Concurrent Systems

V. Elementary Operations

Wolfgang Schröder-Preikschat

November 29, 2016

Subject Matter

- discussion on abstract concepts as to elementary operations at instruction structure set architecture level
  - atomic load/store of a naturally aligned machine word
  - atomic read-modify-write of complex machine instructions
Outline

Preface

Primitive Instructions
  Atomic Operations
  Equivalence

Memory Models
  Properties

Summary

Subject Matter

- discussion on abstract concepts as to elementary operations at instruction structure set architecture level
  - atomic load/store of a naturally aligned machine word
  - atomic read-modify-write of complex machine instructions
- impartation of knowledge on memory models that are relevant to multi-threading on multi/many-core (multi-) processors
  - atomicity, visibility, and ordering of memory operations against the background of UMA, NUMA, and (partly) COMA architectures
  - ordering enforcing hardware such as memory barriers or fences, resp., allowing one to pattern sequential, relaxed, and weak data consistency
- excursion into practice of hardware features that are of importance for the implementation of any synchronisation algorithm

Memory-Operation Semantics

- of particular interest (at this point) are shared-memory operations
  - commonality is the opportunity, at least, for invisible execution

Preface

Primitive Instructions

Atomic Operations

Equivalence

Memory Models

Properties

Summary

© wosch CS (WS 2016, LEC 5)
of particular interest (at this point) are shared-memory operations. A commonality is the opportunity, at least, for indivisible execution.

Note, all memory operations are also divisible in the following respect:

- **Sub-operation**: Processors are word-oriented, but memory is byte-oriented with word size as a multiple of byte size, e.g. $4 \times 8$ bits. Thus, loads/stores will operate on a sequence of bytes.
- **Sub-step**: Processors perform a fetch-execute-cycle to run programs.
- $n$-address machines mean $n$-operand instructions, $n \geq 2$.
- Thus, execution requires a sequence of loads/stores.

---

1 In general $n \geq 0$, but only for $n \geq 2$ becomes the problem apparent.

---

```c
#include <stdint.h>

static int64_t label;

int64_t get_label() {
    return label;
}

void set_label(int64_t value) {
    label = value;
}
```

In logical respect any of these single statements is indivisible, atomic.
- Lines 6 conceals a load and line 10 conceals a store operation.
- Each case forms an ELOP of the *abstract processor* “C”.

---

In physical respect these statements are conditionally atomic, only
- A matter of optimisation options, the CPU, and alignment restrictions.
### Load/Store II

**ASM—Level 4**

```asm
1  get_label:
2     movl label, %eax
3     movl label+4, %edx
4     ret
5
6  set_label:
7     movl 4(%esp), %eax
8     movl 8(%esp), %ecx
9     movl %ecx, label+4
10    movl %eax, label
11    ret
```

**gcc -m32...**

- actions 2-3 and 9-10 are divisible
- any of these 8 mov instructions is **conditionally indivisible**

**gcc -m64...**

- actions 2-3 and 9-10 are divisible
- any of these 8 mov instructions is **conditionally indivisible**

- beware of the processor architecture or the data alignment, resp.
- usually, memory-word loads/stores are indivisible if “word” corresponds to the smallest addressable unit of main memory: byte, nowadays
- on some architectures (e.g., x86) they are indivisible too if the address of the memory operand is **naturally aligned**

---

© wosch CS (WS 2016, LEC 5)  Primitive Instructions–Atomic Operations 8
- **execution cycle** of a machine instruction that involves the ALU²

  - consists of the following individual operation steps:
    - i. load input operands (acc. operation code or addressing mode, resp.)
    - ii. compute result (acc. operation code)
    - iii. store output operand (acc. operation code or addressing mode, resp.)

  - steps (i) and (iii) require the **bus** in case of memory-sensitive operations
    - reusable hardware resource, shareable, allocated per (load/store) step
  - typical **compound action** at instruction set architecture (ISA) level
    - is memory-sensitive only for a complex instruction set computer (CISC)

²*arithmetic-logic unit*, the operation unit of the CPU.
Read-Modify-Write (RMW)

The execution cycle of a machine instruction that involves the ALU consists of the following individual operation steps:

1. Load input operands (acc. operation code or addressing mode, resp.)
2. Compute result (acc. operation code)
3. Store output operand (acc. operation code or addressing mode, resp.)

Steps (i) and (iii) require the bus in case of memory-sensitive operations. Steps (i) and (iii) require the bus in case of memory-sensitive operations.

Typical compound action at instruction set architecture (ISA) level is memory-sensitive only for a complex instruction set computer (CISC).

In a multiprocessor case, the whole cycle is divisible (non-atomic). Merely the individual sub-steps may form indivisible actions (cf. p. 8).

Indivisibility requires a bus lock for the duration of the whole cycle:

- An atomic RMW instruction that implicitly performs the lock or
- A lock prefix that makes the adjacent normal RMW instruction atomic.

Test & Set I (TAS)

Definition (TS, acc. IBM System/370)

The leftmost bit (bit position 0) of the byte located at the second-operand address is used to set the condition code, and then the entire addressed byte is set to all ones. [8, p. 144]

The operation effectively does an unconditional store in main memory.

