local read version clock



Ordinary Transactions

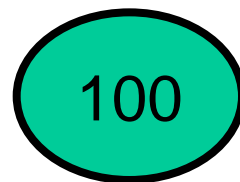
Mem

Locks

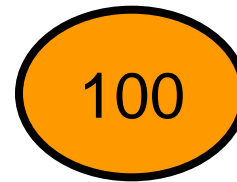


Copy version clock to local read version clock

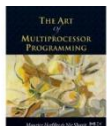
On read/write, check:
Unlocked & version # < RV
Add to R/W set



Shared Version Clock



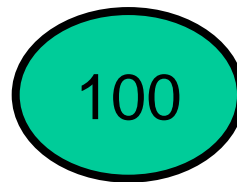
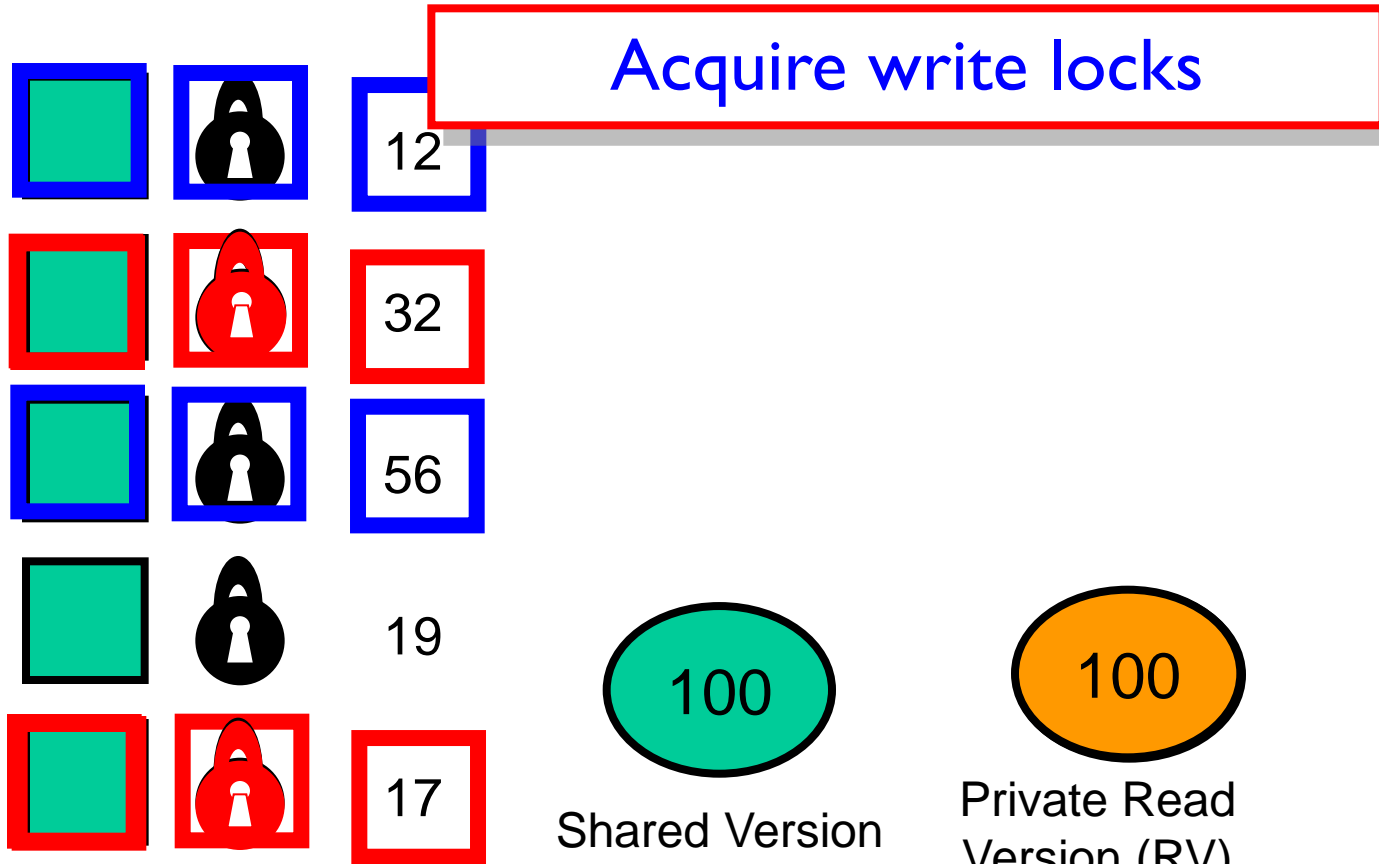
Private Read Version (RV)



