

Clara: a Framework for Statically Evaluating Finite-state Runtime Monitors

Eric Bodden, Patrick Lam and Laurie Hendren



McGill



CASED



TECHNISCHE
UNIVERSITÄT
DARMSTADT

tutorial vs. this talk

Integrate results of three communities

AOP /
AspectJ

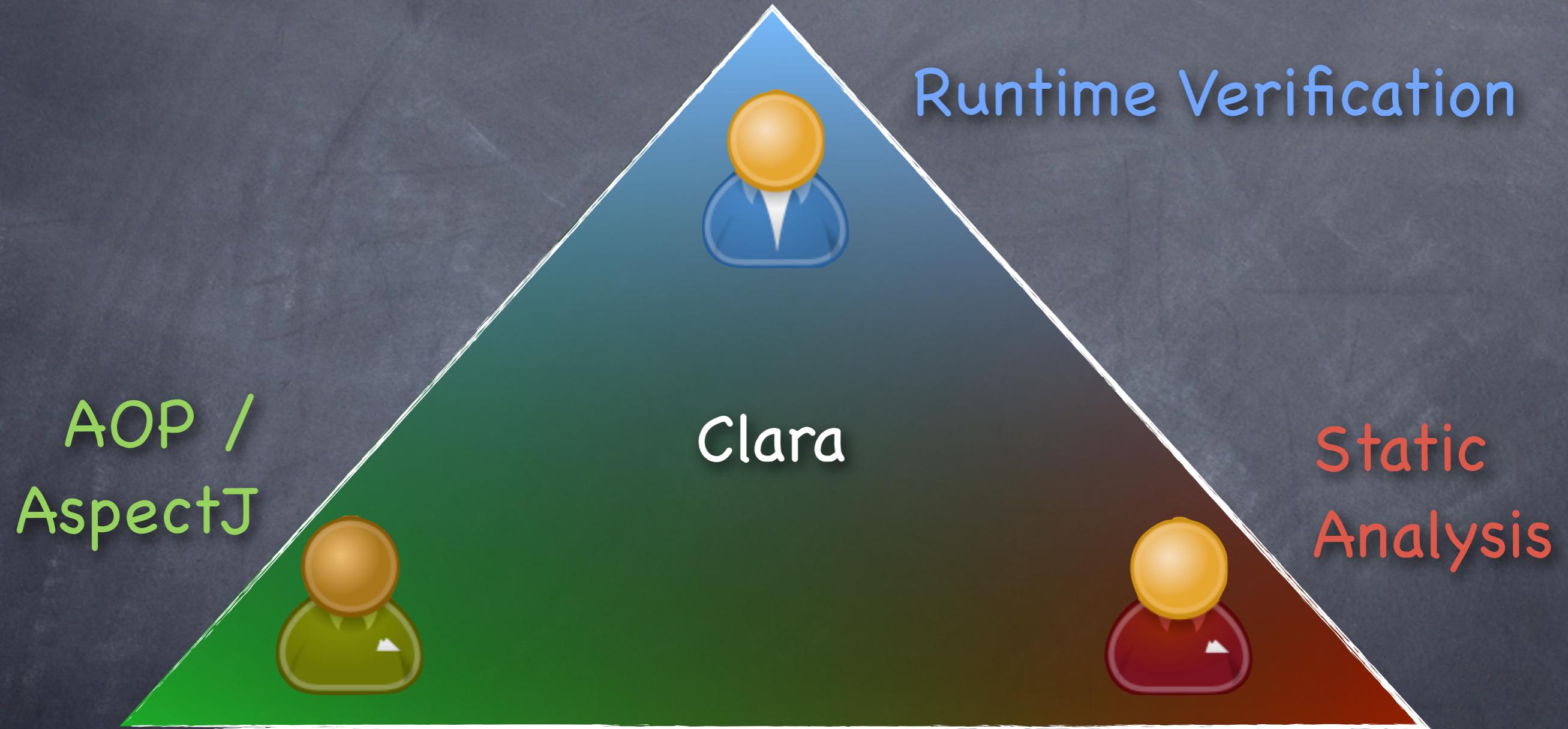


Runtime Verification

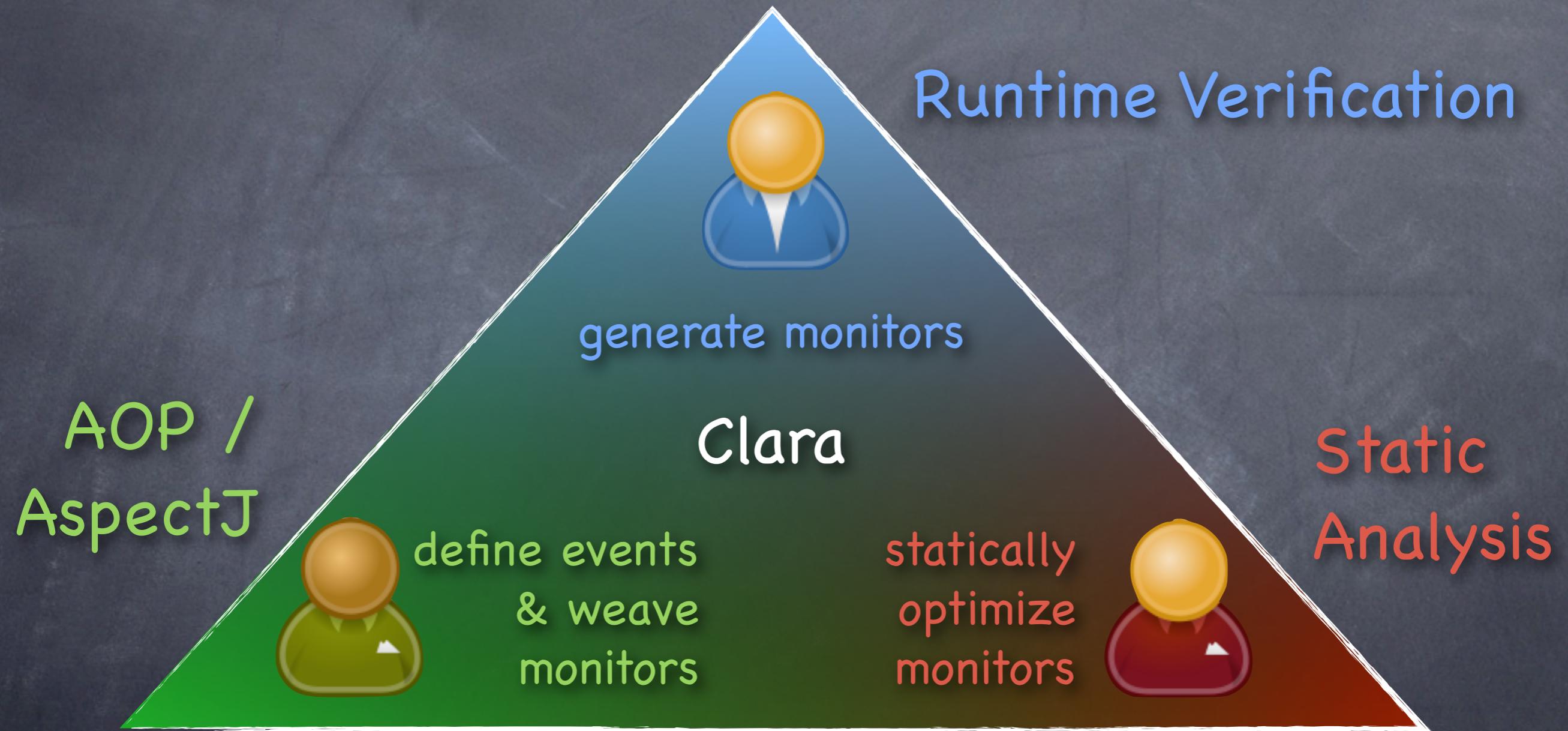
Static
Analysis



Integrate results of three communities



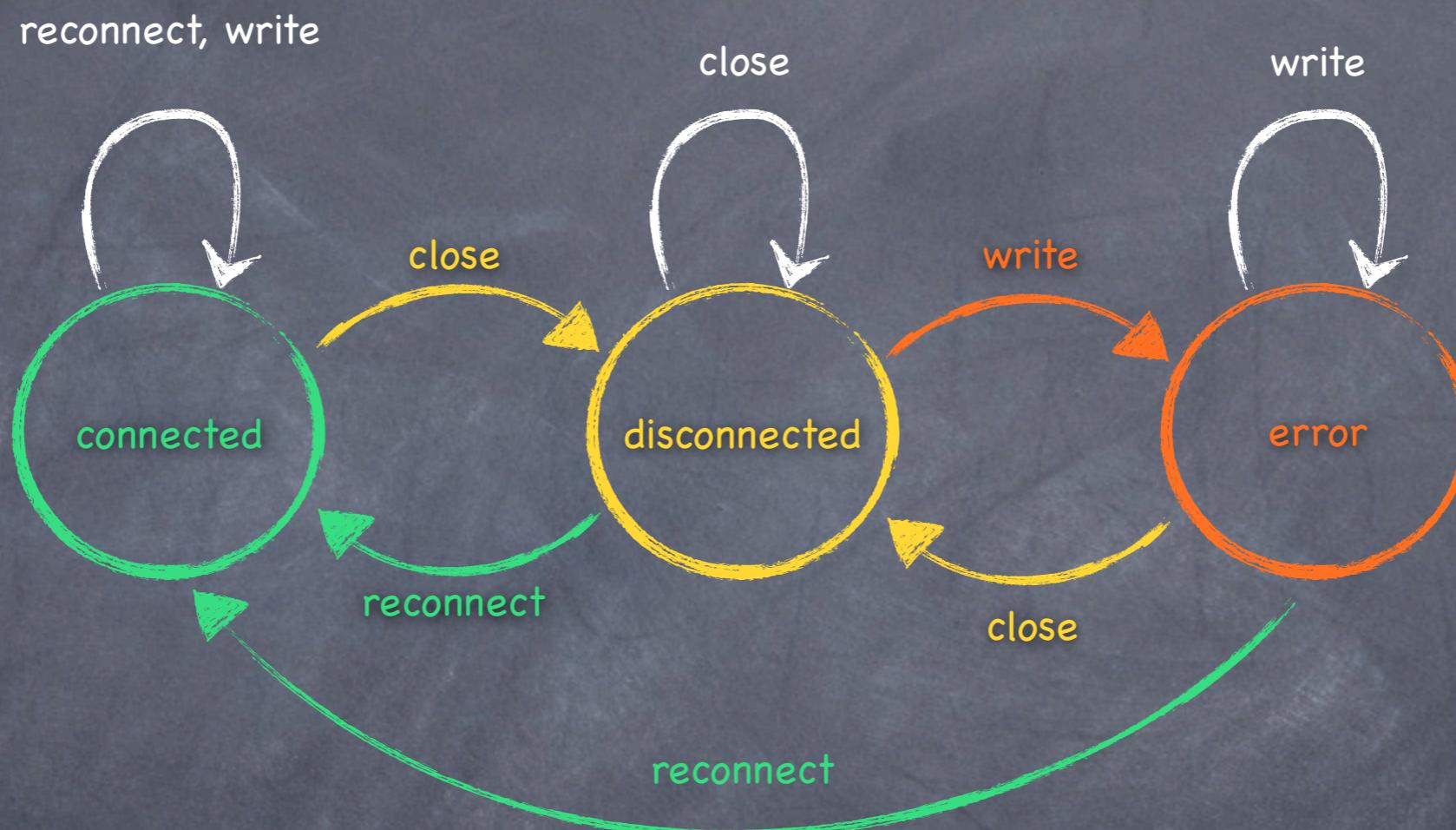
Integrate results of three communities



Finite-state properties

“After closing a connection c ,
don’t write to c until c is reconnected.”

Finite-state properties



“After closing a connection c ,
don’t write to c until c is reconnected.”

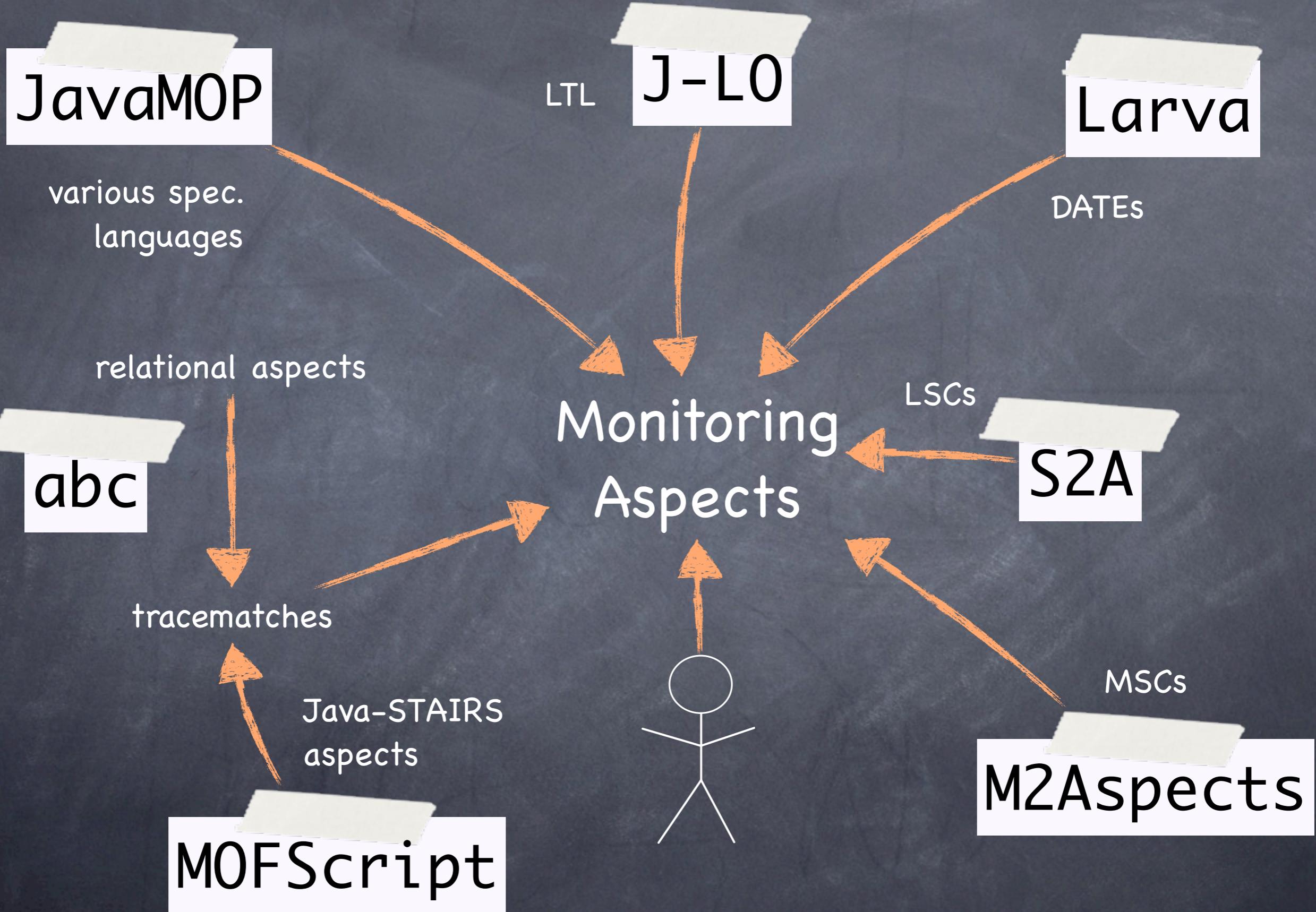
Runtime verification of finite-state properties (in AspectJ)

```
Set closed = new WeakIdentityHashSet();

after(Connection c) returning:
    call(* Connection.close()) && target(c) {
        closed.add(c);
    }

after(Connection c) returning:
    call(* Connection.reconnect()) && target(c) {
        closed.remove(c);
    }

after(Connection c) returning:
    call(* Connection.write(..)) && target(c) {
        if(closed.contains(c))
            error("May not write to "+c+", as it is closed!");
    }
```

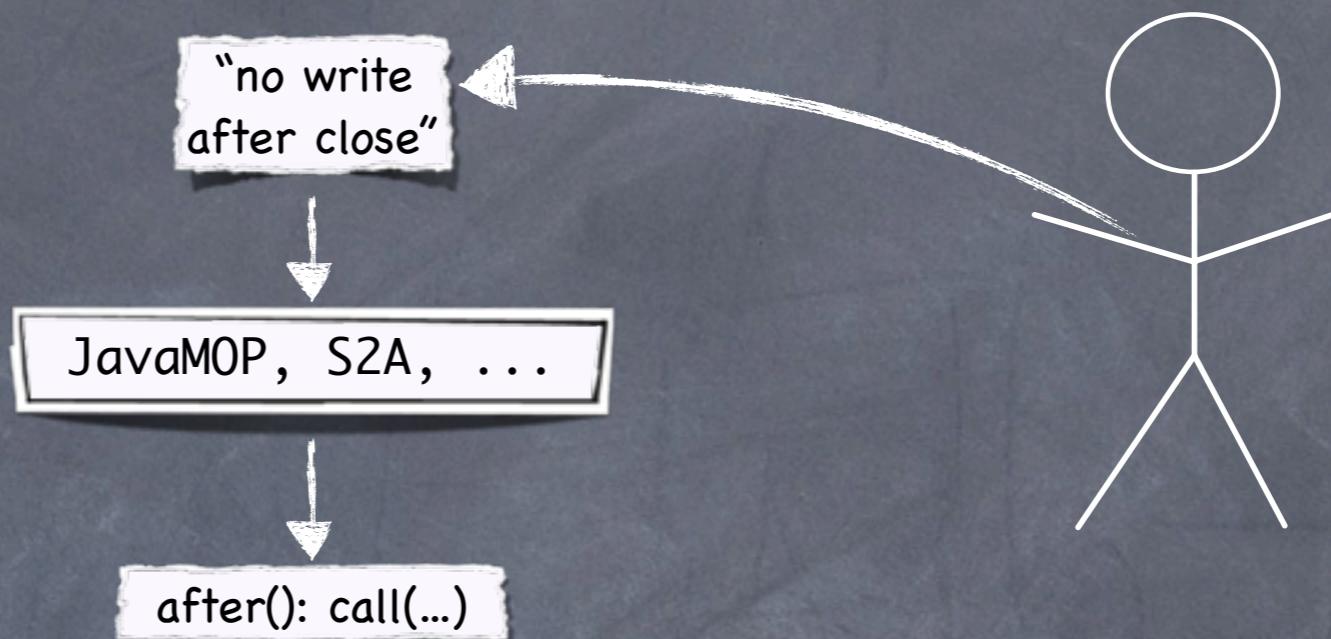


Runtime verification of finite-state properties

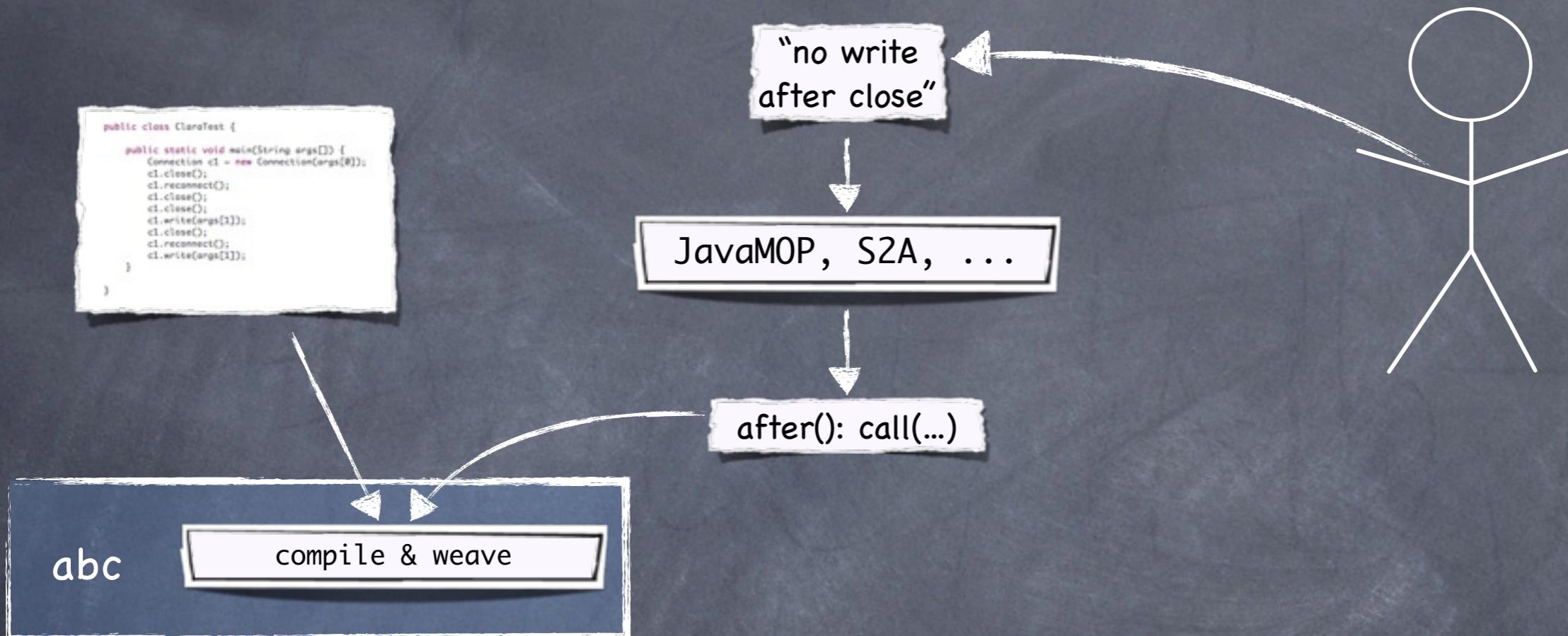
Runtime verification of finite-state properties