- The byte in storage is set to all ones as it is fetched for the testing of bit position 0. [8, p. 144]

A similar effect has ldstub of SPARC V9.

Test & Set I (TAS)

Definition (TS, acc. IBM System/370)

The leftmost bit (bit position 0) of the byte located at the second-operand address is used to set the condition code, and then the entire addressed byte is set to all ones. [8, p. 144]

The byte in storage is set to all ones as it is fetched for the testing of bit position 0. [8, p. 144]

In terms of main memory significance, this translates into the following:

```cpp
bool tas(byte *ref) {
    atomic { bool aux = *ref & 0x1; *ref = 0x11111111; }
    return aux;
}
```

With first and second operand being used to form effective address ref.

A similar effect has ldstub of SPARC V9.
Test & Set I

Definition (TS, acc. IBM System/370)

The leftmost bit (bit position 0) of the byte located at the second-operand address is used to set the condition code, and then the entire addressed byte is set to all ones. [8, p. 144]

Test & Set II

the original copy (IBM System/370) has swapping characteristic

swap(x, y), with x = *ref[0] and y = 11111111[0]

the operation effectively does an unconditional store in main memory

The byte in storage is set to all ones as it is fetched for the testing of bit position 0. [8, p. 144] A similar effect has ldstub of SPARC V9.

in terms of main memory significance, this translates into the following:

    bool tas(byte *ref) {
        atomic { bool aux = *ref & 0x1; *ref = 0x11111111; }
        return aux;
    }

    – with first and second operand being used to form effective address ref

note that TS interlocks against simultaneous main memory accesses

Same holds for the cache line holding the main memory operand is always invalidated this semantic has a deleterious effect for cache-coherent processors

whereby (using GCC atomic built-in functions):

    int tas(any_t *ref) {
        movl 4(%esp), %ecx
        movl $1, %eax
        xchgl %eax, (%ecx)
        ret
    }

Test & Set II

the original copy (IBM System/370) has swapping characteristic

swap(x, y), with x = *ref[0] and y = 11111111[0]

for a contemporary processor (x86), this translates into the following:

    int tas(any_t *ref) {
        movl 4(%esp), %ecx
        movl $1, %eax
        xchgl %eax, (%ecx)
        ret
    }

    #define TAS(ref) __sync_lock_test_and_set(ref, 1)
Test & Set II

1. the original copy (IBM System/370) has **swapping characteristic**
   - swap(x,y), with \( x = \ast \text{ref}[0] \) and \( y = 11111112[0] \)
   - for a contemporary processor (x86), this translates into the following:

   ```
   int tas(\text{any_t} * \text{ref}) {
     return TAS(\text{ref});
   }
   ```

   whereby (using GCC atomic built-in functions):

   9. `#define TAS(\text{ref}) \_\_sync\_lock\_test\_and\_set(\text{ref}, 1)`
   - note that `xchg` interlocks against simultaneous main memory accesses

   ```
   \#define TAS(\text{ref}) \_\_sync\_lock\_test\_and\_set(\text{ref}, 1)
   ```

   4. note that `xchg` interlocks against simultaneous main memory accesses
   
   beware of the unconditional store carried out by both `TS` and `xchg`
   
   this semantic has a **deleterious effect** for cache-coherent processors
   
   the cache line holding the main memory operand is always invalidated
   
   ----

   5. same holds for `TAS` of the M68000 family and `ldstub` of the SPARC family.

Test & Set III

**Definition (Dual-Ported RAM)**

A kind of random access memory (RAM) that supports simultaneous load and store operations from two directions.

```markdown
Definition (Dual-Ported RAM)
A kind of random access memory (RAM) that supports simultaneous load and store operations from two directions.
```
**Test & Set III**

**Definition (Dual-Ported RAM)**

A kind of random access memory (RAM) that supports simultaneous load and store operations from two directions.

- **interlock** is conducted by a “DPRAM monitor” that, e.g. [18]:
  - records the processor that issued the TAS and acquired access
  - notifies processors that, at a time, issue a TAS simultaneously
    - signalling BUSY interrupt, forcing the receiving processor into busy waiting
  - performs the test and then, if and only if the test succeeds:
    i. sets the memory location to the value given by the owning processor and
    ii. releases access to that memory location
- this scheme translates into a **conditional store** as follows:

```c
word tas(word *ref) {
    word aux;
    atomic { if ((aux = *ref) == 0) *ref = 1; }
    return aux;
}
```

© wosch CS (WS 2016, LEC 5) Primitive Instructions–Atomic Operations 12

---

**Compare & Swap I**

**Definition (CS, acc. IBM System/370)**

The first and second operands are compared. If they are equal, the third operand is stored in the second-operand location. If they are unequal, the second operand is loaded into the first-operand location. [8, p. 123]

- the operation effectively performs a **conditional store** in main memory
  - The first and third operands [each are] occupying a general register.
  - The second operand is a word in main storage. [8, p. 123]

```c
atomic word cas(register old, word *ref, register new) {
    word aux;
    return aux = (*ref == old) ? (*ref = new) : (old = *ref);
}
```