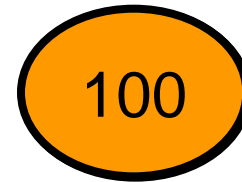
On Commit

Mem

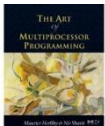
Locks



Shared Version Clock



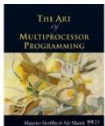
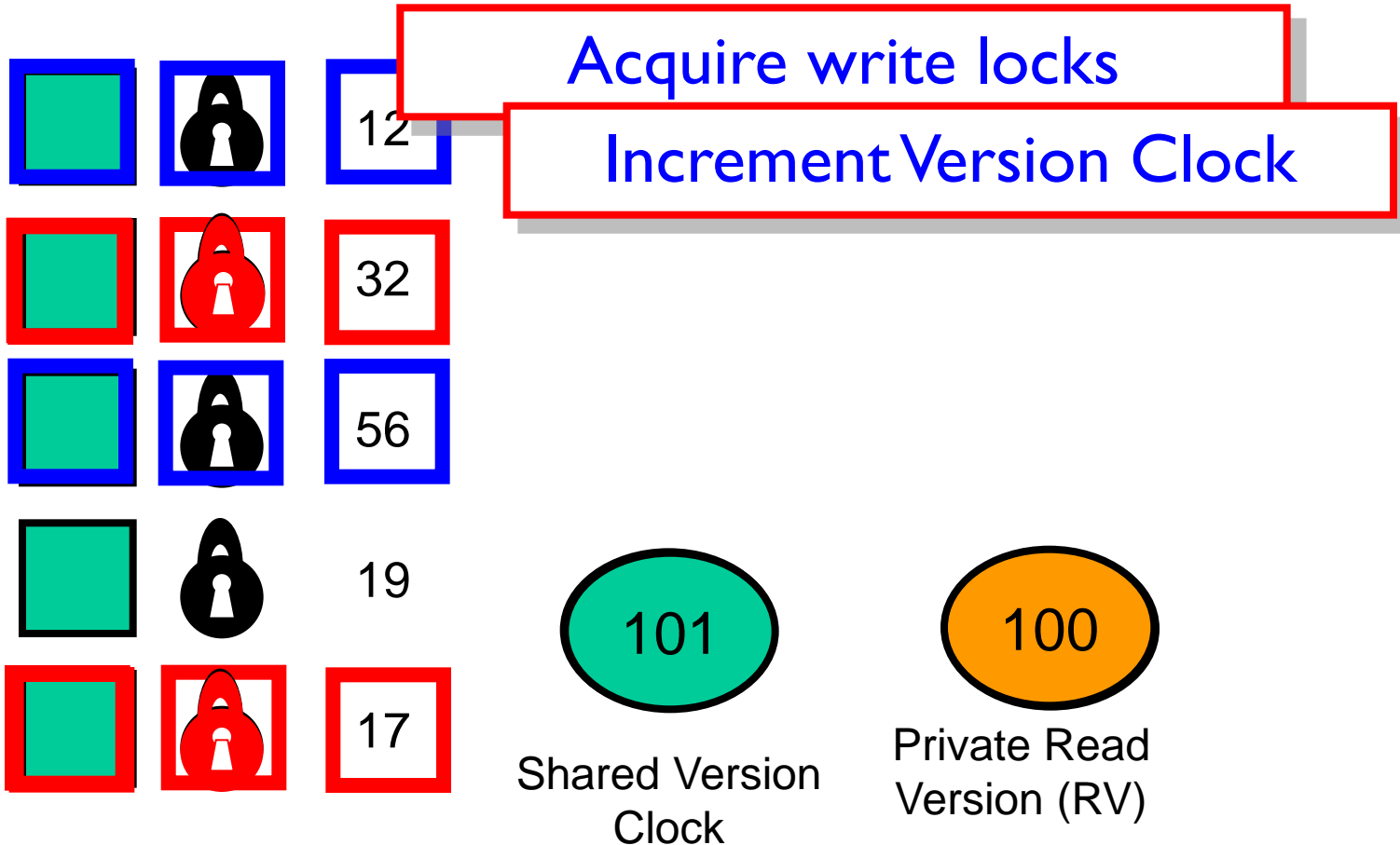
Private Read Version (RV)