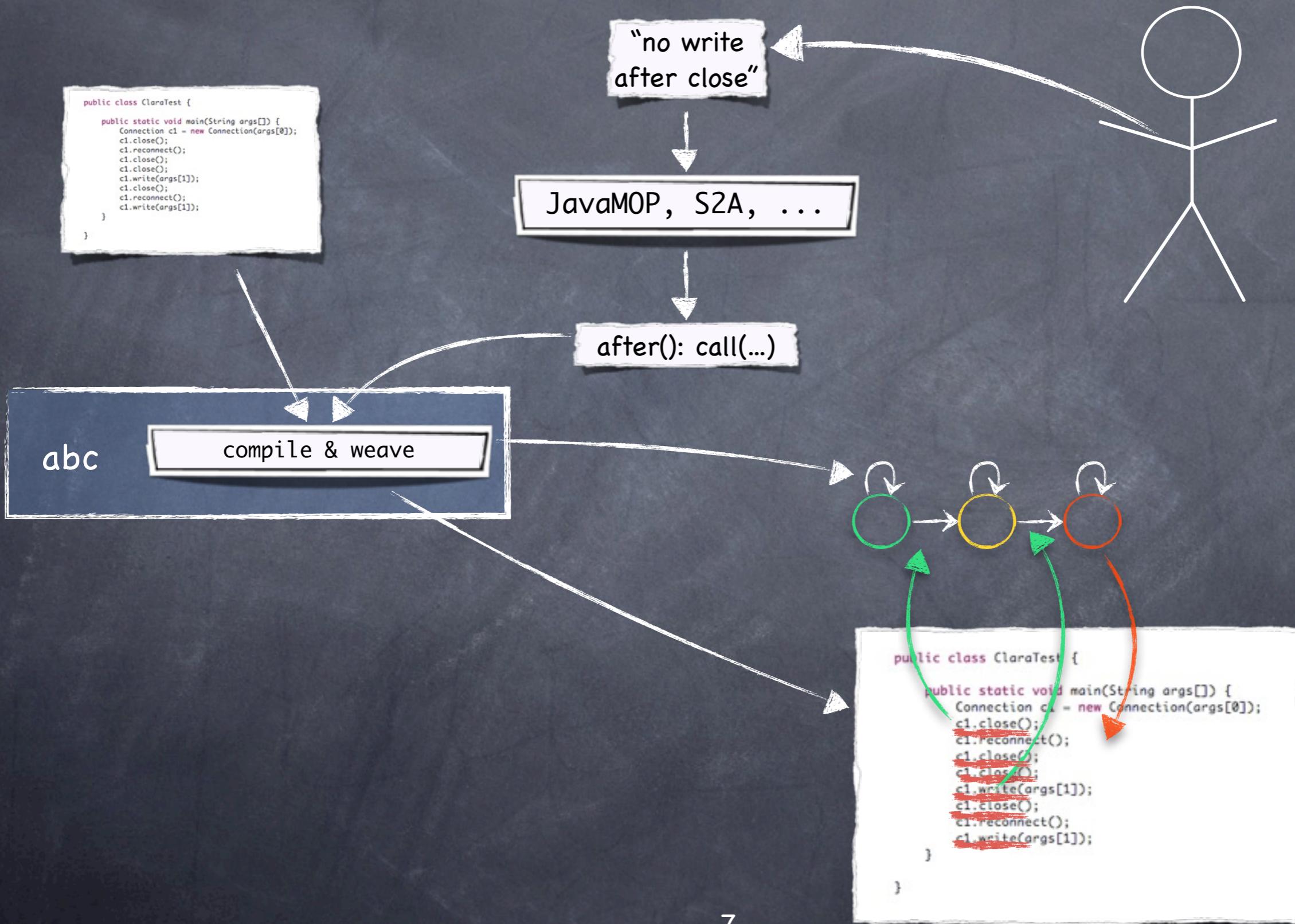
Runtime verification of finite-state properties



Runtime verification of finite-state properties



Runtime verification of finite-state properties



Runtime verification of finite-state properties

Runtime verification of finite-state properties



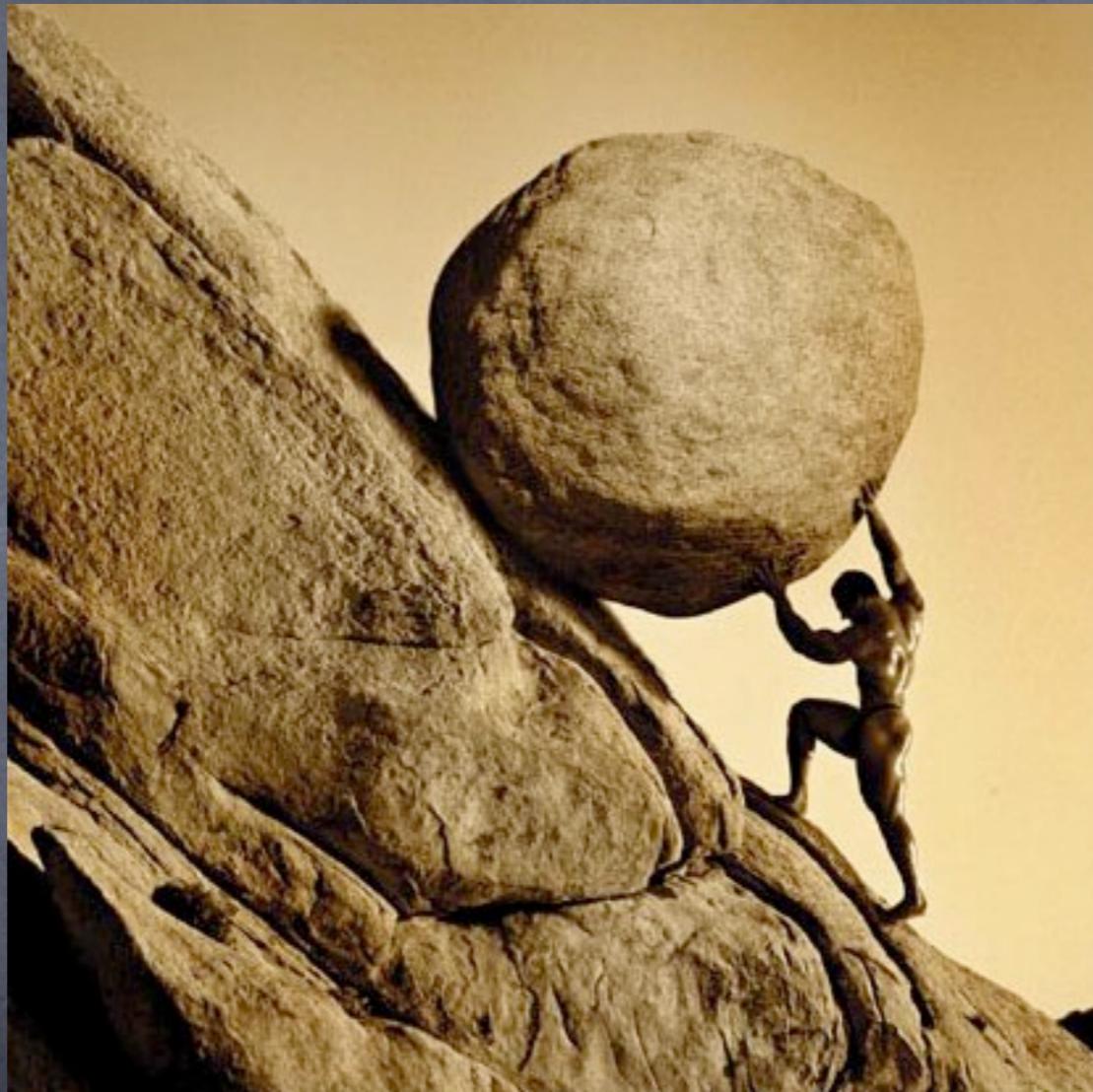
No static guarantees

Runtime verification of finite-state properties



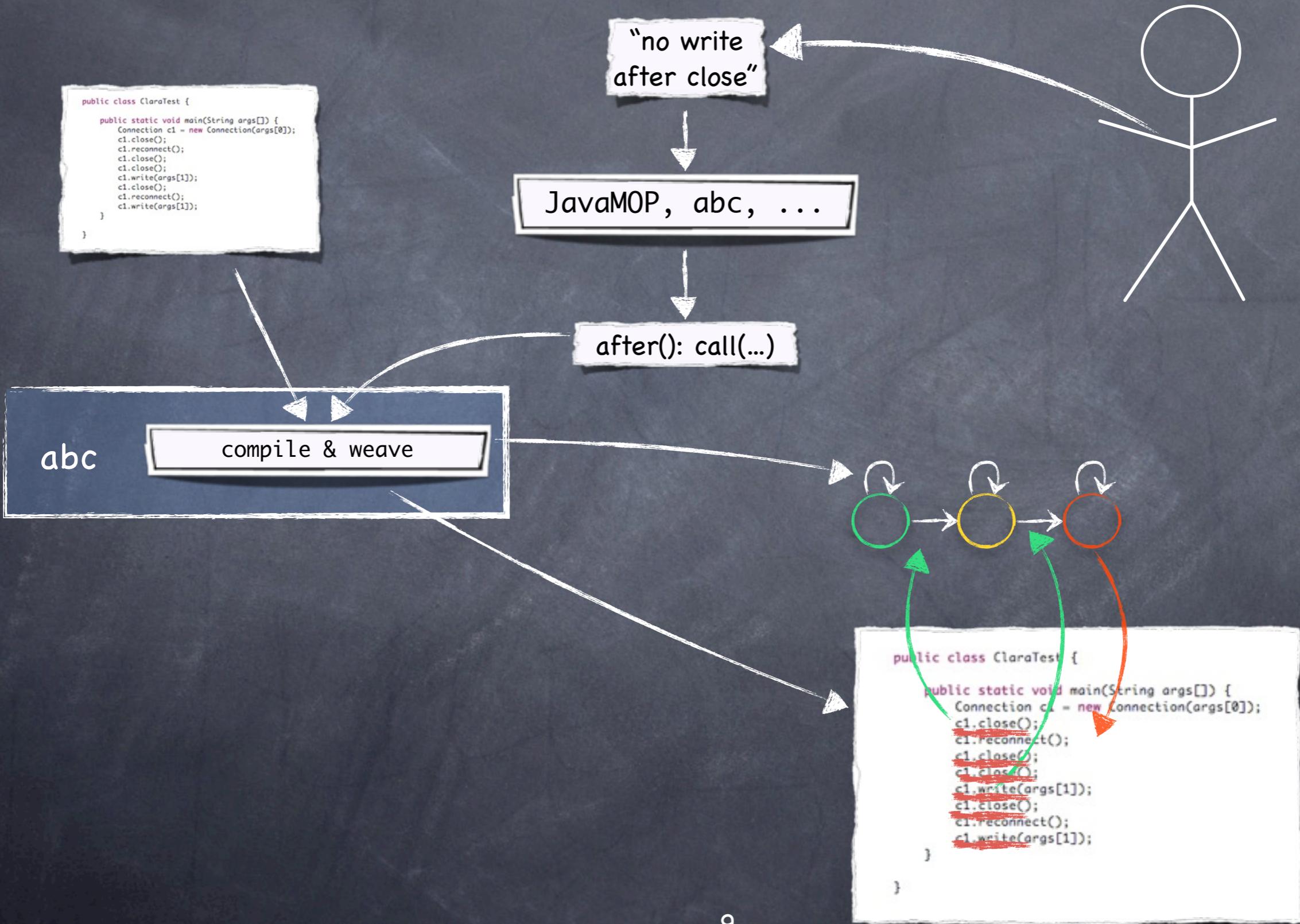
Potentially large runtime overhead

Runtime verification of finite-state properties

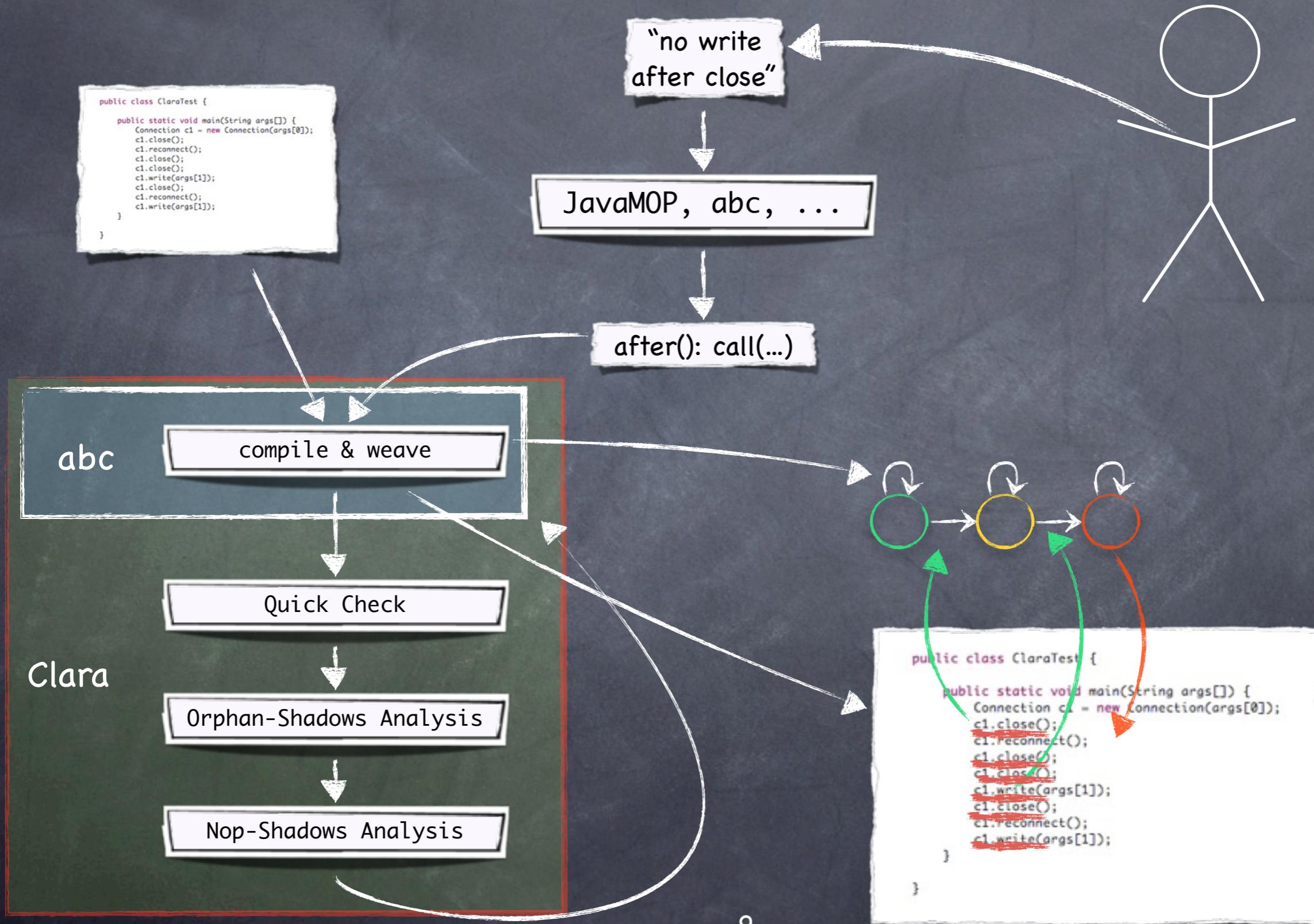


When to finish testing?