© wosch CS (WS 2016, LEC 5) Primitive Instructions–Atomic Operations 13
**Compare & Swap I**

**Definition (CS, acc. IBM System/370)**

The first and second operands are compared. If they are equal, the third operand is stored in the second-operand location. If they are unequal, the second operand is loaded into the first-operand location. [8, p.123]

- the operation effectively performs a **conditional store** in main memory
- The first and third operands [each are] occupying a general register. The second operand is a word in main storage. [8, p.123]
- in terms of **main memory significance**, this translates into the following:

```
1 atomic word cas(register old, word *ref, register new) {
  word aux;
  return aux = (*ref == old) ? (*ref = new) : (old = *ref);
}
- with the actual parameters old and new being kept in general registers
- note that CS interlocks against simultaneous main memory accesses
```

**Compare & Swap II**

**Definition (ABA, also A-B-A)**

The ABA problem is a **false positive** execution of a CAS-based speculation on a shared location $L_i$. [2, p.186]

- when the successful execution of a CAS instruction indicates:
  i that the two operands subject to comparison are equal and, thus, purport the presence of a certain condition (**positive**),
  ii but the condition is not in fact present (**false**)

**Compare & Swap II**

**Ambiguity (cf. also [8, p.125])**

**Definition (ABA, also A-B-A)**

The ABA problem is a **false positive** execution of a CAS-based speculation on a shared location $L_i$. [2, p.186]

- when the successful execution of a CAS instruction indicates:
  i that the two operands subject to comparison are equal and, thus, purport the presence of a certain condition (**positive**),
  ii but the condition is not in fact present (**false**)
- assuming that processes $P_1$ and $P_2$ simultaneously access location $L_i$
  - value $A$ read by $P_1$ from $L_i$ be a sign of a dedicated global state $S_1$, but $P_1$ will be delayed before being able to commit a new value to $L_i$
  - meanwhile $P_2$ changes the value of $L_i$ to $B$ and then back to $A$, defining a new global state $S_2 \neq S_1$
  - $P_1$ resumes, observes that the value of $L_i$ equals $A$ and, thus, acts on the assumption that the global state must be $S_1$—which is no longer true

```
word aux;

1} note that
```
**Load-Locked/Store-Conditional I**

**Definition**
Paired instructions to form a flow of actions without any guarantee of indivisibility but that it succeeds only in case of indivisible operation.

- originated in the MIPS II or R6000, resp., RISC architecture [9]

---

**Load-Locked/Store-Conditional I**

**Definition**
Paired instructions to form a flow of actions without any guarantee of indivisibility but that it succeeds only in case of indivisible operation.

- originated in the MIPS II or R6000, resp., RISC architecture [9]:
  - LL: loads a word from the specified effective memory address
  - makes a reservation on that very address (range)\(^5\)
  - SC: checks for a reservation on the specified effective memory address\(^5\)
  - if the reservation persists, stores the specified word at that address
  - delivers the result of the reservation check

---

**Compare & Swap II**

**Definition (ABA, also A-B-A)**
The ABA problem is a false positive execution of a CAS-based speculation on a shared location \(L_i\). [2, p. 186]

- when the successful execution of a CAS instruction indicates:
  - i. that the two operands subject to comparison are equal and, thus, purport the presence of a certain condition (positive),
  - ii. but the condition is not in fact present (false)
- assuming that processes \(P_1\) and \(P_2\) simultaneously access location \(L_i\)
  - value \(A\) read by \(P_1\) from \(L_i\) be a sign of a dedicated global state \(S_1\), but \(P_1\) will be delayed before being able to commit a new value to \(L_i\)
  - meanwhile \(P_2\) changes the value of \(L_i\) to \(B\) and then back to \(A\), defining a new global state \(S_2 \neq S_1\)
- \(P_1\) resumes, observes that the value of \(L_i\) equals \(A\) and, thus, acts on the assumption that the global state must be \(S_1\)—which is no longer true
- severity of false positive execution depends on the problem (cf. p. 36)

---

\(^5\)The dimension of the reservation depends on the hardware implementation. It may be exact the effective address or a larger address range around.
Load-Linked/Store-Conditional I

**Definition**
Paired instructions to form a flow of actions without any guarantee of indivisibility but that it succeeds only in case of indivisible operation.

- originated in the MIPS II or R6000, resp., RISC architecture [9]:
  - **LL** loads a word from the specified effective memory address
  - makes a reservation on that very address (range)\(^5\)
  - **SC** checks for a reservation on the specified effective memory address\(^5\)
  - if the reservation persists, stores the specified word at that address
  - delivers the result of the reservation check
- reasons for cancellation of a persisting address (range) reservation:
  1. successful execution of SC—hoped for, normally
  2. execution of LL by another processor applying the same address (range)
  3. an exception (trap/interrupt) on the processor holding the reservation
- LL and SC interlock against simultaneous main memory accesses

\(\Phi = \) be

\(\Phi = \max_X \) returns failure of CAS

\(\Phi = \) successfully executes

\(\Phi = \) gets delayed for some reason, thus has not yet executed SC

\(\Phi = \) shares location \(ref\) with \(P_2\), established reservation \(ref_P\) by LL

\(\Phi = \) overlaps \(P_1\), establishes reservation \(ref_P\) and, thus, cancels \(ref_P\)

\(\Phi = \) successfully executes SC ⇒ CAS succeeds

\(\Phi = \) resumes ⇒ SC will fail because reservation \(ref_P\) is invalid

\(\Phi = \) returns failure of CAS ⇒ rolls back, backs up, and retries CAS...

