

The out-of-thin-air problem and a promising solution

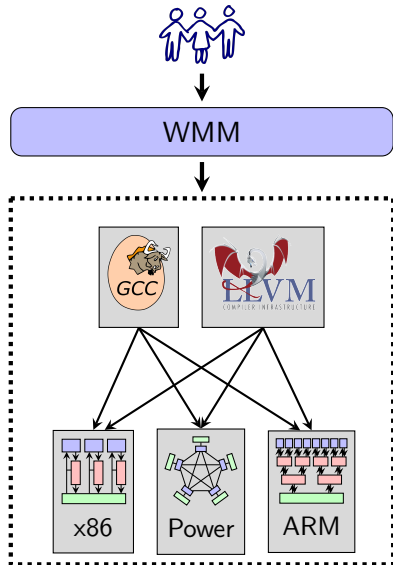
Ori Lahav

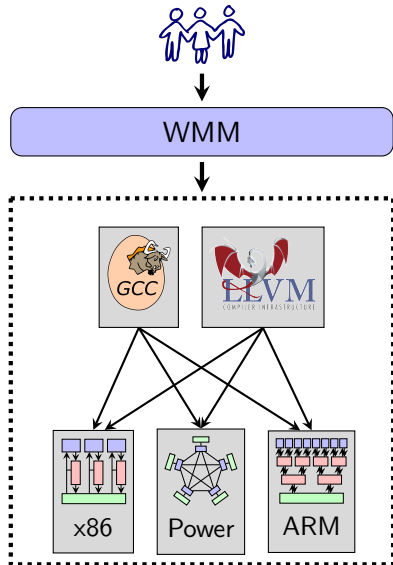
Viktor Vafeiadis

31 August 2017

What is the **right semantics** for
a **concurrent** programming language?

Programming language concurrency semantics





WMM desiderata

1. Mathematically sane
(e.g., monotone)
2. Not too strong
(good for hardware)
3. Not too weak
(allows reasoning)
4. Admits optimizations
(good for compilers)
5. No undefined behavior

The *out-of-thin-air* problem in C11

- ▶ Initially, $x = y = 0$.
- ▶ All accesses are “relaxed”.

Load-buffering

```
a := x; //1 || x := y;  
y := 1;
```

This behavior must be allowed:

Power/ARM allow it

The *out-of-thin-air* problem in C11

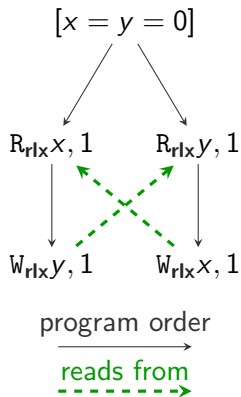
- ▶ Initially, $x = y = 0$.
- ▶ All accesses are “relaxed”.

Load-buffering

```
a := x; //1 || x := y;  
y := 1;
```

This behavior must be allowed:

Power/ARM allow it



The *out-of-thin-air* problem in C11

Load-buffering + data dependency

```
a := x; //1 || x := y;  
y := a;
```

The behavior should be forbidden:

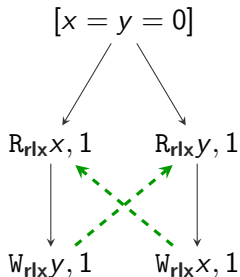
Values appear out-of-thin-air!

The *out-of-thin-air* problem in C11

Load-buffering + data dependency

```
a := x; //1 || x := y;  
y := a;
```

The behavior should be forbidden:
Values appear out-of-thin-air!



Same execution as before!
C11 allows these behaviors

The *out-of-thin-air* problem in C11

Load-buffering + data dependency

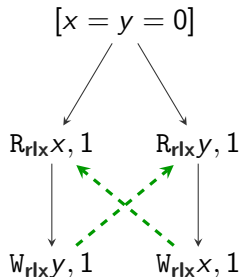
```
a := x; //1 || x := y;  
y := a;
```

The behavior should be forbidden:
Values appear out-of-thin-air!

