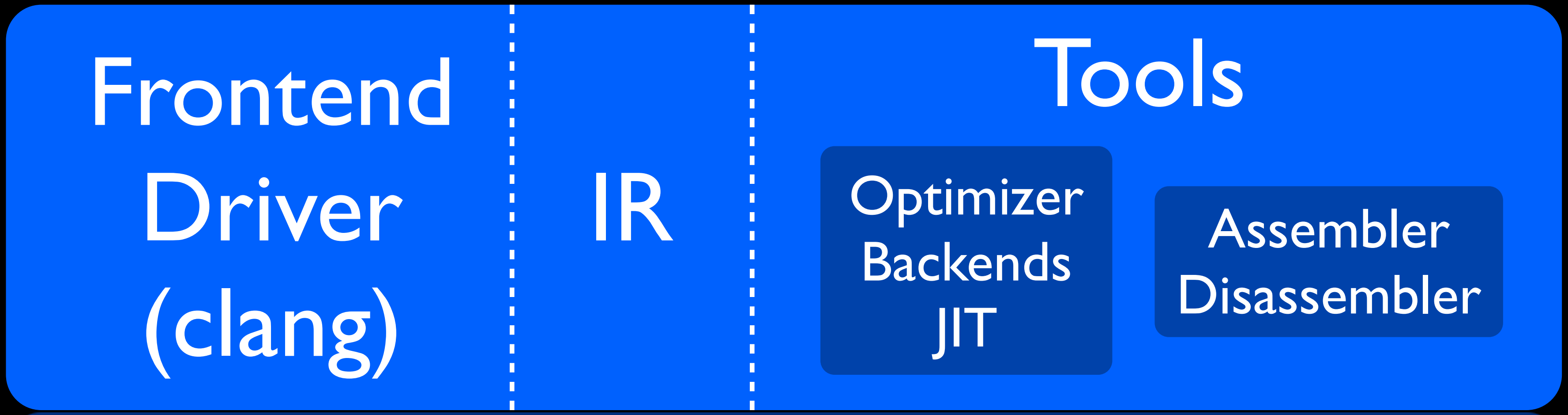
Clang, LLVM, and GNOME

Bruno Cardoso Lopes

What’s LLVM?

What’s LLVM

- Compiler infrastructure



Frontend
Driver
(clang)

IR

Tools

Optimizer
Backends
JIT

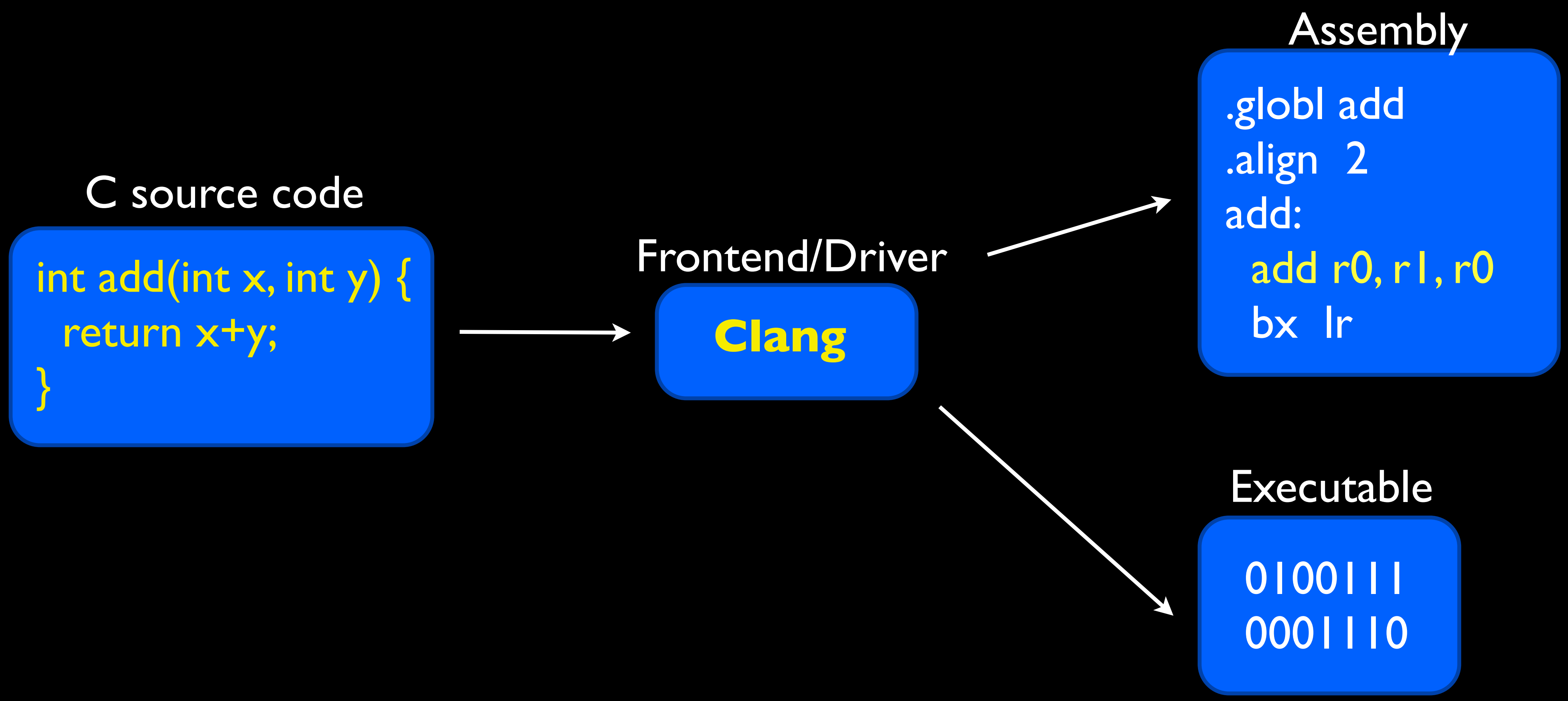
Assembler
Disassembler

What’s LLVM

- Virtual Instruction set (IR)
- Bitcode

What’s LLVM

- Static compilation



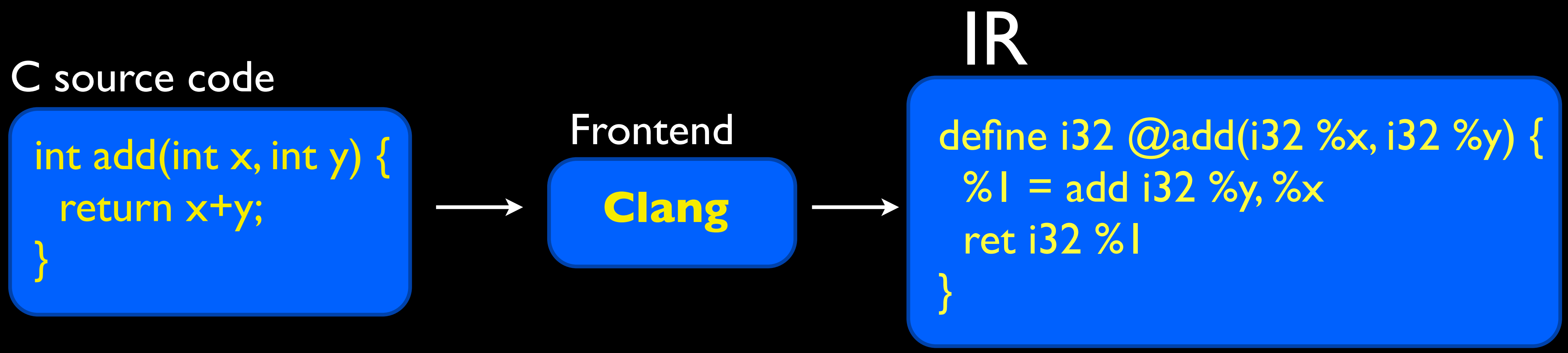
```
int add(int x, int y) {
  return x+y;
}
```

```
.globl add
.align  2
add:
  add r0, r1, r0
  bx  lr
```

```
0100111
0001110
```

What’s LLVM

- Decoupled tools



C source code

```
int add(int x, int y) {
  return x+y;
}
```

Frontend

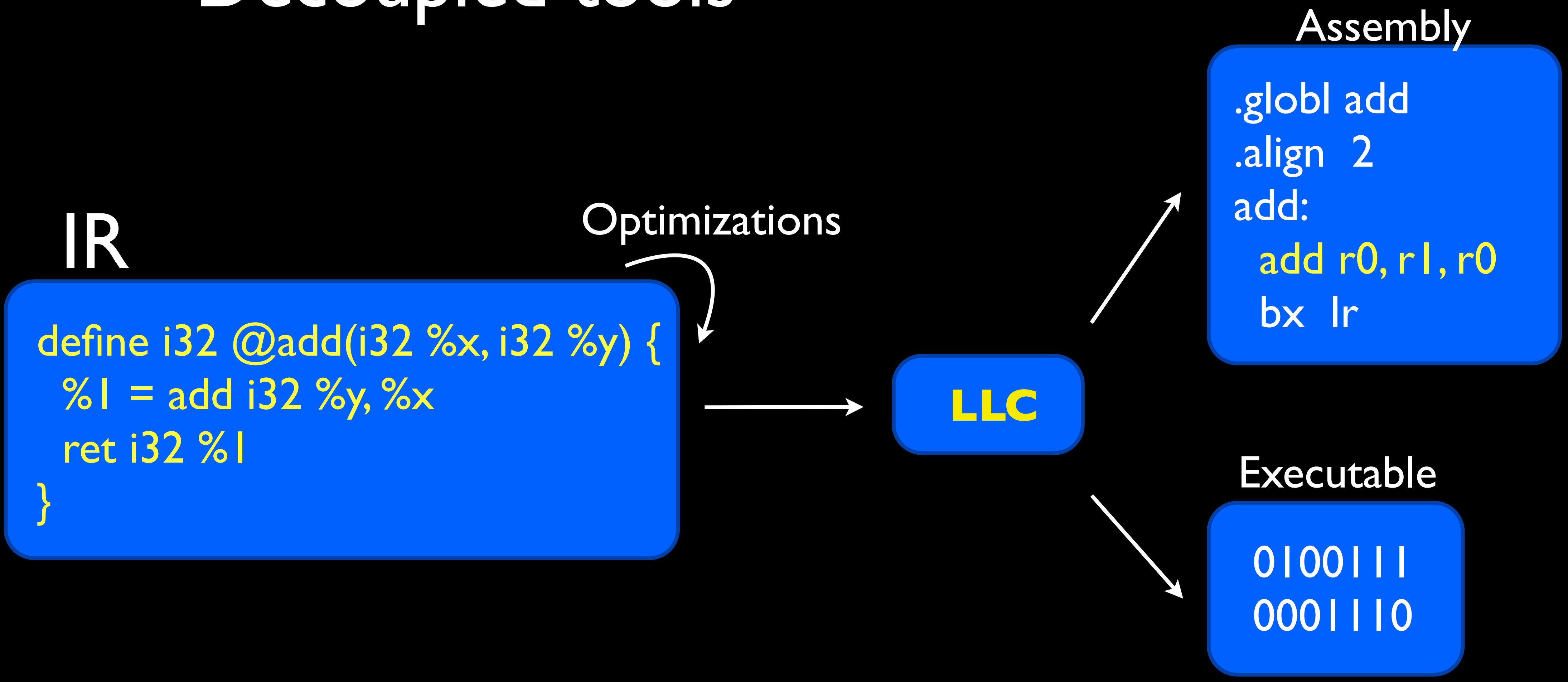
Clang

IR

```
define i32 @add(i32 %x, i32 %y) {
  %1 = add i32 %y, %x
  ret i32 %1
}
```

What’s LLVM

- Decoupled tools



IR

```
define i32 @add(i32 %x, i32 %y) {
  %1 = add i32 %y, %x
  ret i32 %1
}
```

```
.globl add
.align  2
add:
  add r0, r1, r0
  bx  lr
```

```
0100111
0001110
```

Why LLVM?

Why LLVM?

- Open Source
- Active community: easy integration and quick patch review

Why LLVM?