\(\Phi = \) the dimension of the reservation depends on the hardware implementation. It may be exact the effective address or a larger address range around.

Load-Linked/Store-Conditional II

**use of LL/SC to recreate TAS and CAS:**
- in case of TAS, a boolean variable is conditionally set true
- in case of CAS, a memory word is conditionally overwritten

```c
1. int tas(long *ref) {
2.     return (LL(ref) == 0) && SC(ref, 1);
3. }

4. int cas(long *ref, long old, long new) {
5.     return (LL(ref) == old) && SC(ref, new);
6. }
```

Fetch & Add

**Definition (acc. [6, p.17])**
A value-returning instruction that operates on a global (i.e., shared) variable \(G\) and a local variable \(L\).
Fetch & Add

Definition (acc. [6, p. 17])
A value-returning instruction that operates on a global (i.e., shared) variable \(G\) and a local variable \(L\).

- an atomic RMW instruction, inspired by “Replace Add” [3, p. 6]
  - prefix (FAA) or postfix (AAF) form, as to when fetch becomes effective
    prefix – save the old value of \(G\) for return, then add \(L\) to \(G\)
    postfix – add \(L\) to \(G\), then return the new value of \(G\)
  - whereby (cf. p. 39):
    \[
    \begin{align*}
    FAA(G, L) & \equiv AAF(G, L) - L \quad \text{and} \\
    AAF(G, L) & \equiv FAA(G, L) + L
    \end{align*}
    \]

- transferable to any associative binary operation \(\text{fetch-and-} \Phi\)
  - but for noninvertible operations the prefix form is considered more general
  - be \(\Phi = \max\) (i.e., \(X\)): only \(\text{XAF}(G, L) \equiv \max(FAX(G, L), L)\) (cf. p. 40)

Equality of Atomic Operations

Definition (Consensus Number)

\(\text{The consensus number for } X \text{ is the largest } n \text{ for which } X \text{ solves } n\text{-process consensus. If no largest } n \text{ exists, the consensus number is said to be infinite. } [7, \text{p. 130}]\)

- \(n\) processes need to interact to achieve agreement on a single data value
- note that only 1-process consensus requires no interaction
Equality of Atomic Operations

- operations that need consensus number \( n \) cannot have a semantically equivalent implementation by operations of consensus number \( m < n \)

**Definition (Consensus Number)**

The consensus number for \( X \) is the largest \( n \) for which \( X \) solves \( n \)-process consensus. If no largest \( n \) exists, the consensus number is said to be infinite. \([7, \text{p.130}]\)

- \( n \) processes need to interact to achieve agreement on a single data value
- note that only 1-process consensus requires no interaction

Consensus numbers of the elementary operations considered:

- \( \infty \) compare-and-swap, load-linked/store-conditional
- 2 test-and-set, swap, fetch-and-add
- 1 atomic read, atomic write

Properties Relevant to Multi-Threading

- fundamental characteristics that are of particular importance for the implementation of any synchronisation algorithm
Properties Relevant to Multi-Threading

- fundamental characteristics that are of particular importance for the implementation of any synchronisation algorithm:
  - **atomicity**: as to how certain machine instructions are executed
    - differentiates in RISC and CISC machines
    - specific to each ELOP that was discussed before (pp. 7–17)
  - **visibility**: as to when memory-cell changes are observable
    - concerns delays in sensing the most recent memory-word write
    - introduces time factors on the availability of written data
  - **ordering**: as to how memory operations appear to be performed
    - stands for a variant of out-of-order execution
    - reflects on (sequential, relaxed, or weak) consistency models

- these properties are linked with each other, are mutual prerequisites
  - atomicity applies to all other—and to a single machine instruction, only
  - visibility depends on the memory architecture, may cause “jitter”
  - ordering comprises multiple machine instructions, may cause “fencing”

- as to the level of abstraction, they must all be considered together

ordering comprises multiple machine instructions, may cause “fencing”
Properties Relevant to Multi-Threading

- Fundamental characteristics that are of particular importance for the implementation of any synchronisation algorithm:
  - **Atomicity**: as to how certain machine instructions are executed
    - Differentiates in RISC and CISC machines
    - Specific to each ELOP that was discussed before (pp. 7–17)
  - **Visibility**: as to when memory-cell changes are observable
    - Concerns delays in sensing the most recent memory-word write
    - Introduces time factors on the availability of written data
  - **Ordering**: as to how memory operations appear to be performed
    - Stands for a variant of out-of-order execution
    - Reflects on (sequential, relaxed, or weak) consistency models

- These properties are linked with each other, are mutual prerequisites
  - Atomicity applies to all other—and to a single machine instruction, only
  - Visibility depends on the memory architecture, may cause “jitter”
  - Ordering comprises multiple machine instructions, may cause “fencing”
  - As to the level of abstraction, they must all be considered together