On Commit

Mem

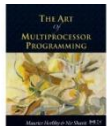
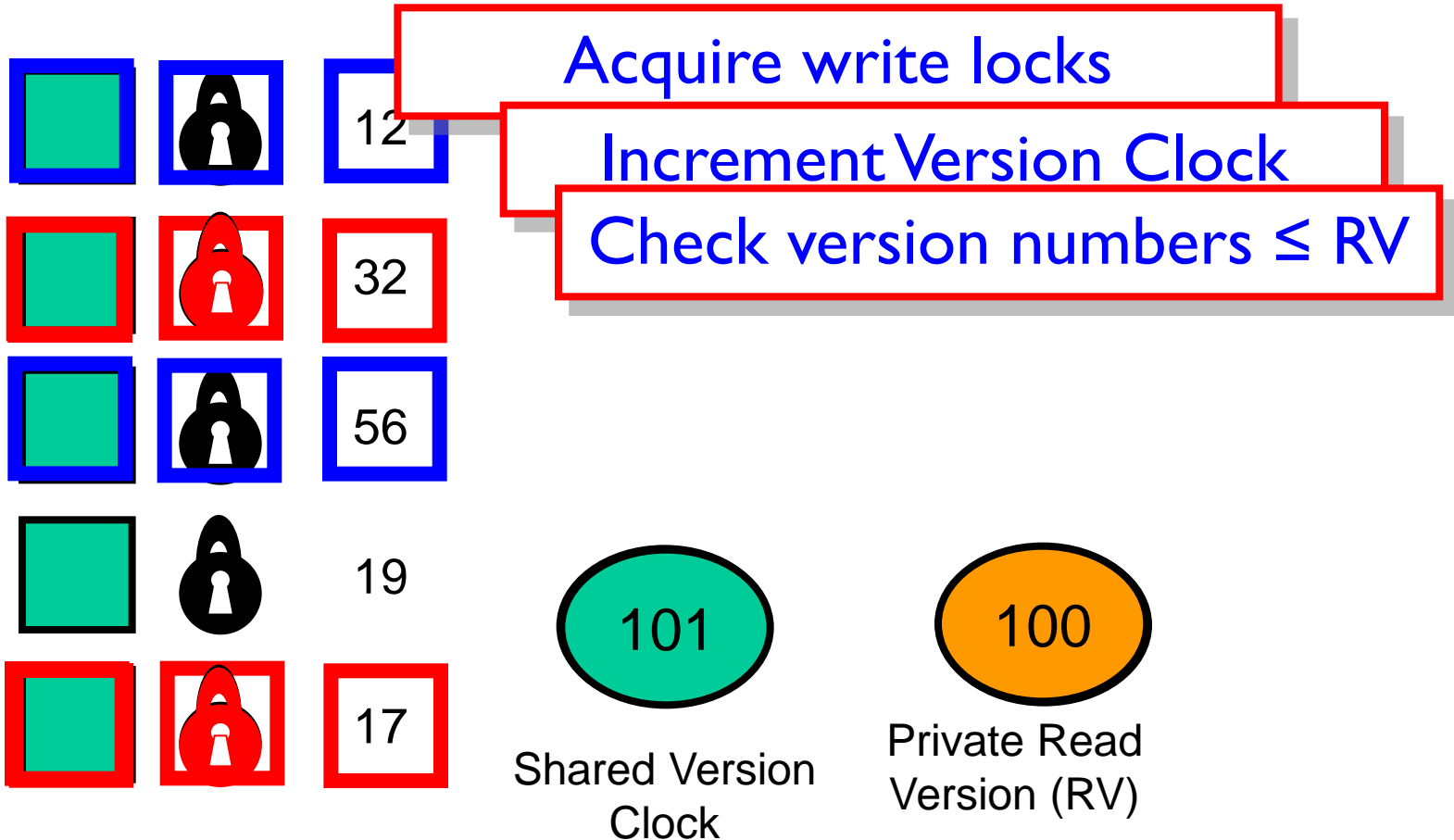
Locks