- Optimization oriented compiler: *compile time*, *link-time* and *run-time*
- More than 30 analysis passes and 60 transformation passes.

Why LLVM?

- Additional tools to check for correctness and bugs

Tools

Tools

- Written in C++
- Modular and composed of several libraries
- Several tools for each part of compilation

Tools

Front-end

- Dragonegg
  - Gcc 4.8 plugin

Tools

Front-end

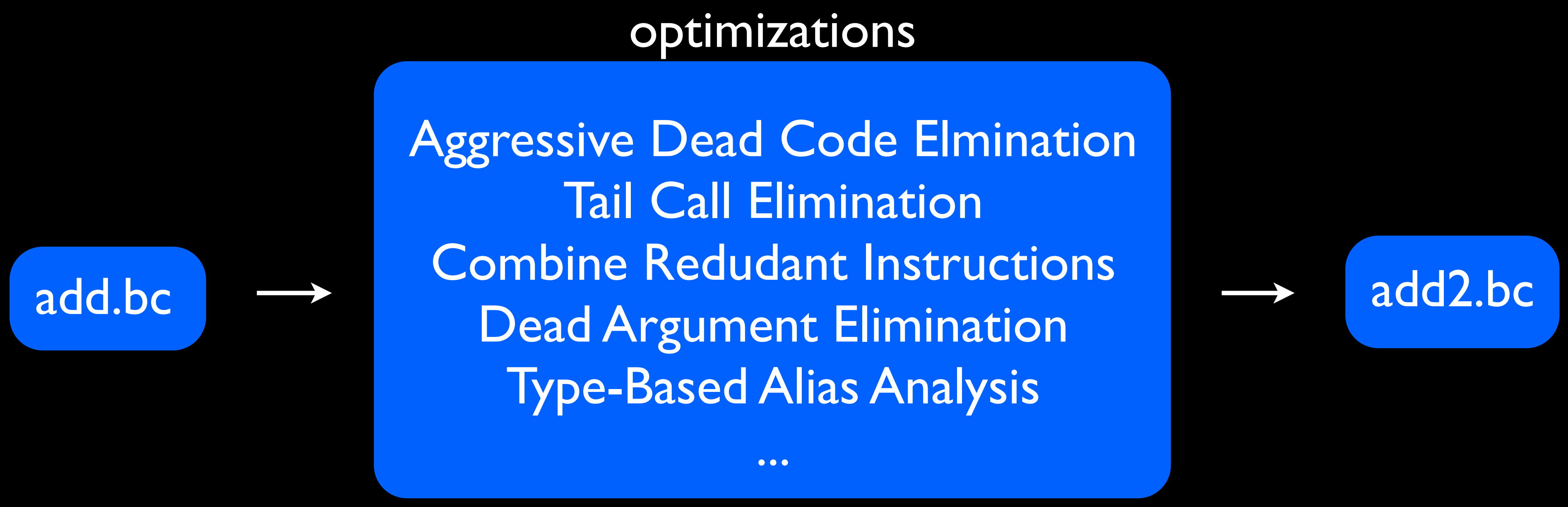
- Clang
  - Library approach
  - No cross-compiler generation needed
  - Good diagnostics
  - Static Analyzer

Tools

Optimizer

- Optimization are applied to the IR
- **opt** tool

```
$ opt -O3 add.bc -o add2.bc
```



optimizations

add.bc →

Aggressive Dead Code Elmination
Tail Call Elimination
Combine Redudant Instructions
Dead Argument Elimination
Type-Based Alias Analysis
...

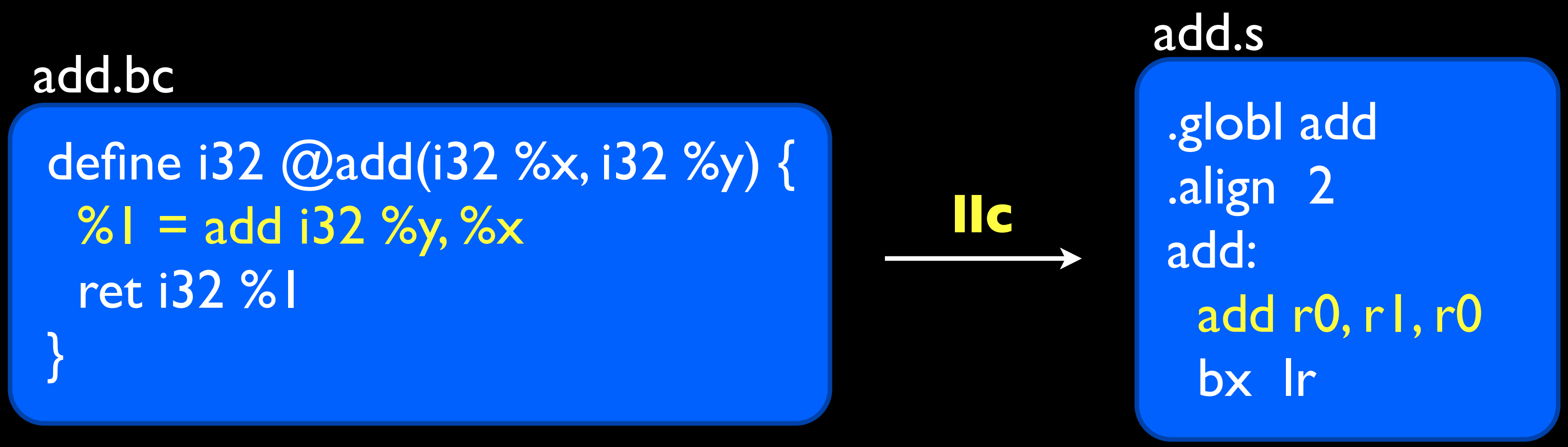
→ add2.bc

Tools

Low Level Compiler

- **llc** tool: invoke the static backends
- Generates assembly or object code

```
$ llc -march=arm add.bc -o add.s
```



add.bc

```
define i32 @add(i32 %x, i32 %y) {
  %1 = add i32 %y, %x
  ret i32 %1
}
```

add.s

```
.globl add
.align  2
add:
  add r0, r1, r0
  bx  lr
```

Tools

Other

- **JIT** compiler
- **libLTO**
- Assemblers and Disassemblers

Gnome

Gnome

- Uses LLVM/Clang libraries in some projects
  - LLVMPipe
  - gedit-code-assistant plugin

Gnome

- Umbrella for several projects
- Building infrastructure: **jhbuild** and **OSTree**
- Buildbots

Gnome

- May benefit from more LLVM features
  - ARM Backend
  - LLVM LTO
  - Static analyzer
  - Address and Memory Sanitizer

ARM Backend

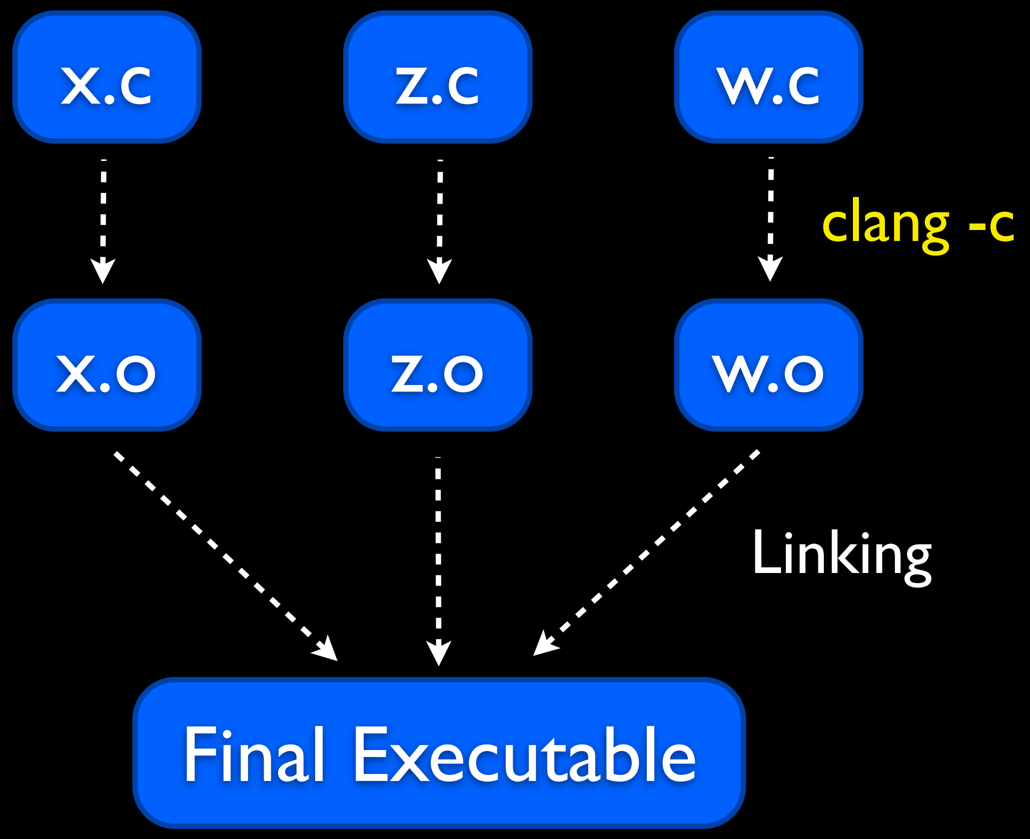
- Very stable
- High quality code generation
- Used to compile iOS applications

ARM Backend

- Support recent ARM processors Cortex-A9, A15, M-3
- Gnome 3 on Nexus 7
- Future: Gnome on GPU-less ARM devices? LLVMPipe tuned for NEON?

