Reversible Debugging

International Training School on Reversible Computation Toruń, Poland

Claudio Antares Mezzina IMT School for Advanced Studies Lucca, Italy

August 30, 2017

Roadmap

Introduction

2 Concurrent Reversible Debugging

Debugging 1/2

Definition

Debugging is the process of finding and resolving misbehaviours of a software system.

It is a well known technique used in software industry at many stages:

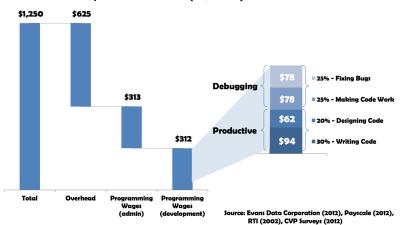
- prototipying
- maintenance after release (e.g., security patches)

Debuggers are integrated in modern IDEs and frameworks

- Eclipse
- .Net
- gdb part of the GNU project, along with gcc

Cost of Debugging

Software development cost structure (US\$ billion)



Study conducted by the Judge Business School of the Cambridge University [1]

Debugging 2/2

50% of programming time is spent in debugging.

• debugging is a costly and time consuming activity

Debugging 2/2

50% of programming time is spent in debugging.

• debugging is a costly and time consuming activity

Classical debugging steps

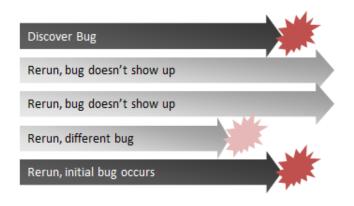
- 1 Trying to reproduce the problem
- 2 Trying to understand the causes (adding printings)
- Trying to highlight the exact line of code which generates the error (breakpoints)

Debugging life¹



¹http://phdcomics.com/

Cyclic Debugging



Cyclic Debugging

- Cumbersome and time consuming activity
- Difficult to understand the real reason of a bug
- Easy to get lost and forget the reason why you are debugging

Cyclic Debugging

- Cumbersome and time consuming activity
- Difficult to understand the real reason of a bug
- Easy to get lost and forget the reason why you are debugging

I WAS TRYING TO FIGURE OUT WHY MY BROWSER WAS ACTING WEIRD

TURNS OUT IT WASN'T THE BROWSER-THE ISSUE WAS WITH MY KEYBOARD DRIVER.



DEBUGGING THAT LED ME TO A MYSTERIOUS ERROR MESSAGE FROM A SYSTEM UTILITY...



ANYWAY, LONG STORY SHORT, I FOUND THE SWORD OF MARTIN THE WARRIOR. I THINK AT SOME POINT THERE YOU SWITCHED PUZZLES.

Record-Replay Debugging

How to reproduce a bug?

One technique to reproduce the bug is to use record-replay

Record-Replay Debugging

How to reproduce a bug?

One technique to reproduce the bug is to use record-replay



- The execution is recorded and replayed
- Execution is still forward
- Cannot step back the execution

Record-Replay Debugging

How to reproduce a bug?

One technique to reproduce the bug is to use record-replay

Discover bug in record mode

Replay bug to debug

- The execution is recorded and replayed
- Execution is still forward
- Cannot step back the execution

Solution?

reversible debugging

What is Reversible Debugging?

Definition [2]

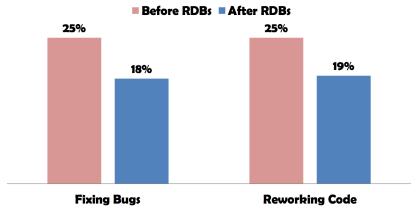
Reverse debugging is the ability of a debugger to stop after a failure in a program has been observed and go back into the history of the execution to uncover the reason for the failure.

Implications

- Ability to execute a program both in forward and backward way.
- Reproduce or keep track of the past of an execution.