- This is especially true for the operating-system machine level (i.e., level 3)

---

Atomicity

- Common are two classes of memory-sensitive operations (cf. p. 25):
  - **L/S**: Atomic load (L) or store (S), resp., as single action
    - Granularity is the **machine word**, i.e., a multiple of a byte
    - With **word-alignment** constraint on the operand address, usually
      - Only word-aligned accesses will be carried out indivisibly
  - **RMW**: Atomic read (R), modify (M), and write (W) as single action
    - Common for CISC and, there, for **two-address machines**
      - Uncommon for RISC, which is characteristic of load/store principle
      - Single- or double-word cycles for 32- or 64-bit architectures, resp.
      - “Double” means “physically consecutive” or “logically interrelated”
      - I.e.: CDS or cmpxchg8b/cmpxchg16b compared to DCAS or CAS2

--

Visibility

- When other interacting processes will notice the changes made by the current process, and whether they will notice them at all.

- Depends on the memory architecture and behaviour of read or write operations to the same memory location
When other interacting processes will notice the changes made by the current process, and whether they will notice them at all.

- depends on the **memory architecture** and behaviour of read or write operations to the same memory location
  - **UMA** - *uniform memory architecture* ⇒ the same access time
    - each address is assigned a fixed home in the global address space
    - no processor uses private (local) memory besides shared memory
  - **NUMA** - *non-uniform memory architecture* ⇒ different access times
    - each address is assigned a fixed home in the global address space
    - each processor (“NUMA node”) uses private (local) memory, too
  - **COMA** - *cache-only memory architecture* ⇒ different access times
    - no address is assigned a fixed home in the global address space
    - each processor uses private (local) memory, only

- orthogonal with it is the **consistency** aspect as to shared information stored in multiple local **caches**
  - *cache-coherent (cc)* v. *non-cache-coherent (ncc)* memory architecture
Memory Architectures at a Glance

UMA (symmetric multiprocessing, SMP)

- **UMA**
  - bus interconnect
  - P P P M

NUMA

- **NUMA node (N)**
- zone of uniform memory characteristic
- **NUMA distance**
- number of (network) hops to distant memory

COMA

- **COMA distance**
- number of (network) hops to distant memory

UMA (symmetric multiprocessing, SMP)

- **NUMA**
  - scalable interconnect
  - M M M M

COMA

- **NUMA node (N)**
- zone of uniform memory characteristic
- **NUMA/COMA distance**
- number of (network) hops to distant memory
- **UMA/NUMA combination**
Ordering

What memory re-orderings are possible for a process, relatively to the order as specified by its program.

- to improve performance, memory-sensitive machine instructions are not executed in the order originally specified by the program
  - on the one hand, the compiler reorders (L3) instructions\(^6\) before run-time
  - on the other hand, the CPU reorders (L2) instructions\(^6\) at run-time
    - it is this aspect of dynamic ordering that is of relevance in the following
  - mainly, dynamic ordering is an issue of non-blocking synchronisation
  - as blocking synchronisation implicitly can take care of “fencing” proper
    - depending on the kind of critical section and type of data dependency
  - but, critical section per se is no guarantee for memory ordering (cf. p. 25)

\(^6\)According to the actual level of abstraction: operating-system machine (L3) or instruction set architecture (L2) level. See also [10] or [17, p. 34].
Dynamic Ordering

assuming that the following function is executed by a single processor, but the global variables are then read by at least one more processor:

```c
int a = 1, b = 2;  // what values of a and b do other processors see
void ab_set() {  // once line 6 has been reached by one processor
    a = 3;
    b = 4;
}
```

writes are not necessarily seen by other processors in the order as specified by the program!

Dynamic Ordering

assuming that the following function is executed by a single processor, but the global variables are then read by at least one more processor:

```c
int a = 1, b = 2;  // what values of a and b do other processors see
void ab_set() {  // once line 6 has been reached by one processor
    a = 3;
    b = 4;
}
```

Dynamic Ordering

assuming that the following function is executed by a single processor, but the global variables are then read by at least one more processor:

```c
int a = 1, b = 2;  // what values of a and b do other processors see
void ab_set() {  // once line 6 has been reached by one processor
    a = 3;
    b = 4;
}
```

Dynamic Ordering

assuming that the following function is executed by a single processor, but the global variables are then read by at least one more processor:

```c
int a = 1, b = 2;  // what values of a and b do other processors see
void ab_set() {  // once line 6 has been reached by one processor
    a = 3;
    b = 4;
}
```

Dynamic Ordering

assuming that the following function is executed by a single processor, but the global variables are then read by at least one more processor:

```c
int a = 1, b = 2;  // what values of a and b do other processors see
void ab_set() {  // once line 6 has been reached by one processor
    a = 3;
    b = 4;
}
```

Dynamic Ordering

assuming that the following function is executed by a single processor, but the global variables are then read by at least one more processor:

```c
int a = 1, b = 2;  // what values of a and b do other processors see
void ab_set() {  // once line 6 has been reached by one processor
    a = 3;
    b = 4;
}
```