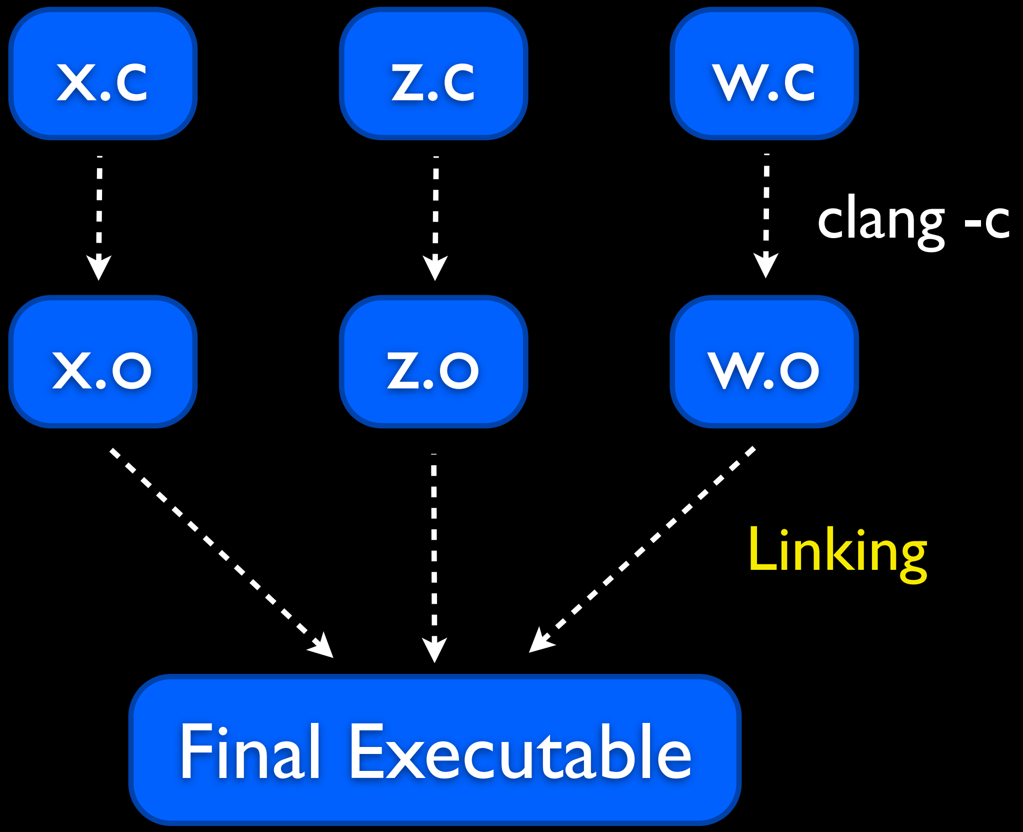
LLVM LTO

Regular linking process



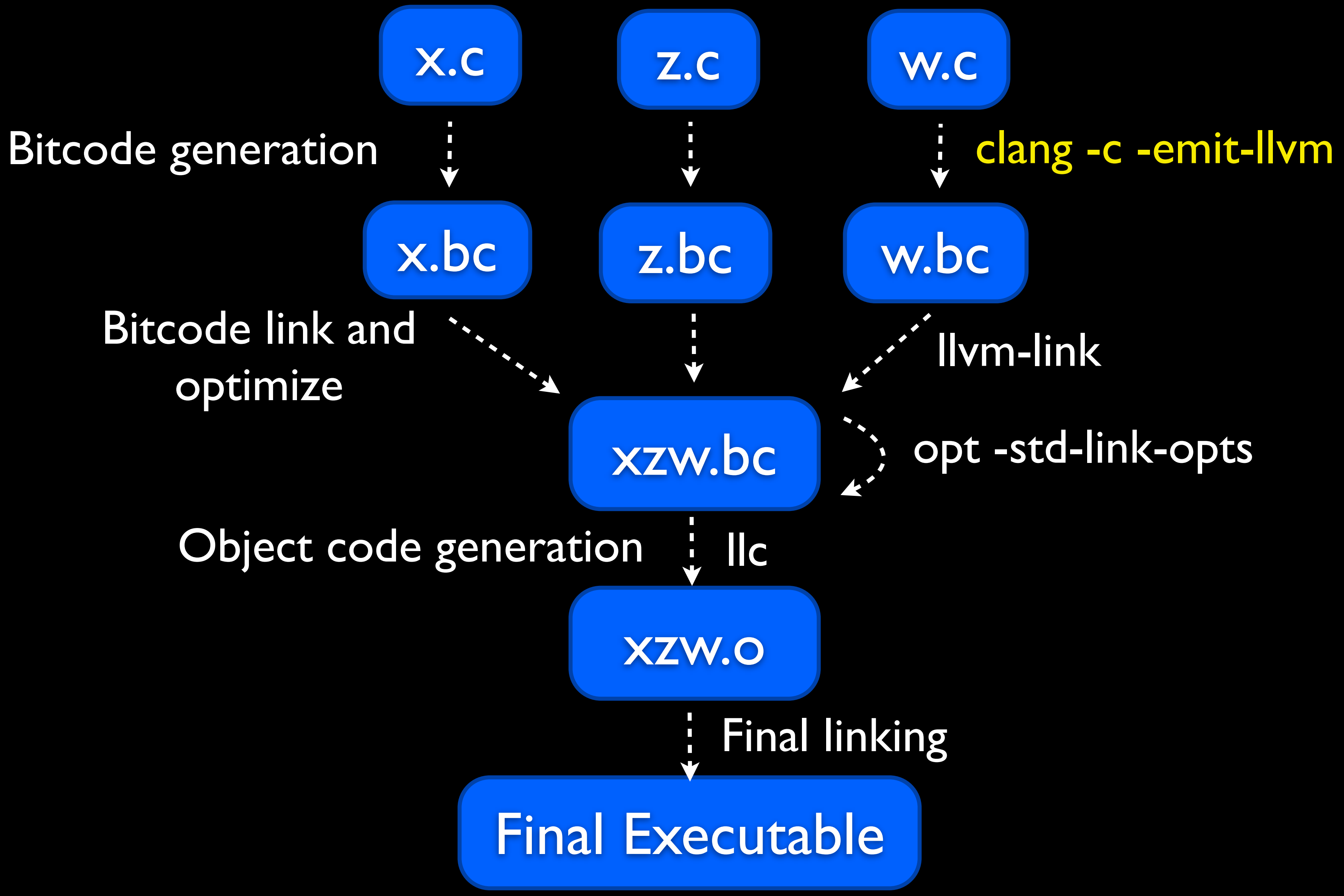
LLVM LTO

Regular linking process



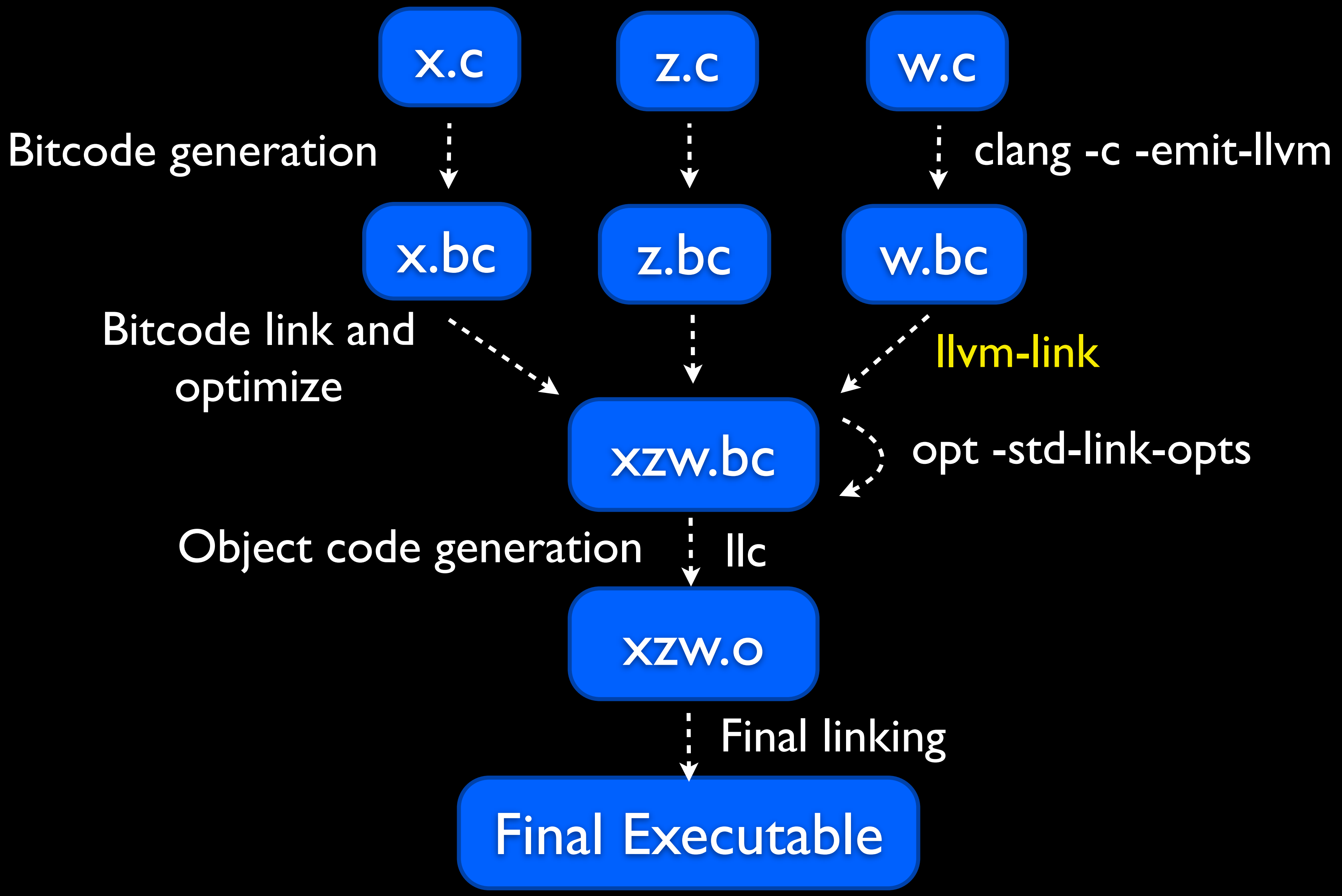
LLVM LTO

LTO linking process



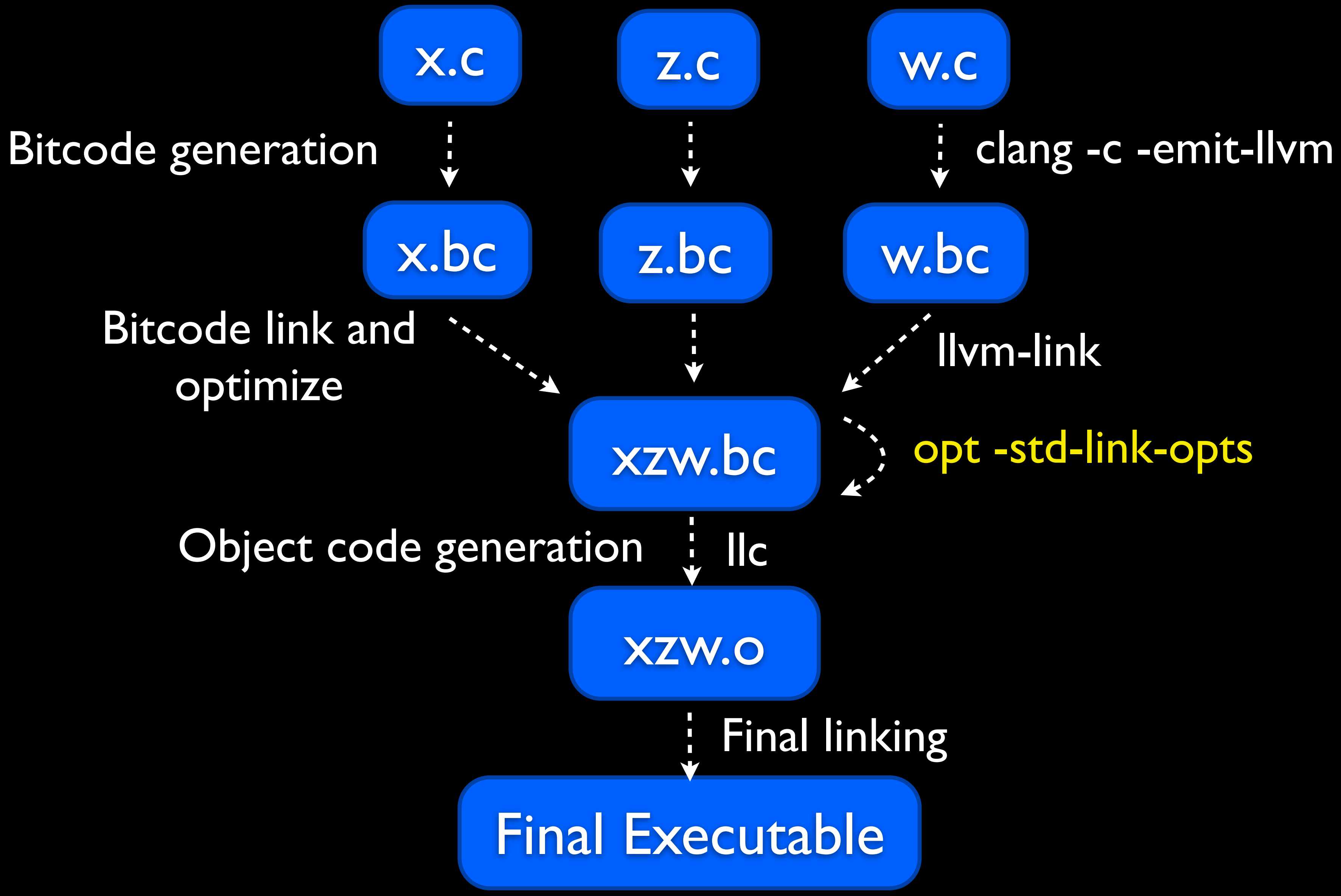
LLVM LTO

LTO linking process



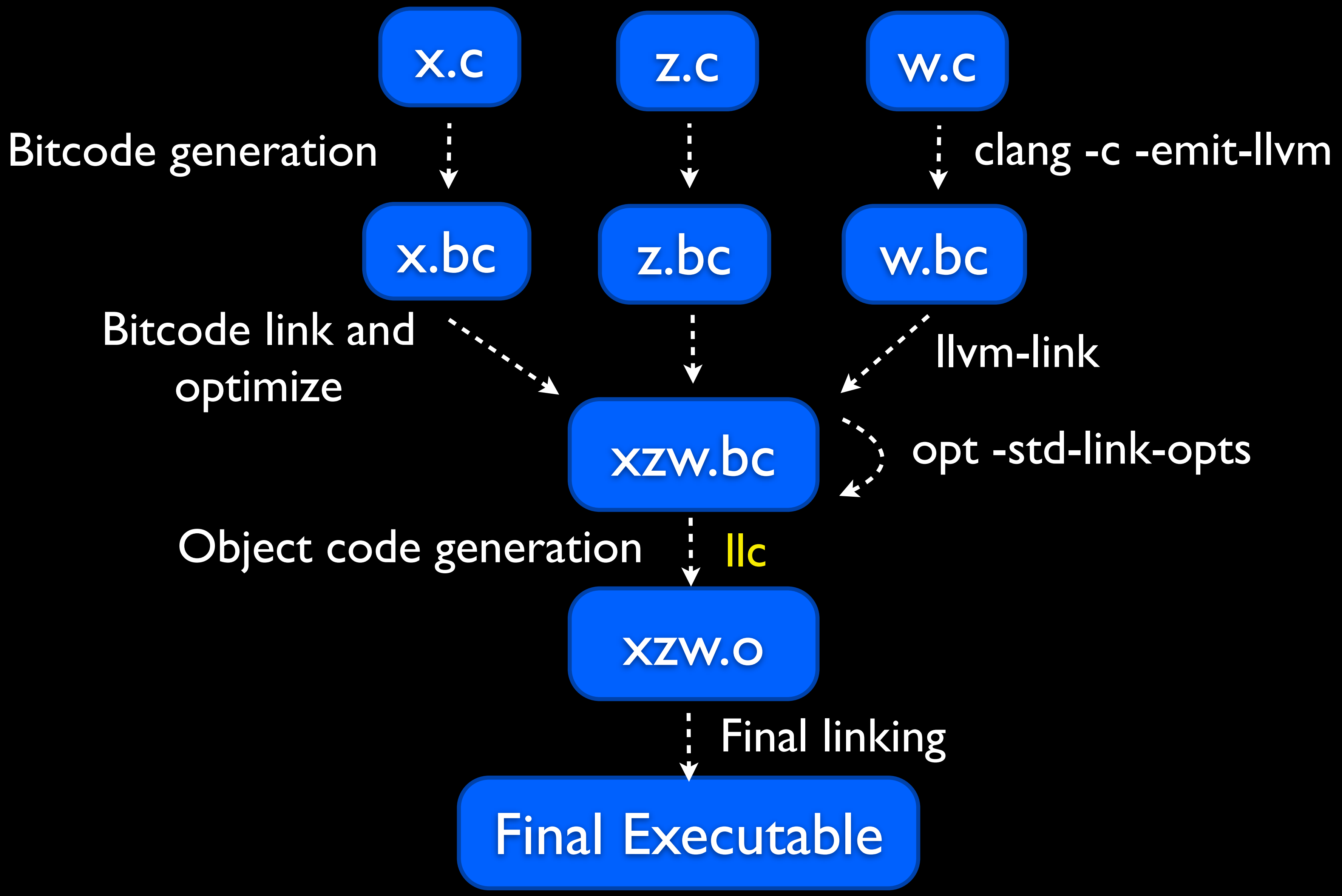
LLVM LTO

LTO linking process



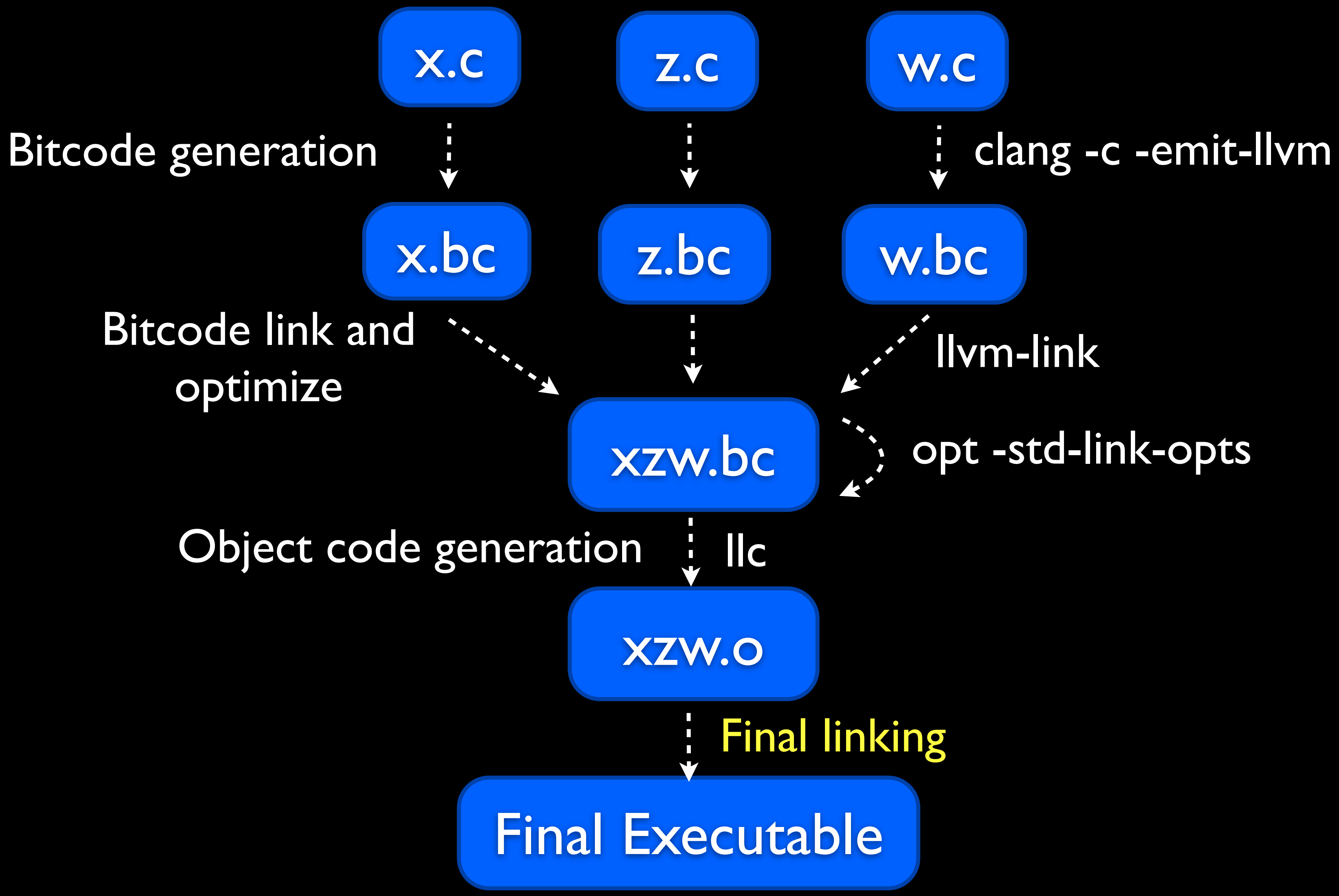
LLVM LTO

LTO linking process



LLVM LTO

LTO linking process



x.c

z.c

w.c

Bitcode generation

clang -c -emit-llvm

x.bc

z.bc

w.bc

Bitcode link and optimize

llvm-link

xzw.bc

opt -std-link-opts

Object code generation

llc

xzw.o

Final linking