On Commit

Mem

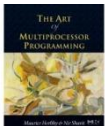
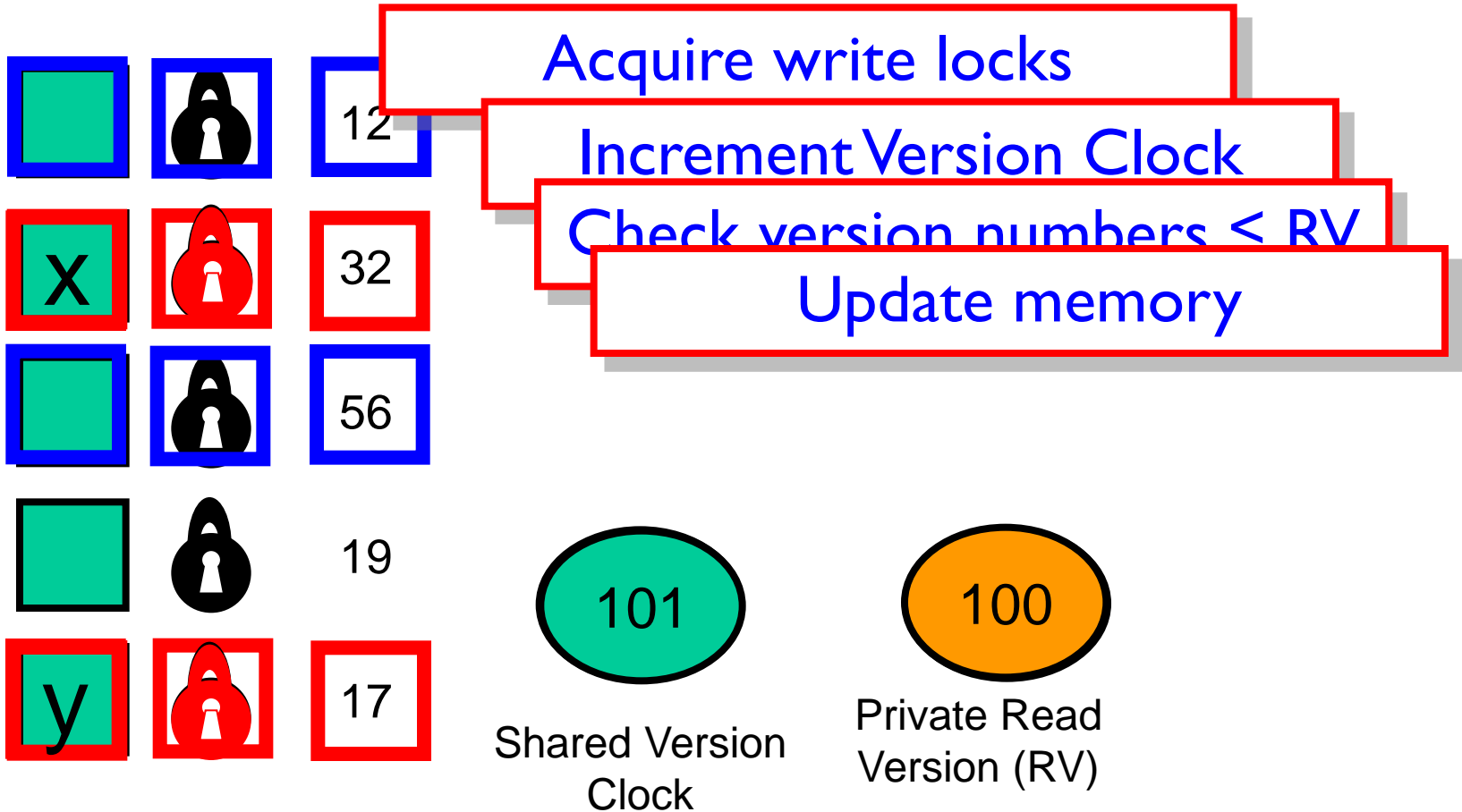
Locks



