

Compiler Construction

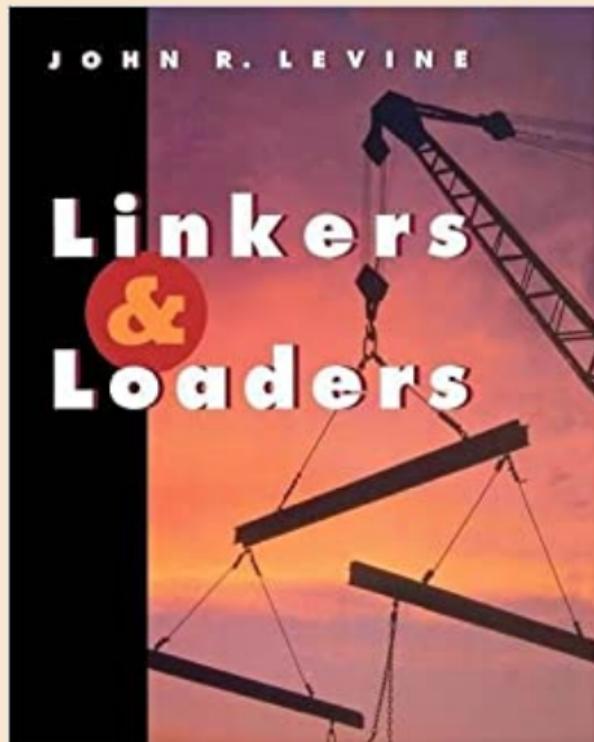
Lecture 15: Linking

Jeremy Yallop

`jeremy.yallop@cl.cam.ac.uk`

Lent 2025

Recommended book



Linkers & Loaders

John Levine

1st edition (October 25, 1999)

ISBN: 1558604960

Application Binary Interfaces

What happens after compilation?