Final Executable

Static Analyzer

- Source code analysis: C, C++, Objective C
- Bug finding automatization
- Good to be integrated into build systems

Static Analyzer

- Developers may improve code quality
- False positives
- Slower compilation

Static Analyzer

- How to use:

```
$ scan-build ../configure ....
```

```
$ scan-build make
```

or

```
$ scan-build gcc x.c -o x
```

- Generates html reports

```
$ scan-view /tmp/result-dir
```

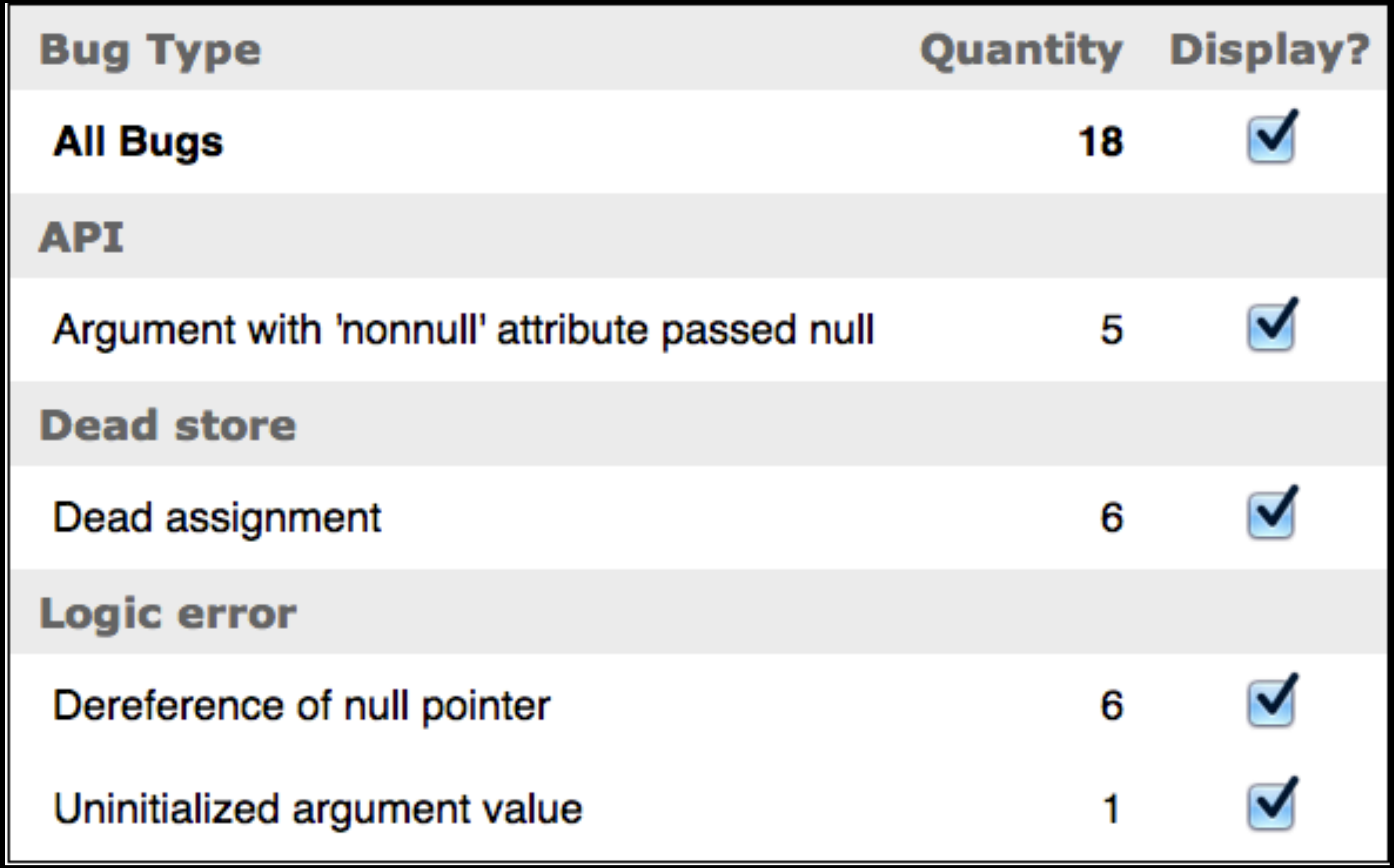
Static Analyzer

Gnome

- Analyzed some projects from gnome codebase https://git.gnome.org/
- eog, evolution, evolution-data-server, gedit, glade, gnome-control-center, gstreamer, gtk+, nautilus.

