

Processes

Abhilash Jindal

Agenda

Agenda

- Vocabulary (OSTEP Ch. 4)
 - What is a process? System calls? Scheduler? Address space?

Agenda

- Vocabulary (OSTEP Ch. 4)
 - What is a process? System calls? Scheduler? Address space?
- Memory management (OSTEP Ch. 13-17)
 - How to manage and isolate memory? What are memory APIs? How are they implemented?

Agenda

- Vocabulary (OSTEP Ch. 4)
 - What is a process? System calls? Scheduler? Address space?
- Memory management (OSTEP Ch. 13-17)
 - How to manage and isolate memory? What are memory APIs? How are they implemented?
- Processes in action (xv6 Ch. 3: system calls, x86 protection, trap handlers)
 - Process control block, user stack<>kernel stack, sys call handling

Agenda

- Vocabulary (OSTEP Ch. 4)
 - What is a process? System calls? Scheduler? Address space?
- Memory management (OSTEP Ch. 13-17)
 - How to manage and isolate memory? What are memory APIs? How are they implemented?
- Processes in action (xv6 Ch. 3: system calls, x86 protection, trap handlers)
 - Process control block, user stack<>kernel stack, sys call handling
- Scheduling (xv6 Ch5: context switching, OSTEP Ch. 6-9)
 - Response time, throughput, fairness

Process is a running program

- Load program from disk to memory
 - Exactly how we loaded OS
- Give control to the process. Jump cs, eip

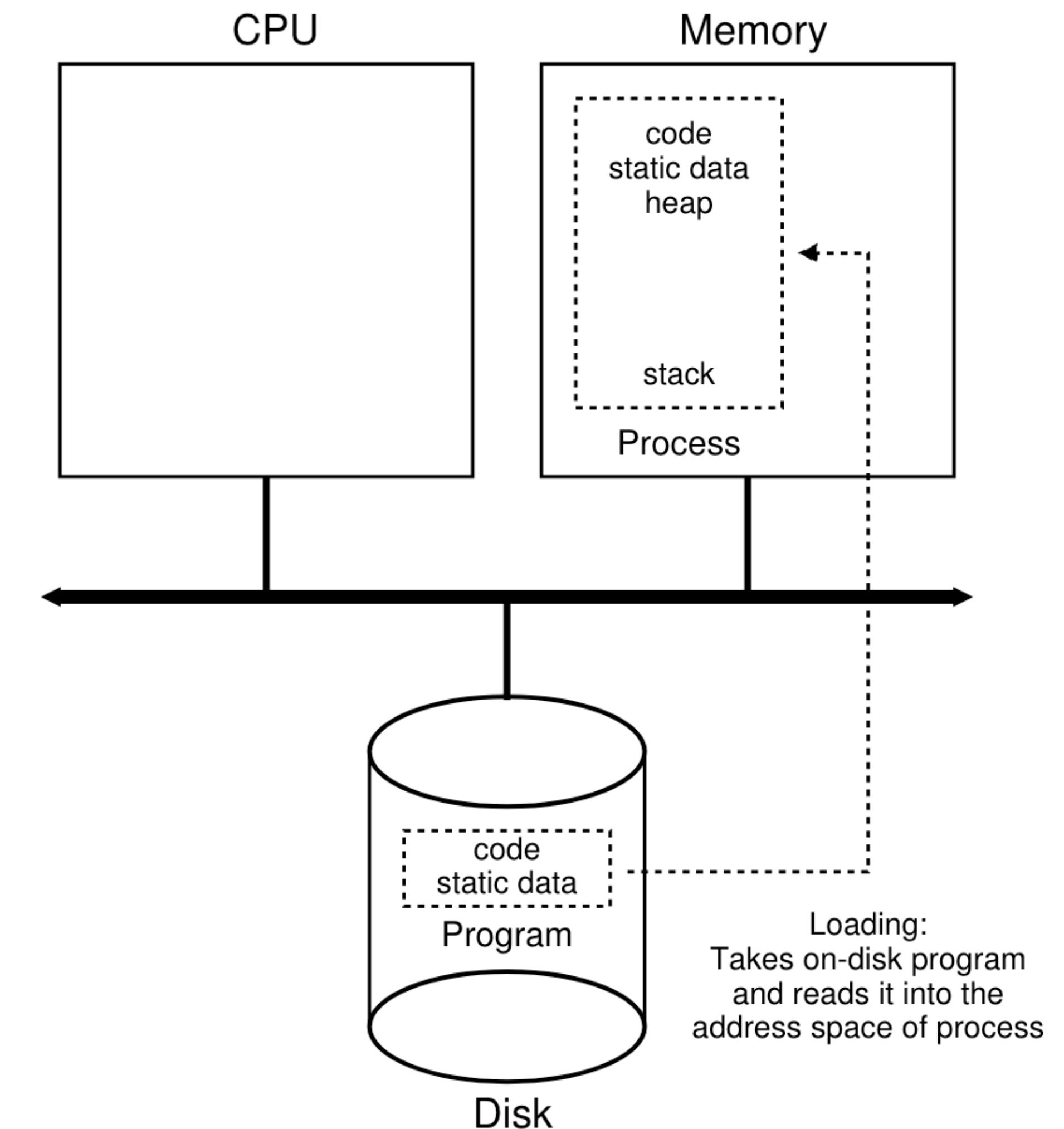


Figure 4.1: Loading: From Program To Process

Processes can ask OS to do work for them

System calls

```
$ strace cat /tmp/foo
...
openat(AT_FDCWD, "/tmp/foo", 0_RDONLY) = 3
read(3, "hi\n", 131072)                = 3
write(1, "hi\n", 3)                    = 3
...
```