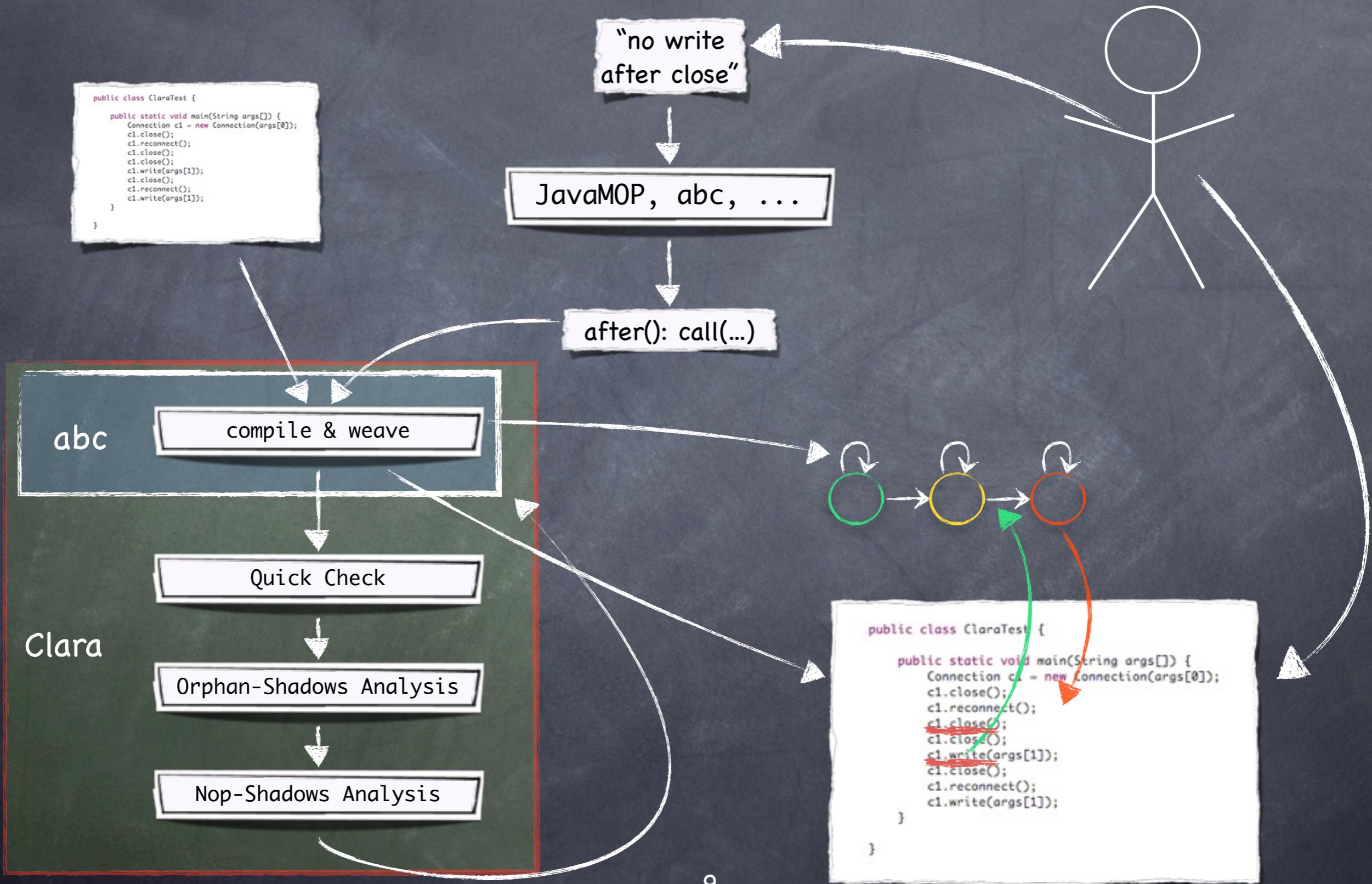
The Clara Framework



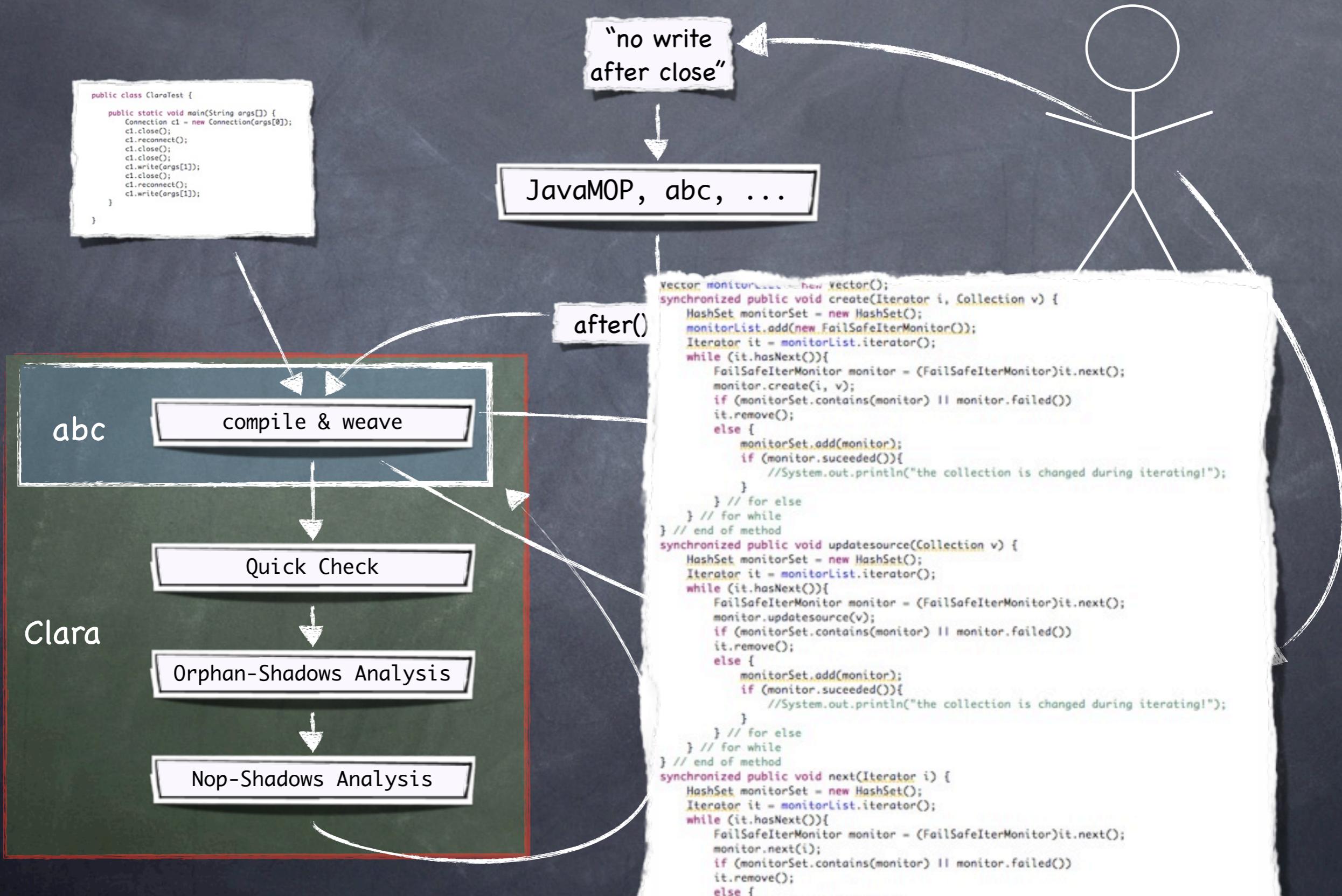
The Clara Framework



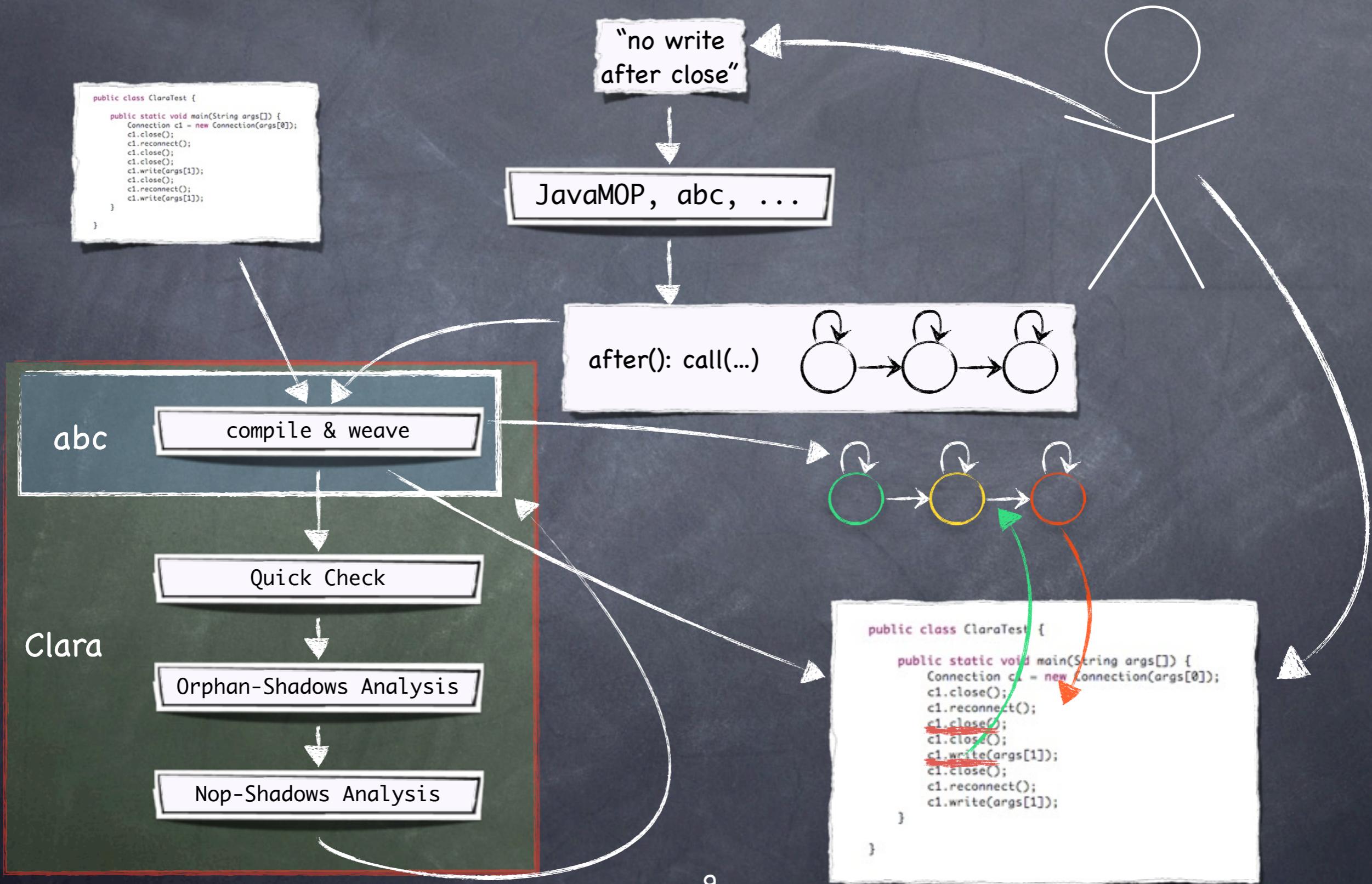
The Clara Framework



The Clara Framework



The Clara Framework



Dependency State Machines

```
Set closed = new WeakIdentityHashSet();

after(Connection c) returning:
    call(* Connection.close()) && target(c) {
        closed.add(c);
    }

after(Connection c) returning:
    call(* Connection.reconnect()) && target(c) {
        closed.remove(c);
    }

after(Connection c) returning:
    call(* Connection.write(..)) && target(c) {
        if(closed.contains(c))
            error("May not write to "+c+", as it is closed!");
    }
```

Dependency State Machines

```
Set closed = new WeakIdentityHashSet();
```

```
dependent after disconnect(Connection c) returning:  
    call(* Connection.close()) && target(c) {  
        closed.add(c);  
    }
```

```
dependent after reconnect(Connection c) returning:  
    call(* Connection.reconnect()) && target(c) {  
        closed.remove(c);  
    }
```

```
dependent after write(Connection c) returning:  
    call(* Connection.write(..)) && target(c) {  
        if(closed.contains(c))  
            error("May not write to "+c+", as it is closed!");  
    }
```

abstract

concrete

```
Set closed = new WeakIdentityHashSet();

dependent after disconnect(Connection c) returning:
    call(* Connection.close()) && target(c) {
        closed.add(c);
    }

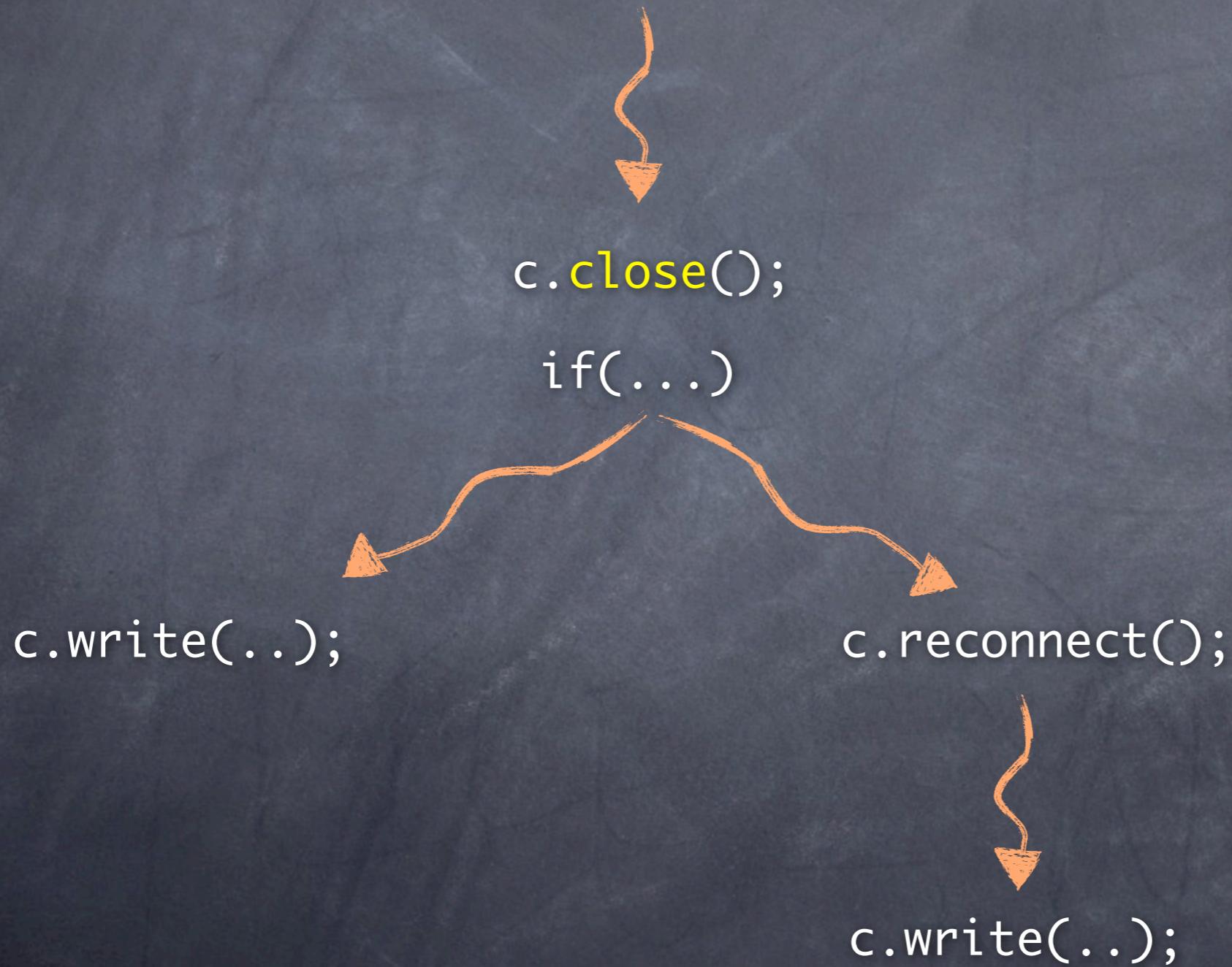
dependent after reconnect(Connection c) returning:
    call(* Connection.reconnect()) && target(c) {
        closed.remove(c);
    }

dependent after write(Connection c) returning:
    call(* Connection.write(..)) && target(c) {
        if(closed.contains(c))
            error("May not write to "+c+", as it is closed!");
    }
```

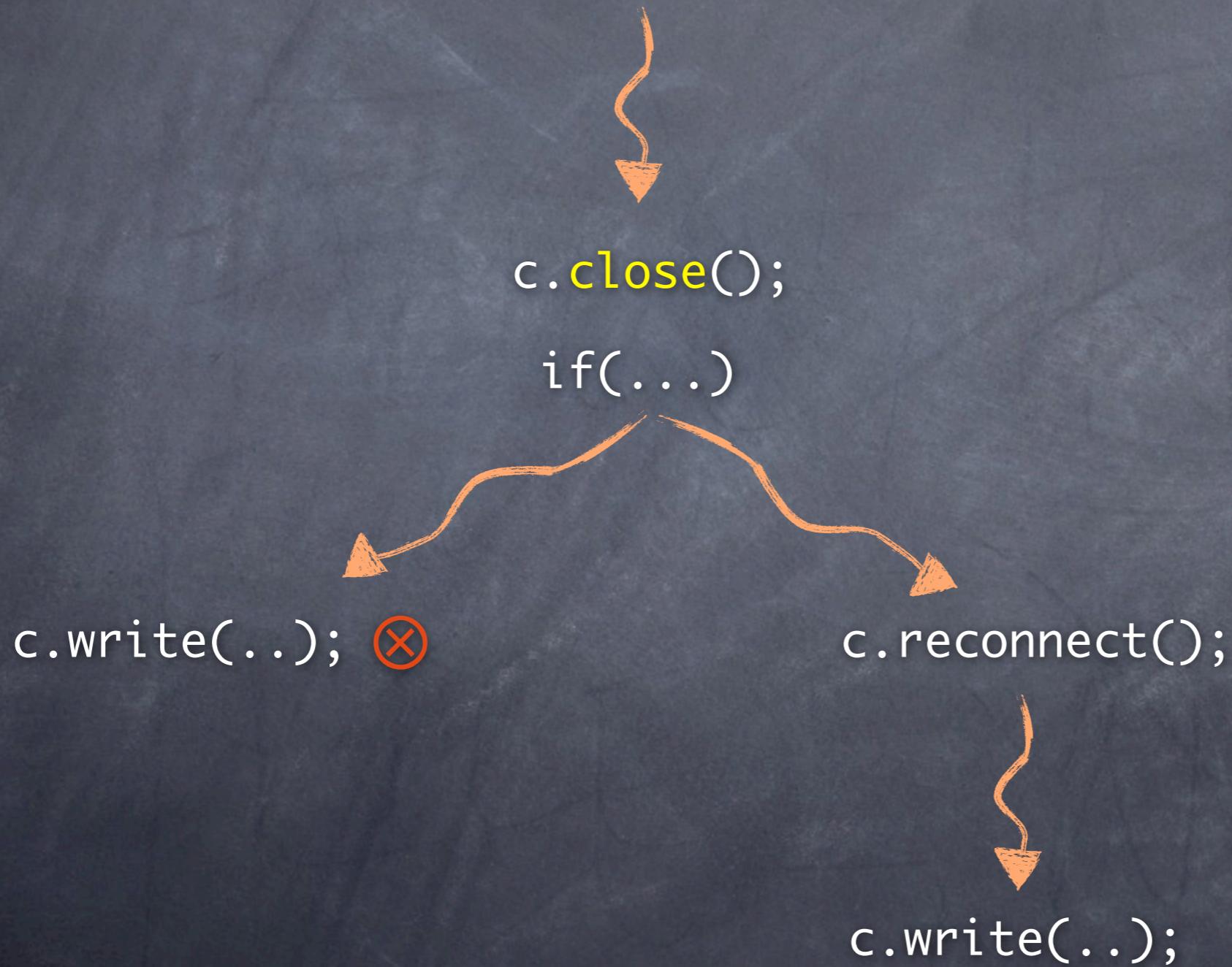
```
dependency{
    disconnect, write, reconnect;
initial   connected: disconnect -> connected,
          write -> connected,
          reconnect -> connected,
          disconnect -> disconnected;
disconnect: disconnect -> disconnected,
           write -> error;
final      error: write -> error;
}
```