Load-buffering + control dependencies

```
a := x; //1 || if (y = 1)  
if (a = 1) || x := 1  
y := 1
```

The behavior should be forbidden:
DRF guarantee is broken!



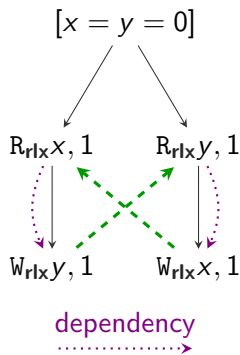
Same execution as before!
C11 allows these behaviors

The hardware solution

Keep track of syntactic dependencies, and forbid “dependency cycles”.

Load-buffering + data dependency

```
a := x; //1      ||  
y := a;          ||      x := y;
```



The hardware solution

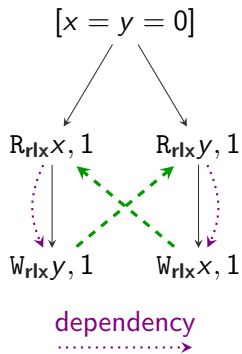
Keep track of syntactic dependencies, and forbid “dependency cycles”.

Load-buffering + data dependency

$a := x;$	$//1$	\parallel	$x := y;$
$y := a;$			

Load-buffering + fake dependency

$a := x;$	$//1$	\parallel	$x := y;$
$y := a + 1 - a;$			



This approach is not suitable for a programming language:
Compilers do not preserve syntactic dependencies.

A “promising” semantics for relaxed-memory concurrency

We will now describe a model that satisfies all these goals, and covers nearly all features of C11.

- ▶ DRF guarantees
- ▶ No “out-of-thin-air” values
- ▶ Avoid “undefined behavior”
- ▶ Efficient implementation on modern hardware
- ▶ Compiler optimizations

Key idea: Start with an operational interleaving semantics, but allow threads to **promise** to write in the future

Store buffering

	$x = y = 0$	
$x := 1;$		$y := 1;$
$a := y; \text{//0}$		$b := x; \text{//0}$

Simple operational semantics for C11's relaxed accesses

Store buffering

```
x = y = 0
┆
▶ x := 1;      │      │      ▶ y := 1;
a := y; //0    │      │      b := x; //0
```

Memory

```
⟨x : 0@0⟩
⟨y : 0@0⟩
```

T_1 's view

x	y
0	0

T_2 's view

x	y
0	0

- ▶ Global memory is a pool of messages of the form

$\langle location : value @ timestamp \rangle$

- ▶ Each thread maintains a *thread-local view* recording the last observed timestamp for every location

Simple operational semantics for C11's relaxed accesses

Store buffering

```
x = y = 0
x := 1;      |      y := 1;
▶ a := y; //0 |      b := x; //0
```

Memory

```
⟨x : 0@0⟩
⟨y : 0@0⟩
⟨x : 1@1⟩
```

T_1 's view

x	y
0	0
1	

T_2 's view

x	y
0	0

- ▶ Global memory is a pool of messages of the form

$\langle location : value @ timestamp \rangle$

- ▶ Each thread maintains a *thread-local view* recording the last observed timestamp for every location

Simple operational semantics for C11's relaxed accesses

Store buffering

```
x = y = 0
x := 1;      y := 1;
▶ a := y; //0  ▶ b := x; //0
```

Memory

```
⟨x : 0@0⟩
⟨y : 0@0⟩
⟨x : 1@1⟩
⟨y : 1@1⟩
```

T_1 's view

x	y
0	0
1	

T_2 's view

x	y
0	1
	1

- ▶ Global memory is a pool of messages of the form

$\langle location : value @ timestamp \rangle$

- ▶ Each thread maintains a *thread-local view* recording the last observed timestamp for every location