Dynamic Ordering

assuming that the following function is executed by a single processor, but the global variables are then read by at least one more processor:

```c
int a = 1, b = 2;  // what values of a and b do other processors see
void ab_set() {  // once line 6 has been reached by one processor
    a = 3;
    b = 4;
}
```

Dynamic Ordering

assuming that the following function is executed by a single processor, but the global variables are then read by at least one more processor:

```c
int a = 1, b = 2;  // what values of a and b do other processors see
void ab_set() {  // once line 6 has been reached by one processor
    a = 3;
    b = 4;
}
```

Dynamic Ordering

assuming that the following function is executed by a single processor, but the global variables are then read by at least one more processor:

```c
int a = 1, b = 2;  // what values of a and b do other processors see
void ab_set() {  // once line 6 has been reached by one processor
    a = 3;
    b = 4;
}
```
### Dynamic Ordering Effects

- assuming that the following function is executed by a single processor, but the global variables are then read by at least one more processor:

```c
int a = 1, b = 2;
void ab_set() {
    a = 3;
    b = 4;
}
```

- what values of `a` and `b` do other processors see once line 6 has been reached by one processor?

```c
void ab_get(int ab[2]) {
    a = 3;
    b = 4;
}
```

- although the assignment to `a` (line 4) was instructed previous to the one of `b`.

### Memory Barrier Instructions

Memory barrier instructions directly control only the interaction of a CPU with its cache, with its write-buffer that holds stores waiting to be flushed to memory, and/or its buffer of waiting loads or speculatively executed instructions. [12]

- `ld`, `LoadLoad`, `ld`: ensures that `a` is read before `b` is accessed\(^7\)
- `st`, `StoreStore`, `st`: ensures that `a` is visible before `b` is flushed\(^7\)
- `ld`, `LoadStore`, `st`: ensures that `a` is visible before `b` is accessed\(^7\)
- `st`, `StoreLoad`, `ld`: ensures that `a` is visible before `b` is accessed\(^7\)

- `st`, `StoreLoad`, `ld`: write to same location by another processor

- CAS and LL/SC typically include a `StoreLoad` barrier on the target
  i.e., not only a general-purpose but also the most expensive fence

\(^7\)Including the execution of all subsequent loads or stores, resp.
Consistency Models

- **data consistency** as close as possible to sequential processes or with optimisation margins for high-latency memory

  **sequential**
  - processors see writes on the same target in the same order
  - but the order may appear different for an "external observer" 
  - two requirements: program order and write atomicity

  **relaxed**
  - in terms of the constraints defined by sequential consistency
  - as to (i) program order, (ii) write atomicity, or (iii) both:
    - i. write to read, write to write, read to read and read to write
    - ii. read other's, write early
    - iii. read own, write early
  - pertaining to (i) different or (ii) same memory locations

  **weak**
  - “limited to hardware-recognized synchronizing variables” [4]
    - implemented by operating system machine level programs
    - usually not provided by the instruction set architecture level

  ---

  © wosch CS (WS 2016, LEC 5) Memory Models–Properties 27

---

Consistency Models

- **data consistency** as close as possible to sequential processes or with optimisation margins for high-latency memory

  **sequential**
  - processors see writes on the same target in the same order
  - but the order may appear different for an "external observer"
  - two requirements: program order and write atomicity

  **relaxed**
  - in terms of the constraints defined by sequential consistency
  - as to (i) program order, (ii) write atomicity, or (iii) both:
    - i. write to read, write to write, read to read and read to write
    - ii. read other's, write early
    - iii. read own, write early
  - pertaining to (i) different or (ii) same memory locations

  **weak**
  - “limited to hardware-recognized synchronizing variables” [4]
    - implemented by operating system machine level programs
    - usually not provided by the instruction set architecture level

  ---

  © wosch CS (WS 2016, LEC 5) Memory Models–Properties 27

---

Consistency Models

- **data consistency** as close as possible to sequential processes or with optimisation margins for high-latency memory

  **sequential**
  - processors see writes on the same target in the same order
  - but the order may appear different for an "external observer"
  - two requirements: program order and write atomicity

  **relaxed**
  - in terms of the constraints defined by sequential consistency
  - as to (i) program order, (ii) write atomicity, or (iii) both:
    - i. write to read, write to write, read to read and read to write
    - ii. read other's, write early
    - iii. read own, write early
  - pertaining to (i) different or (ii) same memory locations

  **weak**
  - “limited to hardware-recognized synchronizing variables” [4]
    - implemented by operating system machine level programs
    - usually not provided by the instruction set architecture level

  ---

  © wosch CS (WS 2016, LEC 5) Memory Models–Properties 27

---

Consistency Models

- **data consistency** as close as possible to sequential processes or with optimisation margins for high-latency memory

  **sequential**
  - processors see writes on the same target in the same order
  - but the order may appear different for an "external observer"
  - two requirements: program order and write atomicity

  **relaxed**
  - in terms of the constraints defined by sequential consistency
  - as to (i) program order, (ii) write atomicity, or (iii) both:
    - i. write to read, write to write, read to read and read to write
    - ii. read other's, write early
    - iii. read own, write early
  - pertaining to (i) different or (ii) same memory locations