http://brunocardoso.cc/guadec13/static/

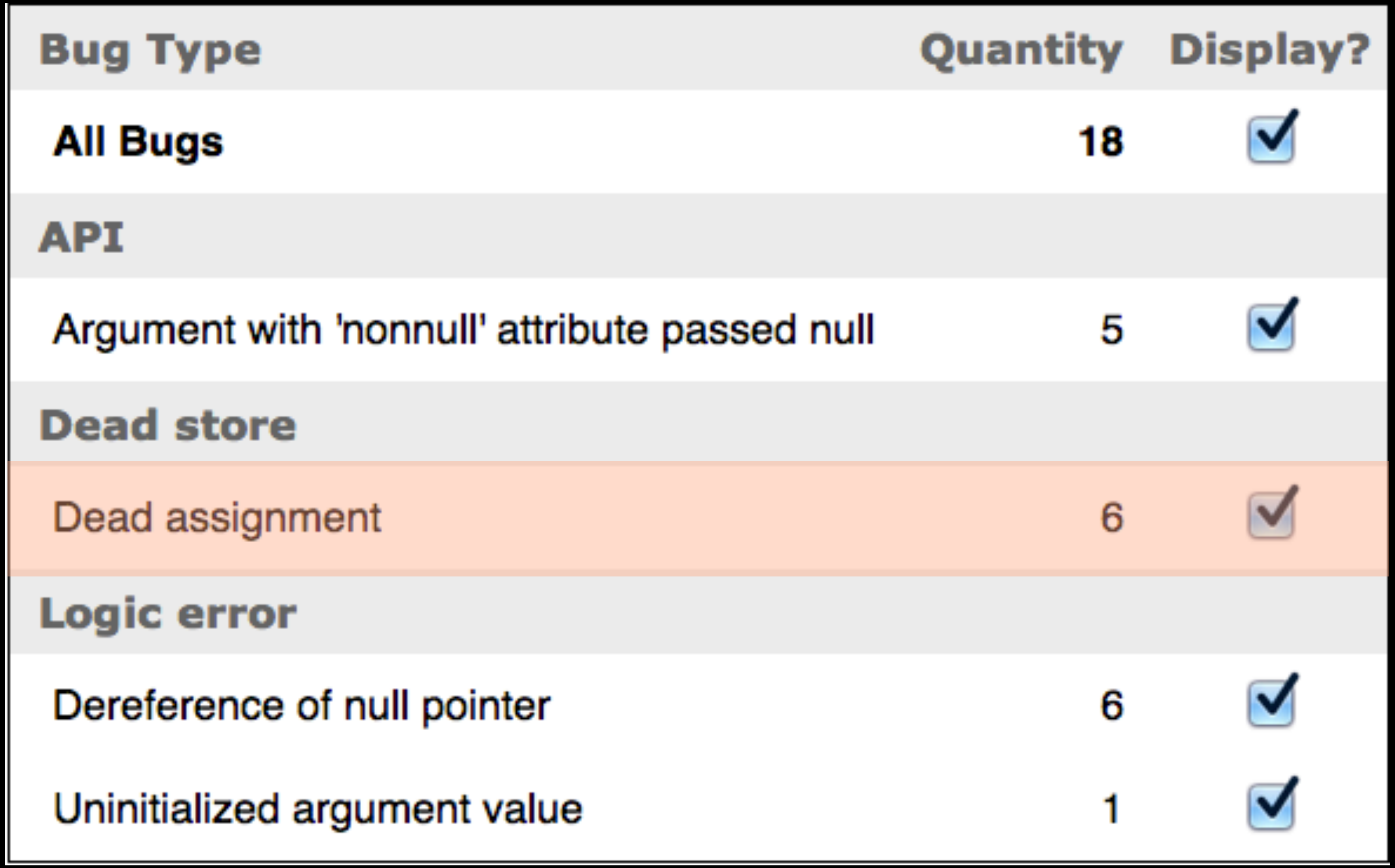
Glade



Bug Type	Quantity	Display?
All Bugs	**18**	☑
API		
Argument with 'nonnull' attribute passed null	5	☑
Dead store		
Dead assignment	6	☑
Logic error		
Dereference of null pointer	6	☑
Uninitialized argument value	1	☑

http://brunocardoso.cc/guadec13/static/

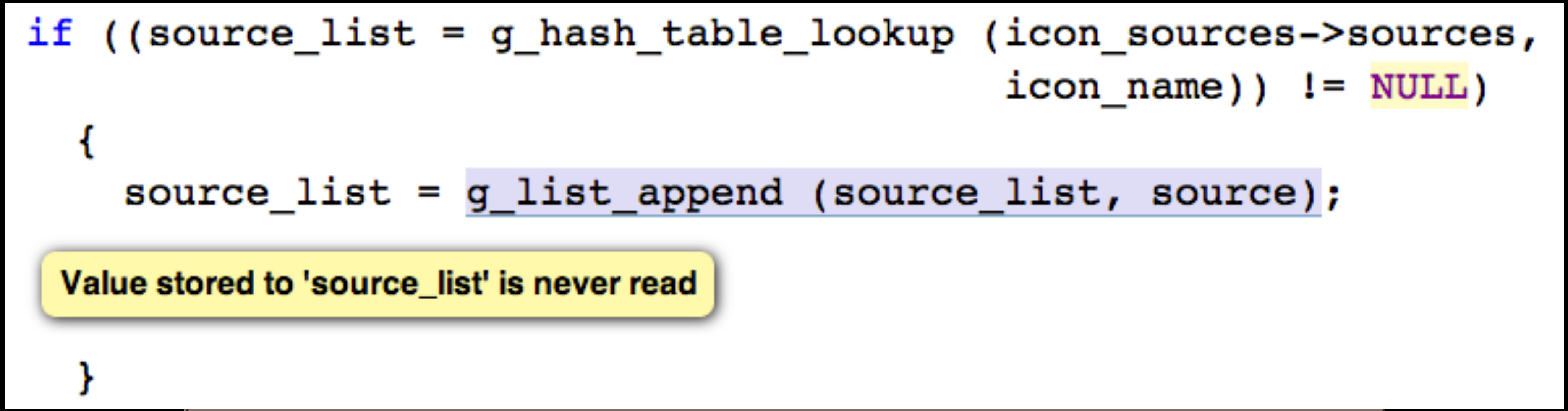
Glade



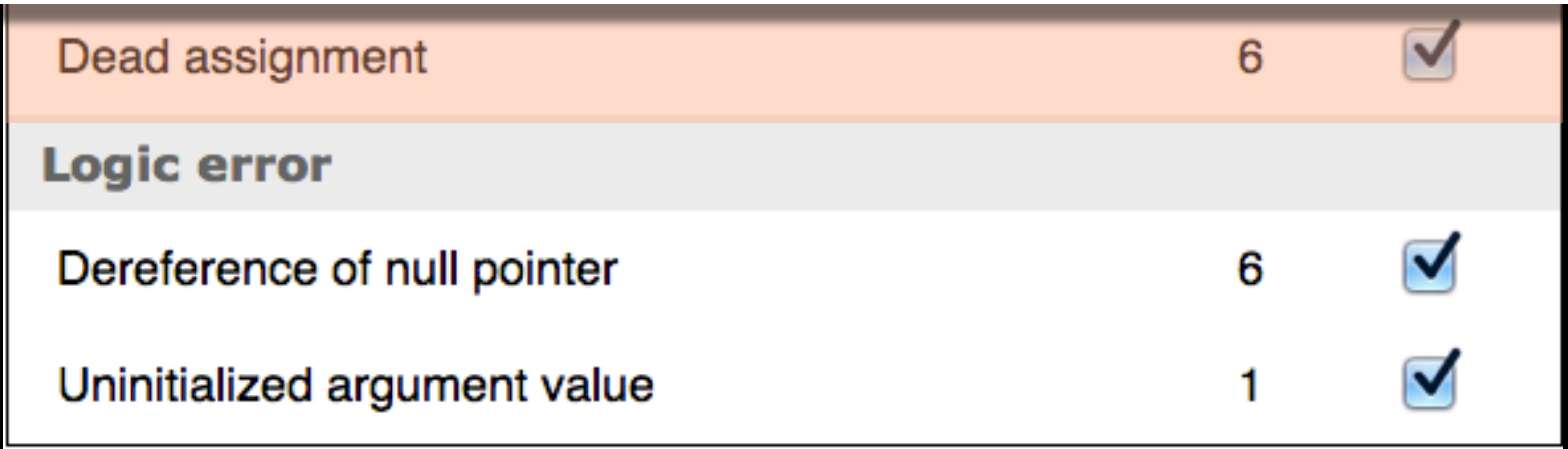
Bug Type	Quantity	Display?
All Bugs	**18**	☑
API		
Argument with 'nonnull' attribute passed null	5	☑
Dead store		
Dead assignment	6	☑
Logic error		
Dereference of null pointer	6	☑
Uninitialized argument value	1	☑

http://brunocardoso.cc/guadec13/static/

Glade



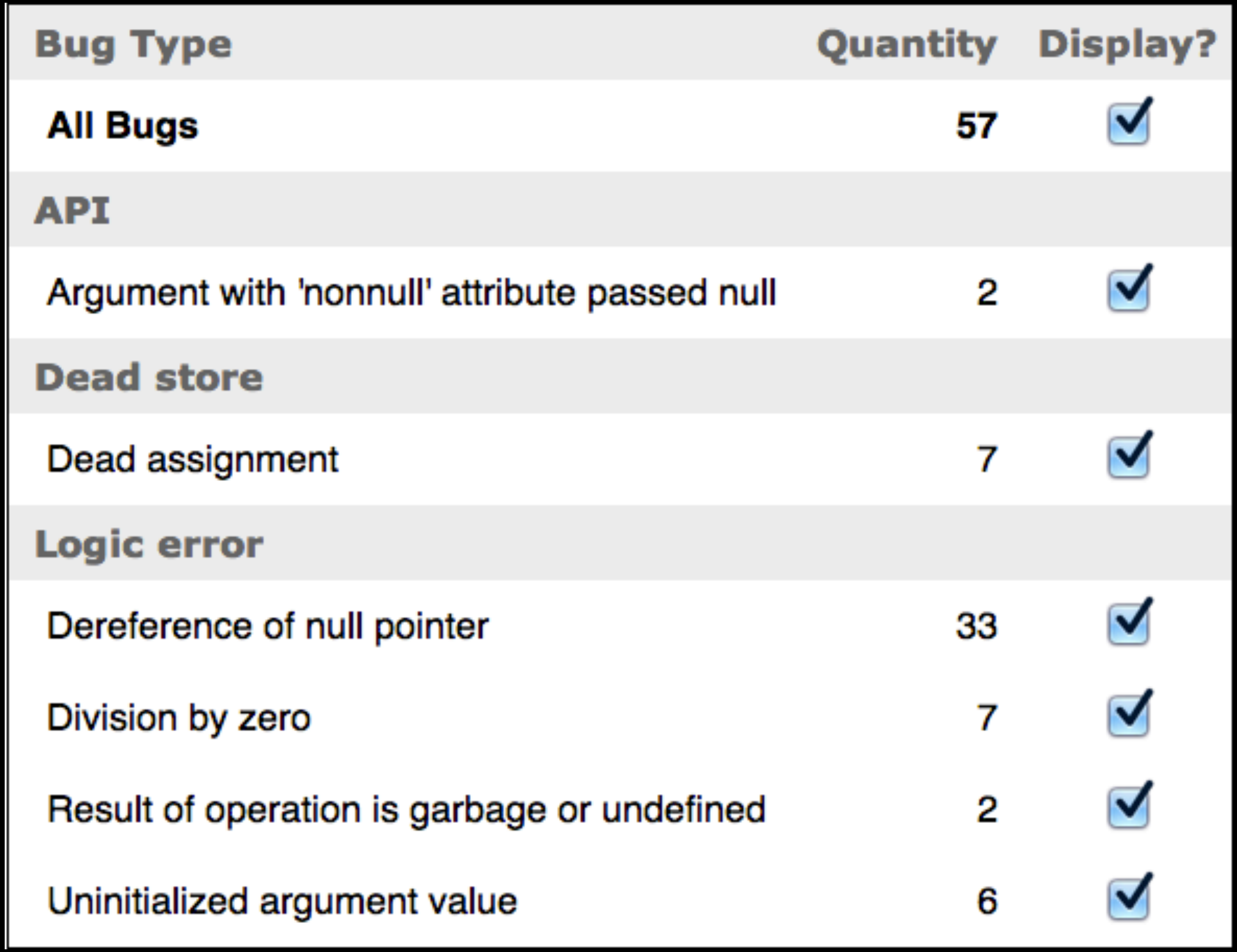
```
if ((source_list = g_hash_table_lookup (icon_sources->sources,
                                         icon_name)) != NULL)
  {
    source_list = g_list_append (source_list, source);

Value stored to 'source_list' is never read

  }
```