Simple operational semantics for C11's relaxed accesses

Store buffering

```
x = y = 0
x := 1;      |      y := 1;
a := y; //0  |      b := x; //0
```

Memory

```
⟨x : 0@0⟩
⟨y : 0@0⟩
⟨x : 1@1⟩
⟨y : 1@1⟩
```

T_1 's view

x	y
0	0
1	

T_2 's view

x	y
0	0
	1

- ▶ Global memory is a pool of messages of the form

$\langle location : value @ timestamp \rangle$

- ▶ Each thread maintains a *thread-local view* recording the last observed timestamp for every location

Simple operational semantics for C11's relaxed accesses

Store buffering

```
x = y = 0
x := 1;      y := 1;
a := y; //0  b := x; //0
```

Memory

```
⟨x : 0@0⟩
⟨y : 0@0⟩
⟨x : 1@1⟩
⟨y : 1@1⟩
```

T_1 's view

x	y
0	0
1	

T_2 's view

x	y
0	0
	1

- ▶ Global memory is a pool of messages of the form

$\langle location : value @ timestamp \rangle$

- ▶ Each thread maintains a *thread-local view* recording the last observed timestamp for every location

Simple operational semantics for C11's relaxed accesses

Store buffering

```
x = y = 0
x := 1;      ||      y := 1;
a := y; //0  ||      b := x; //0
```

Memory

```
⟨x: 0@0⟩
⟨y: 0@0⟩
⟨x: 1@1⟩
⟨y: 1@1⟩
```

T_1 's view

x	y
0	0
1	

T_2 's view

x	y
0	0
	1

Coherence test

```
x = 0
x := 1;      ||      x := 2;
a := x; //2  ||      b := x; //1
```

Simple operational semantics for C11's relaxed accesses

Store buffering

```

x = y = 0
x := 1;      |      y := 1;
a := y; //0  |      b := x; //0
    
```

Memory

```

⟨x: 0@0⟩
⟨y: 0@0⟩
⟨x: 1@1⟩
⟨y: 1@1⟩
    
```

T_1 's view

x	y
0	0
1	

T_2 's view

x	y
0	0
	1

Coherence test

```

x = 0
▶ x := 1;      |      ▶ x := 2;
a := x; //2    |      b := x; //1
    
```

Memory

```

⟨x: 0@0⟩
    
```

T_1 's view

x
0

T_2 's view

x
0

Simple operational semantics for C11's relaxed accesses

Store buffering

```
x = y = 0
x := 1;      ||      y := 1;
a := y; //0  ||      b := x; //0
```

Memory

```
⟨x: 0@0⟩
⟨y: 0@0⟩
⟨x: 1@1⟩
⟨y: 1@1⟩
```

T_1 's view

x	y
0	0
1	

T_2 's view

x	y
0	0
	1

Coherence test

```
x = 0
x := 1;      ||      x := 2;
a := x; //2  ||      b := x; //1
```

Memory

```
⟨x: 0@0⟩
⟨x: 1@1⟩
```

T_1 's view

x
0
1

T_2 's view

x
0

Simple operational semantics for C11's relaxed accesses

Store buffering

```
x = y = 0
x := 1;      ||      y := 1;
a := y; //0  ||      b := x; //0
```

Memory

```
⟨x: 0@0⟩
⟨y: 0@0⟩
⟨x: 1@1⟩
⟨y: 1@1⟩
```

T_1 's view

x	y
0	0
1	

T_2 's view

x	y
0	0
	1

Coherence test

```
x = 0
x := 1;      ||      x := 2;
a := x; //2  ||      b := x; //1
```

Memory

```
⟨x: 0@0⟩
⟨x: 1@1⟩
⟨x: 2@2⟩
```

T_1 's view

x
0
1

T_2 's view

x
0
2

Simple operational semantics for C11's relaxed accesses

Store buffering

```
x = y = 0
x := 1;      |      y := 1;
a := y; //0  |      b := x; //0
```