*finite-state
property*

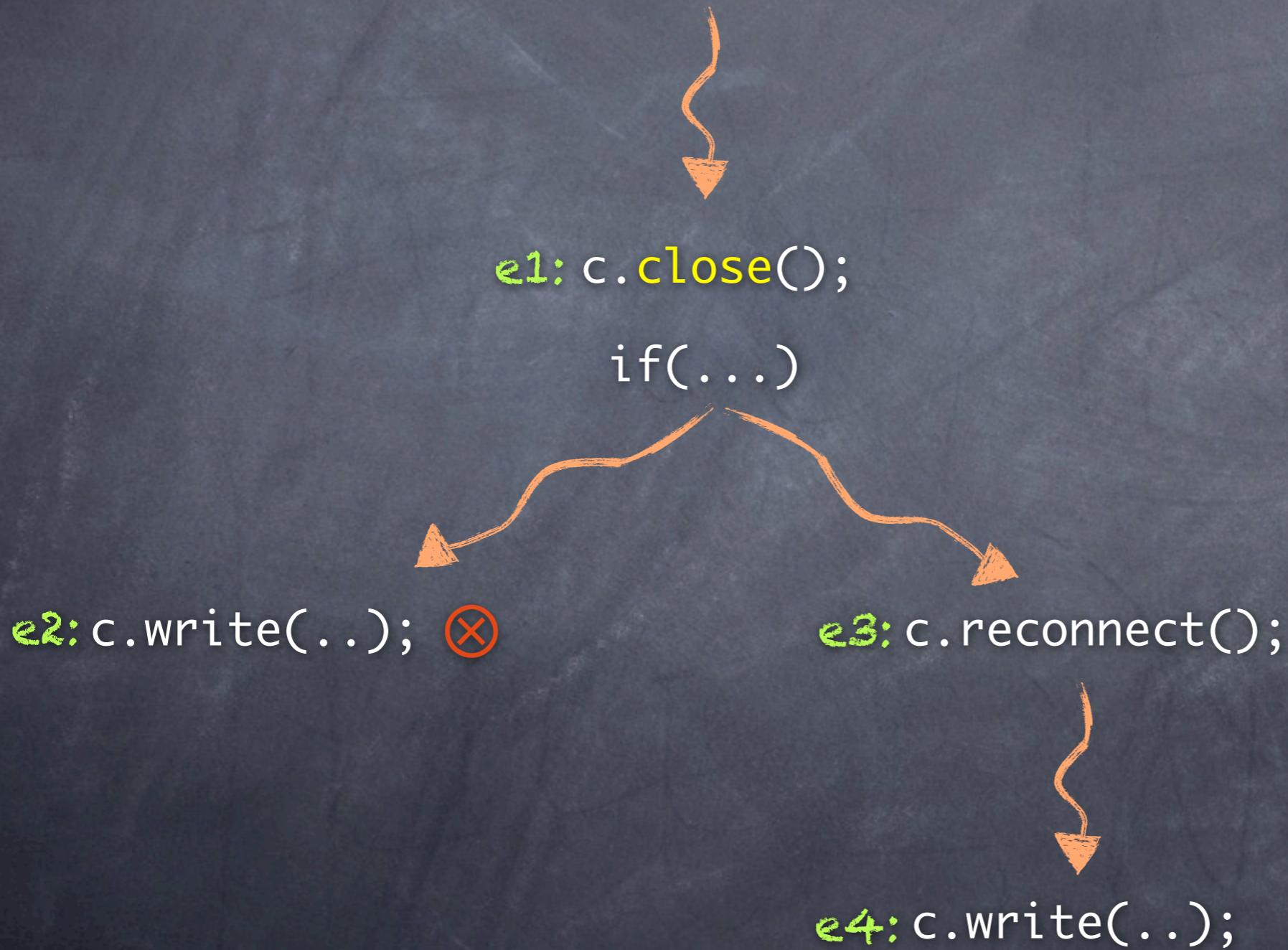
Semantics of Dependency State Machines



Semantics of Dependency State Machines

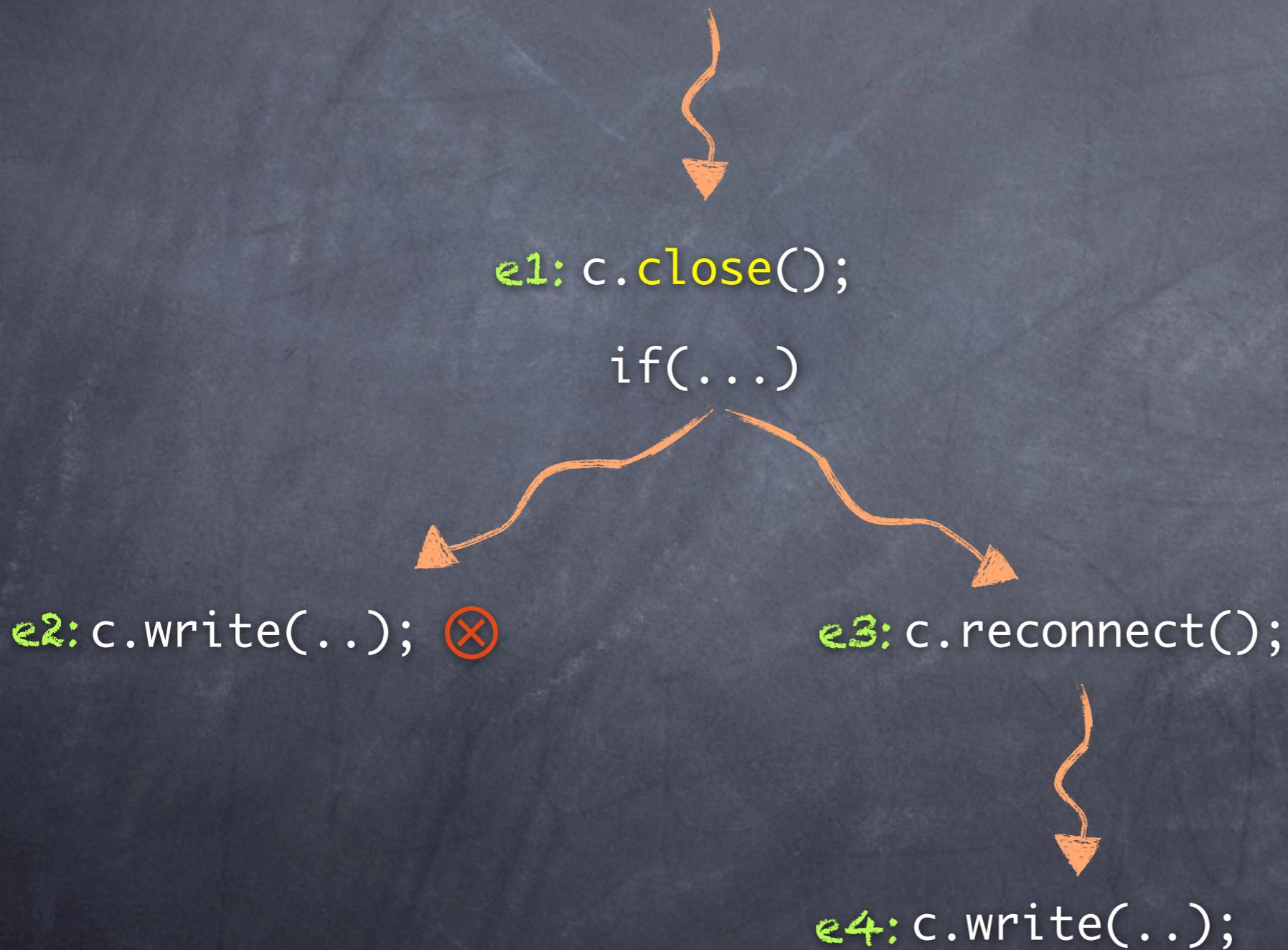


Semantics of Dependency State Machines

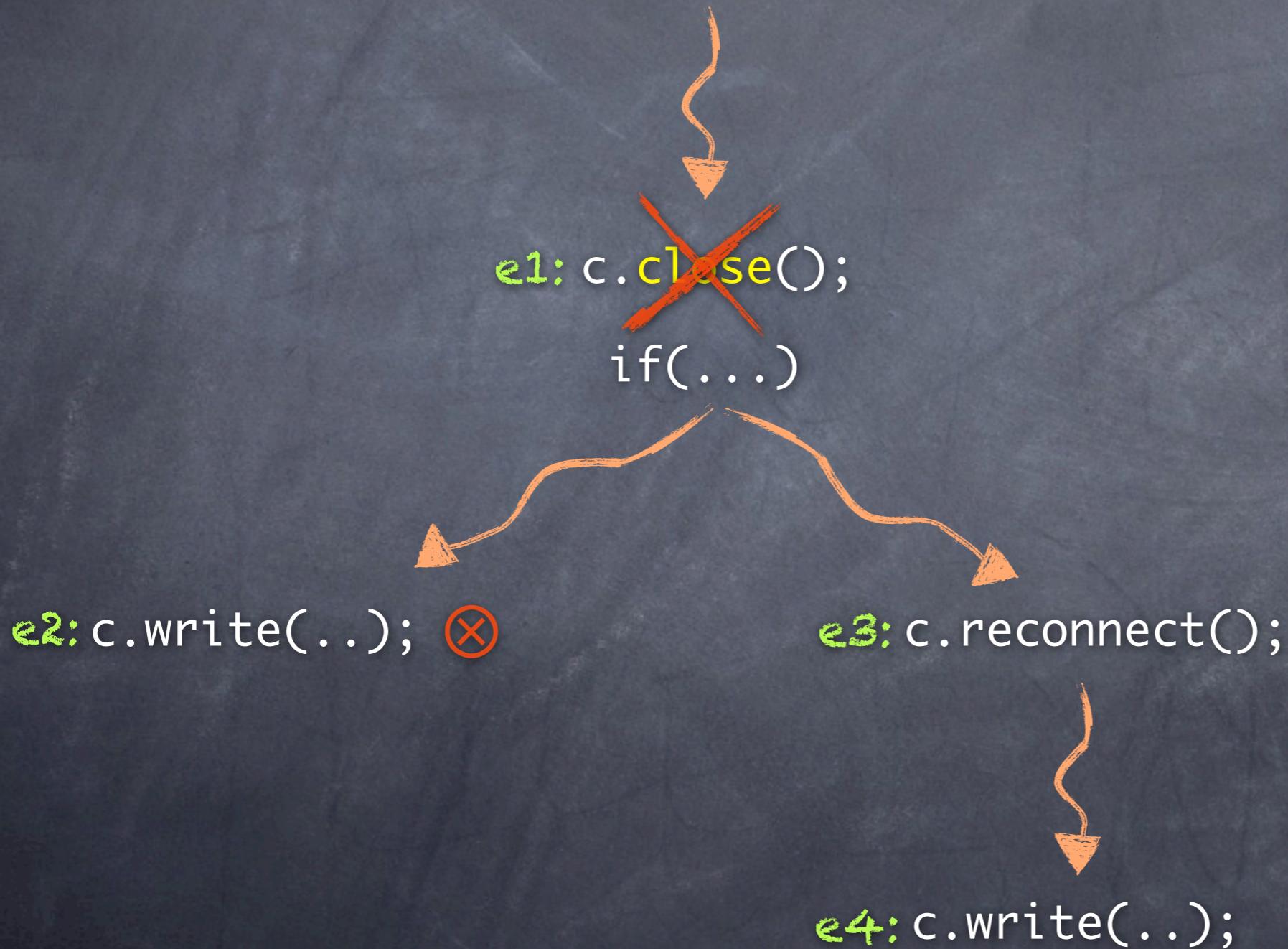


Semantics of Dependency State Machines

advice "close" must
execute at **e1** if

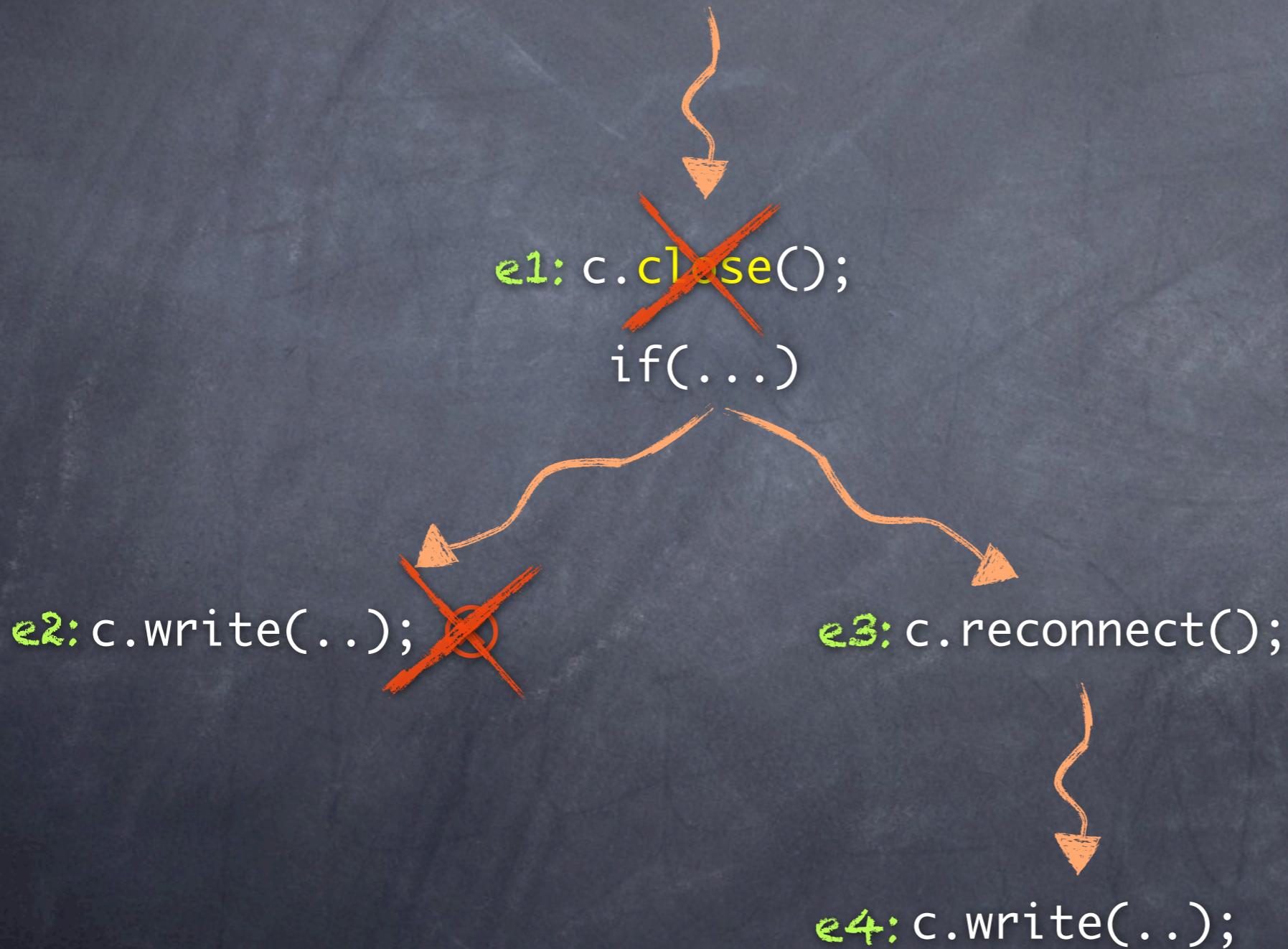


Semantics of Dependency State Machines



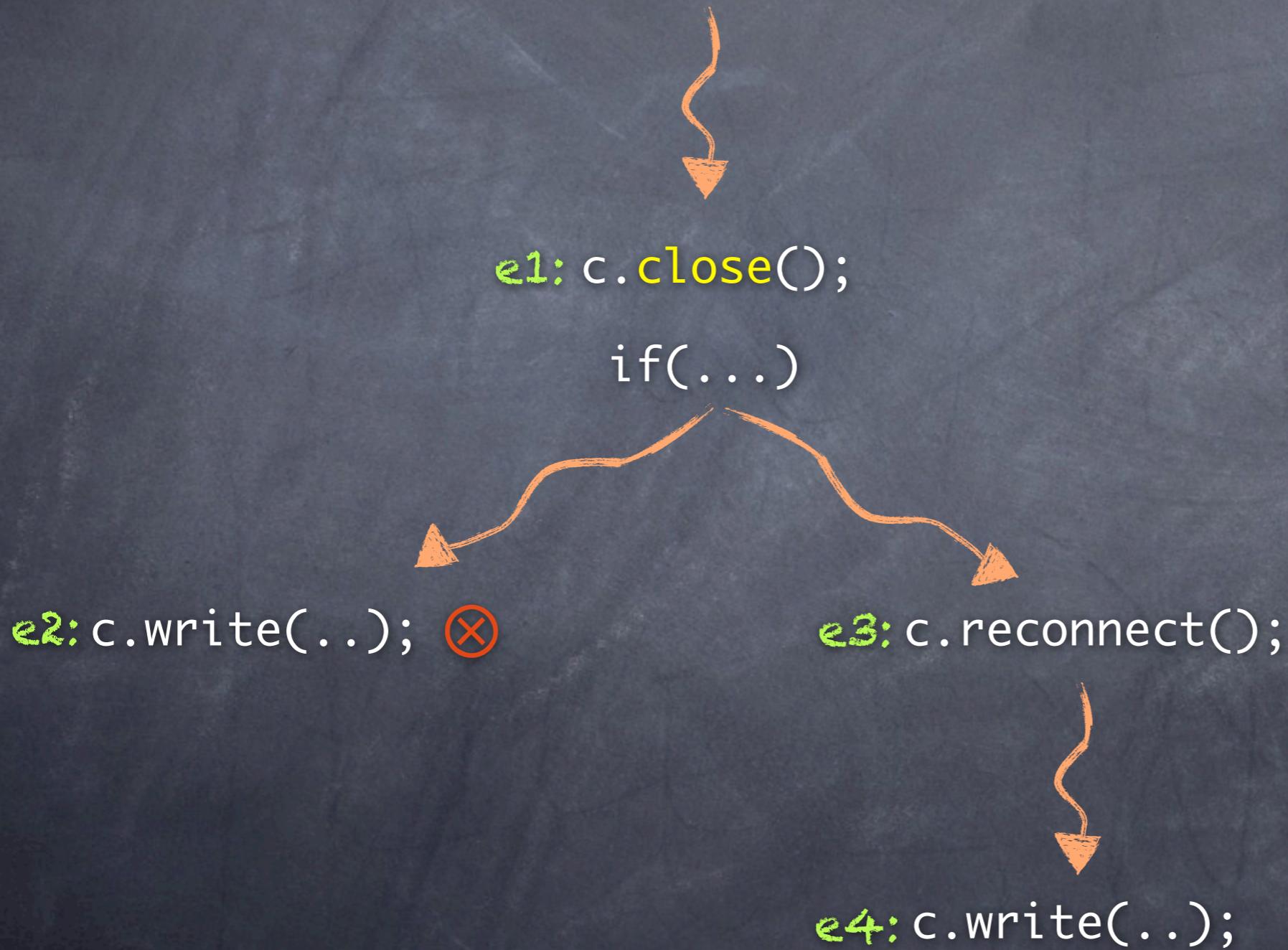
advice "close" must
execute at **e1** if
omitting
"close" at **e1**

Semantics of Dependency State Machines

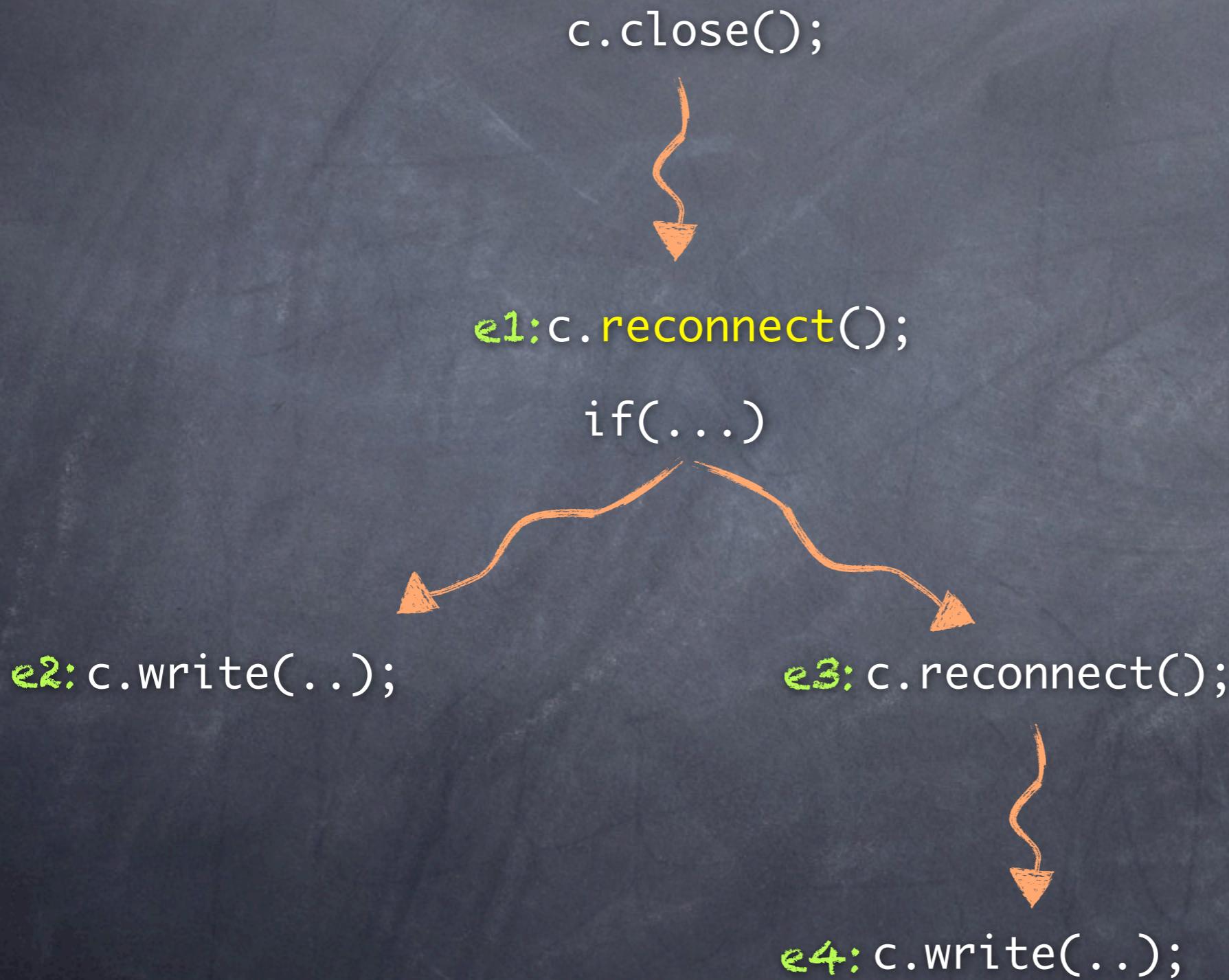


advice "close" must execute at **e1** if omitting "close" at **e1** may change the events at which a DSM reaches an error state