  **weak**
  - “limited to hardware-recognized synchronizing variables” [4]
    - implemented by operating system machine level programs
    - usually not provided by the instruction set architecture level

  ---

  © wosch CS (WS 2016, LEC 5) Memory Models–Properties 27

---
Consistency Models

Relevant Excerpt (cf. [13])

- **data consistency** as close as possible to sequential processes or with optimisation margins for high-latency memory

  **sequential**
  - processors see writes on the same target in the same order
  - but the order may appear different for an "external observer"\(^8\)
  - two requirements: **program order** and **write atomicity** [11]

  **relaxed**
  - in terms of the constraints defined by sequential consistency
  - as to (i) program order, (ii) write atomicity, or (iii) both:
    - write to read, write to write, read to read and read to write
    - read other’s, write early
    - read own, write early
  - pertaining to (i) different or (ii) same memory locations

  **weak**
  - “limited to hardware-recognized synchronizing variables” [4]
    - implemented by operating system machine level programs
    - usually not provided by the instruction set architecture level

- state of the art processors provide relaxed or weak consistency models

\(^8\)Weaker than "strict consistency" that requires a read from a memory location to return the value of the most recent write.

Outline

Preface

Primitive Instructions

Atomic Operations

Equivalence

Memory Models

Properties

Summary

Résumé

- elementary operations at instruction structure set architecture level
  - atomic load/store of a naturally aligned machine (double-) word
  - atomic read-modify-write of complex machine instructions
    - TAS, CAS and FAA or FAΦ, resp., for CISC and LL/SC for RISC
  - equality of atomic operations as to their consensus number
- memory-access properties that are relevant to multi-threading
  - atomicity, visibility, and ordering of memory operations
  - memory architectures of type UMA, NUMA, and COMA
  - **dynamic ordering** at instruction set architecture level
  - memory barriers or fences, resp., to enforce ordering properly
  - sequential, relaxed, and weak data consistency
  - **hardware features** that are of importance for the implementation of any synchronisation algorithm
    - including but not limited to non-blocking synchronisation, especially

© wosch  CS (WS 2016, LEC 5)  Memory Models–Properties

Résumé

- elementary operations at instruction structure set architecture level
  - atomic load/store of a naturally aligned machine (double-) word
  - atomic read-modify-write of complex machine instructions
    - TAS, CAS and FAA or FAΦ, resp., for CISC and LL/SC for RISC
  - equality of atomic operations as to their **consensus number**
- memory-access properties that are relevant to multi-threading
  - atomicity, visibility, and ordering of memory operations
  - memory architectures of type UMA, NUMA, and COMA
  - **dynamic ordering** at instruction set architecture level
  - memory barriers or fences, resp., to enforce ordering properly
  - sequential, relaxed, and weak data **consistency**
  - **hardware features** that are of importance for the implementation of any synchronisation algorithm
    - including but not limited to non-blocking synchronisation, especially

© wosch  CS (WS 2016, LEC 5)  Summary
Résumé

- elementary operations at instruction structure set architecture level
- atomic load/store of a naturally aligned machine (double-) word
- atomic read-modify-write of complex machine instructions
  - TAS, CAS and FAA or FAΦ, resp., for CISC and LL/SC for RISC
- equality of atomic operations as to their consensus number
- memory-access properties that are relevant to multi-threading
  - atomicity, visibility, and ordering of memory operations
  - memory architectures of type UMA, NUMA, and COMA
  - dynamic ordering at instruction set architecture level
  - memory barriers or fences, resp., to enforce ordering proper
  - sequential, relaxed, and weak data consistency
- hardware features that are of importance for the implementation of any synchronisation algorithm
  - including but not limited to non-blocking synchronisation, especially

Reference List I

[1] Bershad, B. N. ; Zekauskas, M. J.:
Midway: Shared Memory Parallel Programming with Entry Consistency for Distributed Memory Multiprocessors / School of Computer Science, Carnewgie Mellon University.
Pittsburgh, PA, USA, 1991 (CMU-CS-91-170). – Forschungsbericht

Understanding and Effectively Preventing the ABA Problem in Descriptor-based Lock-free Designs.

Programming Considerations for Parallel Computers / Courant Institute of Mathematical Sciences.
New York University, Nov. 1967 (IMM 362). – Forschungsbericht

Reference List II

[4] Dubois, M. ; Scheurich, C. ; Briggs, F. A.:
Memory Access Buffering in Multiprocessors.
In: Aiso, H. (Hrsg.):

Memory Consistency and Event Ordering in Scalable Shared-Memory Multiprocessors.

Coordinating Parallel Processors: A Partial Unification.