Memory

```
⟨x : 0@0⟩
⟨y : 0@0⟩
⟨x : 1@1⟩
⟨y : 1@1⟩
```

T_1 's view

x	y
0	0
1	

T_2 's view

x	y
0	0
	1

Coherence test

```
x = 0
x := 1;      |      x := 2;
a := x; //2  |      b := x; //1
```

Memory

```
⟨x : 0@0⟩
⟨x : 1@1⟩
⟨x : 2@2⟩
```

T_1 's view

x
0
1
2

T_2 's view

x
0
2

Simple operational semantics for C11's relaxed accesses

Store buffering

```
x = y = 0
x := 1;      ||      y := 1;
a := y; //0  ||      b := x; //0
```

Memory

```
⟨x : 0@0⟩
⟨y : 0@0⟩
⟨x : 1@1⟩
⟨y : 1@1⟩
```

T_1 's view

x	y
0	0
1	

T_2 's view

x	y
0	0
	1

Coherence test

```
x = 0
x := 1;      ||      x := 2;
a := x; //2  ||      b := x; //1
```

Memory

```
⟨x : 0@0⟩
⟨x : 1@1⟩
⟨x : 2@2⟩
```

T_1 's view

x
0
1
2

T_2 's view

x
0
2

Load-buffering

```
x = y = 0  
a := x; //1  
y := 1; || x := y;
```

- ▶ To model load-store reordering, we allow **“promises”**.
- ▶ At any point, a thread may promise to write a message in the future, allowing other threads to read from the promised message.

Load-buffering

```
x = y = 0
▶ a := x; //1 || ▶ x := y;
  y := 1;
```

Memory

```
⟨x : 0@0⟩
⟨y : 0@0⟩
```

T_1 's view

x	y
0	0

T_2 's view

x	y
0	0

- ▶ To model load-store reordering, we allow **“promises”**.
- ▶ At any point, a thread may promise to write a message in the future, allowing other threads to read from the promised message.

Load-buffering

```
x = y = 0
▶ a := x; //1 || ▶ x := y;
  y := 1;
```

Memory

$\langle x : 0@0 \rangle$

$\langle y : 0@0 \rangle$

$\langle y : 1@1 \rangle$

T_1 's view

x	y
0	0

T_2 's view

x	y
0	0

- ▶ To model load-store reordering, we allow **“promises”**.
- ▶ At any point, a thread may promise to write a message in the future, allowing other threads to read from the promised message.

Load-buffering

```
x = y = 0
▶ a := x; //1 || ▶ x := y;
y := 1;
```

Memory

$\langle x : 0@0 \rangle$

$\langle y : 0@0 \rangle$

$\langle y : 1@1 \rangle$

T_1 's view

x	y
0	0

T_2 's view

x	y
0	0
	1

- ▶ To model load-store reordering, we allow **“promises”**.
- ▶ At any point, a thread may promise to write a message in the future, allowing other threads to read from the promised message.

Load-buffering

```
x = y = 0
▶ a := x; //1 || x := y;
y := 1; ▶
```

Memory

```
⟨x : 0@0⟩
⟨y : 0@0⟩
⟨y : 1@1⟩
⟨x : 1@1⟩
```

T_1 's view

x	y
0	0

T_2 's view

x	y
0	0
1	1

- ▶ To model load-store reordering, we allow **“promises”**.
- ▶ At any point, a thread may promise to write a message in the future, allowing other threads to read from the promised message.

Load-buffering

```
x = y = 0
a := x; //1
y := 1;
x := y;
```

Memory

```
<x : 0@0>
<y : 0@0>
<y : 1@1>
<x : 1@1>
```

T_1 's view

x	y
0	0
1	

T_2 's view

x	y
0	0
1	1

- ▶ To model load-store reordering, we allow **“promises”**.
- ▶ At any point, a thread may promise to write a message in the future, allowing other threads to read from the promised message.