Semantics of Dependency State Machines

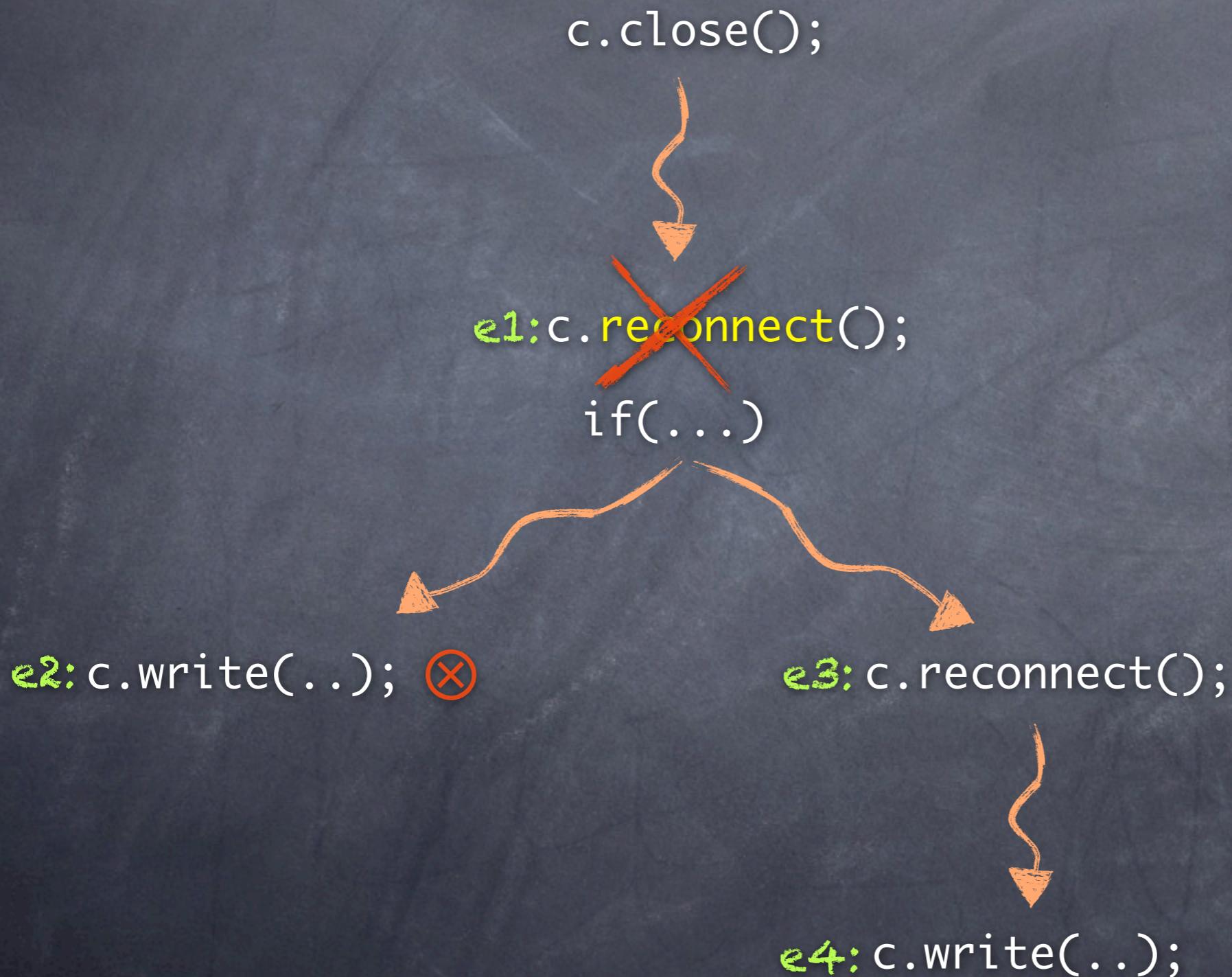


Inverse case: match-preventing events



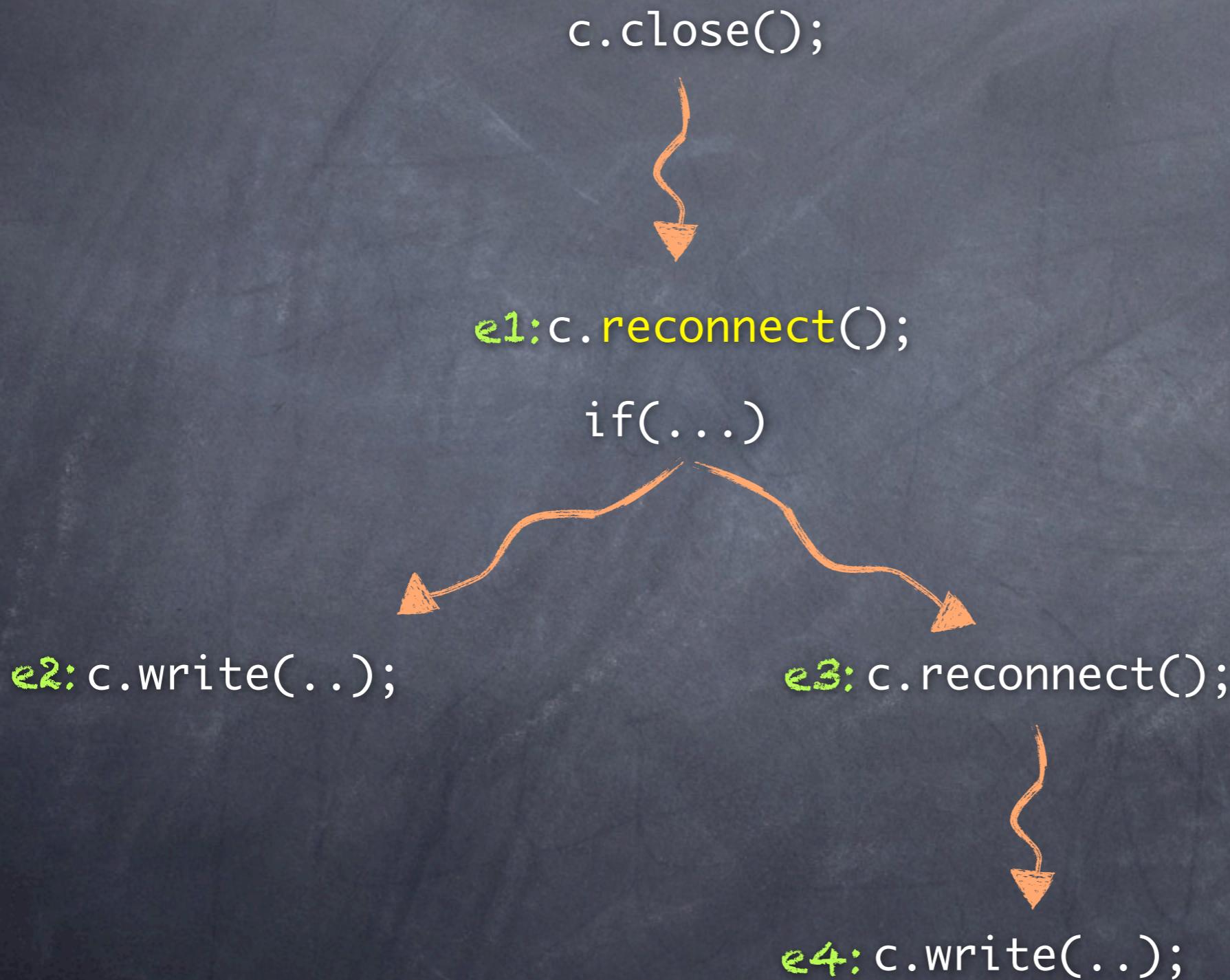
advice “reconnect” must execute at `e1` if omitting “reconnect” at `e1` may change the events at which a DSM reaches an error state

Inverse case: match-preventing events



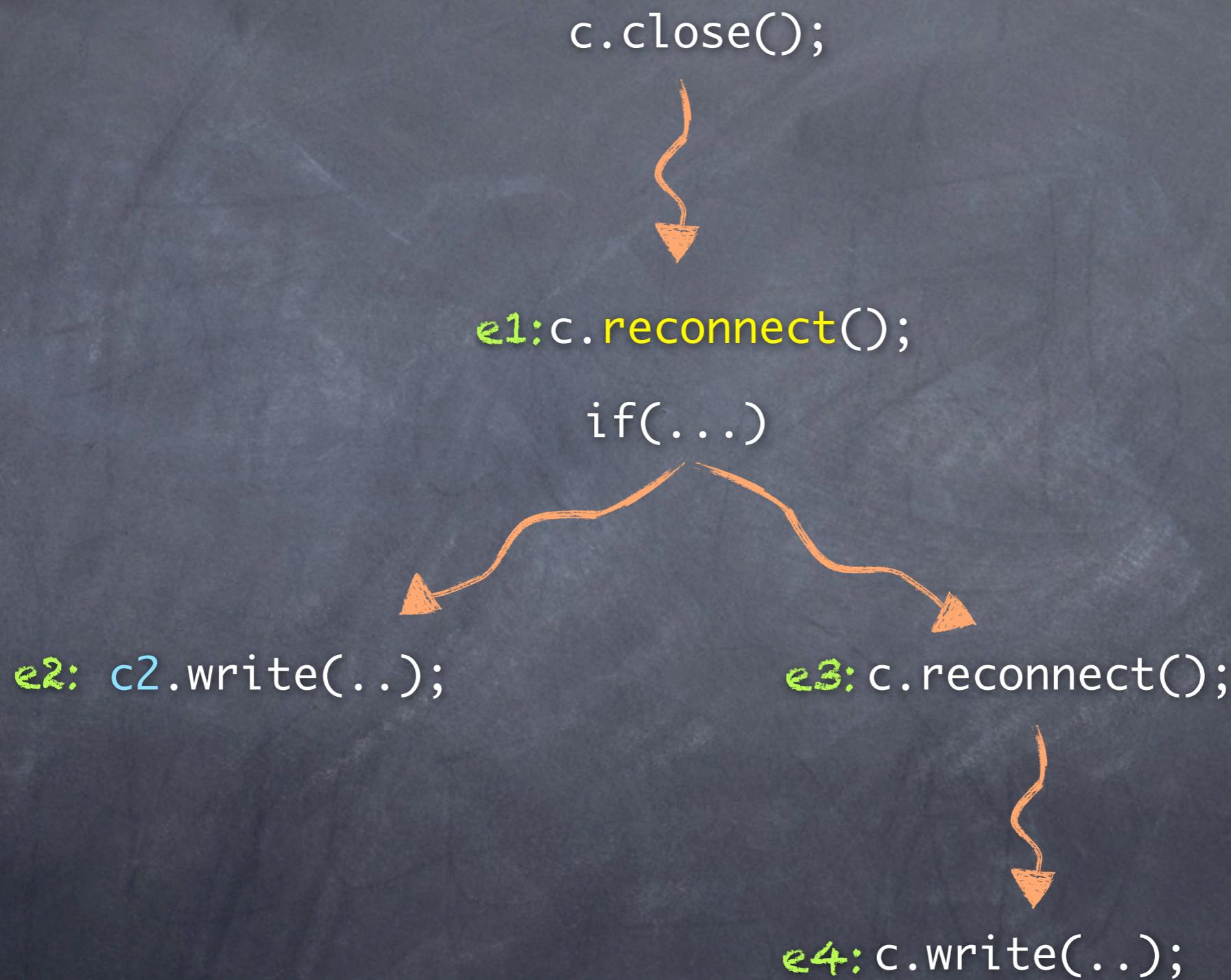
advice “reconnect” must execute at `e1` if omitting “reconnect” at `e1` may change the events at which a DSM reaches an error state

Inverse case: match-preventing events



advice “reconnect” must execute at `e1` if omitting “reconnect” at `e1` may change the events at which a DSM reaches an error state

Variable bindings matter



Variable bindings matter

```
c.close();
```



```
e1:c.reconnect();
```

```
if(...)
```



```
e3:c.reconnect();
```



```
e4:c.write(..);
```

Variable bindings matter

```
c.close();
```

~~e1:c.reconnect();~~

if(...)

e3: c.reconnect();

e4: c.write(..);

AspectJ matching function:

$$match : \mathcal{A} \times \mathcal{J} \rightarrow \{\beta \mid \beta : \mathcal{V} \rightarrow \mathcal{O}\} \cup \{\perp\}.$$

Dependent advice allow family of possible optimized matching functions:

$$stateMatch : \mathcal{A} \times \hat{\mathcal{E}}^* \times \mathbb{N} \rightarrow \{\beta \mid \beta : \mathcal{V} \rightarrow \mathcal{O}\} \cup \{\perp\}$$

$$stateMatch(a, \hat{t}, i) :=$$

let $\beta = match(a, e)$ in

$$\begin{cases} \beta & \text{if } \beta \neq \perp \wedge \exists t \in groundTraces(\hat{t}) \text{ such that } necessaryShadow(a, t, i) \\ \perp & \text{otherwise} \end{cases}$$

Optimization goal: return \perp whenever possible but β whenever necessary

AspectJ matching function:

$$match : \mathcal{A} \times \mathcal{J} \rightarrow \{\beta \mid \beta : \mathcal{V} \rightarrow \mathcal{O}\} \cup \{\perp\}.$$

Dependent advice allow family of possible optimized matching functions:

$$stateMatch : \mathcal{A} \times \hat{\mathcal{E}}^* \times \mathbb{N} \rightarrow \{\beta \mid \beta : \mathcal{V} \rightarrow \mathcal{O}\} \cup \{\perp\}$$

$$stateMatch(a, \hat{t}, i) :=$$

let $\beta = match(a, e)$ in

$$\begin{cases} \beta & \text{if } \beta \neq \perp \wedge \exists t \in groundTraces(\hat{t}) \text{ such that } \\ & \quad necessaryShadow(a, t, i) \\ \perp & \text{otherwise} \end{cases}$$

Optimization goal: return \perp whenever possible but β whenever necessary

$$\text{stateMatch} : \mathcal{A} \times \hat{\mathcal{E}}^* \times \mathbb{N} \rightarrow \{\beta \mid \beta : \mathcal{V} \multimap \mathcal{O}\} \cup \{\perp\}$$
$$\text{stateMatch}(a, \hat{t}, i) :=$$

let $\beta = \text{match}(a, e)$ in

$$\begin{cases} \beta & \text{if } \beta \neq \perp \wedge \exists t \in \text{groundTraces}(\hat{t}) \text{ such that } \text{necessaryShadow}(a, t, i) \\ \perp & \text{otherwise} \end{cases}$$

Any **sound** optimization must satisfy:

$$\forall t = t_1 \dots t_i \dots t_n \in \Sigma^+. \forall i \leq n \in \mathbb{N}.$$
$$\begin{aligned} \text{matches}_{\mathcal{L}(\mathcal{M})}(t_1 \dots \underline{t_{i-1} t_i t_{i+1}} \dots t_n) &\neq \text{matches}_{\mathcal{L}(\mathcal{M})}(t_1 \dots \underline{t_{i-1} t_{i+1}} \dots t_n) \\ \implies \text{necessaryShadow}(t_i, t, i) \end{aligned}$$

More on the semantics

CLARA also supports Collaborative Runtime Verification, which distributes instrumentation overhead among multiple users; and ranking heuristics, which aid programmers in inspecting remaining instrumentation manually [13] [2, Ch. 6 & 7]. Space limitations preclude us from discussing ranking and Collaborative Runtime Verification here.

CLARA is freely available as free software at <http://bodden.cs/CLARA/>, along with extensive documentation, the first author's dissertation [2], which describes CLARA in detail, and benchmarks and benchmark results.

We next describe the syntax and semantics of Dependency State Machines, the key abstraction of CLARA. This abstraction allows CLARA to decouple runtime monitor implementations from static analyses.

3 Syntax and Semantics of Dependency State Machines

Dependency State Machines extend the AspectJ language to include semantic information about relationships between different pieces of advice. Runtime verification tools which generate AspectJ aspects can use this extension to produce augmented aspects. CLARA can reason about the augmented aspects to prove that programs never violate monitored properties or to generate optimized code.

3.1 Syntax

Our extensions modify the AspectJ grammar in two ways: they add syntax for defining Dependent Advice [14] and Dependency State Machines. The idea of Dependent Advice is that pieces of monitoring advice are often inter-dependent in the sense that the execution of one piece of advice only has an effect when executing before or after another piece of advice, on the same objects. Dependency State Machines allow programmers to make these dependencies explicit so that static analyses can exploit them. Our explanations below refer to the ConnectionClosed example in Figure 2.

The `dependent` modifier flags advice to CLARA for potential optimization; such advice may be omitted from program locations at which it provably has no effect on the state of the runtime monitor. Dependent advice must be named. Lines 4, 7 and 10 all define dependent advice.

The Dependency State Machines extension enables users to specify state machines which relate different pieces of dependent advice. Dependency State Machine declarations define state machines by including a list of edges between states and an alphabet; each edge is labelled with a member of the alphabet. CLARA infers the set of states from the declared edges. Line 16 declares the state machine's alphabet: `{disconn, write, reconn}`. Every symbol in the alphabet references dependent advice from the same aspect. Lines 17–19 enumerate, for each state, a (potentially empty) list of outgoing transitions. An entry `"s1: t -> s2"` means "there exists a `t`-transition from `s1` to `s2`". Users can also mark states as initial or final (error states). Final states denote states

VERIFYING FINITE-STATE PROPERTIES
OF LARGE-SCALE PROGRAMS

by
Eric Bodden

School of Computer Science
McGill University, Montréal

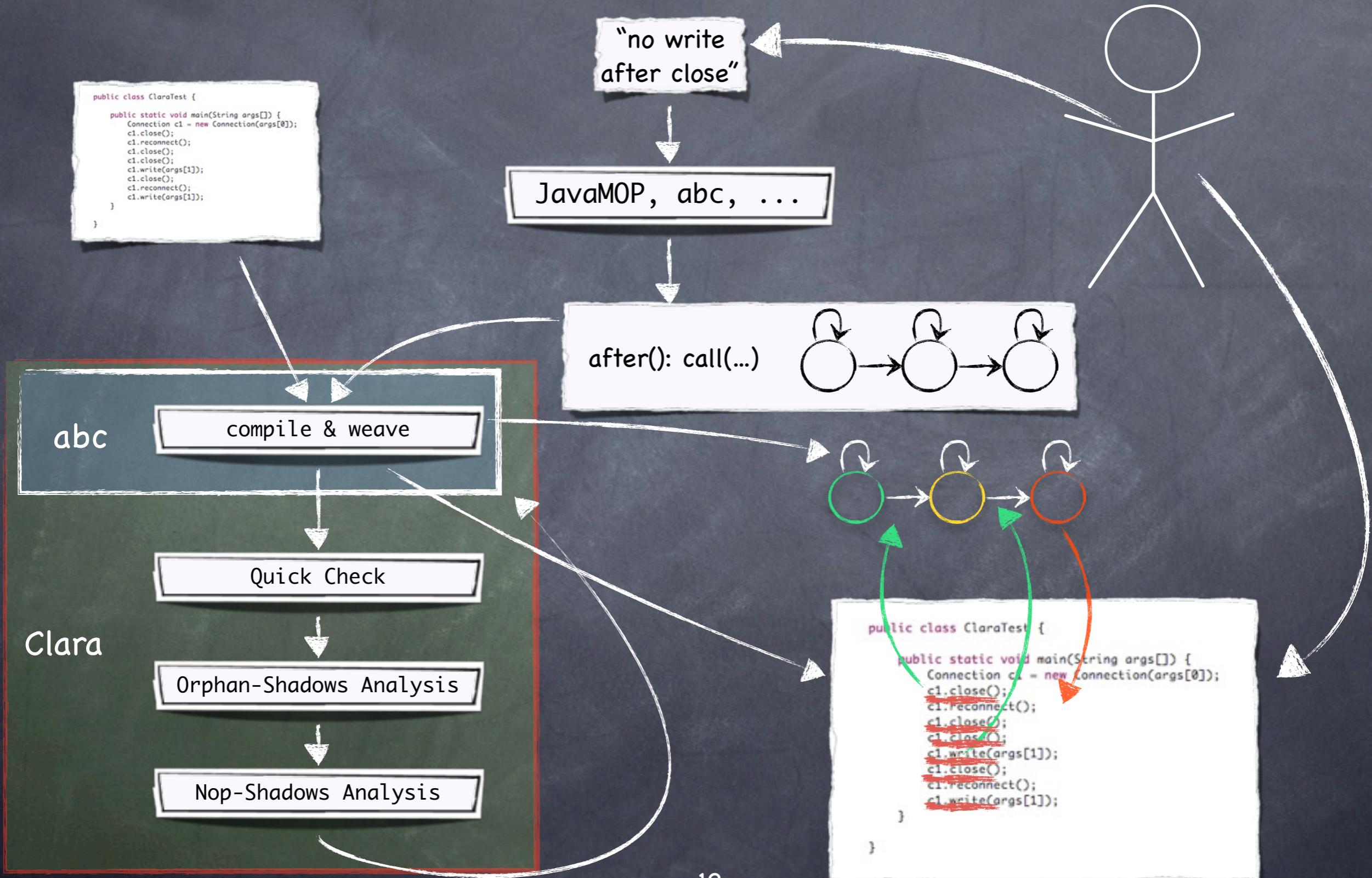
June 2009

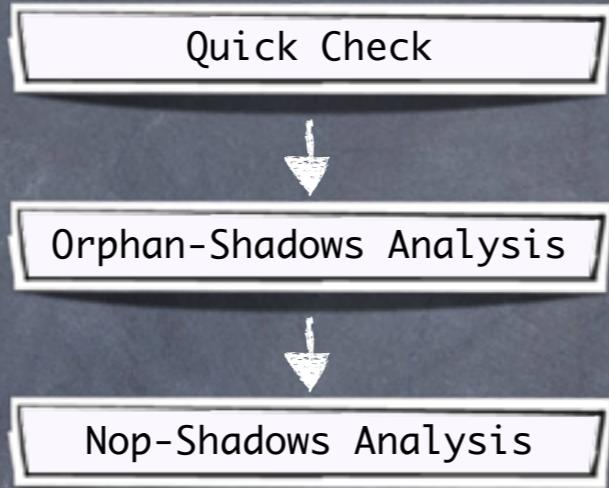
A THESIS SUBMITTED TO MCGILL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF
DOCTOR OF PHILOSOPHY