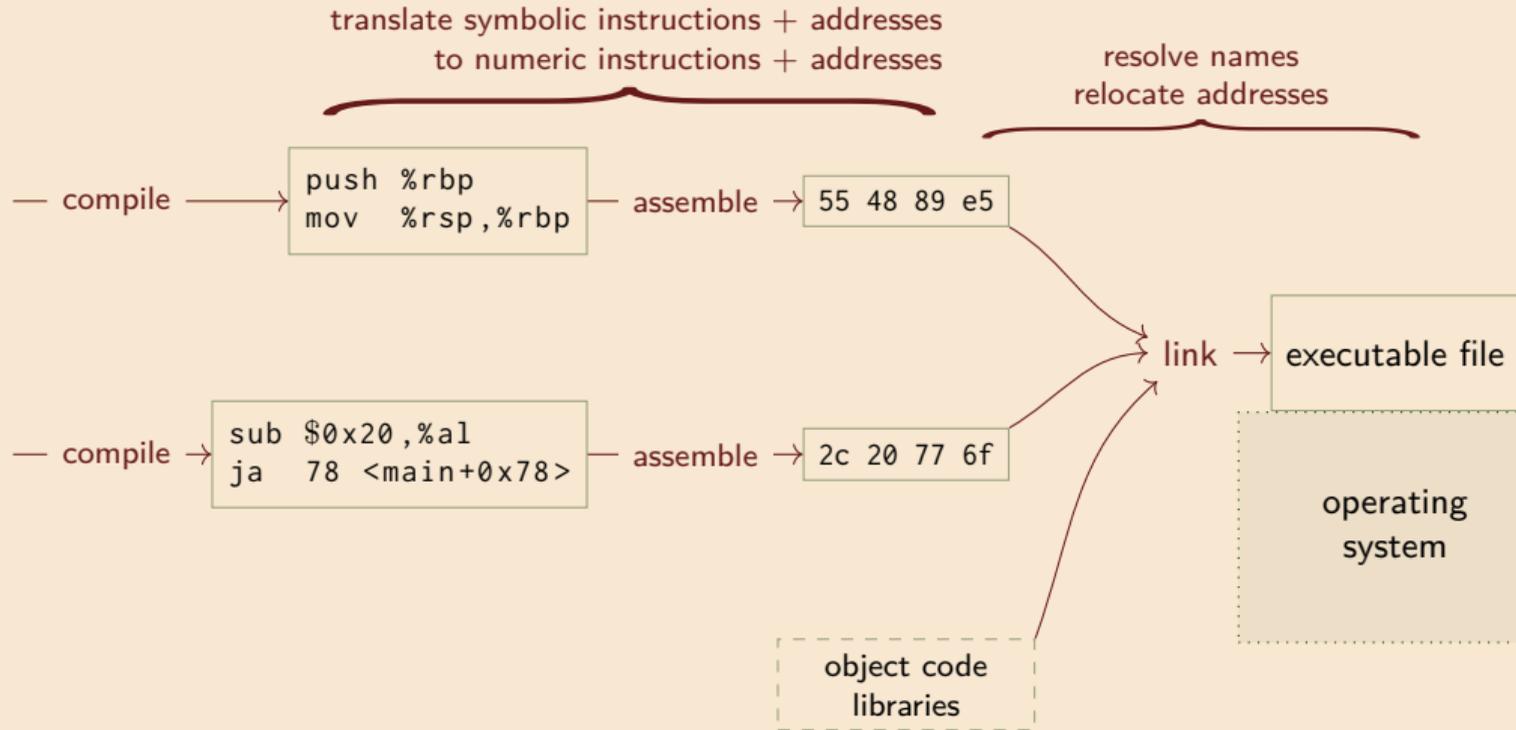
ABI



Object files

Linking

Runtime



Application Binary Interface (ABI)

ABI



Object files

Linking

Runtime

ABI: conventions that programs on a particular OS must follow:

- set of system calls (open, read, write, etc.)
- procedure for invoking the system calls
- what memory addresses a program can use
- how registers are used (e.g. passing parameters, returning results)
- stack frame layout
- data layout: endianness, alignment, etc.
- object file layout (e.g. ELF)
- linking, loading, name mangling



From *System V Application Binary Interface: AMD64 Architecture Processor Supplement*:

The control bits of the `MXCSR` register are callee-saved (preserved across calls), while the status bits are caller-saved (not preserved). The x87 status word register is caller-saved, whereas the x87 control word is callee-saved.

3.2.2 The Stack Frame

In addition to registers, each function has a frame on the run-time stack. This stack grows downwards from high addresses. Figure 3.3 shows the stack organization.

The end of the input argument area shall be aligned on a 16 (32 or 64, if `__m256` or `__m512` is passed on stack) byte boundary.¹¹ In other words, the stack needs to be 16 (32 or 64) byte aligned immediately before the call instruction is executed. Once control has been transferred to the function entry point, i.e. immediately after the return address has been pushed, `%rsp` points to the return address, and the value of $(\%rsp + 8)$ is a multiple of