Load-buffering

```
x = y = 0
a := x; //1
y := 1;
x := y;
```

Memory

```
<x : 0@0>
<y : 0@0>
<y : 1@1>
<x : 1@1>
```

T_1 's view

x	y
0	0
1	1

T_2 's view

x	y
0	0
1	1

- ▶ To model load-store reordering, we allow **“promises”**.
- ▶ At any point, a thread may promise to write a message in the future, allowing other threads to read from the promised message.

Promises

Load-buffering

```
x = y = 0
a := x; //1 || x := y;
y := 1;    ▶
```

Memory

```
⟨x : 0@0⟩
⟨y : 0@0⟩
⟨y : 1@1⟩
⟨x : 1@1⟩
```

T_1 's view

x	y
0	0
1	1

T_2 's view

x	y
0	0
1	1

Load-buffering + dependency

```
a := x; //1 || x := y;
y := a;
```

Must not admit
the same execution!

Load-buffering

```
x = y = 0  
a := x; //1  
y := 1;  || x := y;  
▶       ▶
```

Load-buffering + dependency

```
a := x; //1  
y := a;  || x := y;
```

Key Idea

A thread can only promise if it can perform the write anyway (even without having made the promise)

Thread-local certification

A thread can promise to write a message, if it can *thread-locally certify* that its promise will be fulfilled.

Thread-local certification

A thread can promise to write a message, if it can *thread-locally certify* that its promise will be fulfilled.

Load-buffering

$$\begin{array}{l} a := x; \text{ // } 1 \\ y := 1; \end{array} \parallel x := y;$$

T_1 **may promise** $y = 1$, since it is able to write $y = 1$ by itself.

Load buff. + fake dependency

$$\begin{array}{l} a := x; \text{ // } 1 \\ y := a + 1 - a; \end{array} \parallel x := y;$$

T_1 **may NOT promise** $y = 1$, since it is not able to write $y = 1$ by itself.

Load buffering + dependency

$$\begin{array}{l} a := x; \text{ // } 1 \\ y := a; \end{array} \parallel x := y;$$

Is this behavior possible?

```
a := x; //1  
x := 1;
```

Is this behavior possible?

```
a := x; //1  
x := 1;
```

No.

Suppose the thread promises $x = 1$. Then, once $a := x$ reads 1, the thread view is increased and so the promise cannot be fulfilled.

Is this behavior possible?

```
a := x; //1 ||| y := x; ||| x := y;  
x := 1;
```

Is this behavior possible?

```
a := x; //1 ||| y := x; ||| x := y;  
x := 1;
```

Yes. And the ARM-Flowing model allows it!

Is this behavior possible?

$$\begin{array}{l} a := x; \\ x := 1; \end{array} \parallel \begin{array}{l} //1 \\ y := x; \end{array} \parallel x := y;$$

Yes. And the ARM-Flowing model allows it!

This behavior can be also explained by sequentialization:

$$\begin{array}{l} a := x; \\ x := 1; \end{array} \parallel \begin{array}{l} //1 \\ y := x; \end{array} \parallel x := y; \quad \rightsquigarrow \quad \begin{array}{l} a := x; \\ x := 1; \\ y := x; \end{array} \parallel \begin{array}{l} //1 \\ x := y; \end{array}$$

But, note that sequentialization is generally unsound in our model:

$$\begin{array}{l}
 a := x; \text{ // } 1 \\
 \text{if } a = 0 \text{ then} \\
 \quad x := 1;
 \end{array}
 \parallel
 \begin{array}{l}
 y := x; \\
 x := y;
 \end{array}
 \rightsquigarrow
 \begin{array}{l}
 a := x; \text{ // } 1 \\
 \text{if } a = 0 \text{ then} \\
 \quad x := 1; \\
 \quad y := x;
 \end{array}
 \parallel
 \begin{array}{l}
 x := y;
 \end{array}$$

- ▶ Atomic updates (e.g., CAS, fetch-and-add)
- ▶ Release/acquire fences and accesses
- ▶ Release sequences
- ▶ SC fences (no SC accesses)
- ▶ Plain accesses (C11's non-atomics & Java's normal accesses)

To achieve all of this we enrich our timestamps, messages, and thread views.