OS maintains process states

Scheduler switches between processes

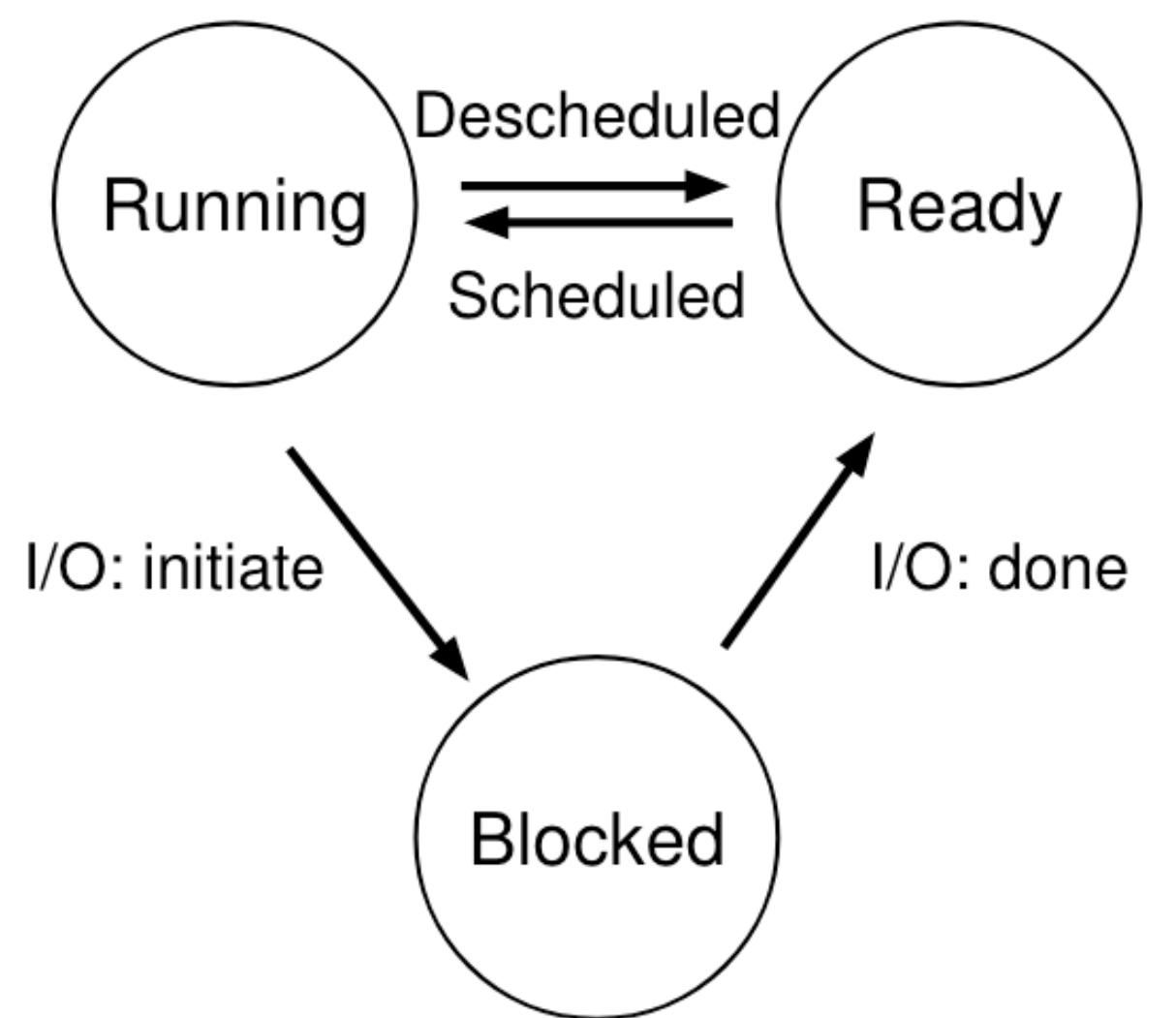


Figure 4.2: Process: State Transitions

OS maintains process states

Scheduler switches between processes

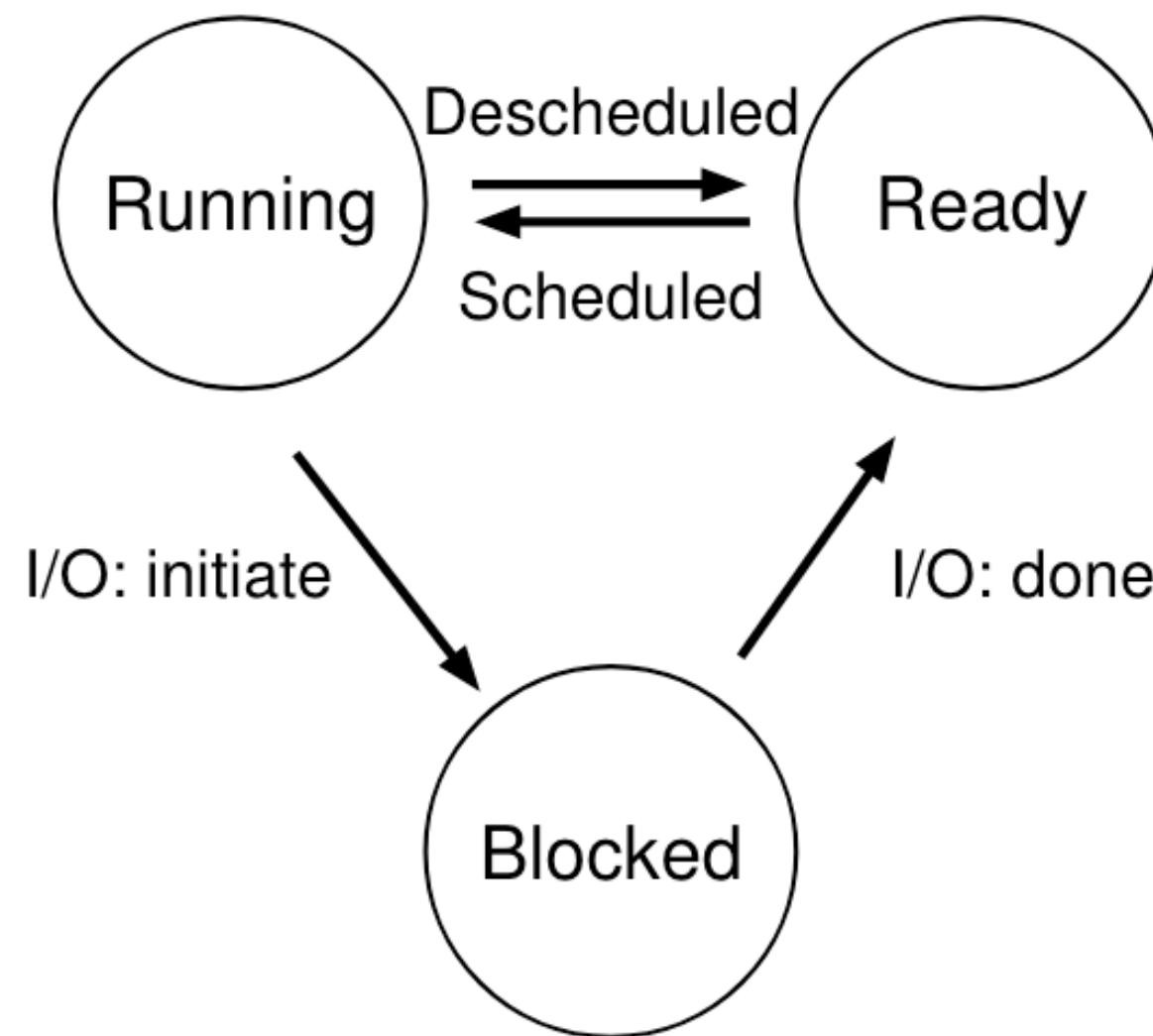


Figure 4.2: Process: State Transitions

Time	Process ₀	Process ₁	Notes
1	Running	Ready	
2	Running	Ready	
3	Running	Ready	
4	Blocked	Running	Process ₀ initiates I/O
5	Blocked	Running	Process ₀ is blocked, so Process ₁ runs
6	Blocked	Running	
7	Ready	Running	I/O done
8	Ready	Running	Process ₁ now done
9	Running	—	
10	Running	—	Process ₀ now done

Figure 4.4: Tracing Process State: CPU and I/O

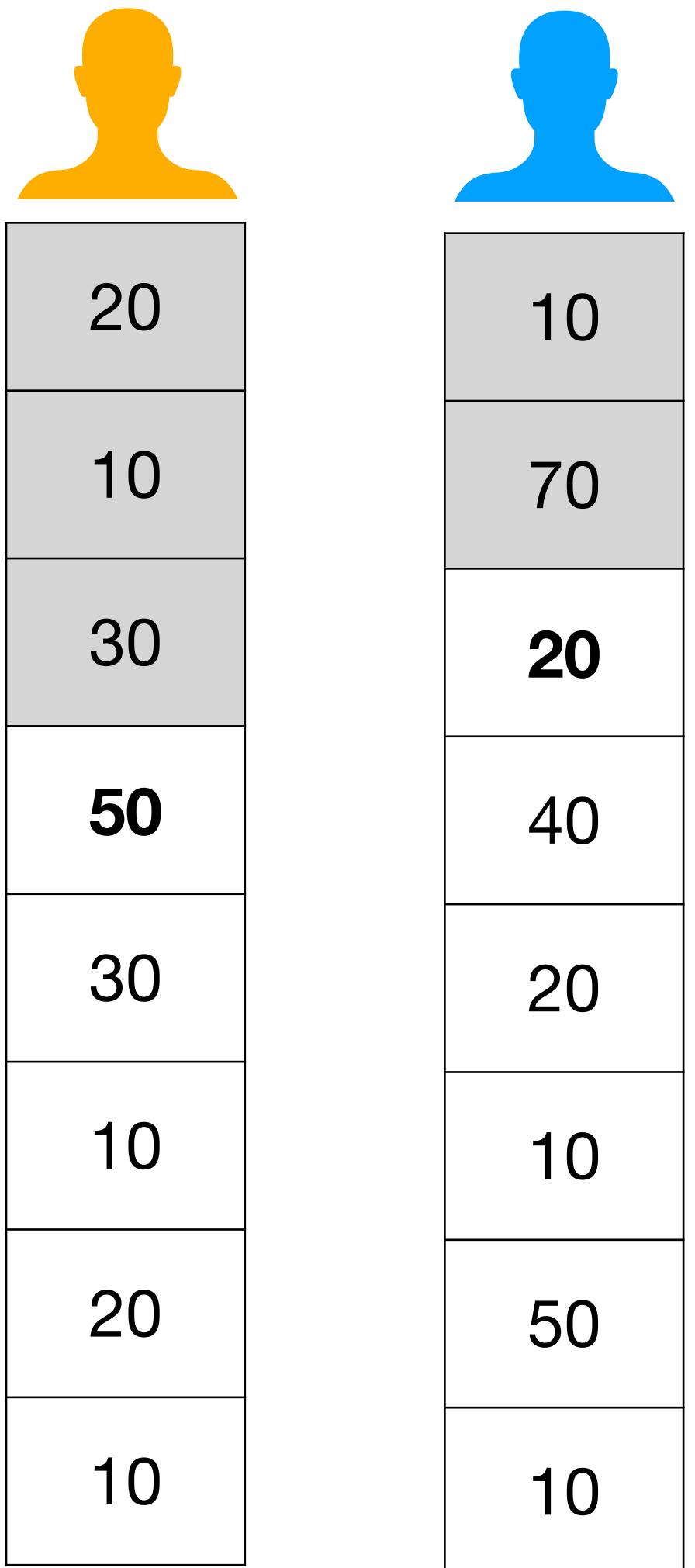
Calculator analogy: Computing long sum



20
10
30
50
30
10
20
10

- $2 \ 0 =$ (move pointer to 10)
- $+ 1 \ 0 =$ (move pointer to 30)
- $+ 3 \ 0 =$ (move pointer to 50)
- $+ 5 \ 0 =$ (move pointer to 30)
- $+ 3 \ 0 =$ (move pointer to 10)
- $+ 1 \ 0 =$ (move pointer to 20)
- $+ 2 \ 0 =$ (move pointer to 10)

Sharing the calculator



Sharing the calculator

20	10
10	70
30	20
50	40
30	20
10	10
20	50
10	10

- Steps to share the calculator:
 - $20 + 10 = 30 + 30 = 60$
 - Write 60 in notebook, remember that we were done till 30, give calculator
 - $10 + 70 = 80$
 - Write 80 in notebook, remember that we were done till 70, give the calculator back