Dead assignment	6	☑
Logic error		
Dereference of null pointer	6	☑
Uninitialized argument value	1	☑

http://brunocardoso.cc/guadec13/static/

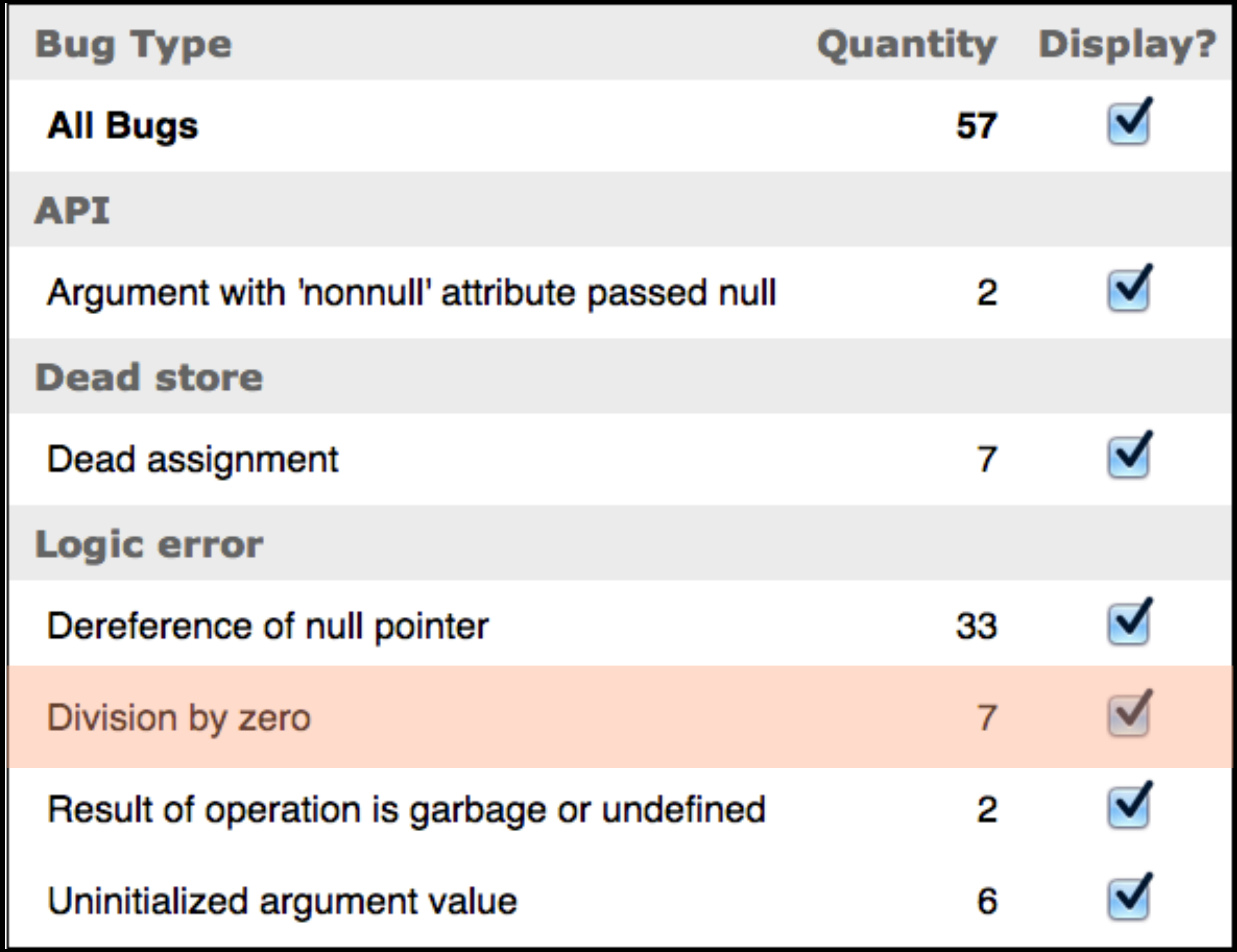
gstreamer



Bug Type	Quantity	Display?
All Bugs	**57**	☑
API		
Argument with 'nonnull' attribute passed null	2	☑
Dead store		
Dead assignment	7	☑
Logic error		
Dereference of null pointer	33	☑
Division by zero	7	☑
Result of operation is garbage or undefined	2	☑
Uninitialized argument value	6	☑

http://brunocardoso.cc/guadec13/static/

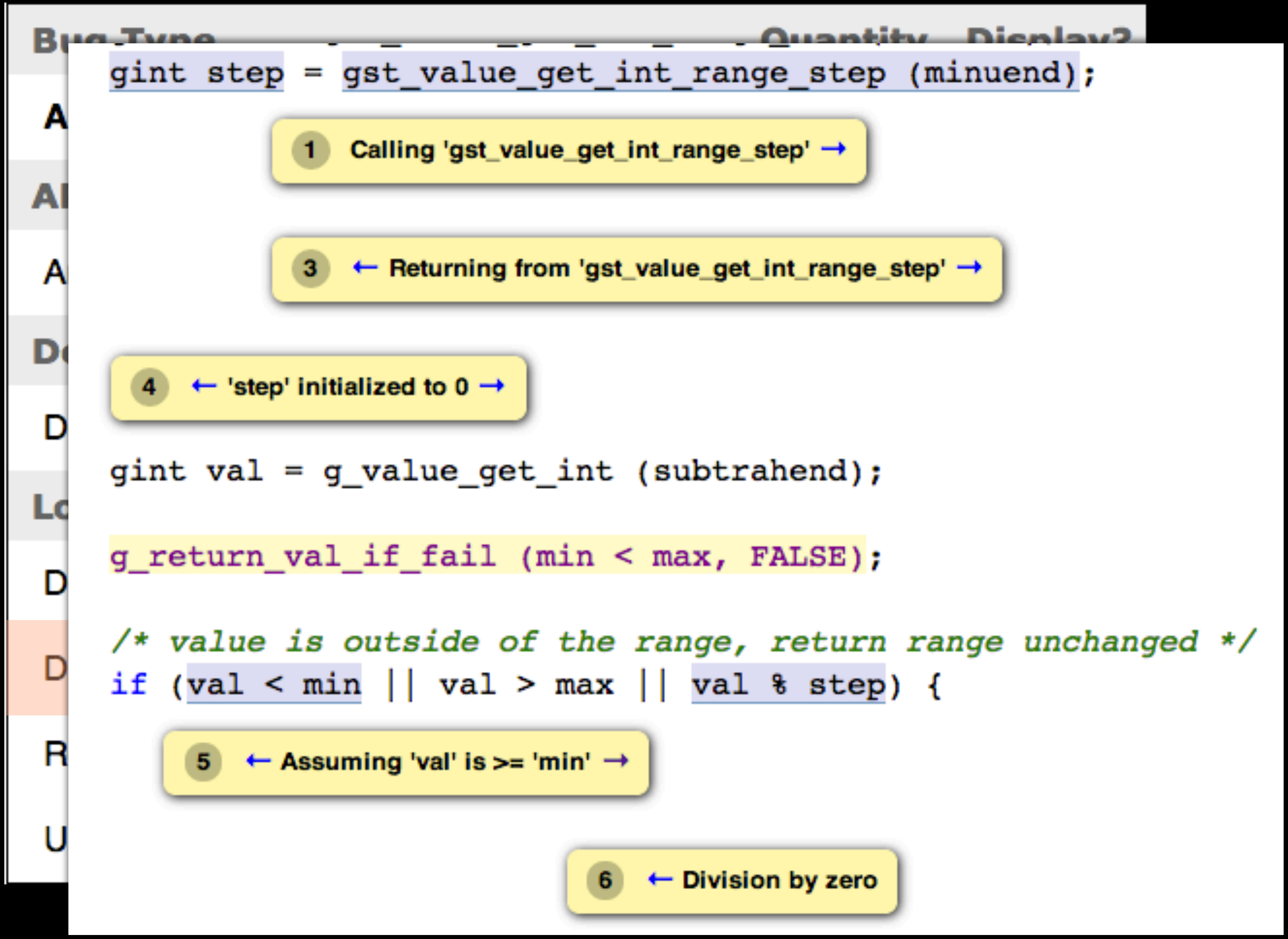
gstreamer



Bug Type	Quantity	Display?
All Bugs	**57**	☑
API		
Argument with 'nonnull' attribute passed null	2	☑
Dead store		
Dead assignment	7	☑
Logic error		
Dereference of null pointer	33	☑
Division by zero	7	☑
Result of operation is garbage or undefined	2	☑
Uninitialized argument value	6	☑

http://brunocardoso.cc/guadec13/static/

gstreamer



```
gint step = gst_value_get_int_range_step (minuend);
```

1 Calling 'gst_value_get_int_range_step' →

3 ← Returning from 'gst_value_get_int_range_step' →

4 ← 'step' initialized to 0 →

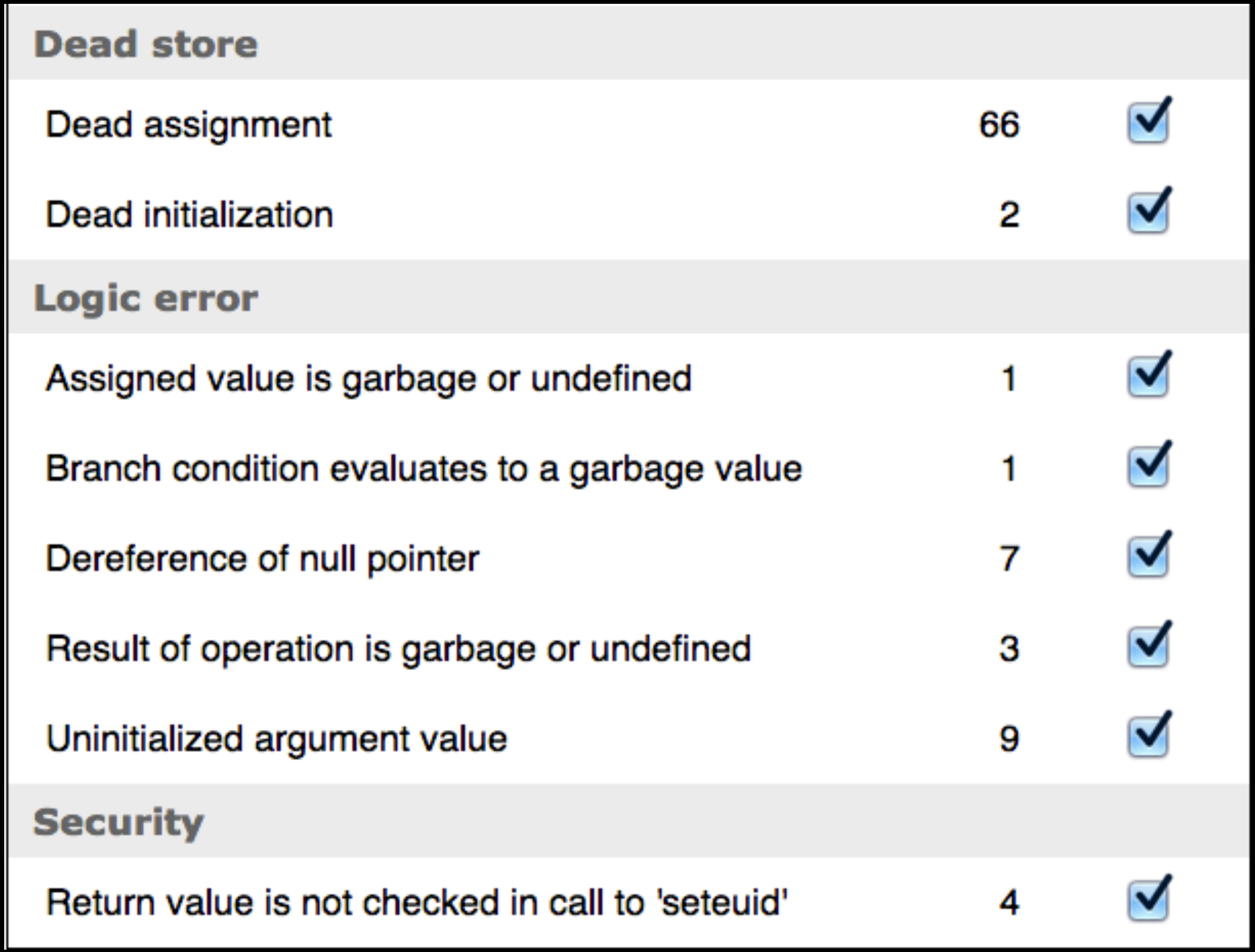
```
gint val = g_value_get_int (subtrahend);

g_return_val_if_fail (min < max, FALSE);

/* value is outside of the range, return range unchanged */
if (val < min || val > max || val % step) {
```