On Commit

Mem

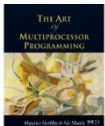
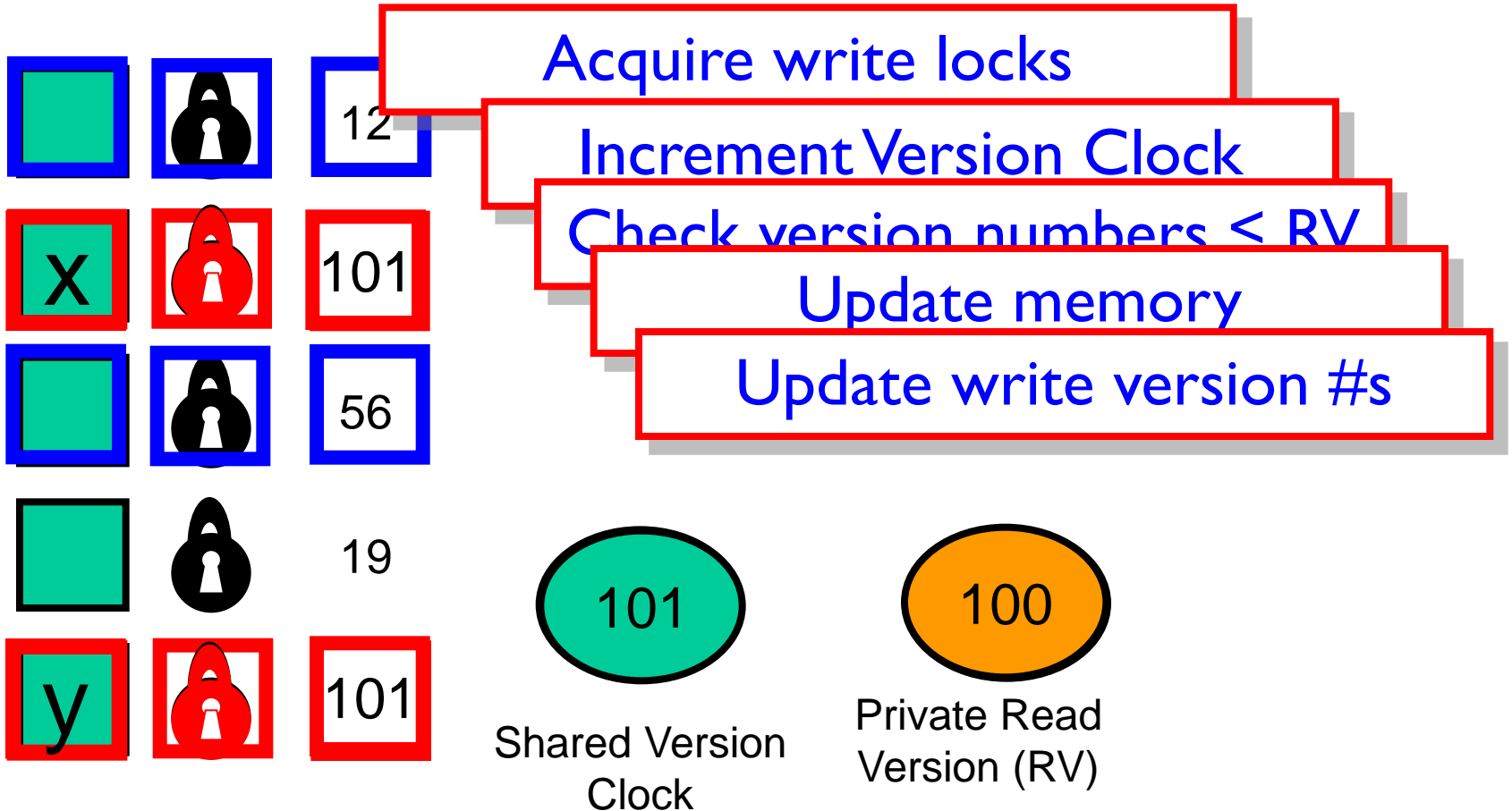
Locks