Message-passing

```
x = y = 0
x := 1;      ||      a := yacq; //1
y :=rel 1;  ||      b := x; //1
```

Message-passing

$x = y = 0$

▶ $x := 1;$ $y :=_{\text{rel}} 1;$	 	▶ $a := y_{\text{acq}}; //1$ $b := x; //1$
---------------------------------------	------	---

Memory $\langle x : 0@0 \rangle$ $\langle y : 0@0 \rangle$ **T_1 's view**

x	y
0	0

 T_2 's view

x	y
0	0

Message-passing

$x = y = 0$

$x := 1;$		$\blacktriangleright a := y_{\text{acq}}; //1$
$\blacktriangleright y :=_{\text{rel}} 1;$		$b := x; //1$

Memory $\langle x : 0@0 \rangle$ $\langle y : 0@0 \rangle$ $\langle x : 1@1 \rangle$ **T_1 's view**

x	y
0	0
1	

 T_2 's view

x	y
0	0

Message-passing

```

x := 1;
y :=rel 1;
      |
x = y = 0
      |
      ▶ a := yacq; //1
        b := x; //1
  
```

Memory

```

⟨x: 0@0⟩
⟨y: 0@0⟩
⟨x: 1@1⟩
⟨y: 1@1 x@1⟩
  
```

 T_1 's view

x	y
0	0
1	1

 T_2 's view

x	y
0	0

Message-passing

```

x := 1;
y :=rel 1;
▶

```

$x = y = 0$

```

||
a := yacq; //1
▶ b := x; //1

```

Memory

 $\langle x : 0@0 \rangle$ $\langle y : 0@0 \rangle$ $\langle x : 1@1 \rangle$ $\langle y : 1@1 \ x@1 \rangle$ T_1 's view

x	y
x	x
1	1

 T_2 's view

x	y
x	x
1	1

Message-passing

$x = y = 0$		
$x := 1;$		$a := y_{\text{acq}}; //1$
$y :=_{\text{rel}} 1;$		$b := x; //1$
▶		▶

Memory $\langle x : 0@0 \rangle$ $\langle y : 0@0 \rangle$ $\langle x : 1@1 \rangle$ $\langle y : 1@1 \quad x@1 \rangle$ **T_1 's view**

x	y
x	x
1	1

 T_2 's view

x	y
x	x
1	1

Key lemma for DRF

Races only on RA under promise-free semantics

⇒ only promise-free behaviors

$w :=_{\text{rel}} 1;$	$\left\ \begin{array}{l} \text{if } w_{\text{acq}} = 1 \text{ then} \\ \quad z := 1; \\ \text{else} \\ \quad y :=_{\text{rel}} 1; \\ \quad a := x; \text{ //1} \\ \quad \text{if } a = 1 \text{ then} \\ \quad \quad z := 1; \end{array} \right\ $	$\begin{array}{l} \text{if } y_{\text{acq}} = 1 \text{ then} \\ \quad \text{if } z = 1 \text{ then} \\ \quad \quad x := 1; \end{array}$
------------------------	---	---

Theorem (Invariant-Based Program Logic)

Fix a global invariant J . Hoare logic where all assertions are of the form $P \wedge J$, where P mentions only local variables, is sound.

- ▶ Useful for proving absence of OOTA.

Load-buffering + data dependency

$$\begin{array}{l} \{J\} \\ a := x; \\ \{J \wedge (a = 0)\} \\ y := a; \\ \{J\} \end{array} \parallel \begin{array}{l} \{J\} \\ x := y; \\ \{J\} \end{array} \quad J \triangleq (x = 0) \wedge (y = 0)$$