Discover and debug immediately

% of programming time spent debugging



- stats taken from [1]
- 26% decrease in debugging time
- \$81.1 billion software development costs saved per year

History

Reversible debugging has a long history

- 1973 Reversible Execution [3]
 - first mention of Reversible Debugging
 - forward-only record-replay for single-threaded progs
- 1987 debugging parallel programs with instant replay [4]
 - assumes deterministic computation by enforcing the same execution order
- 1988 IGOR: a system for program debug via reversible execution [5]
 - debugging single threaded programs
 - record-replay

History

- 1996 Replay for Concurrent Non-Deterministic Shared-Memory Applications [6]
 - record-replay of multi-threaded programs
- 2000 Efficient Algorithms for Bidirectional Debugging [7]
 - single-threaded programs
 - first to introduce backward breakpoints
- 2003 Using events to debug Java programs backwards in time [8]
 - first (and only one) reversible degugger for Java
 - first working debugger for a VM
 - same techniques applied later for MS .Net IntelliTraces
- 2007 Reversible Debugging (added to gdb in 2009) [9]

State of the art: sequential debugging

- Reversible debuggers exist: GDB, UndoDB
- Many reversible debuggers deal only with sequential programs
- Some of them allow one to debug concurrent programs
 - They register scheduler events
 - The same scheduling is used when the program is replayed
 - Program events are linearized
 - Linearized execution can be wind and unwind like a tape

Sequential Reversible Debugging

- Take an execution containing a failure and move backward and forward along it looking for the bug
- The exact same execution can be explored many times forward and backward
 - Even bugs related to concurrency can always be replayed

GDB

- is the GNU debugger
- since version 7 (2009) includes reversible debugging
- uses record and replay technique
 - one has to enable the recording phase (and decide the tape length)
 - executes forward
 - 3 can explored the recorded session back and forth
 - 4 when going back and forth action are not re-executed

GDB Commands

Like the forward commands (step, next, continue), but in the backward direction

- reverse-step: goes back to the last instruction
- reverse-next: goes back to the last instruction, does not go inside functions
- reverse-continue: runs back till a breakpoint/watchpoint triggers
- breakpoints and watchpoints can be used also in the backward direction

gcc file.c -ggdb

Why RDs are not widely used²?

Why is reverse debugging rarely used? [closed]



gdb implemented support for reverse debugging in 2009 (with gdb 7.0). I never heard about it until 2012. Now I find it extremely useful for certain types of debugging problems. I wished that I heard of it before.



Correct me if I'm wrong but my impression is that the technique is still rarely used and most people don't know that it exists. Why?

For one, running in debug mode with recording on is **very** expensive compared to even normal debug mode; it also consumes a lot more memory.

It is easier to decrease the granularity from line level to function call level. For example, the standard debugger in eclipse allows you to "drop to frame," which is essentially a jump back to the start of the function with a reset of all the parameters (nothing done on the heap is reverted, and finally blocks are not executed, so it is not a true reverse debugger; be careful about that).



²Taken from stackowerflow

Why RDs are not widely used: the case of GDB

- Recording phase in GDB is not optimized and adds a huge overehead in terms of both space and time
- Overhead in times may hide under the carpet synchronisation/race condition bugs

UndoDB a commercial debugger

- From UndoSoftware, Cambridge, UK http://undo-software.com/
- A main company in the field of reversible debugging
- Improves GDB recording by using efficient incremental check-poiting techniques
- Available for Linux (intel/amd processors) and Android (ARM processors)
- Allows reversible debugging for programs in C/C++
- Allows to write on a file a recording session
 - useful for replaying the session on different machines
 - a client can send the recorded version to the developer

UndoDB success

Comparison with GDB, on recording gzipping a 16MB file

	Native	UndoDB	GDB
Time	1.49 s	2.16 s (1.75 x)	21 h (50000 x)
Space	-	17.8 MB	63 GB

Roadmap

Introduction

2 Concurrent Reversible Debugging

Reversibility and Debugging

Question:

When a misbehaviour is detected, how one should proceed in order to retrace the steps that led to the bug?

- Sequential setting: recursively undo the last action.
- Concurrent setting: there is not a clear understanding of which the last action is.
- Techniques to undo a concurrent execution in literature:
 - non-deterministic replay
 - deterministic replay/reverse

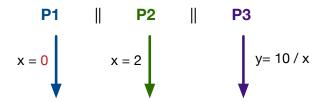
Non-deterministic replay

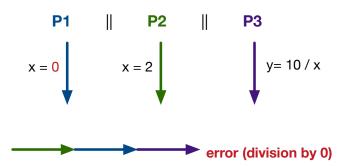
Non-deterministic replay

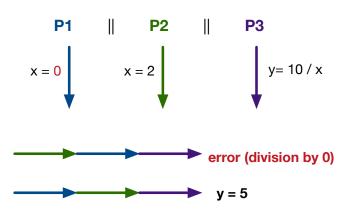
The execution is replayed non deterministically from the start (or from a previous checkpoint) till the desired point.