5 ← Assuming 'val' is >= 'min' →

6 ← Division by zero

http://brunocardoso.cc/guadec13/static/

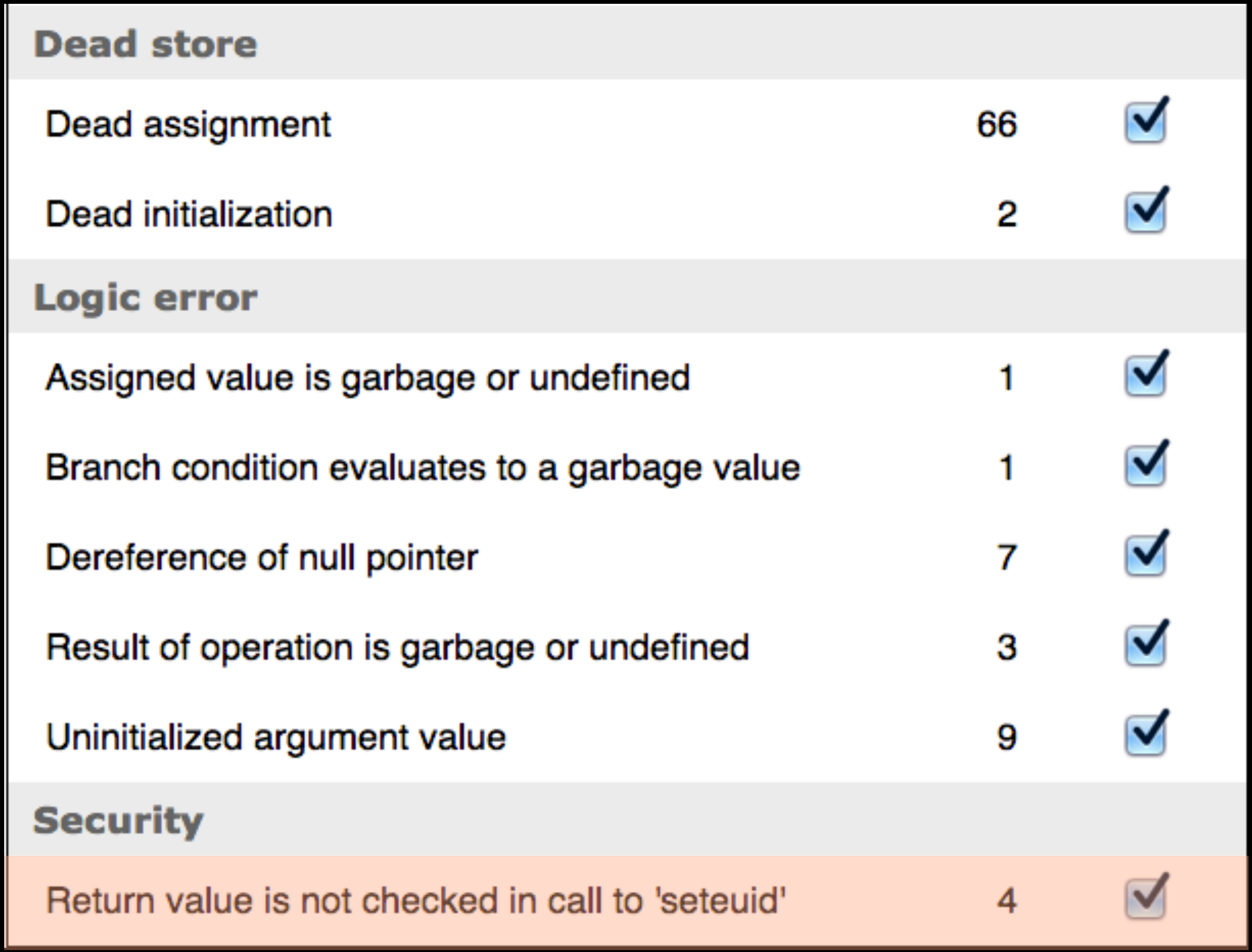
evolution-data-server



Dead store		
Dead assignment	66	☑
Dead initialization	2	☑
Logic error		
Assigned value is garbage or undefined	1	☑
Branch condition evaluates to a garbage value	1	☑
Dereference of null pointer	7	☑
Result of operation is garbage or undefined	3	☑
Uninitialized argument value	9	☑
Security		
Return value is not checked in call to 'seteuid'	4	☑

http://brunocardoso.cc/guadec13/static/

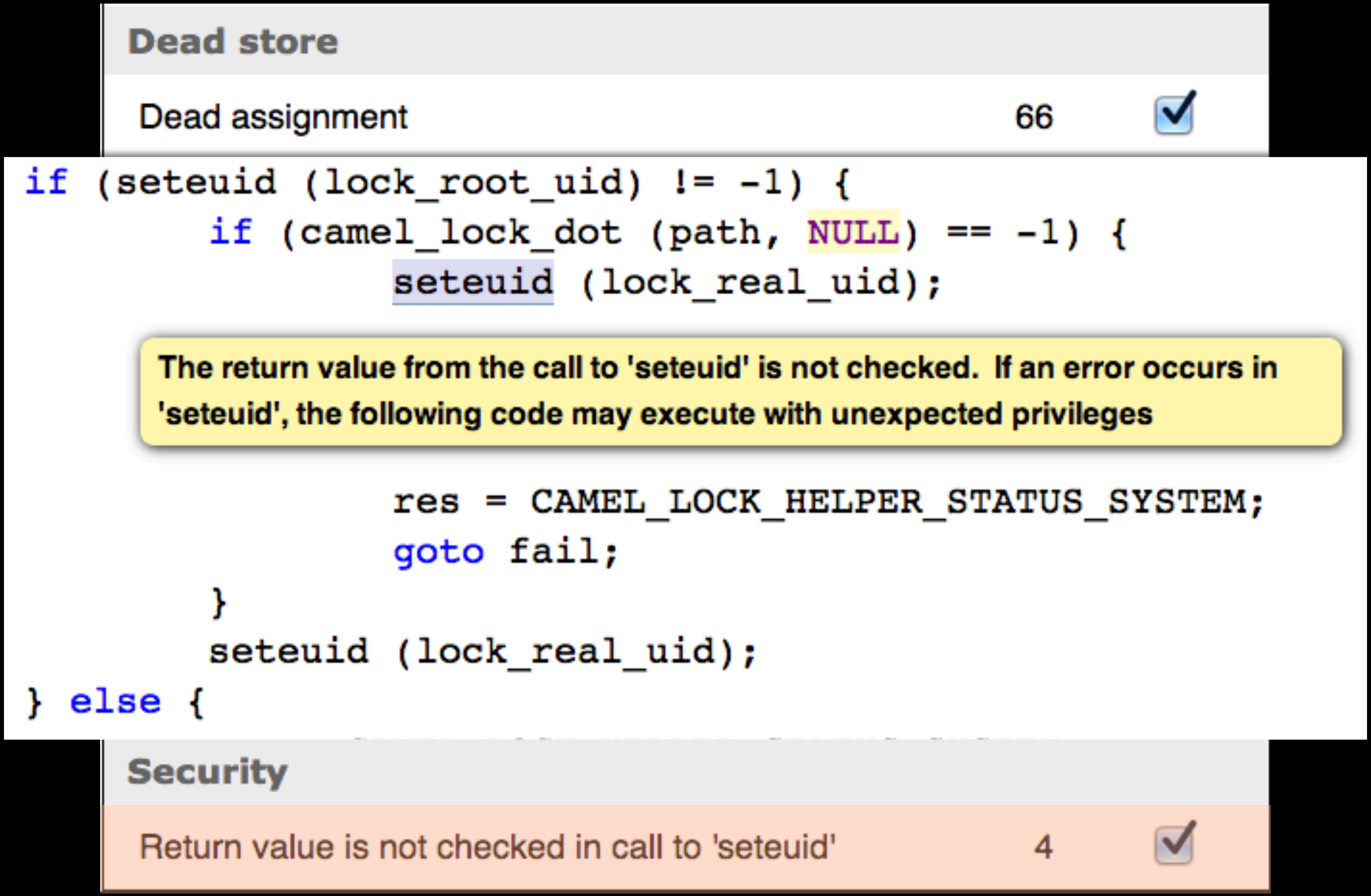
evolution-data-server



Dead store		
Dead assignment	66	☑
Dead initialization	2	☑
Logic error		
Assigned value is garbage or undefined	1	☑
Branch condition evaluates to a garbage value	1	☑
Dereference of null pointer	7	☑
Result of operation is garbage or undefined	3	☑
Uninitialized argument value	9	☑
Security		
Return value is not checked in call to 'seteuid'	4	☑

http://brunocardoso.cc/guadec13/static/

evolution-data-server



Dead store

Dead assignment 66 ☑

```
if (seteuid (lock_root_uid) != -1) {
        if (camel_lock_dot (path, NULL) == -1) {
                seteuid (lock_real_uid);
```

The return value from the call to 'seteuid' is not checked. If an error occurs in 'seteuid', the following code may execute with unexpected privileges

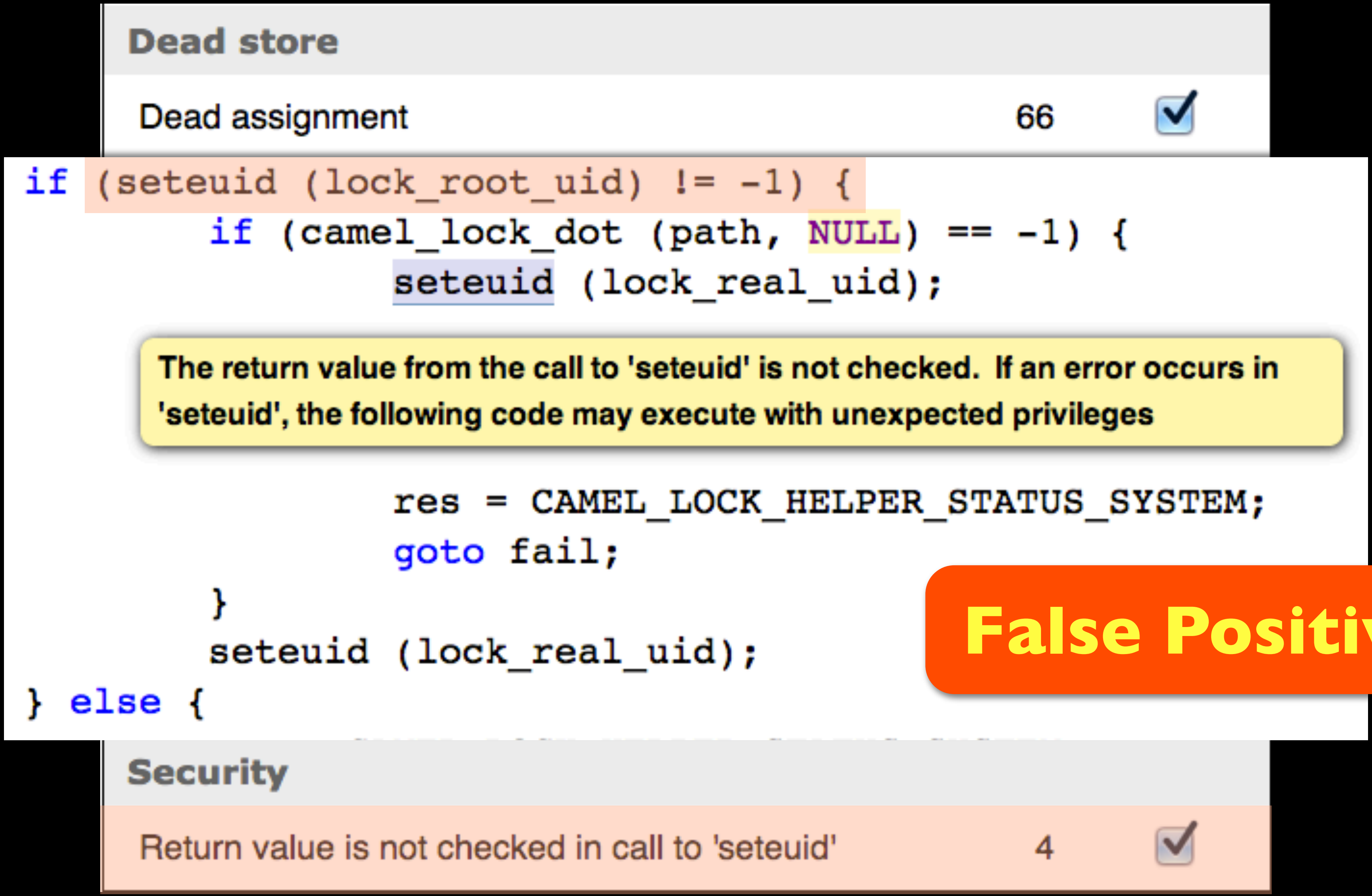
```
                res = CAMEL_LOCK_HELPER_STATUS_SYSTEM;
                goto fail;
        }
        seteuid (lock_real_uid);
} else {
```