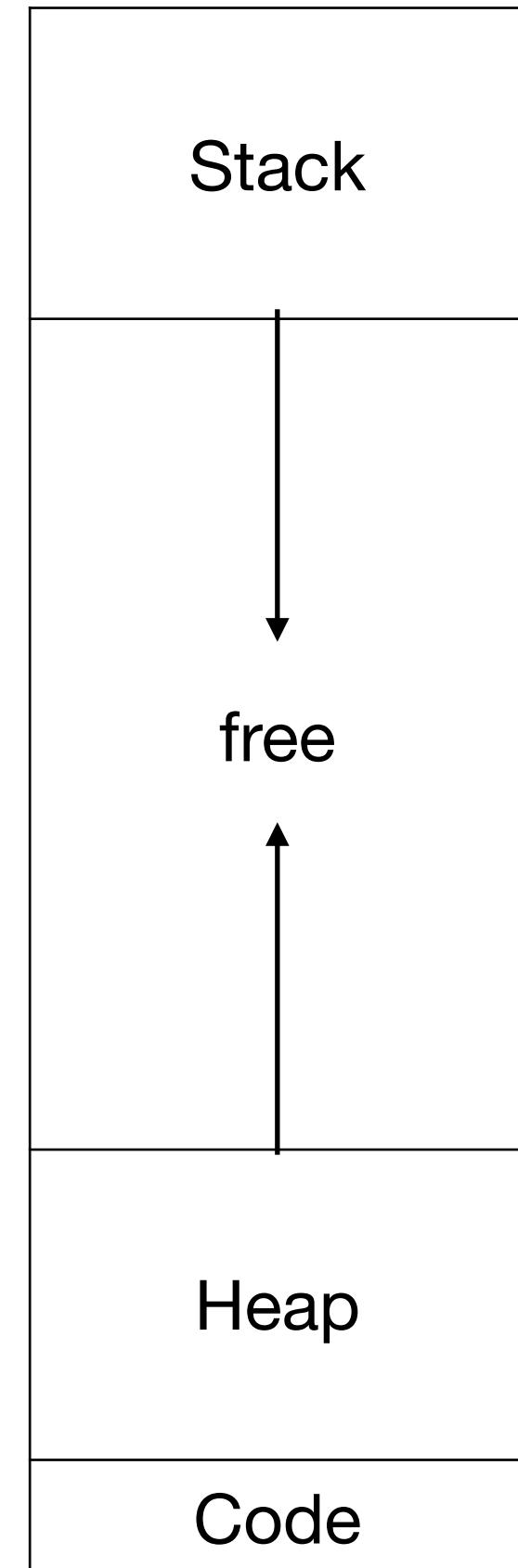
Memory isolation and management

OSTEP Ch. 13-17

Abhilash Jindal

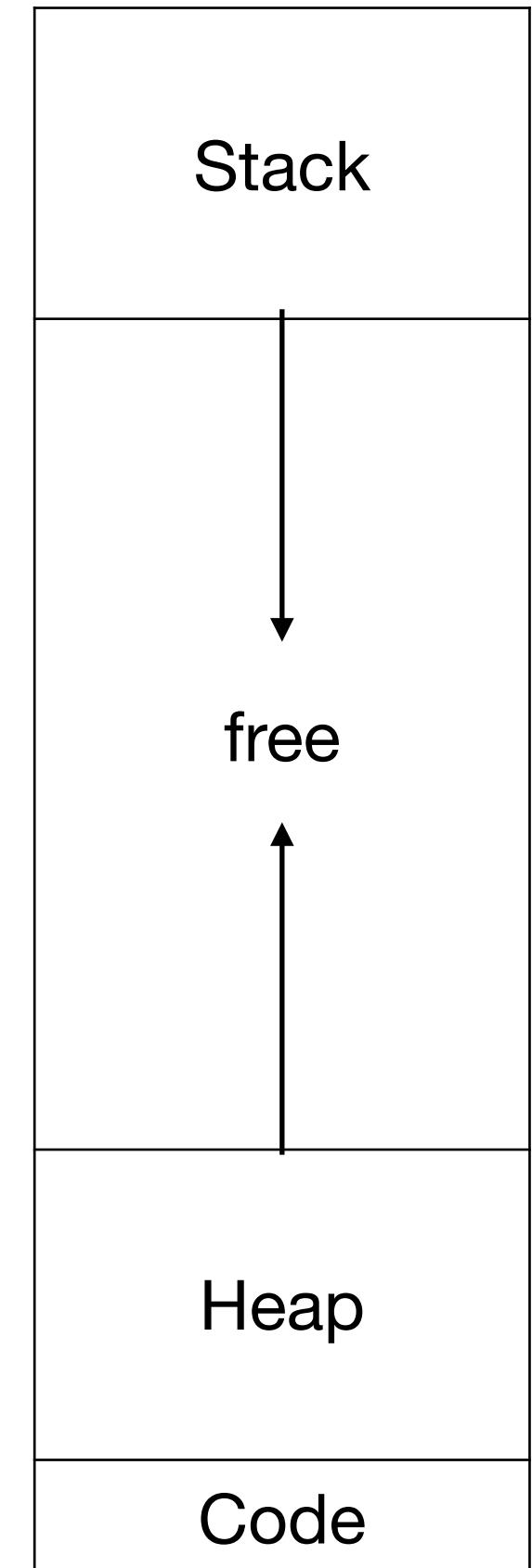
Process Address Space

Code, Heap, Stack



Process Address Space

Code, Heap, Stack



Process address space

Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

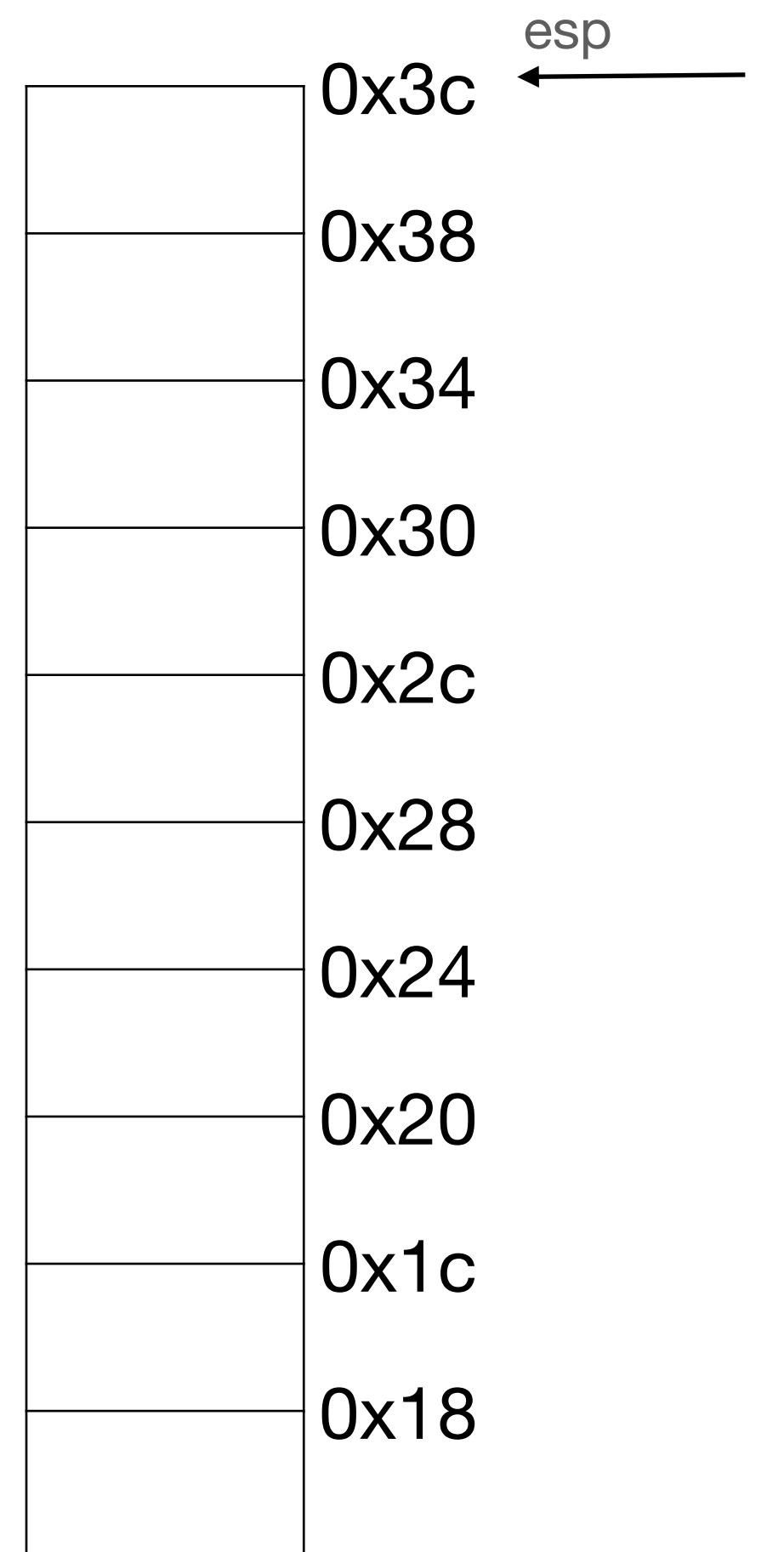
eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp

ebp →



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

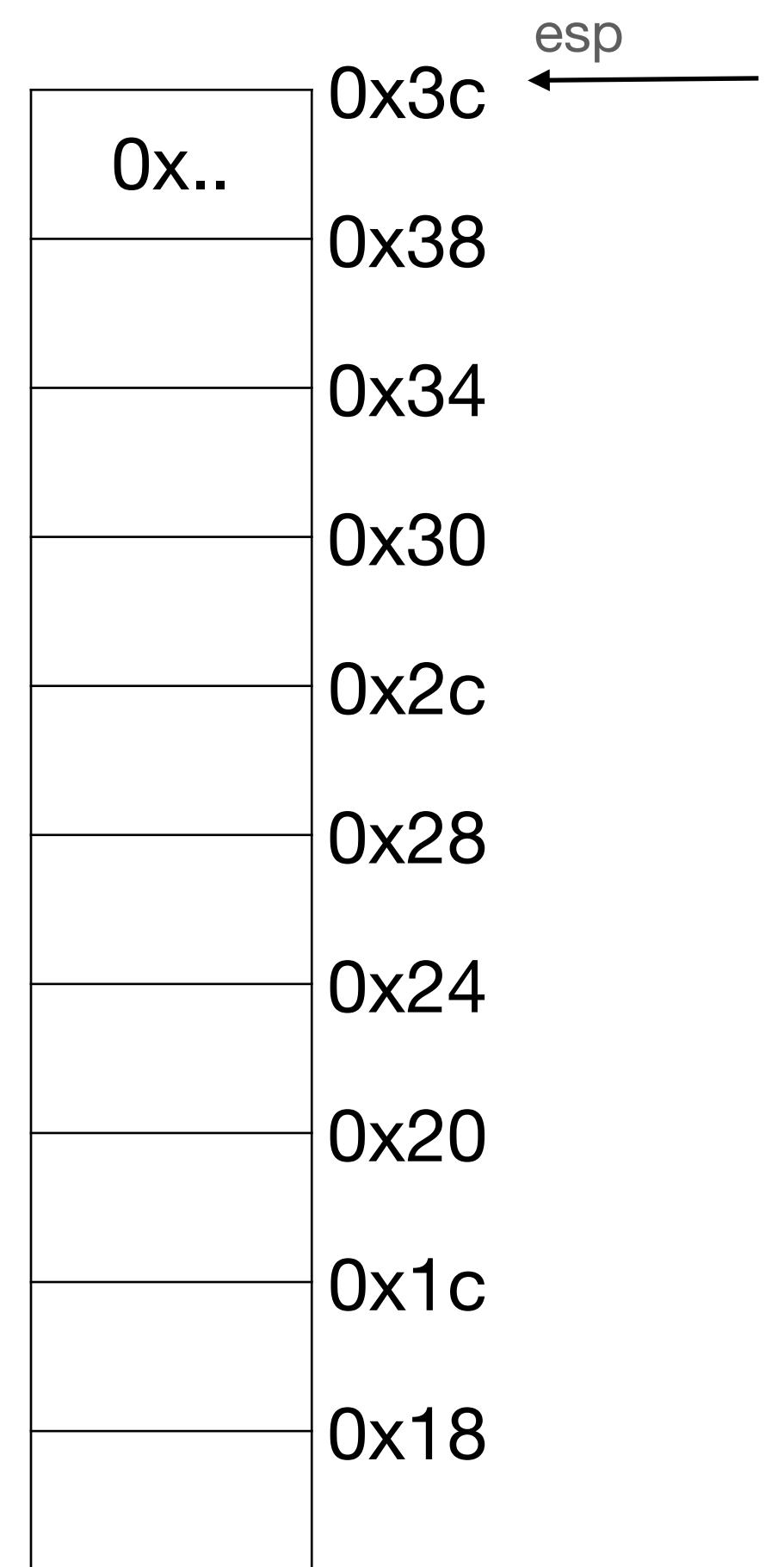
eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp

ebp →



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

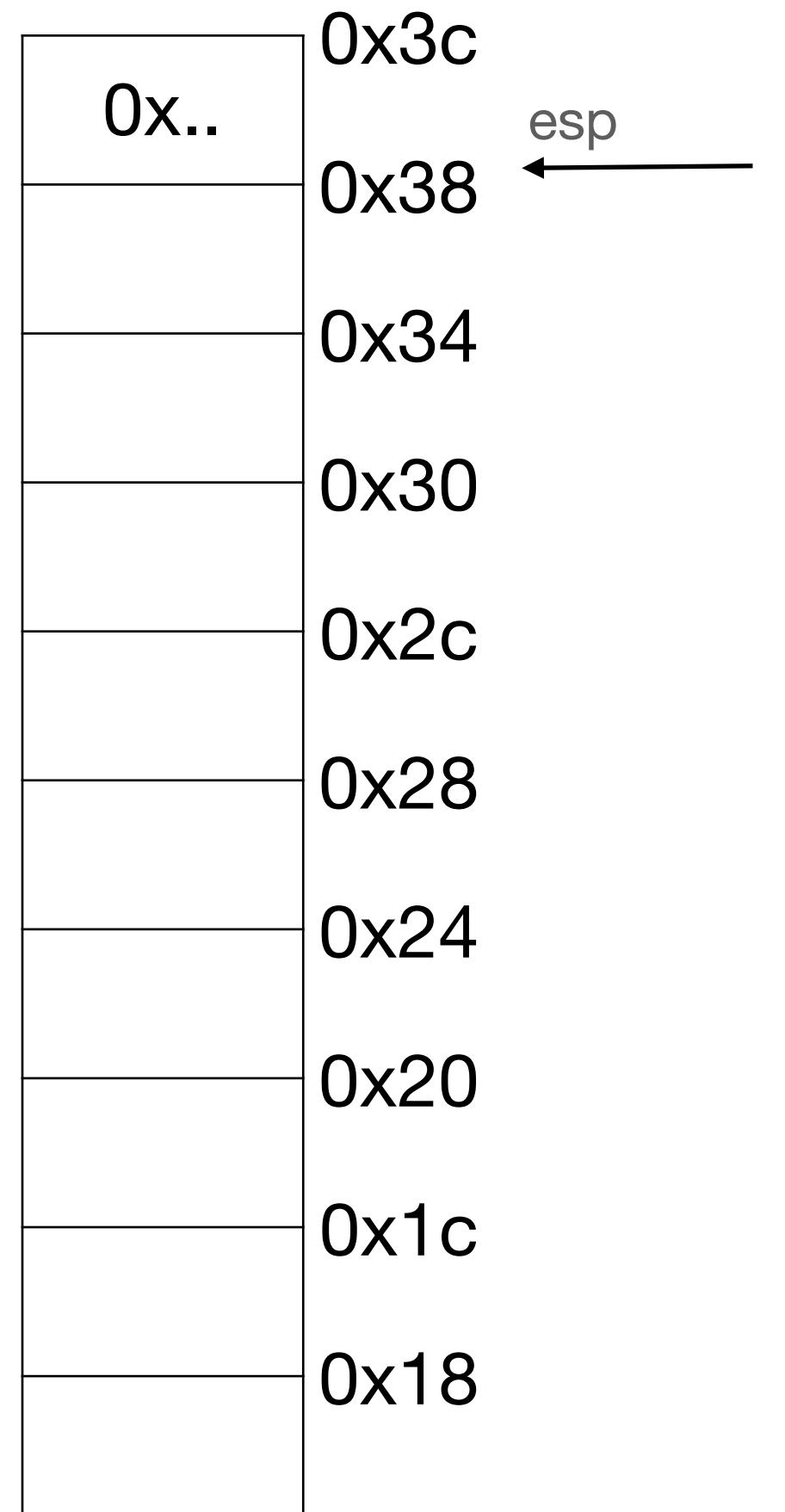
eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp

ebp →



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

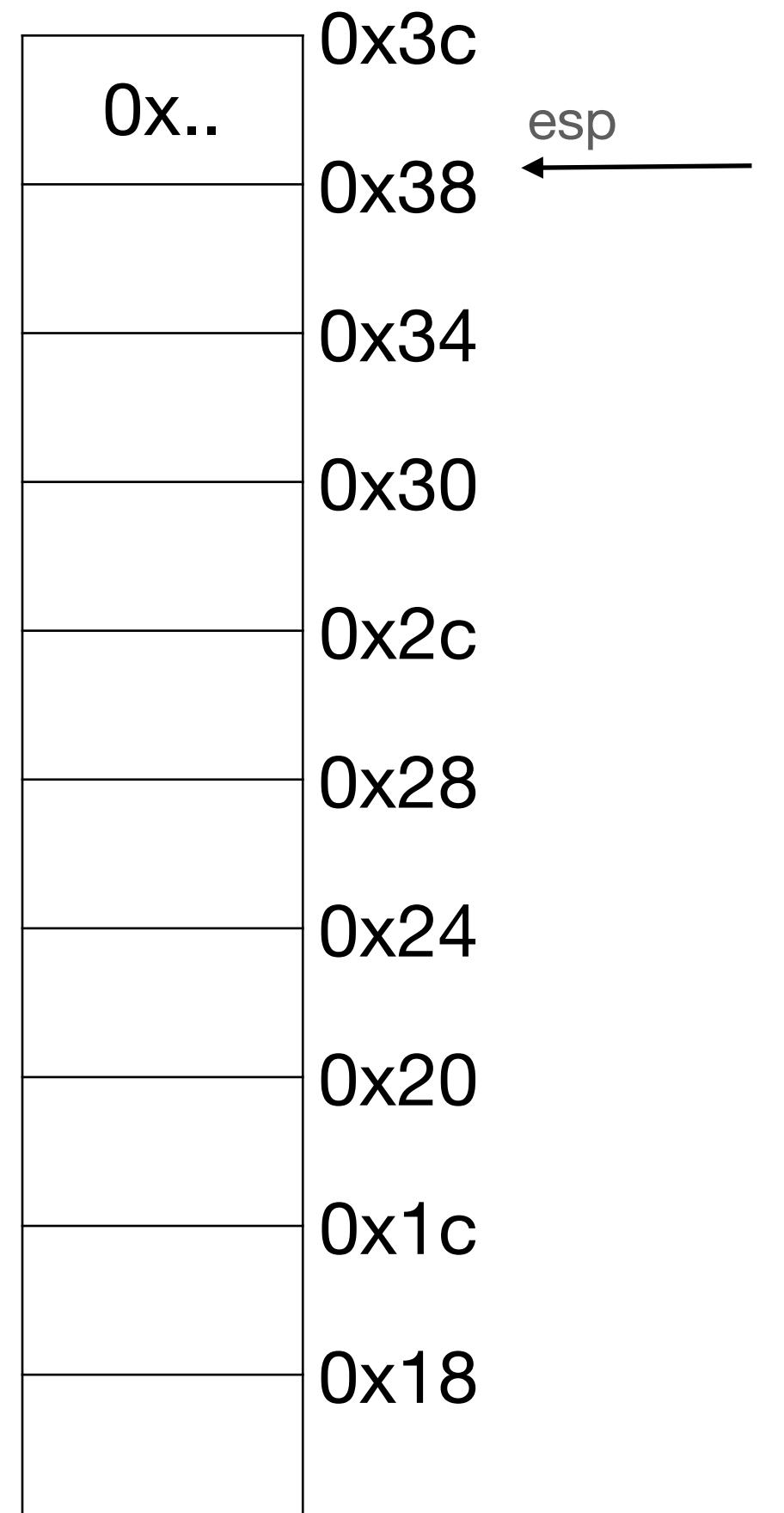
eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp

ebp →



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

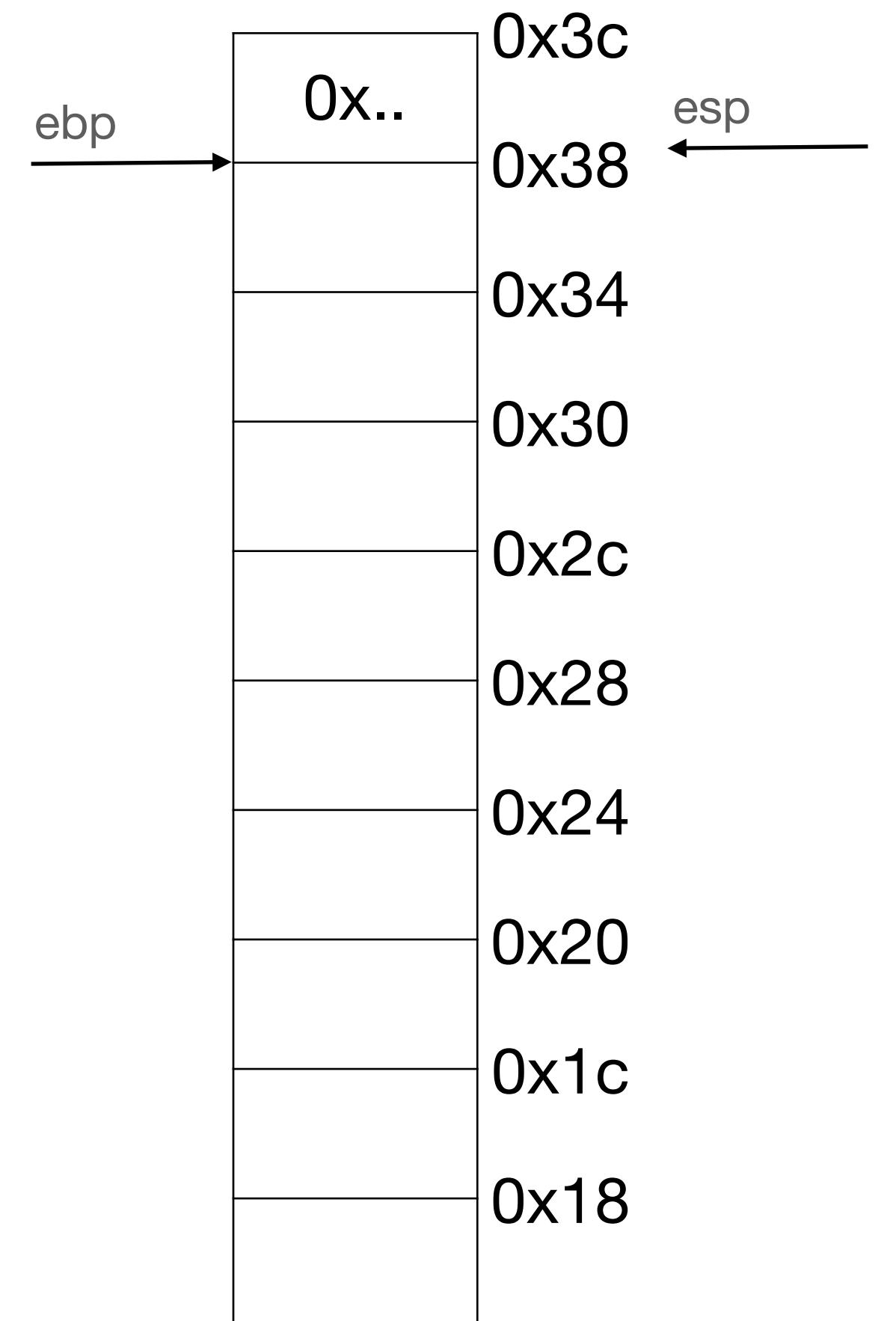
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

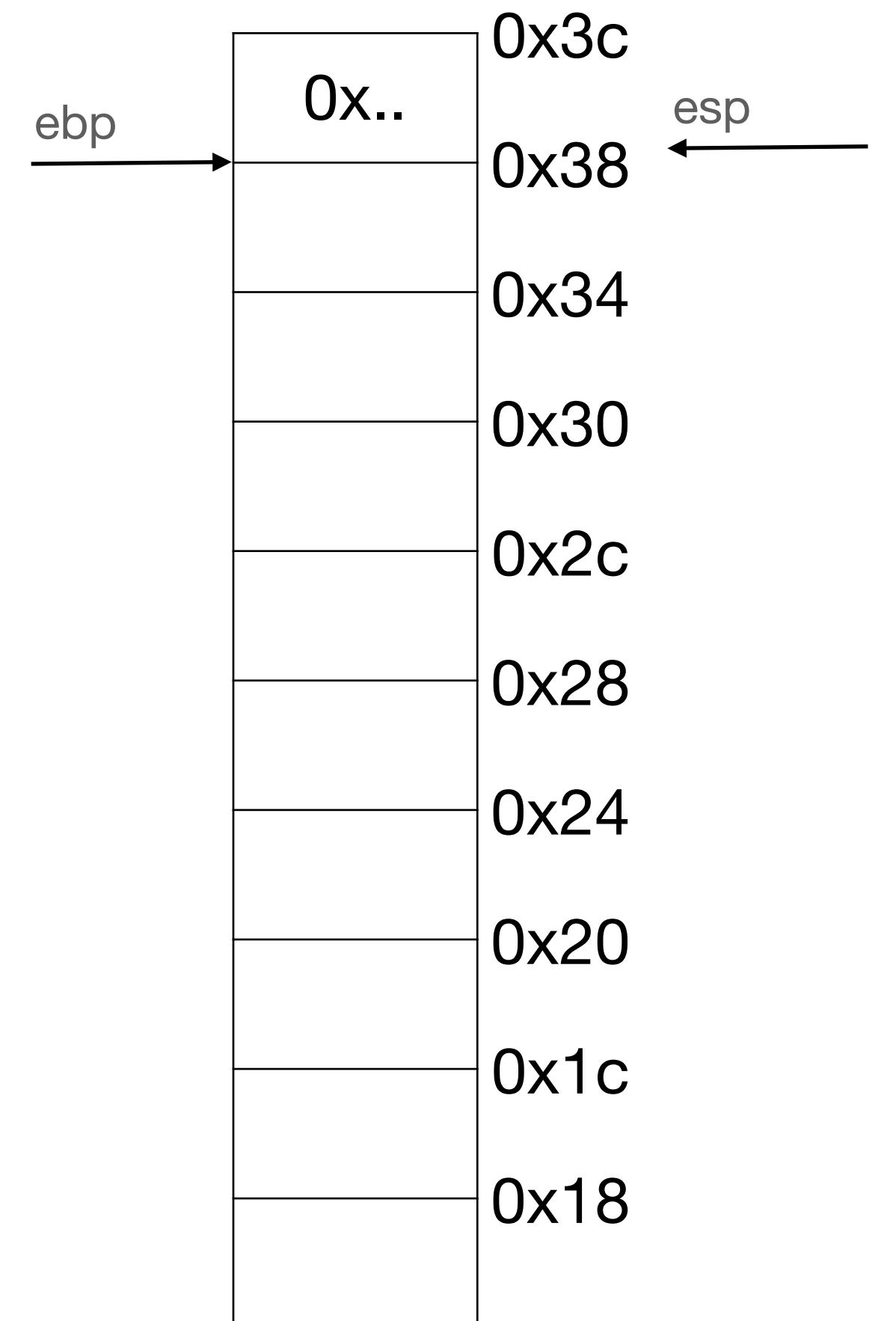
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