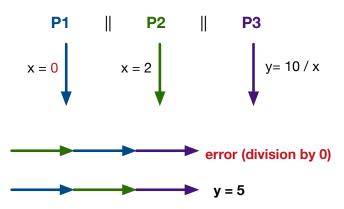
Non-deterministic replay:

- Actions could get scheduled in a different order and hence the bug may not be reproduced.
- Particularly difficult to reproduce concurrency problems (e.g. race conditions).









We need to store information about the threads execution

Replay/Reverse

replay/reverse execution

A log of the scheduling among threads is kept and then actions are reversed or replayed accordingly.

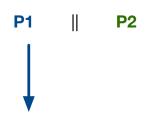
Deterministic replay/reverse execute:

- Also actions in threads not related to the bug may be undone.
- If one among several independent threads causes the bug, and this thread has been scheduled first, then one has to undo the entire execution to find the bug.

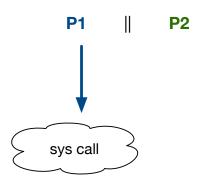
Deterministic Replay: example

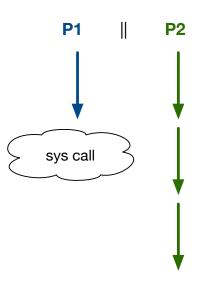


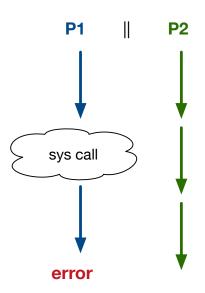
Replay

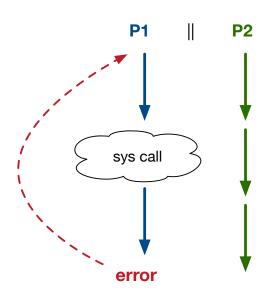


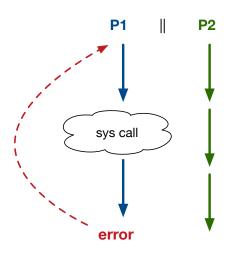
Deterministic Replay: example







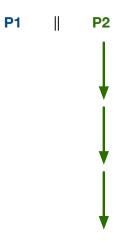








Wanted execution



We need to less information about the threads execution

Causal-Consistent Reversibility

Actions are reversed respecting the causes:

- only actions that have caused no successive actions can be undone;
- concurrent actions can be reversed in any order;
- dependent actions are reversed starting from the consequences.

Benefits:

The programmer can easily individuate and undo the actions that caused a given misbehaviour.

A simple debugger for a simple language

- a subset of the Oz language [10]
- Functional Language
 - thread-based concurrency
 - asynchronous communication via channels (ports)
- well-know stack-based abstract machine

Syntax

Semantics

- The semantics is defined as a reduction relation, noted \rightarrow , between configurations of the form (U, σ) .
- To follow Oz notation, the relation \to is defined by a set of rules of the form below, specifying that (U,σ) reduces to (U',σ') if condition G is satisfied

$$\begin{array}{c|c} U & U' \\ \hline \sigma & \sigma' \end{array} \text{ if } G$$

- programs written as stacks of instructions
- variables are always created fresh and never modified
- sent values are variables names, not their contents

Semantics 1/2

Semantics 2/2

$$\begin{array}{c|c|c|c} \text{nth} & \frac{\langle \operatorname{thread} S \operatorname{end} T \rangle & T \parallel \langle S \, \langle \rangle \rangle}{0} \\ \\ \text{pc} & \frac{\langle \{\,x\,x_1\ldots x_n\,\}\,T \rangle & \langle S\{^{x_1}\!/y_1\}\ldots \{^{x_n}\!/y_n\}\,T \rangle}{x\,=\,\xi\,\parallel} \\ \xi : \operatorname{proc}\,\{\,y_1\ldots y_n\,\}\,S \operatorname{end} & \xi : \operatorname{proc}\,\{\,y_1\ldots y_n\,\}\,S \operatorname{end} \\ \\ \text{snd} & \frac{\langle \{\,\operatorname{Send}\,x\,y\,\}\,T \rangle & T}{x\,=\,\xi\,\parallel\,\xi : Q} & x\,=\,\xi\,\parallel\,\xi : y;Q \\ \\ \text{rcv} & \frac{\langle \operatorname{let}\,x=\,\{\,\operatorname{Receive}\,y\,\}\operatorname{in}\,S\operatorname{end}\,T \rangle & \langle S\{^{x'}\!/_x\}\,T \rangle}{y\,=\,\xi\,\parallel} \\ \xi : Q;z\,\parallel\,z=w & \xi : Q \parallel z=w \parallel \\ \xi : Q;z\,\parallel\,z=w & x'=w \end{array}$$