Security

Return value is not checked in call to 'seteuid' 4 ☑

http://brunocardoso.cc/guadec13/static/

evolution-data-server



Dead store

Dead assignment	66	☑

```
if (seteuid (lock_root_uid) != -1) {
        if (camel_lock_dot (path, NULL) == -1) {
                seteuid (lock_real_uid);
```

The return value from the call to 'seteuid' is not checked. If an error occurs in 'seteuid', the following code may execute with unexpected privileges

```
                res = CAMEL_LOCK_HELPER_STATUS_SYSTEM;
                goto fail;
        }
        seteuid (lock_real_uid);
} else {
```

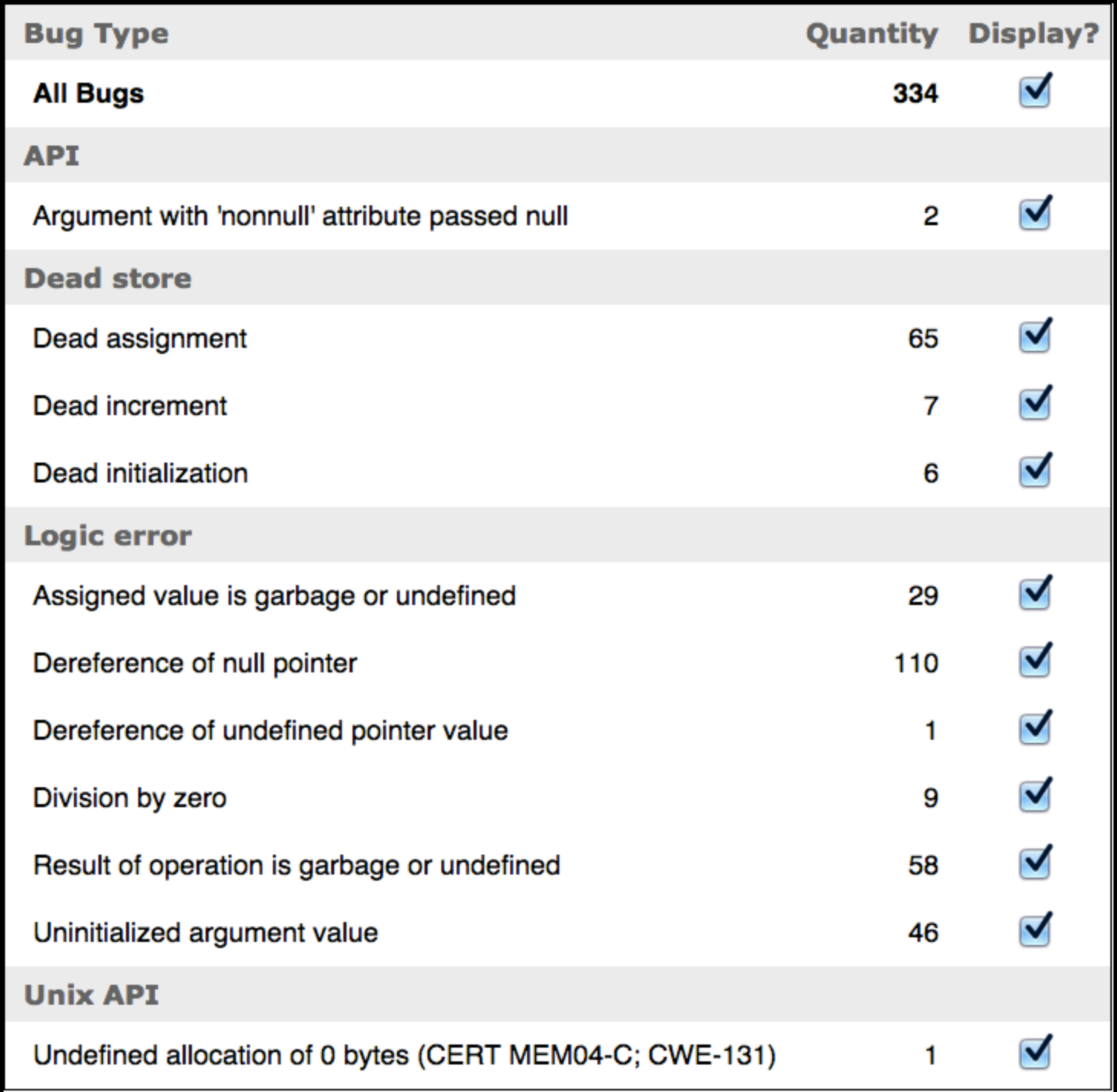
Security

Return value is not checked in call to 'seteuid'	4	☑

False Positive

http://brunocardoso.cc/guadec13/static/

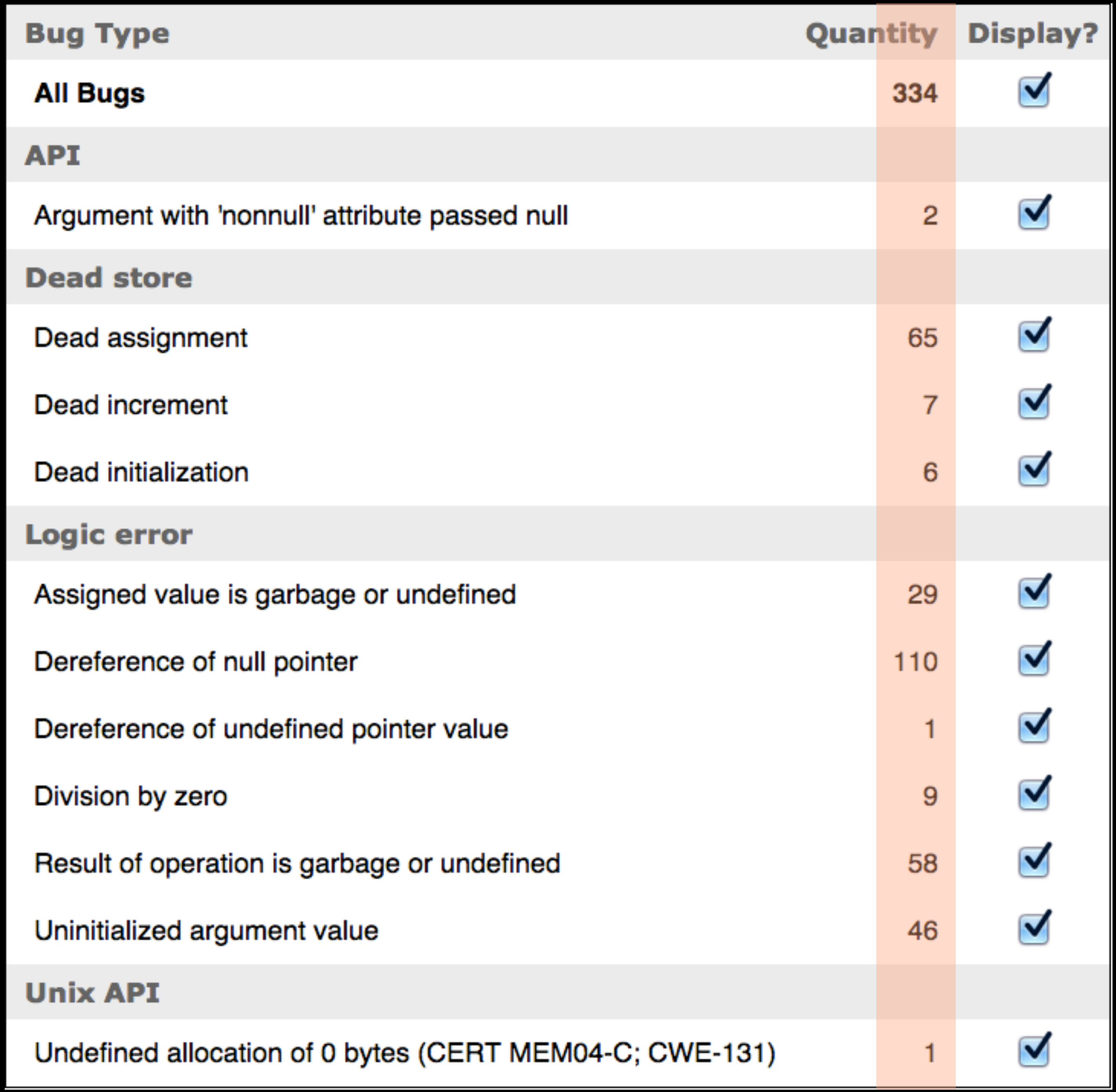
gtk+



Bug Type	Quantity	Display?
All Bugs	**334**	☑
API		
Argument with 'nonnull' attribute passed null	2	☑
Dead store		
Dead assignment	65	☑
Dead increment	7	☑
Dead initialization	6	☑
Logic error		
Assigned value is garbage or undefined	29	☑
Dereference of null pointer	110	☑
Dereference of undefined pointer value	1	☑
Division by zero	9	☑
Result of operation is garbage or undefined	58	☑
Uninitialized argument value	46	☑
Unix API		
Undefined allocation of 0 bytes (CERT MEM04-C; CWE-131)	1	☑

http://brunocardoso.cc/guadec13/static/

gtk+



Bug Type	Quantity	Display?
All Bugs	334	☑
API		
Argument with 'nonnull' attribute passed null	2	☑
Dead store		
Dead assignment	65	☑
Dead increment	7	☑
Dead initialization	6	☑
Logic error		
Assigned value is garbage or undefined	29	☑
Dereference of null pointer	110	☑
Dereference of undefined pointer value	1	☑
Division by zero	9	☑
Result of operation is garbage or undefined	58	☑
Uninitialized argument value	46	☑
Unix API		
Undefined allocation of 0 bytes (CERT MEM04-C; CWE-131)	1	☑

http://brunocardoso.cc/guadec13/static/

Sanitizers

- Memory
- Address
- Instrumentation modules
- Runtime libraries

Address Sanitizer

- Out-of-bounds accesses to heap, stack and globals
- Use-after-free, Use-after-return (to some extent), double-free, invalid free
- No reports in Glade and Gedit

```
$ clang -fsanitize=address ...
```

2x slowdown

Memory Sanitizer

- Detector of uninitialized reads
- 3x slowdown
- In tested projects ‘make’ fails - uninitialized reads from glib

```
$ clang -fsanitize=memory
```

3x slowdown

Questions?