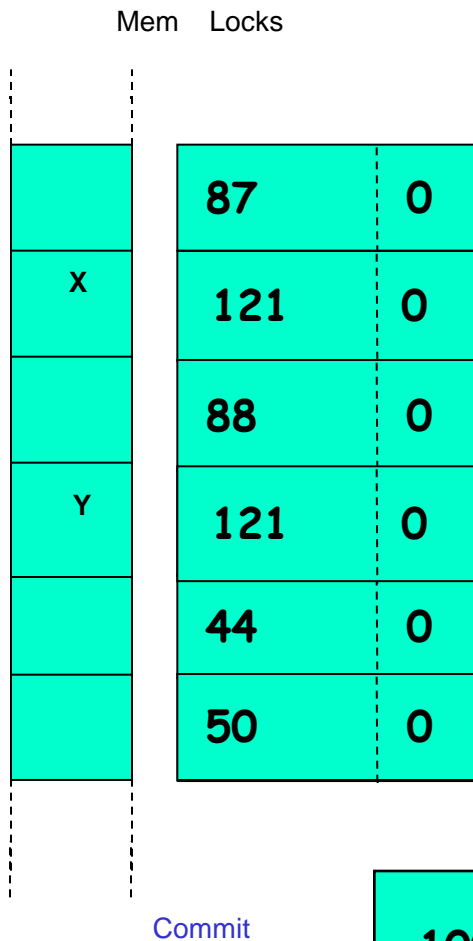
On Commit

Mem

Locks



Example: Writing Trans

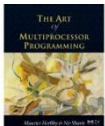


121

Shared Version Clock

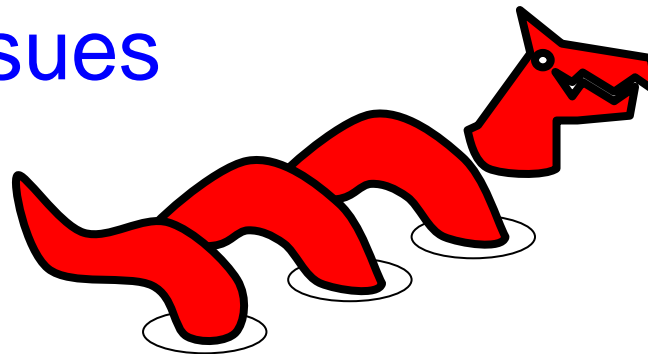
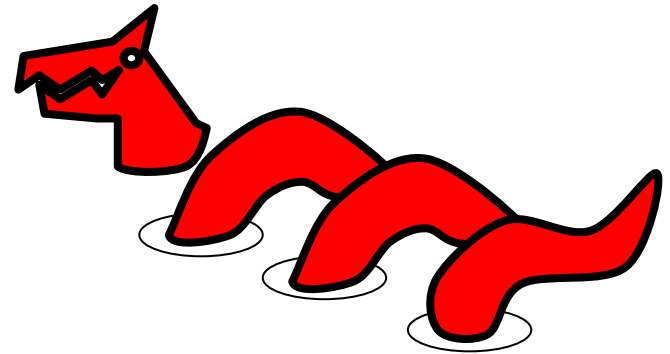
1. $RV \leftarrow$ Shared Version Clock
2. On Read/Write: check unlocked and $v\# \leq RV$ then add to Read/Write-Set
3. Acquire Locks
4. $WV = F\&I(VClock)$
5. Validate each $v\# \leq RV$
6. Release locks with $v\# \leftarrow WV$

Reads+Inc+Writes
=serializable



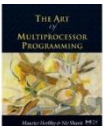
TM Design Issues

- Implementation choices
- Language design issues
- Semantic issues



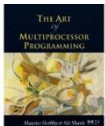
Granularity

- Object
 - managed languages, Java, C#, ...
 - Easy to control interactions between transactional & non-trans threads
- Word
 - C, C++, ...
 - Hard to control interactions between transactional & non-trans threads



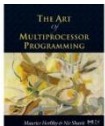
Direct/Deferred Update

- *Deferred*
 - modify private copies & install on commit
 - Commit requires work
 - Consistency easier
- *Direct*
 - Modify in place, roll back on abort
 - Makes commit efficient
 - Consistency harder



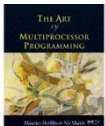
Conflict Detection

- Eager
 - Detect before conflict arises
 - “Contention manager” module resolves
- Lazy
 - Detect on commit/abort
- Mixed
 - Eager write/write, lazy read/write ...



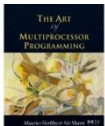
Conflict Detection

- Eager detection may abort transactions that could have committed.
- Lazy detection discards more computation.