Reversing the language

- unique thread identifiers
- threads endowed with a history
- syntactic delimiters to statements, to delimit their scope
- queues with histories

Syntax Modifications

unique thread id and past history

- unique thread id and past history
- history include the new action

- unique thread id and past history
- history include the new action
- scope delimiter
 - identify the scope of the statement to be reversed
 - usefull to reverse procedure calls

Forward Semantics

Forward Semantics

$$\begin{array}{c|c|c|c} \text{nth} & \frac{t[H]\langle \operatorname{thread} S \ \operatorname{end} C \rangle}{0} & \frac{t[H \ * t']C \ \| \ t'[\bot]\langle S \ \langle \rangle \rangle}{0} \ \operatorname{if} \ t' \ \operatorname{fresh} \\ \hline \\ b & t[H]\langle \{ \ x \ (x_i)_1^n \ \} \ C \rangle & \frac{t[H \ \{ \ x \ (x_i)_1^n \ \}]\langle S(\{^{x_i}/_{y_i}\})_1^n \ \langle \operatorname{esc} \ C \rangle \rangle}{x = \xi \ \| & x = \xi \ \| & x = \xi \ \| \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} & \xi : \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n \ \} \ S \ \operatorname{end} \\ \hline \\ s \cdot \operatorname{proc} \ \{ \ (y_i)_1^n$$

Backward Semantics (excerpt)

$$\frac{t[H]\langle \{ \text{ Send } x \ y \ \} \ C \rangle \quad \| \quad t[H \ \uparrow x]C}{x = \xi \parallel \xi : K|K_h \quad \| \quad x = \xi \parallel \xi : t : y; K|K_h}$$

$$\text{snd}^{-1}$$

$$\frac{t[H \ \uparrow x]C \quad \| \quad t[H]\langle \{ \text{ Send } x \ y \ \} \ C \rangle}{x = \xi \parallel \xi : t : y; K|K_h \quad \| \quad x = \xi \parallel \xi : K|K_h}$$

rcv
$$\frac{t[H]\langle \text{let } y = \{ \text{ Receive } x \} \text{ in } S \text{ end } C \rangle \quad \left\| \begin{array}{c} t[H \downarrow x(y')] \langle S\{^{y'}\!/y\} \ \langle \text{esc } C \rangle \rangle \\ \theta \parallel \xi : K; t' : z | K_h & \parallel \theta \parallel \xi : K | t' : z, t; K_h \parallel y' = w \\ \text{if } y' \text{ fresh } \wedge \ \theta \triangleq x = \xi \parallel z = w \end{array}$$

$$\mathsf{rcv}^{-1} \qquad \frac{t[H \downarrow x(z)] \langle S \; \langle \mathsf{esc} \; C \rangle \rangle \qquad \parallel \; t[H] \langle \mathsf{let} \; z = \{ \; \mathsf{Receive} \; x \; \} \; \mathsf{in} \; S \; \mathsf{end} \; C \rangle}{z = w \; \parallel x = \xi \; \parallel \xi : K|t':y,t;K_h \; \parallel \qquad x = \xi \; \parallel \xi : K;t':y|K_h }$$

more details in: A reversible abstract machine and its space overhead [11]

Debugger Commands

	forth (f) t	(forward execution of one step of thread t)
control	run	(runs the program)
	rollvariable (rv) id	(c-c undo of the creation of variable id)
	rollsend (rs) id n	(c-c undo of last n send to port id)
	rollreceive (rr) id n	(c-c undo of last n receive from port id)
	rollthread (rt) t	(c-c undo of the creation of thread t)
	roll (r) t n	(c-c undo of n steps of thread t)
	back (b) t	(bk execution of one step of t (if possible))
explore	list (I)	(displays all the available threads)
	store (s)	(displays all the ids contained in the store)
	print (p) id	(shows the state of a t, c, or v)
	history (h) id	(shows thread/channel computational history)

Example of execution

```
let a = true in
                                                                           (1)
  let b = false in
                                                                           (2)
     let x = port in
                                                                           (3)
                                                                           (4)
        thread \{\text{send x a}\}; \text{skip}; \{\text{send x b}\} \text{ end};
        let y = \{\text{receive } x\} in skip end
                                                                           (5)
     end
                                                                           (6)
  end
                                                                           (7)
                                                                           (8)
end
```

- ullet at line (4) thread t_1 is created from thread t_0
- ullet t_1 fully executes, then t_0 fully executes
- what should be the shape of t_0 (and of the port) if t_1 rolls of 3 steps?