Object files

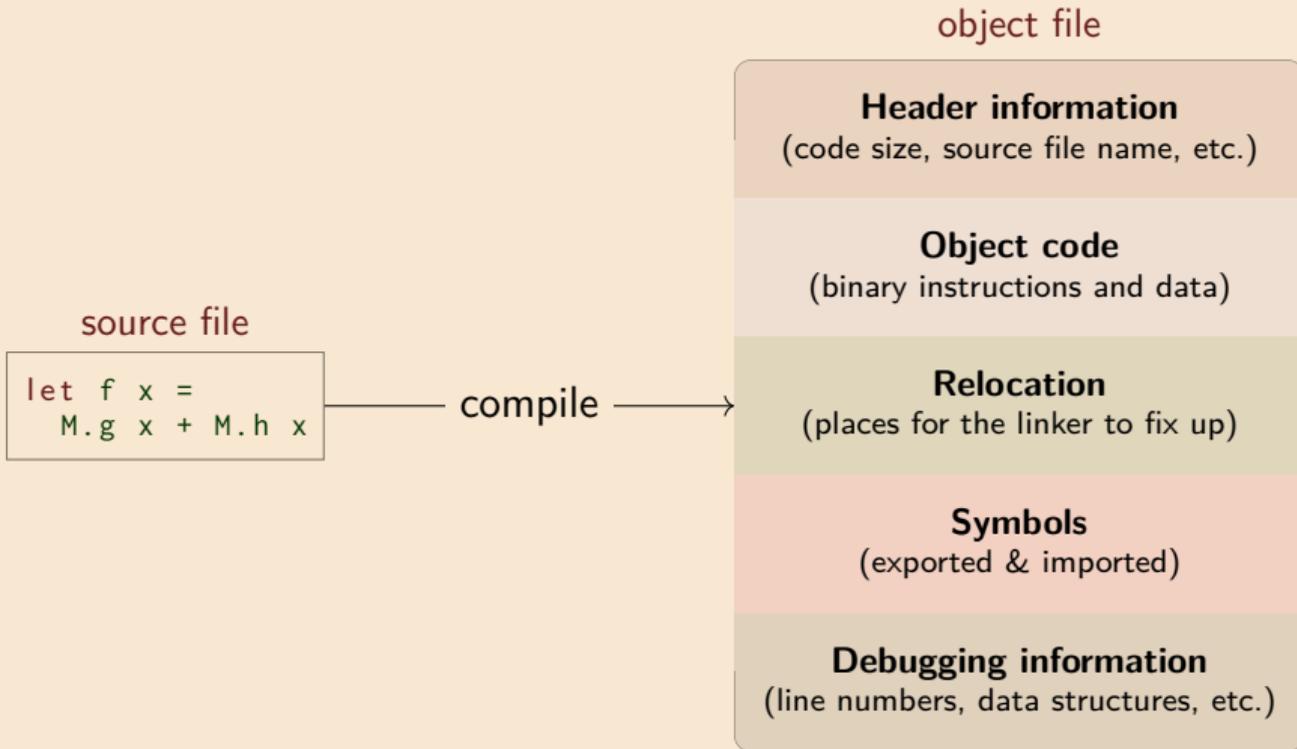
ABI

Object files



Linking

Runtime



ELF (executable and linkable format)

ABI

ELF is a common format for both linker input and output. Sections (not complete):

ELF header	
<code>.text</code>	code segment
<code>.data</code>	writable global data
<code>.rodata</code>	read-only global data
<code>.bss</code>	uninitialized data size
<code>.sym</code>	symbol table
<code>.rel.text</code> <code>.rel.data</code> <code>.rel.rodata</code>	relocation tables: (offset, symbol) pairs
<code>.line</code>	maps source lines to object code locations
<code>.debug</code>	debugging information
<code>.strtab</code>	string names of symbols

Object files



Linking

Runtime

ABI

Object files



Linking

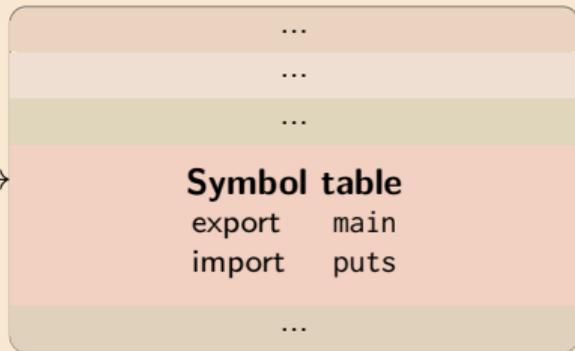
Runtime

The symbol table in an object file may include various types of symbols:

- Global symbols defined (& perhaps referenced) in the module
- Global symbols referenced, but not defined
- Segment names
- Optional: non-global symbols, line number information (for debugging purposes)

```
#include <stdio.h>
int main() {
    puts("Hello, world\n");
}
```