Wait-Free Synchronization.
In: ACM Transactions on Programming Languages and Systems 11 (1991), Jan., Nr. 1, S. 124–149
Unconditional Store: Workaround

- "textbook semantics" of TAS has a deleterious effect for the cache:

```c
1 word tas(word *ref) {
2     atomic { word aux = *ref; *ref = 1; }
3     return aux;
4 }
```

- same is true when using the GCC atomic built-in function (x86, cf. p11):

```c
5 #define TAS(ref) __sync_lock_test_and_set(ref, 1)
```

© wosch CS (WS 2016, LEC 5) Summary-Bibliography 35
Unconditional Store: Workaround

- “textbook semantics” of TAS has a deleterious effect for the cache:
  ```c
  word tas(word *ref) {
    atomic { word aux = *ref; *ref = 1; }
    return aux;
  }
  ```
- same is true when using the GCC atomic built-in function (x86, cf. p11):
  ```c
  #define TAS(ref) __sync_lock_test_and_set(ref, 1)
  ```
- use of CAS, with `#define CAS __sync_bool_compare_and_swap`
  ```c
  int tas(long *ref) {
    return CAS(ref, 0, 1);
  }
  ```

ABA Exemplified

- given a LIFO list (i.e., stack) of following structure: head $\diamond A \diamond B \diamond C$
  - with head stored at location $L_i$ shared by processes $P_1$ and $P_2$
  - `push` (cf. [16, p.11]) and `pull` adding or removing, resp., list items:
    ```c
    chain_t *cas_pull(stack_t *this) {
      chain_t *node;
      do if ((node = this->head.link) == 0) break;
      while (!CAS(&this->head.link, node, node->link));
      return node;
    }
    ```
- assuming that the following sequence of actions will take place:
  
  **$P_1$**
  - reads head item $A$ followed by $B$ on the list, gets delayed at line 4
  - remembers node $= A$, but has not yet done CAS: head $\diamond A \diamond B \diamond C$

  **$P_2$**
  - pulls head item $A$ from the list: head $\diamond B \diamond C$
  - pulls head item $B$ from the list: head $\diamond C$
  - pushes item $A$ back to the list, now followed by $C$: head $\diamond A \diamond C$

  **$P_1$**
  - resumes, CAS realises head $= A$ (followed by $B$): head $\diamond B \diamond \circ$
  - list state head $\diamond A \diamond C$ as left behind by $P_2$ is lost...

ABA Design Risc Reduction

- prevalent approach is to add a change number to the “control word” [8, p.125], i.e., to practice some kind of versioning
- this number increments at each CAS attempt on the control word
- appropriate techniques depend on the change-number parameters
prevalent approach is to add a **change number** to the “control word” [8, p. 125], i.e., to practice some kind of **versioning**

- this number increments at each CAS attempt on the control word

appropriate techniques depend on the change-number parameters

- the values margin has a whole word size available
  - both the control and change-number word must be updated, indivisibly
  - **compare double and swap** (CDS, [8, p. 124]) of two consecutive words\(^9\)
  - **double compare and swap** (DCAS, also CAS2 [14, p. 4-66]) of any two words

b. the values margin utilizes fully unused bits in the control word itself

- CAS facilitates indivisible updates of control word including change number
- workaround, especially suitable for handling aligned data-structure **pointers**
- gimmick is in data-structure padding for an object size of a power of two

\(^9\)See also `cmpxchg8b` or `cmpxchg16b`, in case of x86.

\[ n \] low-order bits then will be used as a **change-number tag**

\[ n \] low-order bits always 0

**for pointer operations**, the change-number tag is temporary neutralised

but the ABA problem never disappears, it only gets more improbable
FAA Exemplified

```c
#define FAA __sync_fetch_and_add
int faa(int *p, int v) {
  return FAA(p, v);
}
```

```c
#define AAF __sync_add_and_fetch
int aaf(int *p, int v) {
  return AAF(p, v);
}
```

Noninvertible Operation

```c
word fax(word *ref, word val) {
  word aux;
  atomic { if ((aux = *ref) < val) *ref = val; }
  return aux;
}
```

Linguistic Devices for LL/SC

- as GCC does not provide atomic built-in functions for this case:

```asm
INLINE
long LL(long *ref) {
  long aux;
  asm volatile(
    "lwarx %0, 0, %1"
  ): "=r" (aux) : "r" (ref)
  return aux;
}

INLINE
long SC(long *ref, long val) {
  long ccr;
  asm volatile(
    "mfcr %0"
  ): "=r" (ccr) : "r" (ccr)
  return ccr & 0x2;
}
```

- with "#define INLINE extern inline" for GCC to ensure that stand-alone object code is never emitted for in-line functions

10 Use "#define INLINE inline" for C99, for the same reason.
Noninvertible Operation

**fetch-and-Φ**, with \( \Phi = \text{max} \)

- safe-load of global variable \( G \) and conditional-store of \( \text{max}(G, L) \) at \( G \)

```c
1 word fax(word *ref, word val) {
2     word aux;
3     atomic { if ((aux = *ref) < val) *ref = val; }
4     return aux;
5 }
```

- conditional-store of \( \text{max}(G, L) \) at \( G \) and return of \( \text{max}(G, L) \)

```c
6 word xaf(word *ref, word val) {
7     atomic { word aux = (*ref > val) ? *ref : *ref = val; }
8     return aux;
9 }
```

- assuming that \( G = 42 \) and \( L = 4711 \):
  - \( XAF(G, L) \equiv \text{max}(FAX(G, L), L) \): both terms result in 4711
  - \( FAX(G, L) \not\equiv \text{max}(XAF(G, L), L) \): \( FAX \) may result in 42 < 4711