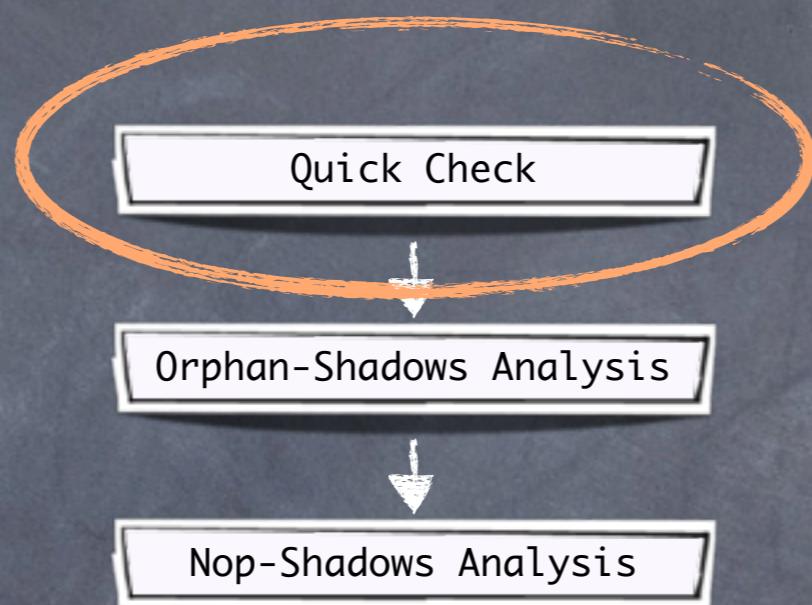
Copyright © 2009 Eric Bodden

This Paper
+ upcoming Journal Paper
Thesis:
proves that analyses
obey "necessaryShadow"

The Clara Framework



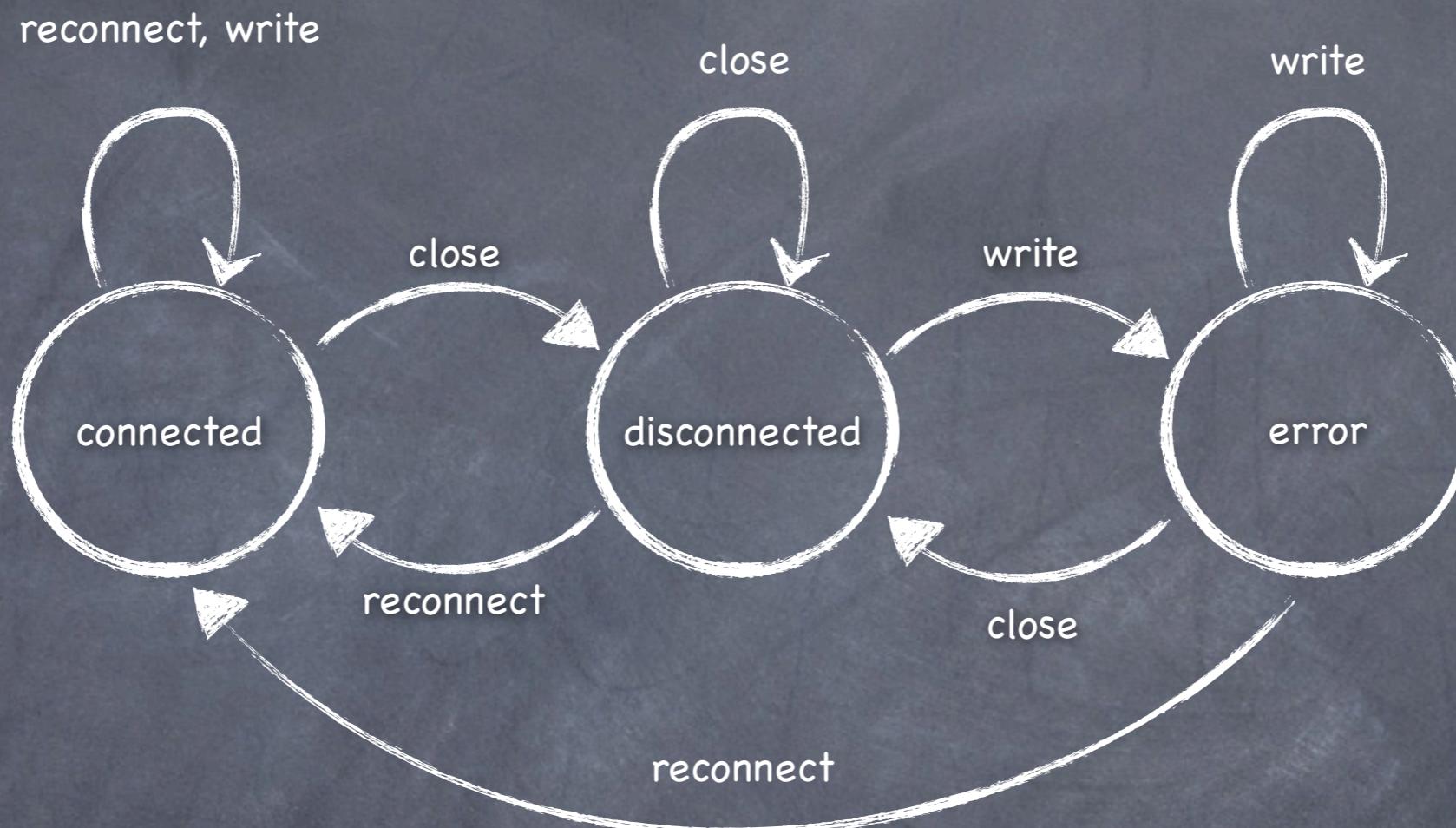




Quick Check

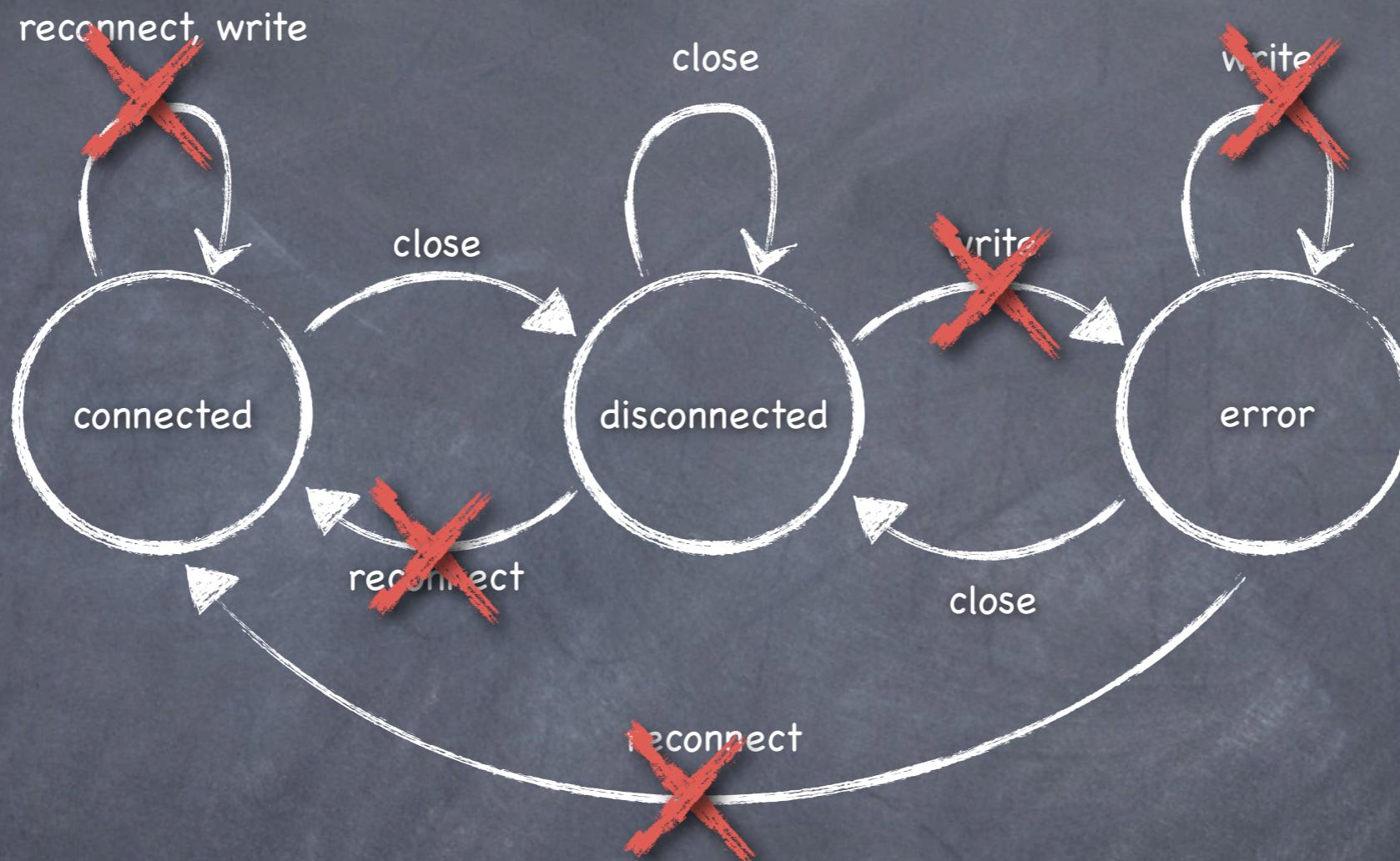
```
public static void main(String[] args) {  
    Connection c = new Connection();  
    while(c.hasMoreData()) {  
        System.err.println(c.read());  
    }  
    c.close();  
}
```

Quick Check



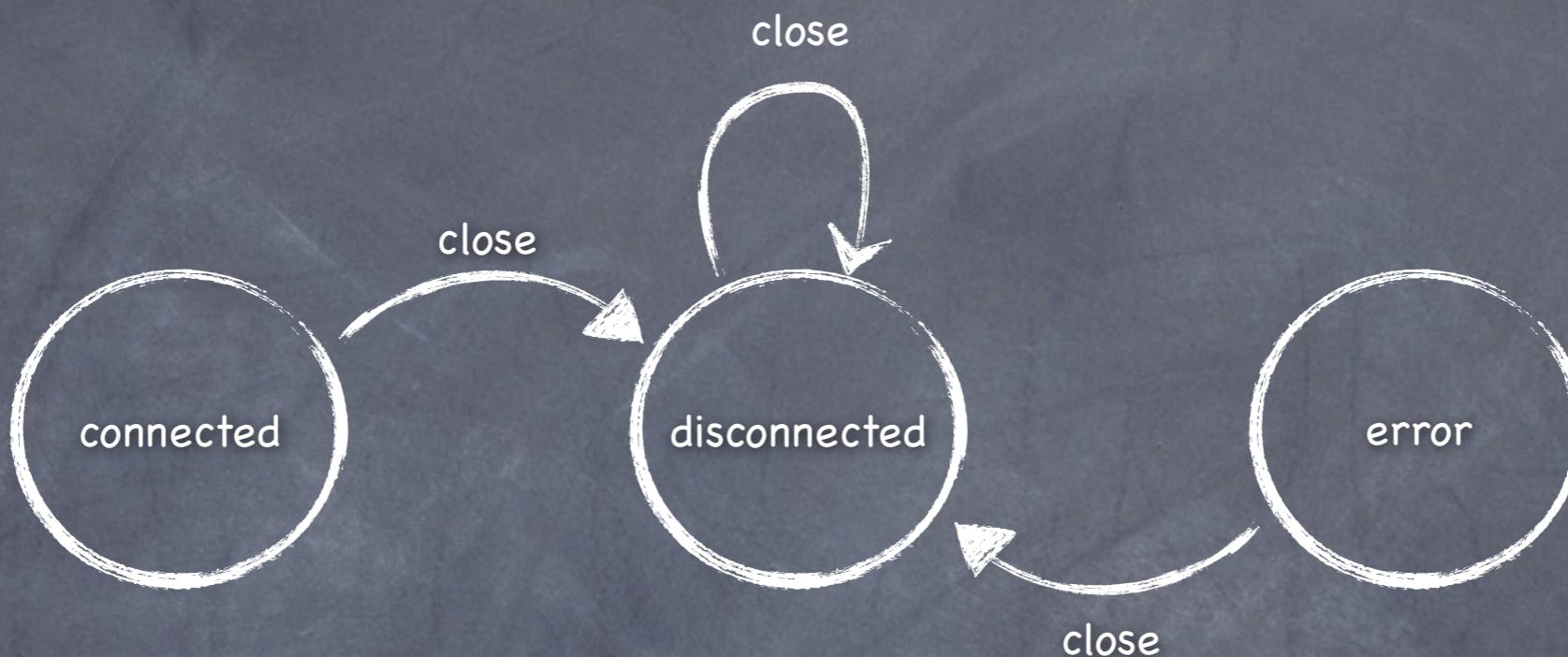
```
public static void main(String[] args) {  
    Connection c = new Connection();  
    while(c.hasMoreData()) {  
        System.err.println(c.read());  
    }  
    c.close();  
}
```

Quick Check



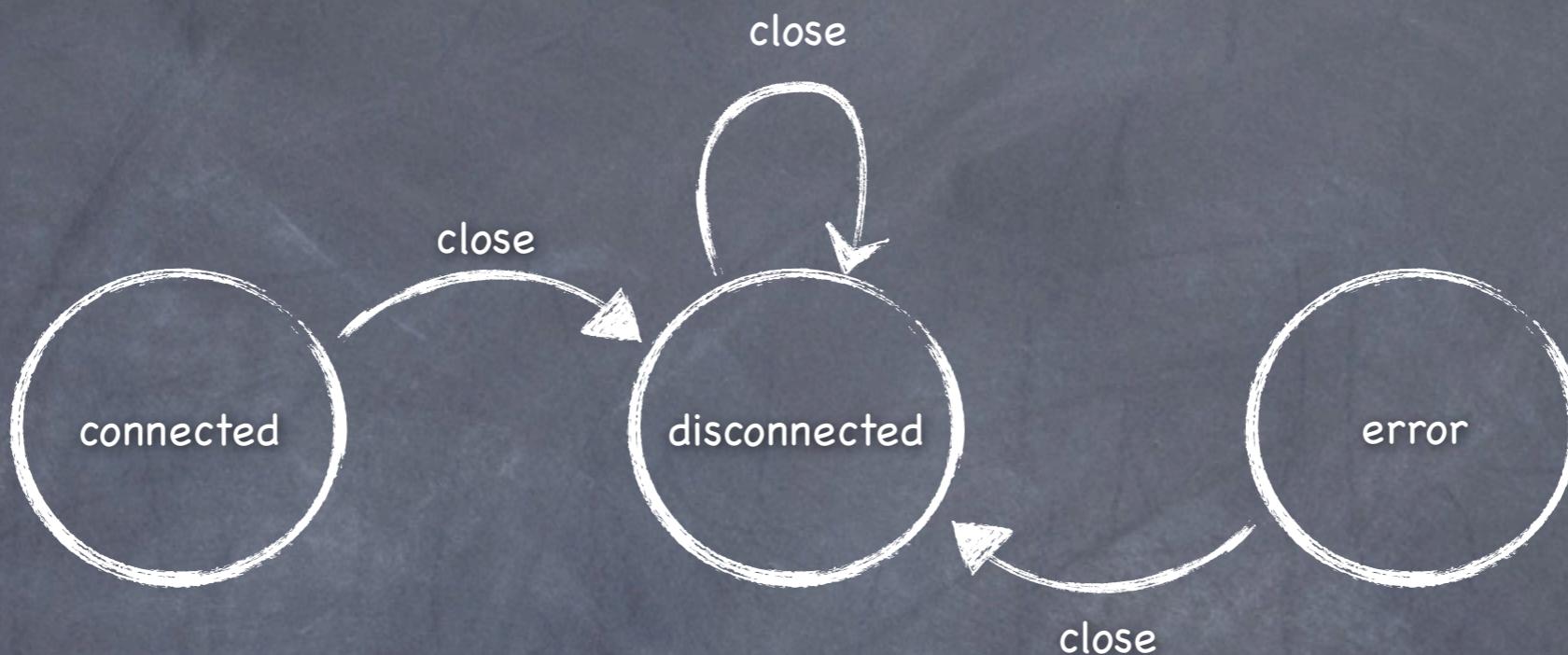
```
public static void main(String[] args) {  
    Connection c = new Connection();  
    while(c.hasMoreData()) {  
        System.err.println(c.read());  
    }  
    c.close();  
}
```

Quick Check



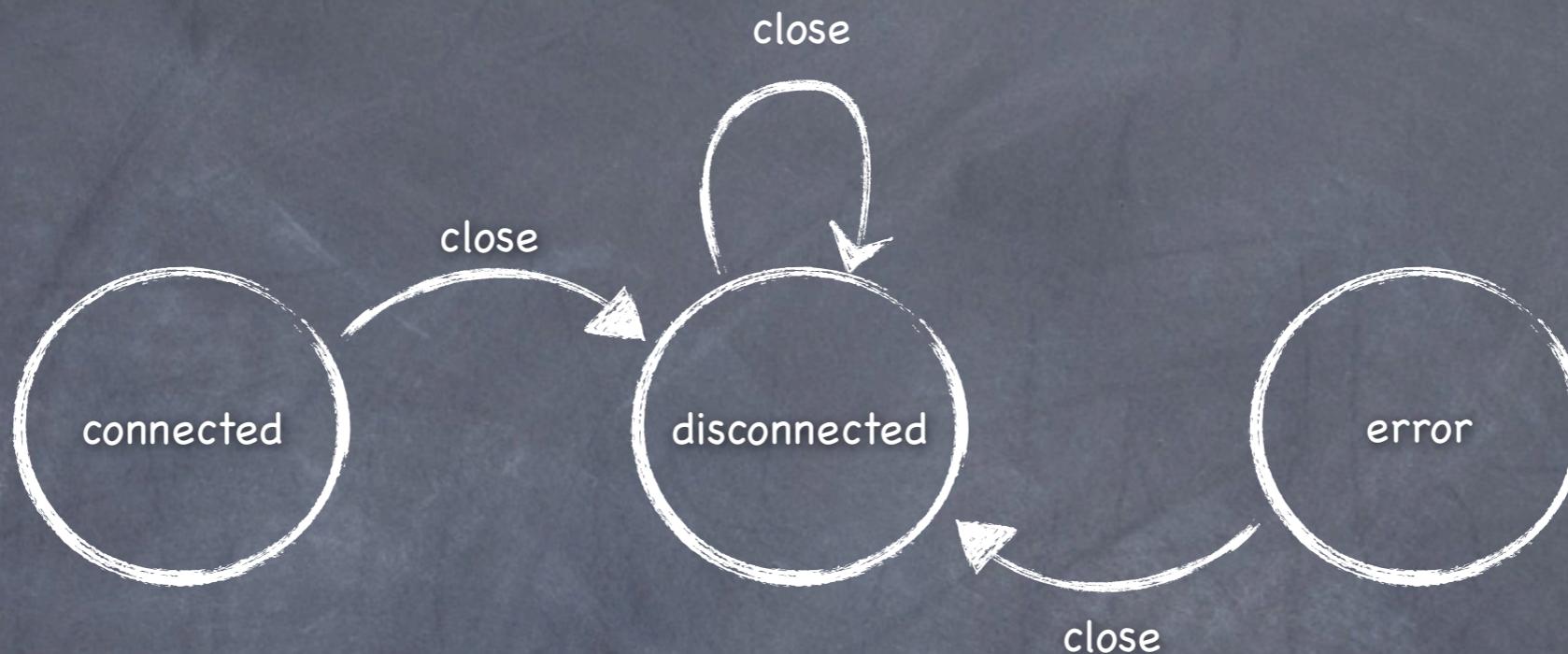
```
public static void main(String[] args) {  
    Connection c = new Connection();  
    while(c.hasMoreData()) {  
        System.err.println(c.read());  
    }  
    c.close();  
}
```

Quick Check



```
public static void main(String[] args) {  
    Connection c = new Connection();  
    while(c.hasMoreData()) {  
        System.err.println(c.read());  
    }  
    c.close();  
}
```

Quick Check



```
public static void main(String[] args) {  
    Connection c = new Connection();  
    while(c.hasMoreData()) {  
        System.err.println(c.read());  
    }  
    c.close();  
}
```

Proving QuickCheck sound

necessaryTransitionQC(a, t, i) := $a \in \text{symbolsThatNeedMonitoring}.$

symbolsThatNeedMonitoring:

- \emptyset if QuickCheck succeeds
- Σ otherwise

$\forall t = t_1 \dots t_i \dots t_n \in \Sigma^+. \forall i \leq n \in \mathbb{N}.$

$\text{matches}_{\mathcal{L}(\mathcal{M})}(t_1 \dots t_{i-1} t_i t_{i+1} \dots t_n) \neq \text{matches}_{\mathcal{L}(\mathcal{M})}(t_1 \dots t_{i-1} t_{i+1} \dots t_n)$
 $\implies \text{necessaryShadow}(t_i, t, i)$

Proving QuickCheck sound

necessaryTransitionQC(a, t, i) := $a \in \text{symbolsThatNeedMonitoring}$.