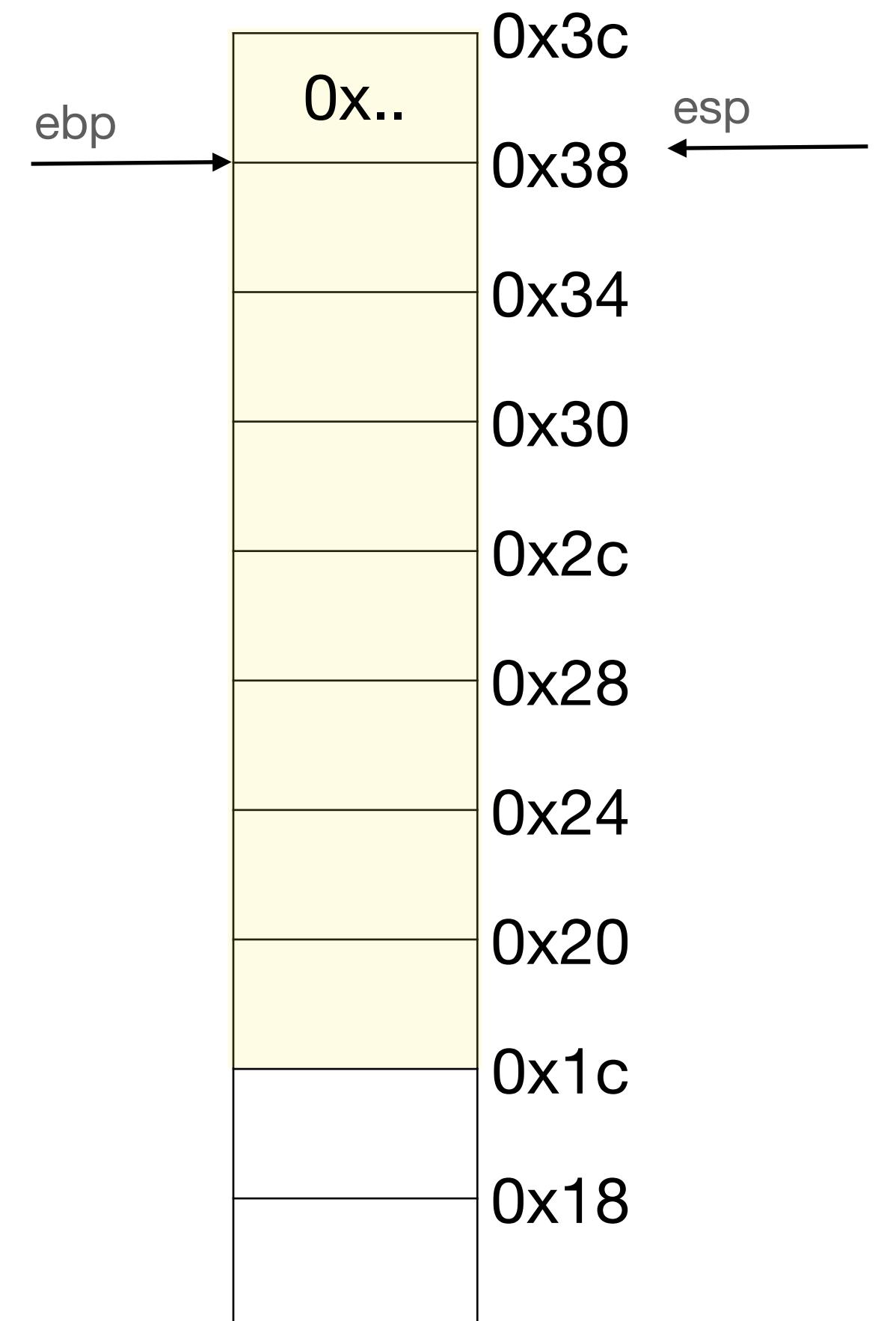
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

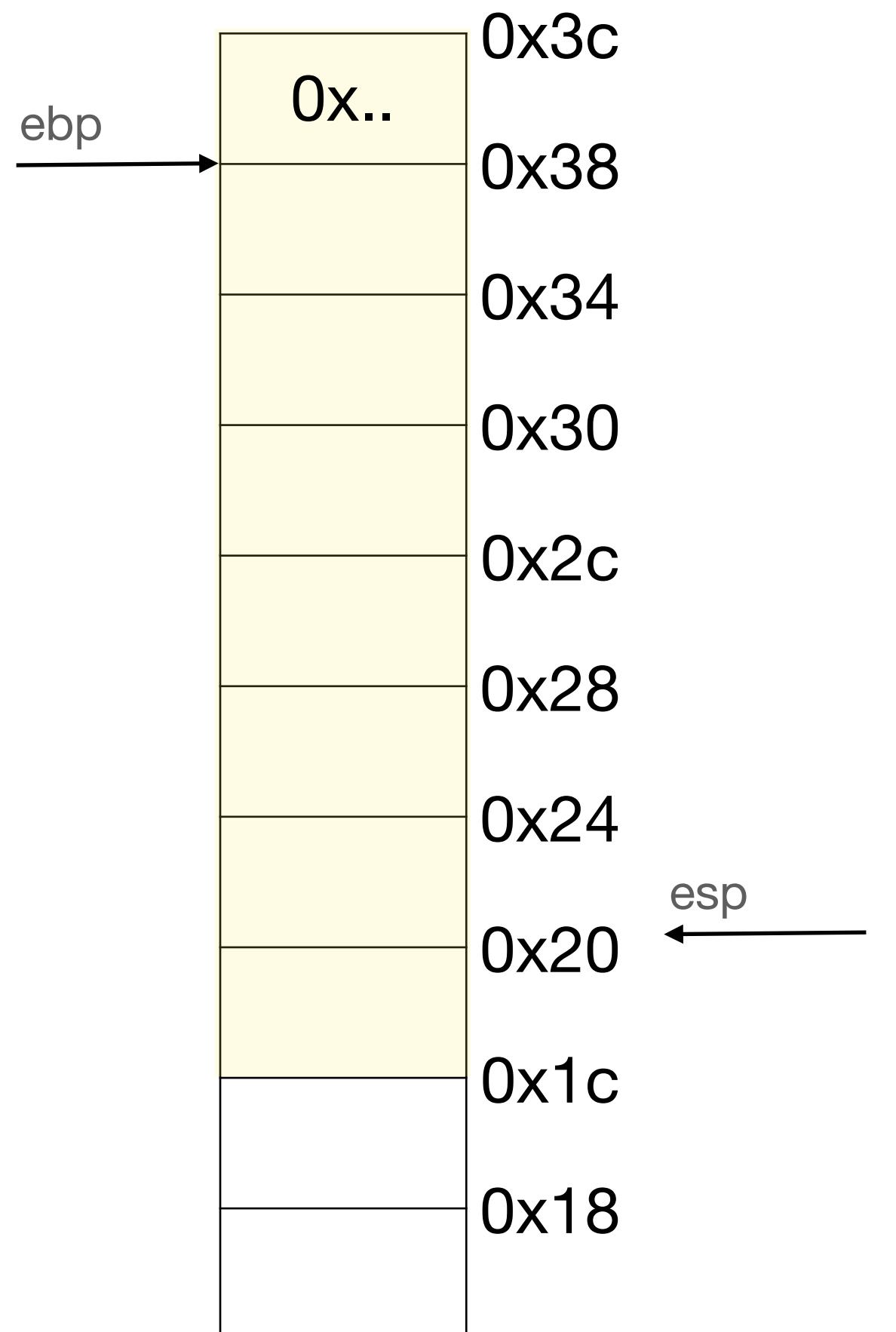
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

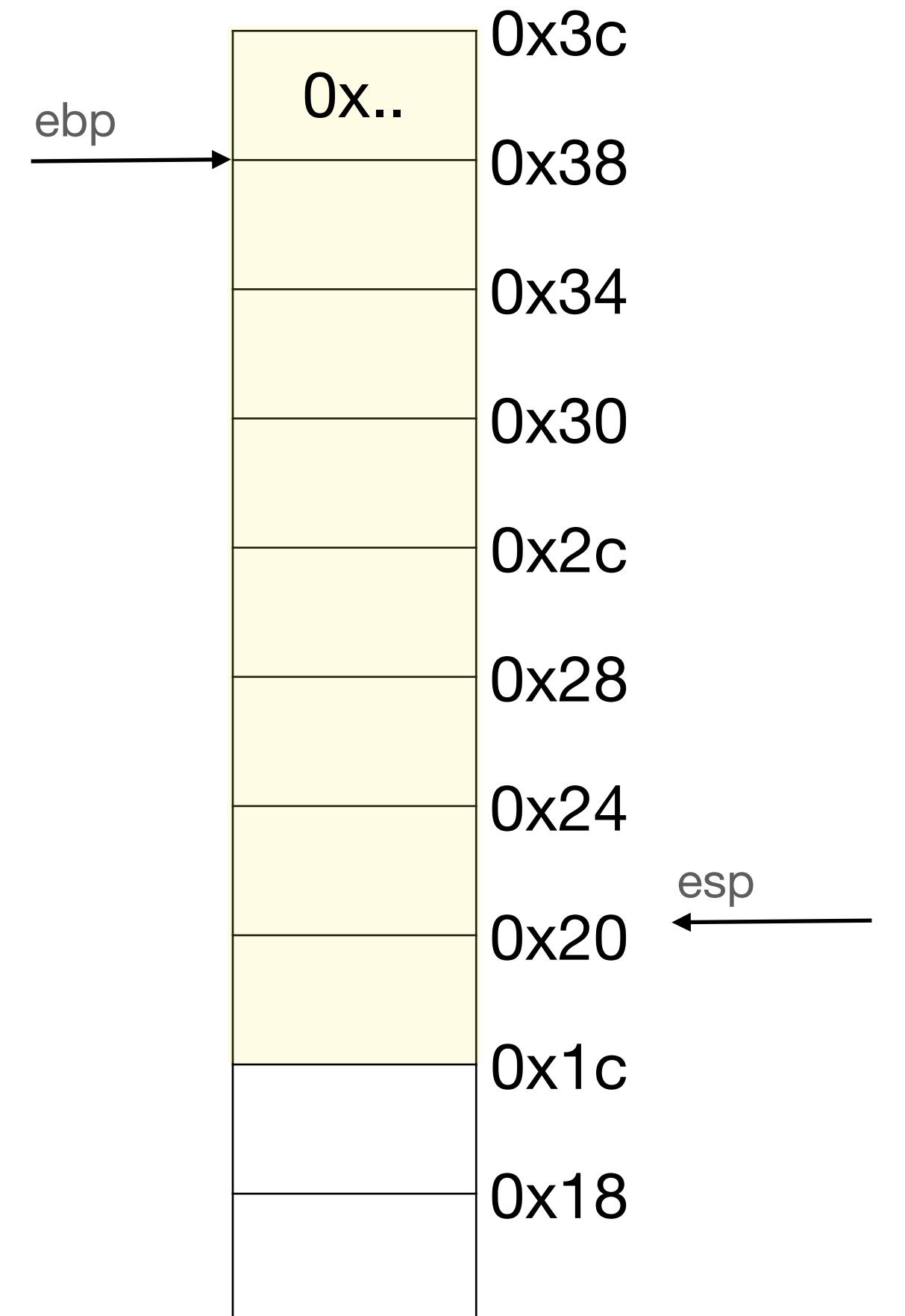
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