Contention Management & Scheduling

- How to resolve conflicts?
- Who moves forward and who rolls back?
- Lots of empirical work but formal work in infancy



Contention Manager Strategies

- Exponential backoff
- Priority to
 - Oldest?
 - Most work?
 - Non-waiting?
- None Dominates
- But needed anyway



Judgment of Solomon

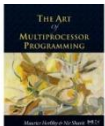
I/O & System Calls?

- Some I/O revocable
 - Provide transaction-safe libraries
 - Undoable file system/DB calls
- Some not
 - Opening cash drawer
 - Firing missile



I/O & System Calls

- One solution: make transaction irrevocable
 - If transaction tries I/O, switch to irrevocable mode.
- There can be only one ...
 - Requires serial execution
- No explicit aborts
 - In irrevocable transactions



Exceptions

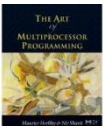


```
int i = 0;  
try {  
    atomic {  
        i++;  
        node = new Node();  
    }  
} catch (Exception e) {  
    print(i);  
}
```

Exceptions

Throws OutOfMemoryException!

```
int i = 0;
try {
    atomic {
        i++;
        node = new Node();
    }
} catch (Exception e) {
    print(i);
}
```

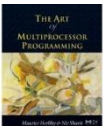


Exceptions

Throws OutOfMemoryException!

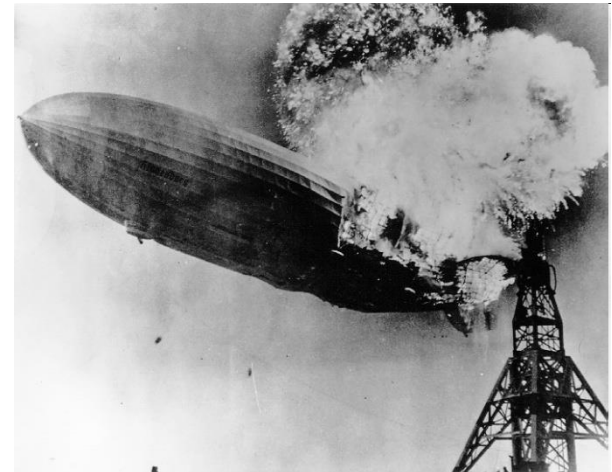
```
int i = 0;
try {
    atomic {
        i++;
        node = new Node();
    }
} catch (Exception e) {
    print(i);
}
```

What is
printed?



Unhandled Exceptions

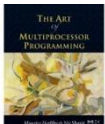
- Aborts transaction
 - Preserves invariants
 - Safer
- Commits transaction
 - Like locking semantics
 - What if exception object refers to values modified in transaction?



Nested Transactions

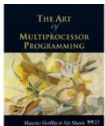
```
atomic void foo() {  
    bar();  
}
```

```
atomic void bar() {  
    ...  
}
```



Nested Transactions

- Needed for modularity
 - Who knew that `cosine()` contained a transaction?
- Flat nesting
 - If child aborts, so does parent
- First-class nesting
 - If child aborts, partial rollback of child only



Hatin' on TM

STM is too inefficient

software transactional memory

- Interaction with code that as a result of either commu
- Livelock, or the system gu

make progress even in the f

In addition to the intrin:

also implementation-specifi

very high transactional overh

rotations for excluding F

the nondeterminism introd

tions complicates debuggin

executed and aborted on co

repeatabe behavior. Both o

argument for the programmer to

implementations.

Given all these issues, wi

yet matured to the point wi

→ why is it only a research toy?

degradation when overflow occurs, and proposals for

managing overflows (for example, signatures¹⁷) incur false

positives that add complexity to the programming model.

Therefore, from an industrial perspective, HTM designs

provide more benefits for the cost on a more

workloads (with varying transactional char-

Hatin' on TM

The problem with STM: your languages still suck

Published by [Brian](#) at 12:13 PM under [To Be Categori](#)

So, I'm reading [Brandon Werner's post](#) caught on yet. There are three Act, allowing many to Java, C#,

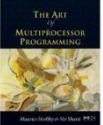
- Interaction with code that as a result of either commu a requirement barring spec
- Livelock, or the system gu make progress even in the f
- In addition to the intrin: implementation-specifi
- Transactional overh for excluding F
- Mechanism introd debuggin
- on ca
- at tc

Requires radical change in programming style

→ why is it o

degradation when overflow - managing overflows (for exam, positives that add complexity to L. therefore, from an industrial persp provide more benefits for the c. workloads (with varying tran.