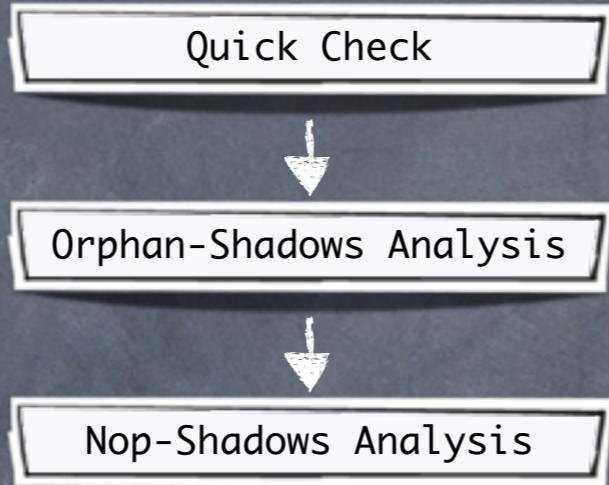
symbolsThatNeedMonitoring:

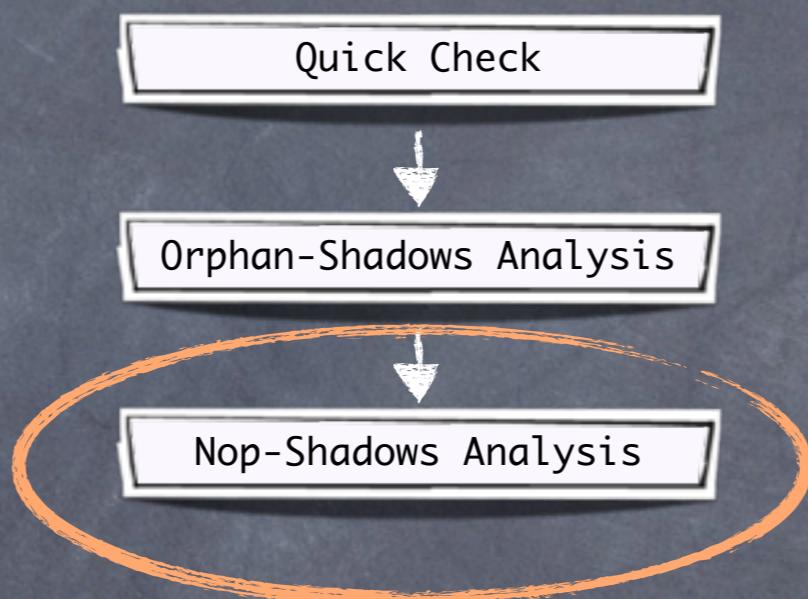
- \emptyset if QuickCheck succeeds
- Σ otherwise

$\forall t = t_1 \dots t_i \dots t_n \in \Sigma^+. \forall i \leq n \in \mathbb{N}.$

$\text{matches}_{\mathcal{L}(\mathcal{M})}(t_1 \dots t_{i-1} t_i t_{i+1} \dots t_n) \neq \text{matches}_{\mathcal{L}(\mathcal{M})}(t_1 \dots t_{i-1} t_{i+1} \dots t_n)$

$\implies a \in \text{symbolsThatNeedMonitoring}$





More details...

Efficient Hybrid Typestate Analysis by Determining Continuation-Equivalent States (Eric Bodden)

In ICSE '10: International Conference on Software Engineering, pages 5–14, ACM, 2010.

Efficient Hybrid Typestate Analysis by Determining Continuation-Equivalent States

Eric Bodden^{*}
Software Technology Group
Department of Computer Science
Technische Universität Darmstadt, Germany
bodden@acm.org

ABSTRACT

Typestate analysis determines whether a program violates a set of finite-state properties. Because the typestate-analysis problem is statically undecidable, researchers have proposed a hybrid approach that uses residual monitors to signal property violations at runtime.

We present an efficient novel static typestate analysis that is flow-sensitive, partially context-sensitive, and that generates residual runtime monitors. To gain efficiency, our analysis uses precise, flow-sensitive information on an intra-procedural level only, and models the remainder of the program using a flow-insensitive pointer abstraction. Unlike previous flow-sensitive analyses, our analysis uses an additional backward analysis to partition states into equivalence classes. Code locations that transition between equivalent states are irrelevant and require no monitoring. As we show in this work, this notion of equivalent states is crucial to obtaining sound runtime monitors.

We proved our analysis correct, implemented the analysis in the CLARA framework for typestate analysis, and applied it to the DuCapo benchmark suite. In half of the cases, our analysis determined exactly the property-violating program points. In many other cases, the analysis reduced the number of instrumentation points by large amounts, yielding significant speed-ups during runtime monitoring.

Categories and Subject Descriptors

D.4 [Software Engineering]: Software/Program Verification—Validation

General Terms

Algorithms, Experimentation, Performance, Verification

Keywords

typestate analysis, static analysis, runtime monitoring

^{*}Eric conducted most of this research as a Ph.D. student at McGill University, under supervision of Laurie Hendren. This work was supported by NSERC and CASED (www.cased.de).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICSE '10, May 2–8 2010, Cape Town, South Africa
Copyright 2010 ACM 978-1-60558-719-6/10/05 ...\$10.00.

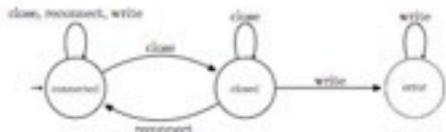


Figure 1: Finite-state machine for “Connection” property

1. INTRODUCTION

A typestate property [22] describes which operations are available on an object or even a group of inter-related objects, depending on this object’s or group’s internal state, the typestate. For instance, programmers must not write to a connection handle that is currently in its “closed” state. Figure 1 shows a non-deterministic finite-state machine for this property. It monitors a connection’s “close”, “reconnect” and “write” events and signals an error at its accepting state.

Typestate properties aid program understanding, and one can even define type systems [5, 14] that prevent programmers from causing typestate errors, or derive static typestate analyses [17] that try to determine whether a given program violates typestate properties. Unfortunately, the typestate-analysis problem is generally undecidable. Researchers have therefore proposed a hybrid approach [9, 10, 16] that uses static-analysis results to generate a residual runtime monitor. This monitor captures actual property violations as they occur, but only updates its internal state at relevant statements, as determined through static analysis.

A correct runtime monitor must observe events like “close” and “write” that can cause a property violation, but also events like “reconnect” that may prevent the violation from occurring. Missing the former causes false negatives while missing the latter causes false positives, i.e., false warnings. Either is unacceptable, as runtime monitors must handle property violations exactly when they occur. A correct static analysis must therefore determine program locations that can trigger either kind of such “relevant” events.

In this work we present an efficient novel static typestate-analysis algorithm called Nep-shadows Analysis¹ that uses a forward and a backward pass to identify provably irrelevant code locations. For every program statement s of interest, the forward analysis determines the possible typestates that can reach s . The additional backward analysis partitions these states into equivalence classes. A program location

¹The aspect-oriented-programming community uses the term “shadow” [18] to refer to instrumentation points.

Implementation

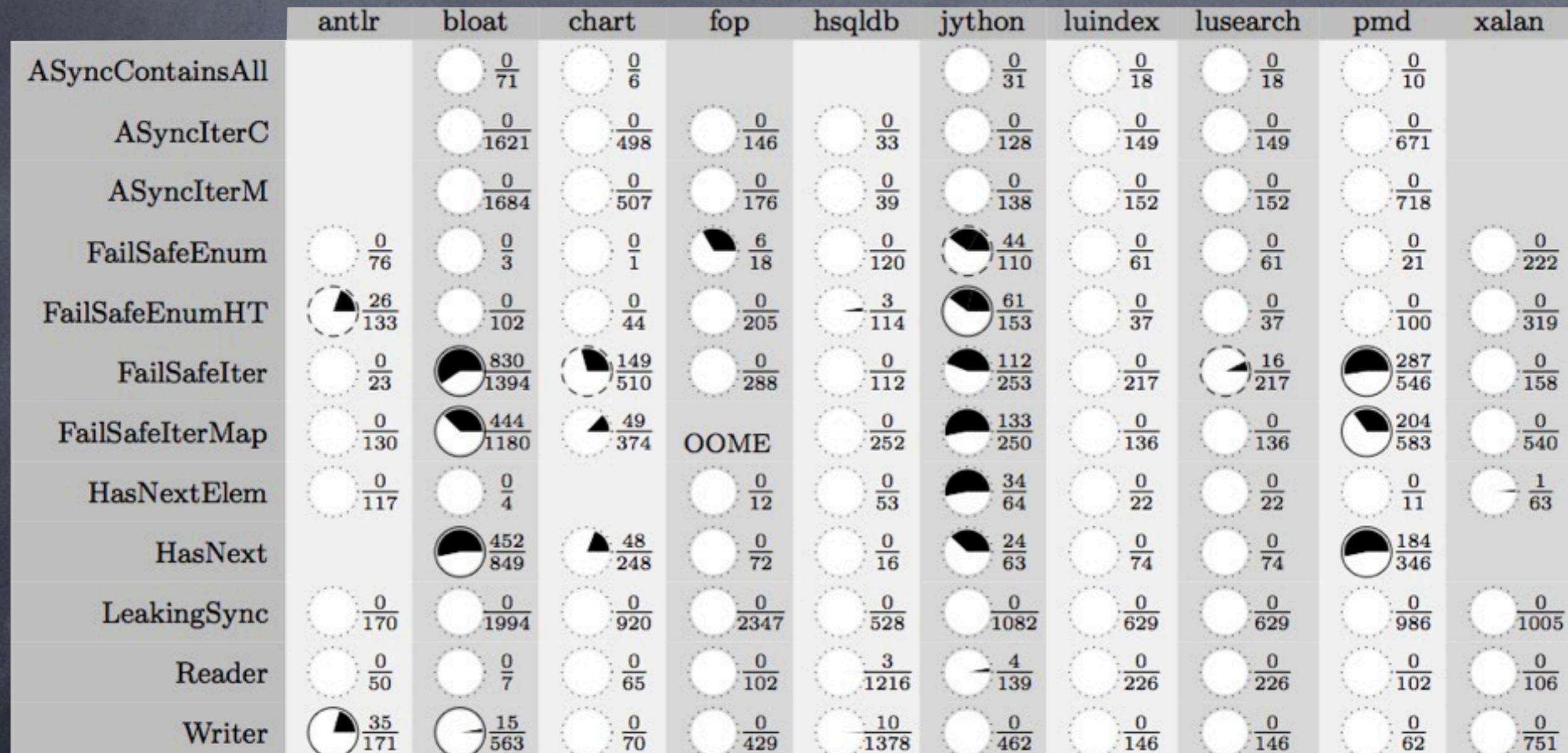
- Clara is an extension to the AspectBench Compiler www.aspectbench.org
 - Builds on Soot Program-Analysis Framework www.sable.mcgill.ca/soot/
 - Is extensible: may implement your own static analysis
 - Can partially evaluate any* AspectJ-based runtime monitor
- *DSM annotation required

Extending Clara

The screenshot shows the Clara project website. At the top, there's a navigation bar with links for Overview, Documentation, Mailing List, Downloads (which is highlighted), Benchmarks, Publications, and Contact. Below the navigation is a banner with the text "Finding bugs through Hybrid Typestate Analysis" and a small image of a caterpillar on a leaf. The main content area is titled "Downloads". It starts with a note that Clara is hosted on Google Code. There are two download options: "clara-1.0.0-complete.jar" (described as containing Clara and all its dependencies in compiled form) and "clara-1.0.0-emptyext.zip" (described as an empty extension for researchers). Both download links are preceded by a green downward arrow icon.

<http://www.bodden.de/clara/downloads/>

Overall success

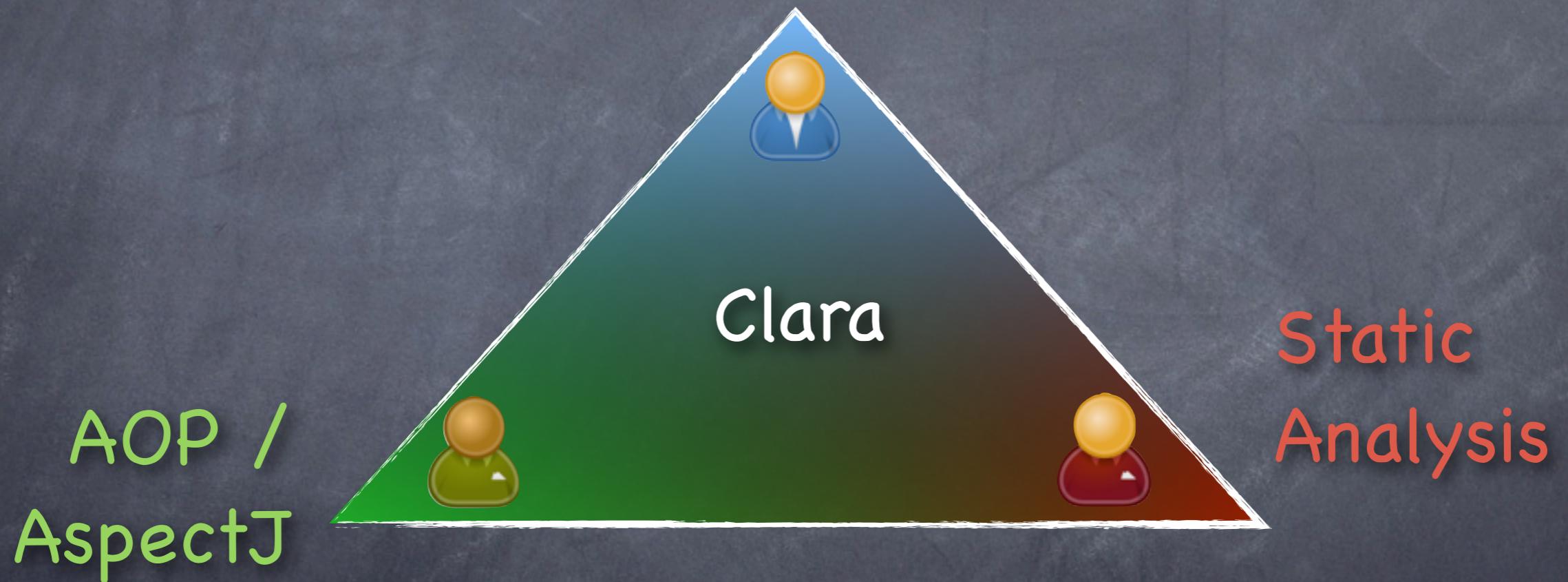


Why so effective?

- Very precise abstractions:
 - Resolve aliasing using three different alias analyses (some context sensitive, others flow sensitive)
- Analysis is path sensitive
- Program properties:
 - Most objects only accessed in few methods
 - Most programs are mostly correct!

Related and previous work

Runtime Verification



Related and previous work

Runtime Verification

Tracecuts (Walker & Viggers, FSE 04)

PTQL (Goldsmith et al., OOPSLA 05)

PQL (Martin et al., OOPSLA 05)

Larva (Colombo et al., SEFM 09)

M2Aspects (Krüger et al., SCESM 06)

S2A (Maoz & Harel, FSE 06)

J-LO (Bodden & Stolz, RV 05)

JavaMOP (Chen et al., OOPSLA 07)

Tracematches (Allan et al., OOPSLA 05)

AOP /
AspectJ

Clara

Hybrid Race Detection with AJ
(Bodden & Havelund, ISSTA 08)

Static
Analysis

Static Typestate Analysis (Fink et al., ISSTA 06)

SCoPE (Aotani & Masuhara, AOSD 07)

Optimizing cflow (Avgustinov et al., PLDI 05)

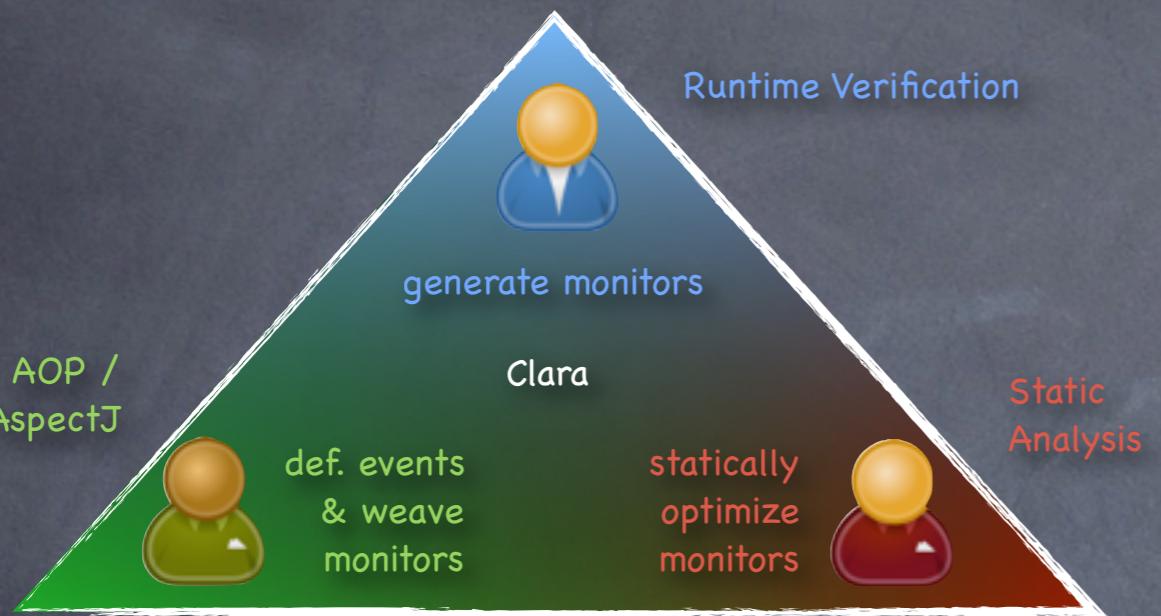
Statically optimizing tracematches
(Naeem & Lhotak, OOPSLA 08; Bodden et al. ECOOP 07, FSE 08)