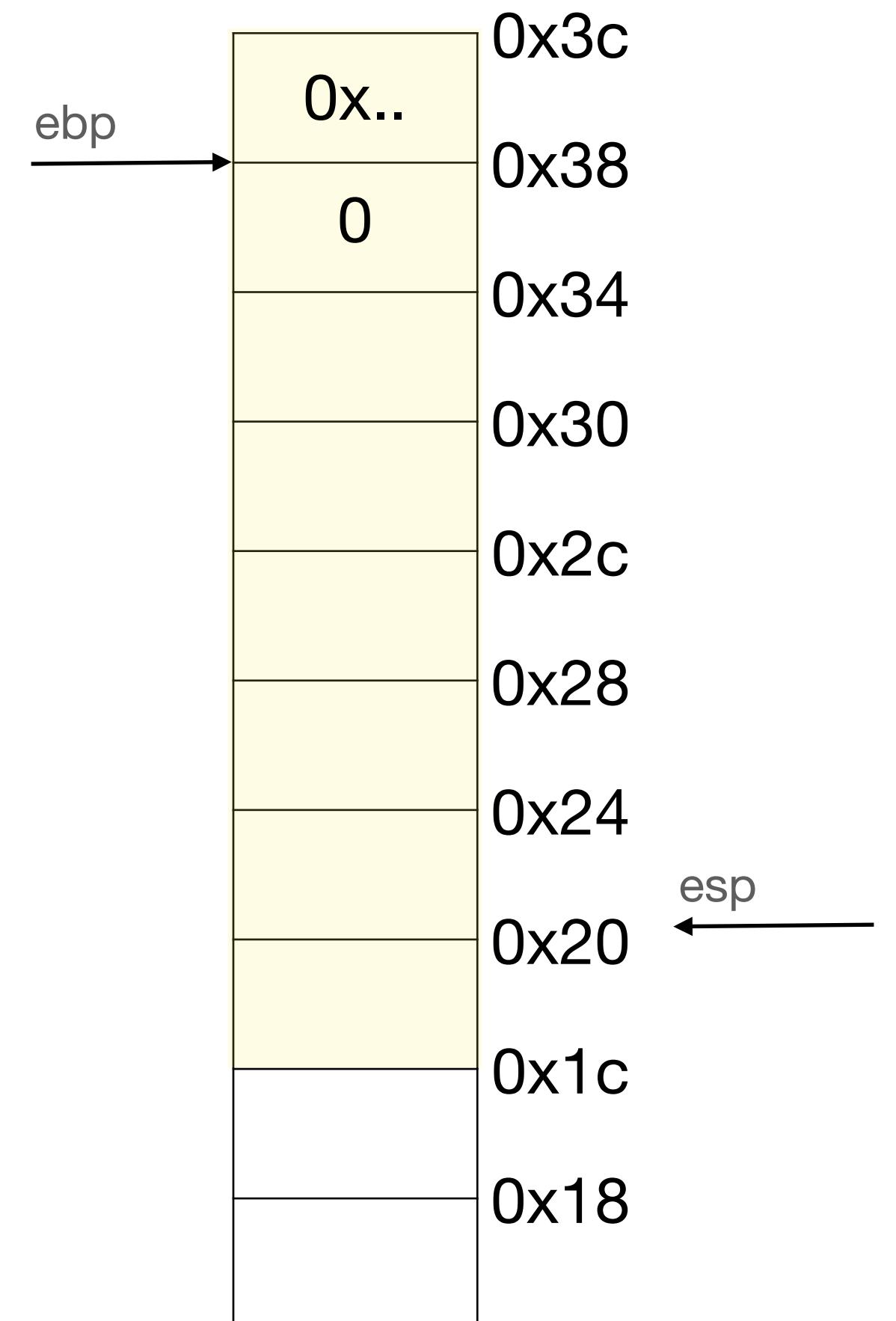
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

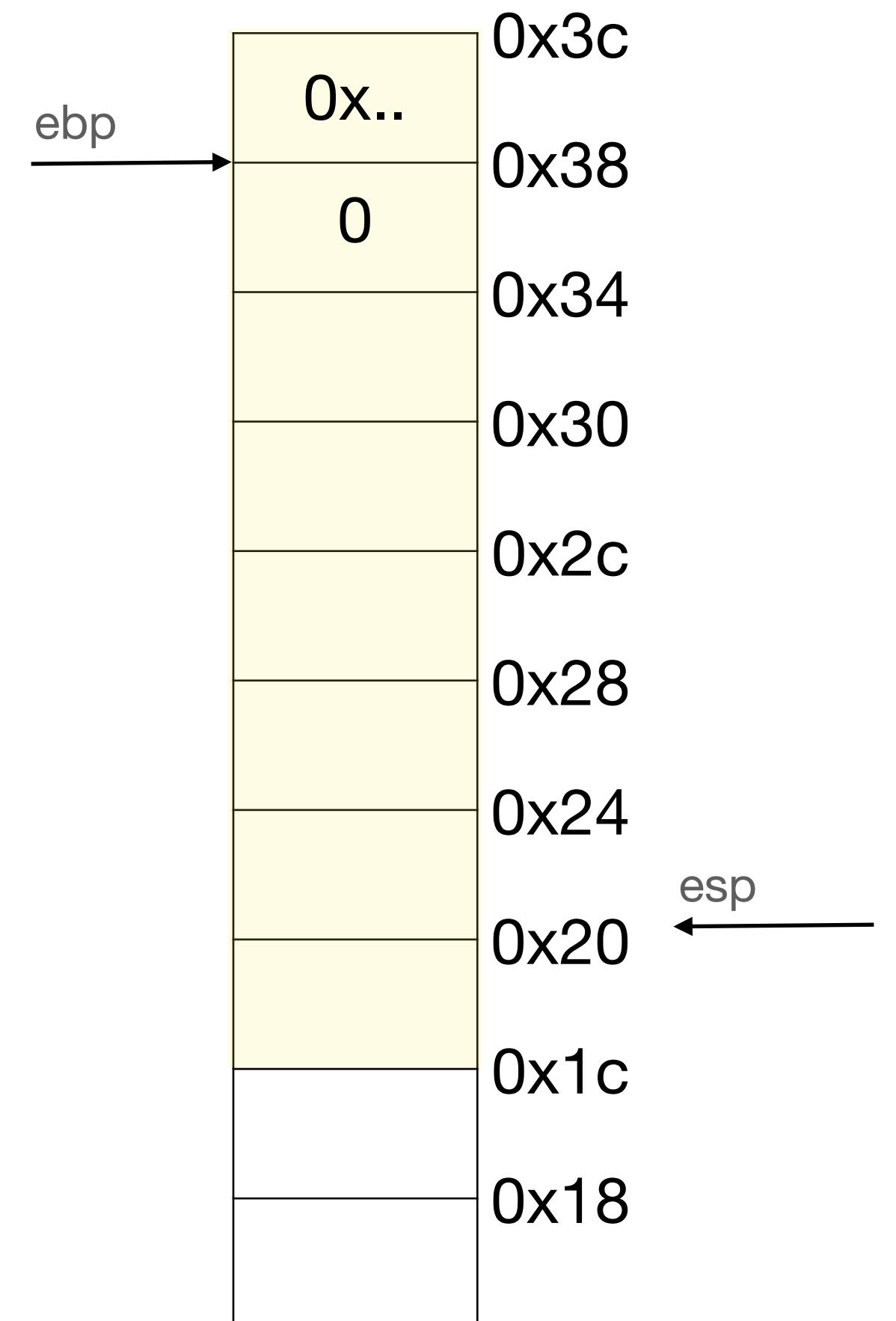
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