compile →



Linking

(Static) linking

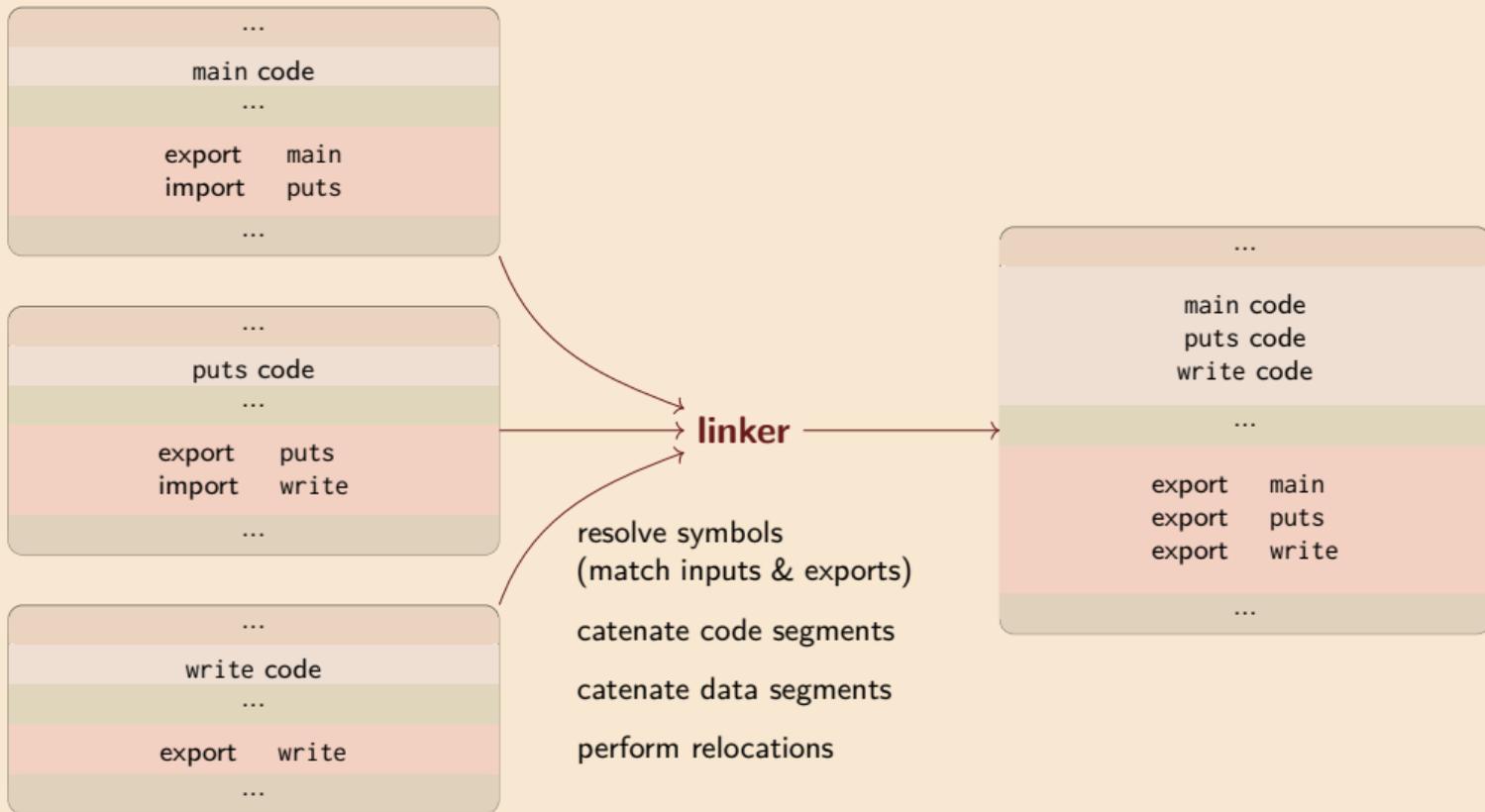
ABI

Object files

Linking



Runtime



Static vs dynamic linking

ABI

Linking may be **static** (compile-time) or **dynamic** (run-time).

Dynamic linking: object files contain stubs; the OS links the code on demand.

Static linking

- Executables are larger
- Libraries can't easily be changed (e.g. for bug fixes)
- + Libraries can't change unexpectedly (silently updating program behaviour)
- + Loading (starting) programs is faster

Dynamic linking

- + Executables are smaller
- + Libraries can easily be changed (e.g. for bug fixes)
- Libraries can change unexpectedly (silently updating program behaviour)
- Loading (starting) programs is slower

Linking



Runtime

Object files

Runtime systems

ABI

Runtime system: a library needed to run compiled code
Provides support for a particular language (“the OCaml runtime”)
Implemented for a particular operating system

Object files

The runtime system may offer:

- an interface between the language and the operating system (**system calls**)
- an interface to other languages (**foreign function interface**)
- efficient implementations of **primitive operations**
- runtime type checking, method lookup, security checks, &c.

Linking

Runtime



Targeting a VM vs targeting a platform

ABI

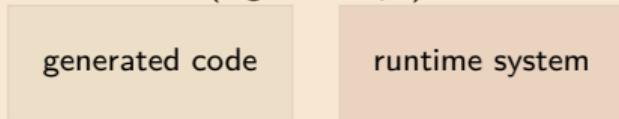
Targeting a VM

(e.g. ocamlc)

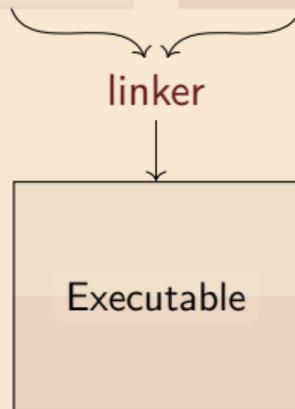
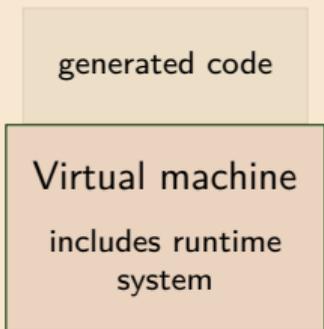
Targeting a platform

(e.g. ocamlpt)

Object files



Linking



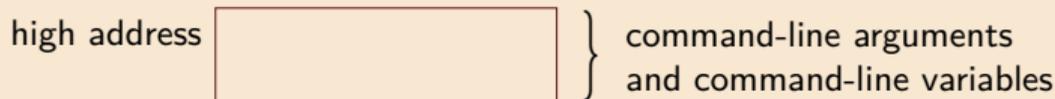
Runtime

In both cases: compiler & runtime implementers must agree on low-level details (memory layout, data representation)



Typical memory layout (UNIX)

ABI



stack



heap

uninitialized data
(bss)

} initialized to
zero by exec

initialized data

} read from
program file
by exec

low address

text

Object files

Linking

Runtime



(Adapted from *Advanced Programming in the Unix Environment*, W. Richard Stevens)

Next time: bootstrapping