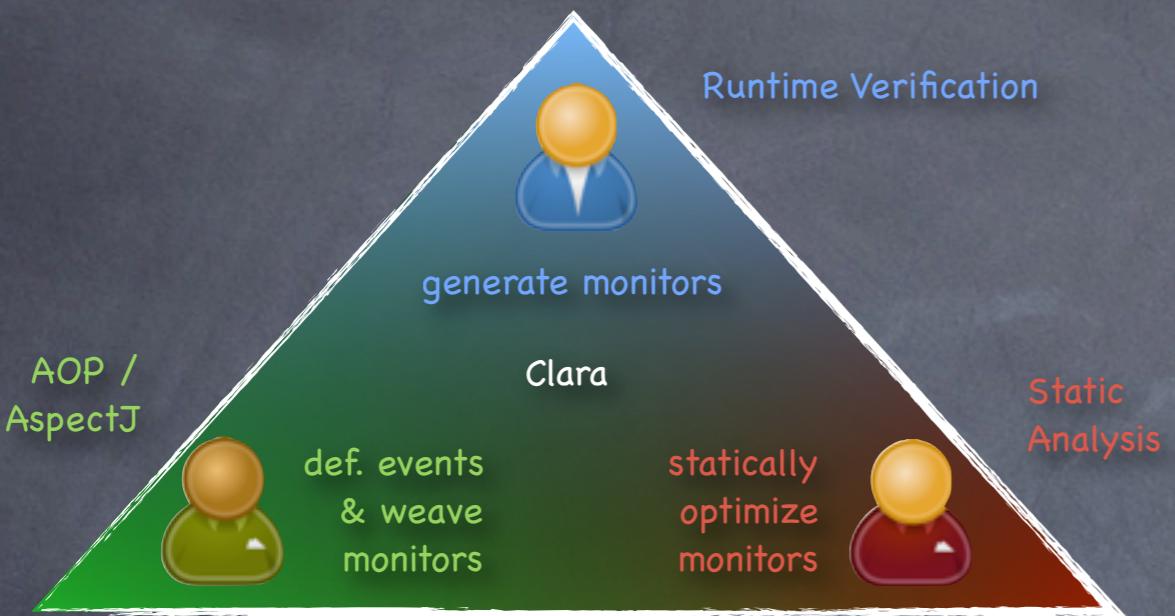
Related and previous work



More:

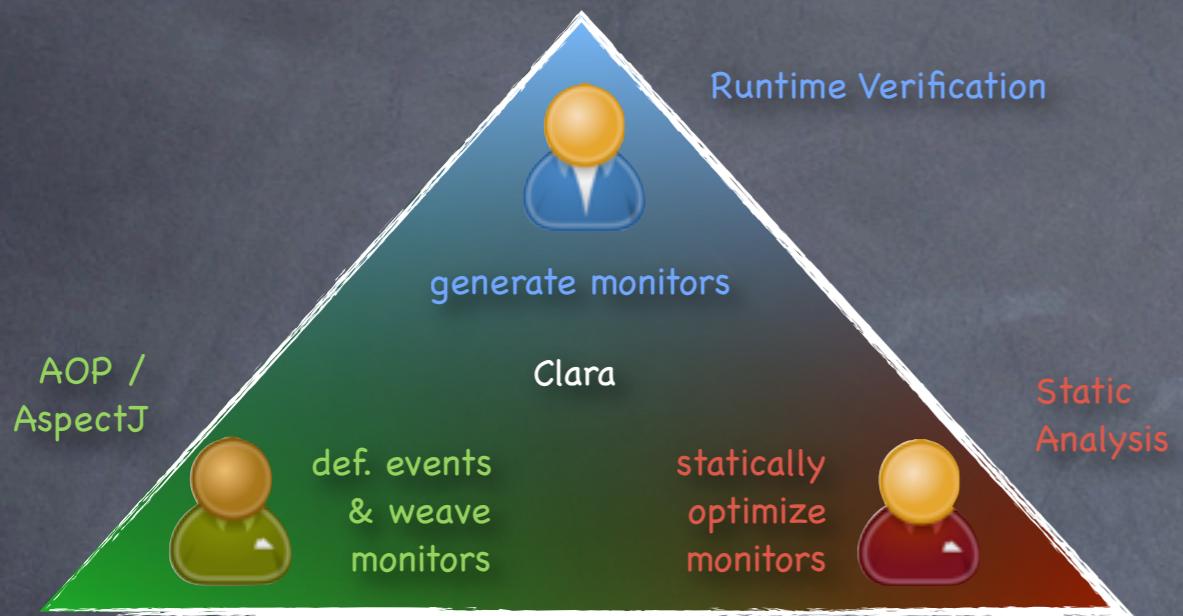
Tutorial paper,
pages 183–197





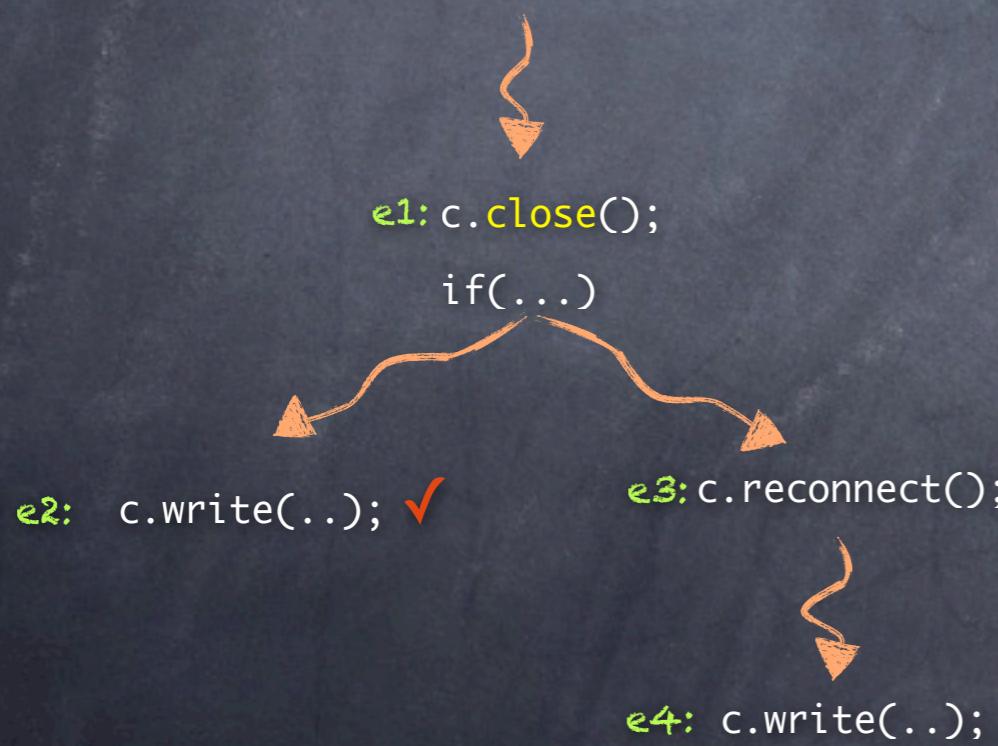
```

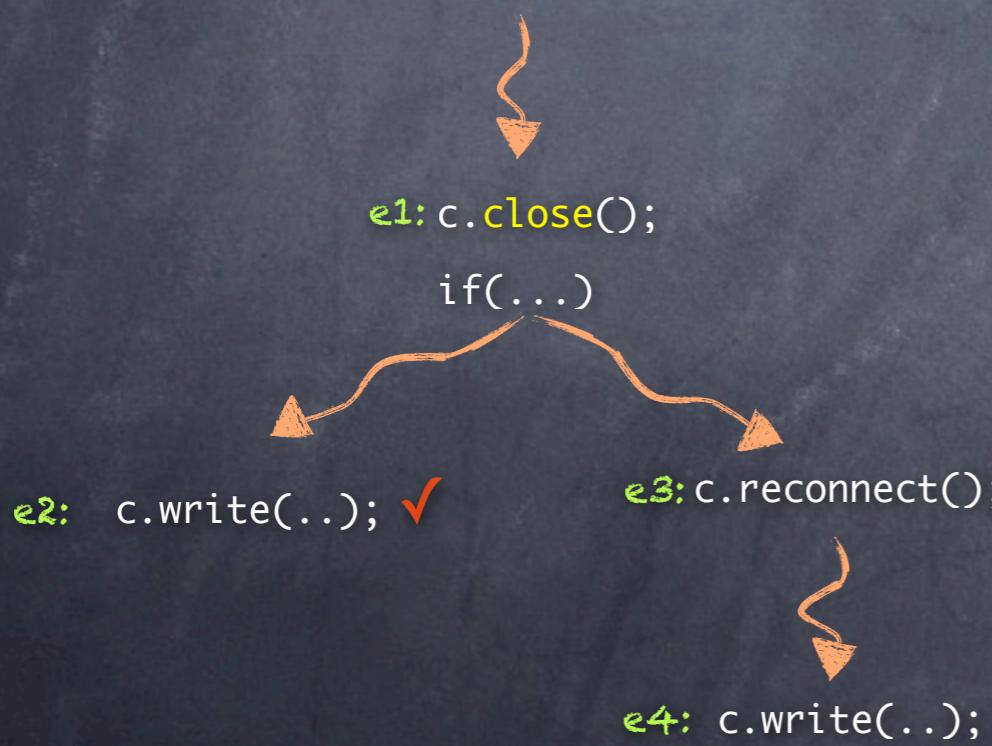
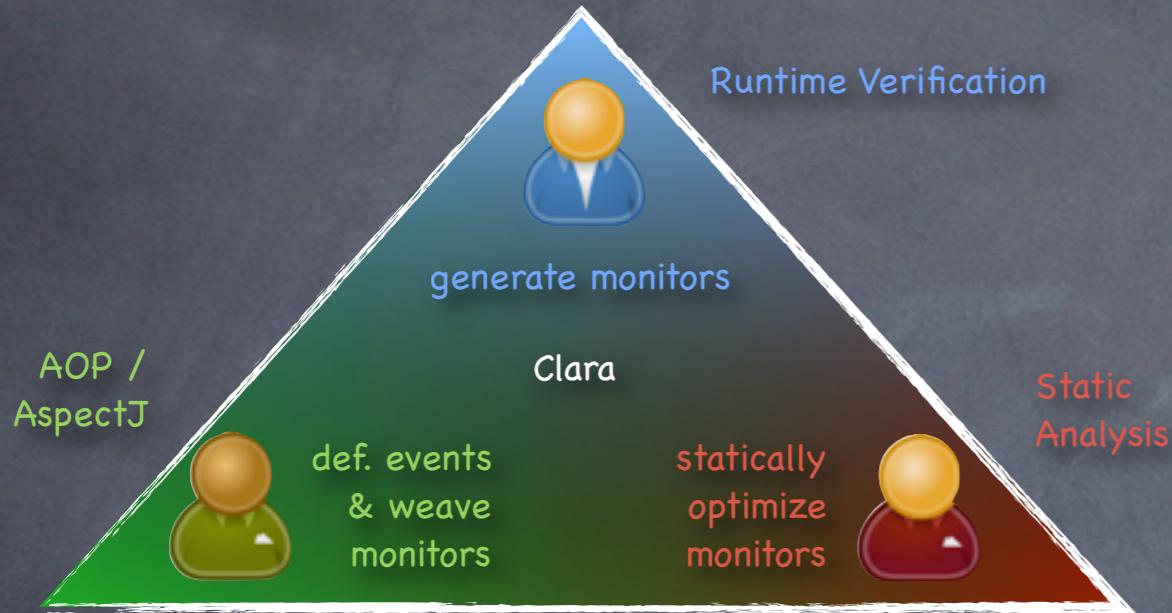
dependency{
    disconnect, write, reconnect;
initial connected: disconnect -> connected,
    write -> connected,
    reconnect -> connected,
    disconnect -> disconnected;
    disconnect: disconnect -> disconnected,
    write -> error;
error: write -> error;
}
final
}
```



```

dependency{
    disconnect, write, reconnect;
    initial connected: disconnect -> connected,
        write -> connected,
        reconnect -> connected,
        disconnect -> disconnected;
    disconnect: disconnect -> disconnected,
        write -> error;
    error: write -> error;
}
final
}
```

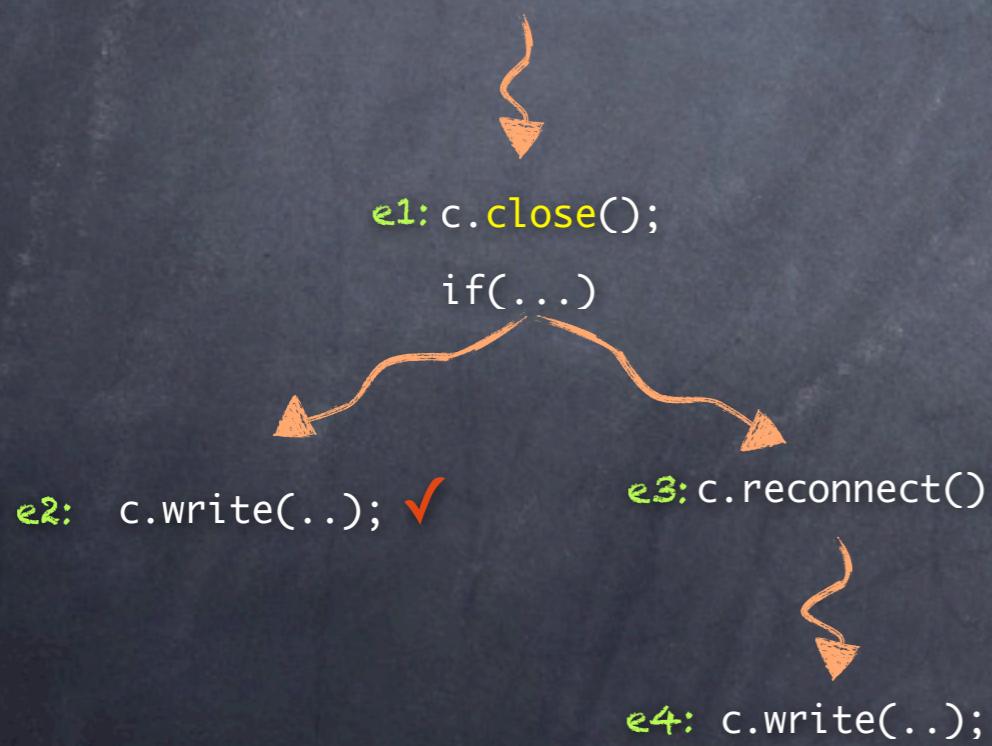
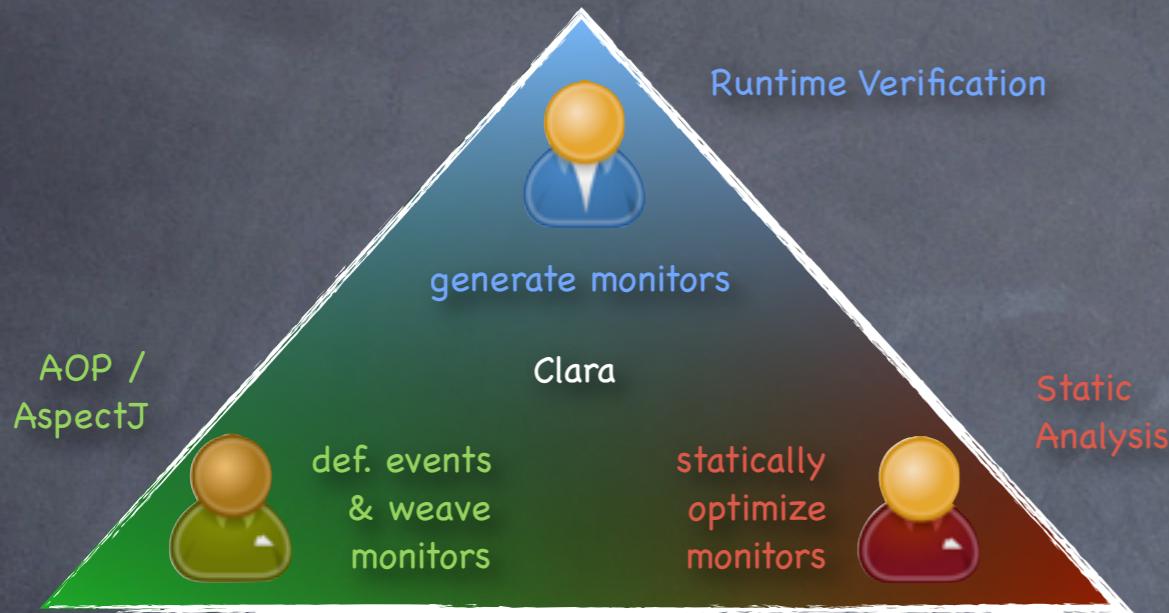




```

dependency{
    disconnect, write, reconnect;
    initial connected: disconnect -> connected,
    write -> connected,
    reconnect -> connected,
    disconnect -> disconnected;
    disconnect: disconnect -> disconnected,
    write -> error;
    error: write -> error;
}
final }
  
```

	antlr	bloat	chart	fop	hsqldb	jython	luindex	lusearch	pmd	xalan
ASyncContainsAll	0 71	0 6	0 6	0 146	0 33	0 128	0 149	0 149	0 671	0 10
ASyncIterC	0 1621	0 498	0 498	0 146	0 33	0 128	0 149	0 149	0 671	0 10
ASyncIterM	0 1684	0 507	0 507	0 176	0 39	0 138	0 152	0 152	0 718	0 10
FailSafeEnum	0 76	0 3	0 1	6 18	0 120	44 110	0 61	0 61	0 21	0 222
FailSafeEnumHT	26 133	0 102	0 44	0 205	-3 114	61 153	0 37	0 37	0 100	0 319
FailSafeIter	0 23	830 1394	149 510	0 288	0 112	112 253	0 217	16 217	287 546	0 158
FailSafeIterMap	0 130	444 1180	49 374	OOME	0 252	133 250	0 136	0 136	204 583	0 540
HasNextElem	0 117	0 4	0 4	0 12	0 53	34 64	0 22	0 22	0 11	0 63
HasNext	452 849	48 248	0 72	0 16	24 63	0 74	0 74	0 74	184 346	0 100
LeakingSync	0 170	0 1964	0 920	0 2347	0 528	0 1082	0 629	0 629	0 986	0 1005
Reader	0 50	0 7	0 65	0 102	3 1210	4 139	0 226	0 226	0 102	0 106
Writer	35 171	15 563	0 70	0 429	10 1378	462 462	0 146	0 146	0 62	0 751



```
dependency{
    disconnect, write, reconnect;
    initial connected: disconnect -> connected,
    write -> connected,
    reconnect -> connected,
    disconnect -> disconnected;
    disconnect: disconnect -> disconnected,
    write -> error;
    error: write -> error;
}
```

final }

	antlr	bloat	chart	fop	hsqldb	jython	luindex	lusearch	pmd	xalan
ASyncContainsAll	0 71	0 6	0 6	0 146	0 33	0 128	0 149	0 149	0 671	0 10
ASyncIterC	0 1621	0 498	0 498	0 146	0 33	0 128	0 149	0 149	0 671	0 10
ASyncIterM	0 1684	0 507	0 507	0 176	0 39	0 138	0 152	0 152	0 718	0 10
FailSafeEnum	0 76	0 3	0 1	6 18	0 120	44 110	0 61	0 61	0 21	0 222
FailSafeEnumHT	26 133	0 102	0 44	0 205	-3 114	61 153	0 37	0 37	0 100	0 319
FailSafeIter	0 23	830 1394	149 510	0 288	0 112	112 253	0 217	16 217	287 546	0 158
FailSafeIterMap	0 130	444 1180	49 374	OOME	0 252	133 250	0 136	0 136	204 583	0 540
HasNextElem	0 117	0 4	0 4	0 12	0 53	34 64	0 22	0 22	0 11	0 63
HasNext	452 849	48 248	0 72	0 16	24 63	0 74	0 74	0 74	184 346	0 63
LeakingSync	0 170	0 1964	0 920	0 2347	0 528	0 1082	0 629	0 629	0 986	0 1005
Reader	0 50	0 7	0 65	0 102	3 1210	4 139	0 226	0 226	0 102	0 106
Writer	35 171	15 563	0 70	0 429	10 1378	462 462	0 146	0 146	0 62	0 751