Software Transactional Memory has succeed in implementing STM for "mainstream" languages l that make it a CS Researcher Full Employment



Hatin' on TM

Wrong Programming Model

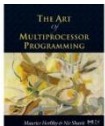
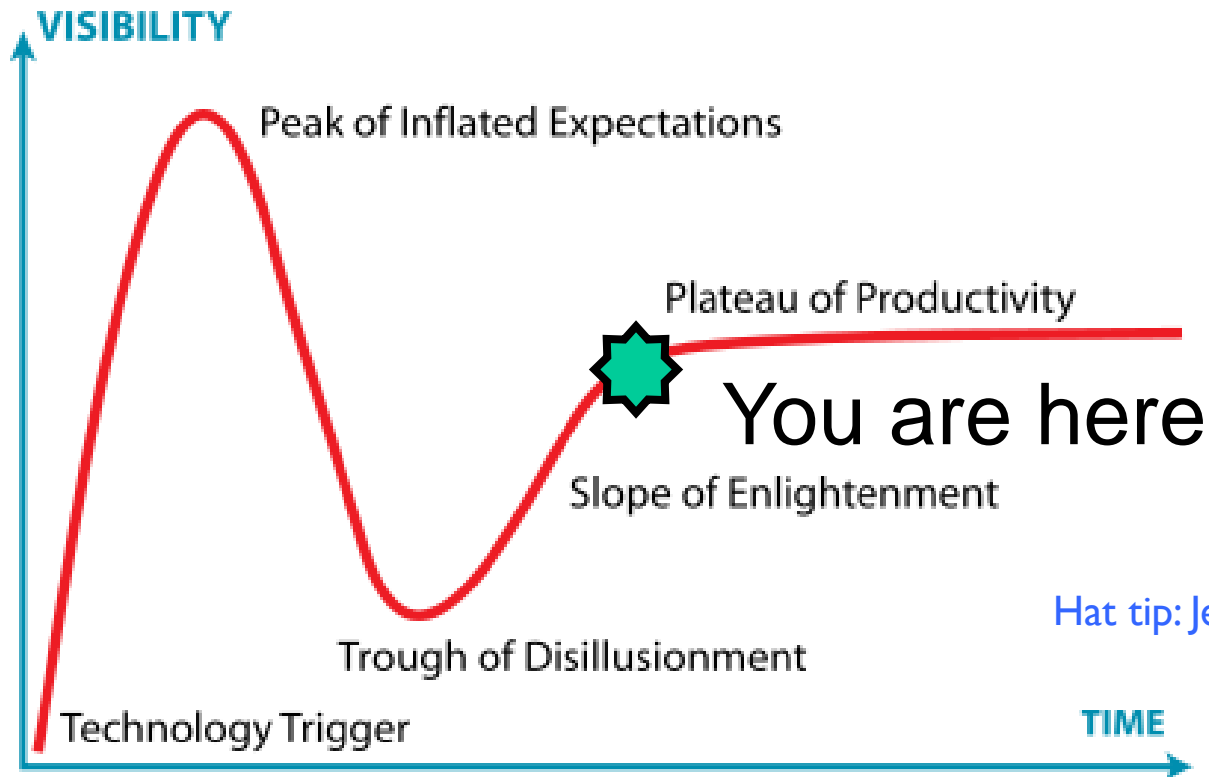
Erlang-style shared nothing only true path to salvation

→ why is it ok

degradation when overflow -
managing overflows (for exam,
positives that add complexity to L
therefore, from an industrial persp
provide more benefits for the c
workloads (with varying tran



Gartner Hype Cycle



I, for one, Welcome our new Multicore Overlords ...

- Multicore forces us to rethink almost everything



I, for one, Welcome our new Multicore Overlords ...

- Multicore forces us to rethink almost everything
- Standard approaches too complex



I, for one, Welcome our new Multicore Overlords ...

- Multicore forces us to rethink almost everything
- Standard approaches won't scale
- Transactions might make life simpler...



I, for one, Welcome our new Multicore Overlords ...

- Multicore forces us to rethink almost everything
- Standard approaches won't scale
- Transactions might ...
- Multicore programming
Plenty more to do...
Maybe *you* will save us...



Thanks ! תודה