Desired execution

```
t_0 let y = \{ \mathbf{receive} \ x \} in skip end

t_1 \{ \mathbf{send} \ \mathbf{x} \ \mathbf{a} \}; \mathbf{skip}; \{ \mathbf{send} \ \mathbf{x} \ \mathbf{b} \}

x \perp
```

- ullet t_0 is rolled-back enough in order to free the read value
- ullet No domino effect, causing t_0 to fully roll-back

Building up a debugger [12]

- Java based Interpreter of the Oz reversible semantics
 - forward and backward steps
 - roll as controlled sequence of backward steps
 - rollvariable, rollthread, rollsend, rollreceive are based on roll
- It keeps history and causality information to enable reversibility

http://www.cs.unibo.it/caredeb

Implementations Details

Computation information:

- The history of each thread
- The history of each channel, containing:
 - elements of the form (t0, i, a, t1, j)
 - t0 sent a value a which has been received by t1
 - i and j are pointers to t0 and t1 send/receive instructions

Also the debugger maintains the following mappings:

- var_name → (thread_name, i) pointing to the variable creator (for rollvar)
- thread_name \rightarrow (thread_name, i) pointing to the thread creator (for rollthread)
- could be retrieved by inspecting histories, but storing them is much more efficient

Reversing: code snippet

```
private static void rollTill(HashMap < String, Integer > map)
 //map contains pairs <thread_name,i>
  Iterator < String > it = map.keySet().iterator();
  while(it.hasNext())
  {
     String id = it.next();
     int gamma = map.get(id);
     //getGamma retrieves the next gamma in the history
     while(gamma <= getGamma(id))</pre>
     ſ
        trv {
              stepBack(id);
        } catch (WrongElementChannel e) {
              rollTill(e.getDependencies());
        } catch (ChildMissingException e) {
              rollEnd(e.getChild());
```

Demo Time



Conclusions



Conclusions

We are not re-inventing the wheel, just improving it!

- causality information helps to find out the root of a bug
- causality information helps in saving the right order of events

Conclusions

We are not re-inventing the wheel, just improving it!

- causality information helps to find out the root of a bug
- causality information helps in saving the **right** order of events

Cakewalk? (Problems)

- Increasing the expressiveness of the language will modify the notion of causality
- system calls, interrupts, shared memory, file descriptors will make the entire setting more complex
 - some of them cannot be reverted
 - should be recorder and replayed as they are

References I

Tom Britton, Lisa Jeng, Graham Carver, and Paul Cheak. Reversible debugging software "quantify the time and cost saved using reversible debuggers".

J. Engblom.

A review of reverse debugging.

In Proceedings of the 2012 System, Software, SoC and Silicon Debug Conference, pages 1–6, 2012.

M. V. Zelkowitz.

Reversible execution.

Commun. ACM, 16(9):566-, September 1973.

Thomas J. LeBlanc and John M. Mellor-Crummey. Debugging parallel programs with instant replay. *IEEE Trans. Computers*, 36(4):471–482, 1987.

References II



Igor: A system for program debugging via reversible execution. In WPDD 98, pages 112–123. ACM, 1988.

Mark Russinovich and Bryce Cogswell.

Replay for concurrent non-deterministic shared-memory applications.

SIGPLAN Not., 31(5):258-266, May 1996.

Bob Boothe.

A fully capable bidirectional debugger. ACM SIGSOFT Software Engineering Notes, 25(1):36–37, 2000.

Bil Lewis and Mireille Ducassé.
Using events to debug java programs backwards in time.
In OOPSLA 2003,, pages 96–97. ACM, 2003.

References III



Peter Van Roy and Seif Haridi.

Concepts, Techniques, and Models of Computer Programming.

MIT Press, 2004.

Michael Lienhardt, Ivan Lanese, Claudio Antares Mezzina, and Jean-Bernard Stefani.

A reversible abstract machine and its space overhead. In *FMOODS / FORTE*, volume 7273 of *Lecture Notes in Computer Science*, pages 1–17. Springer, 2012.

References IV



Elena Giachino, Ivan Lanese, and Claudio Antares Mezzina. Causal-consistent reversible debugging.

In *FASE 2014*, volume 8411 of *LNCS*, pages 370–384. Springer, 2014.