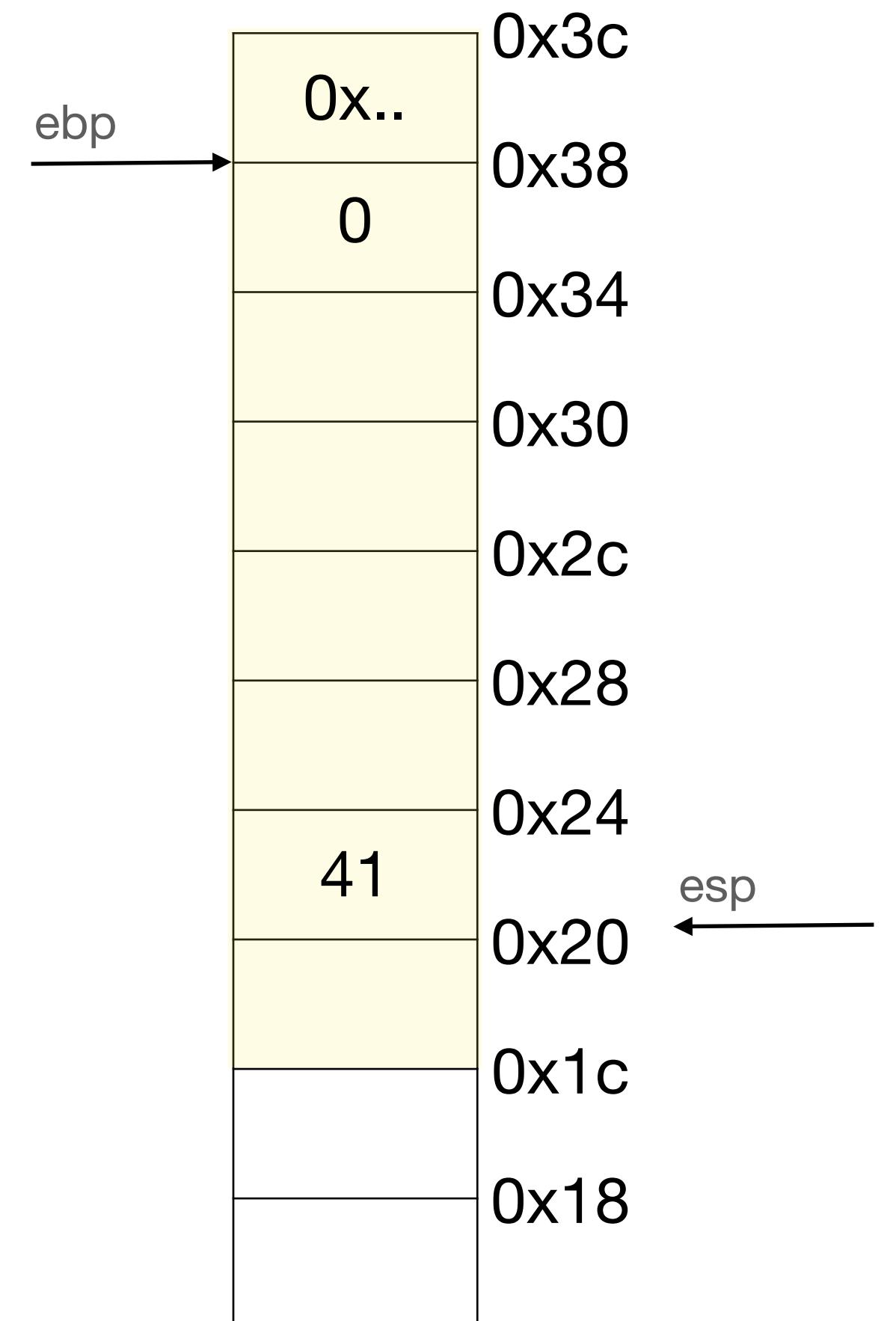
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

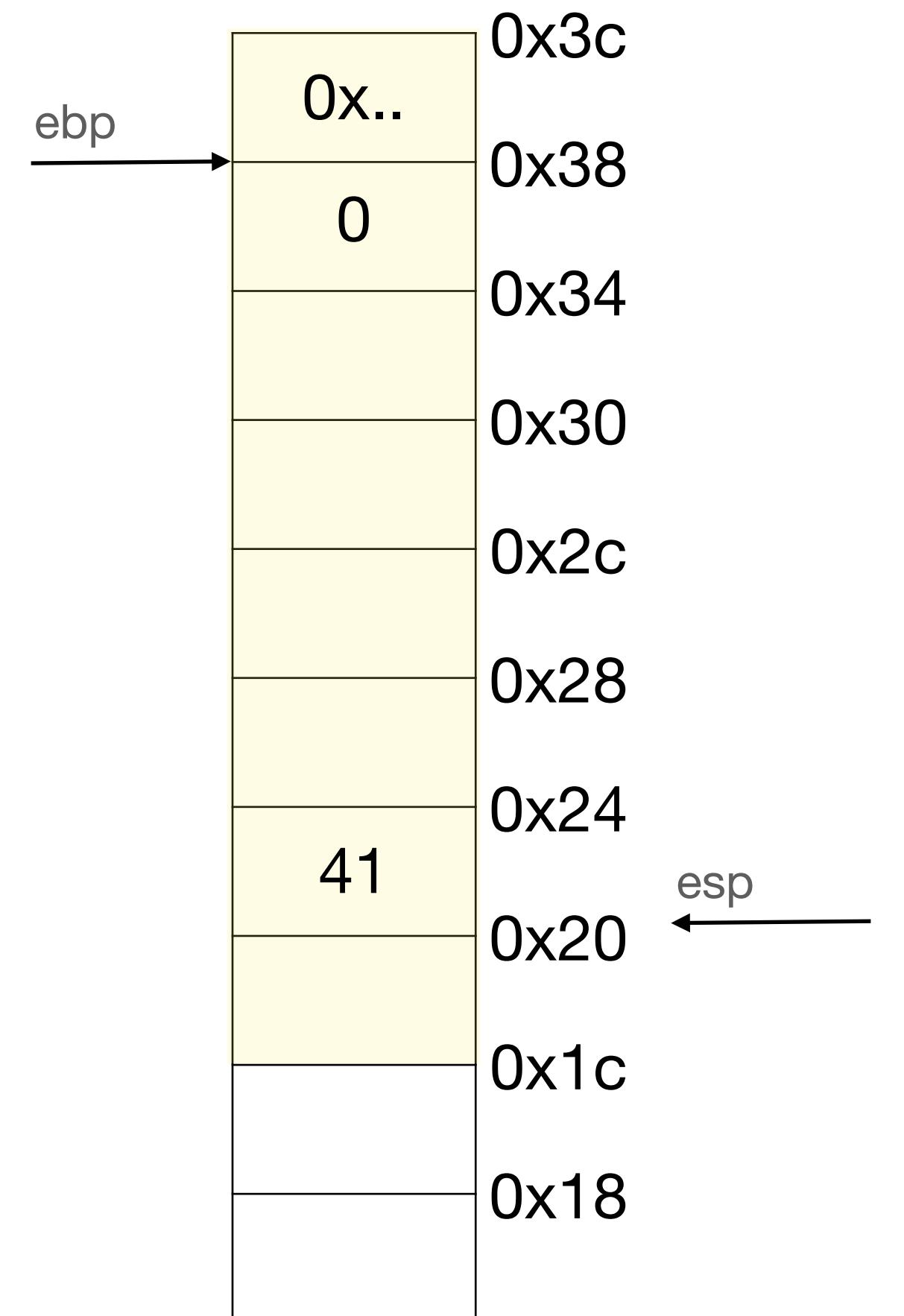
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

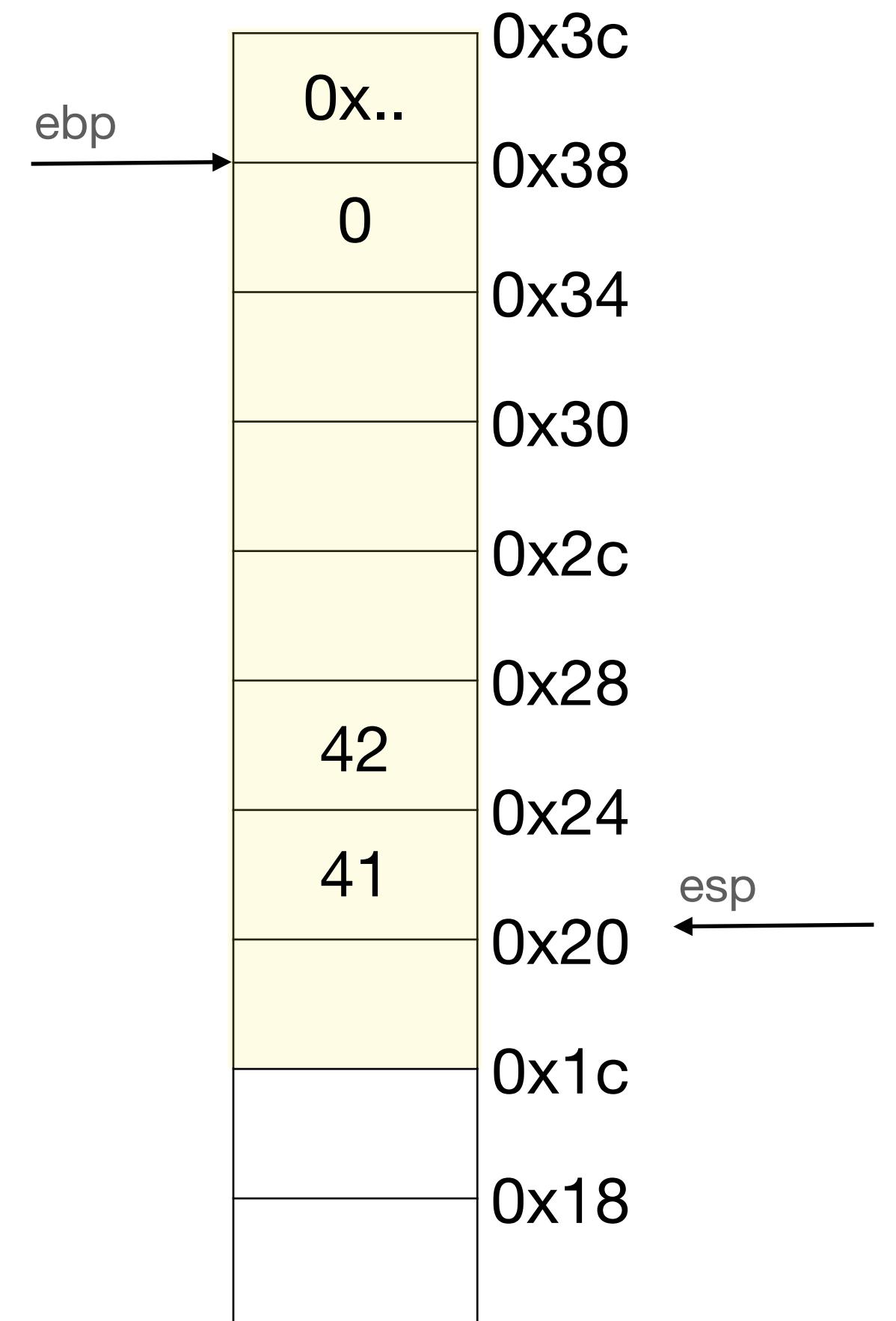
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

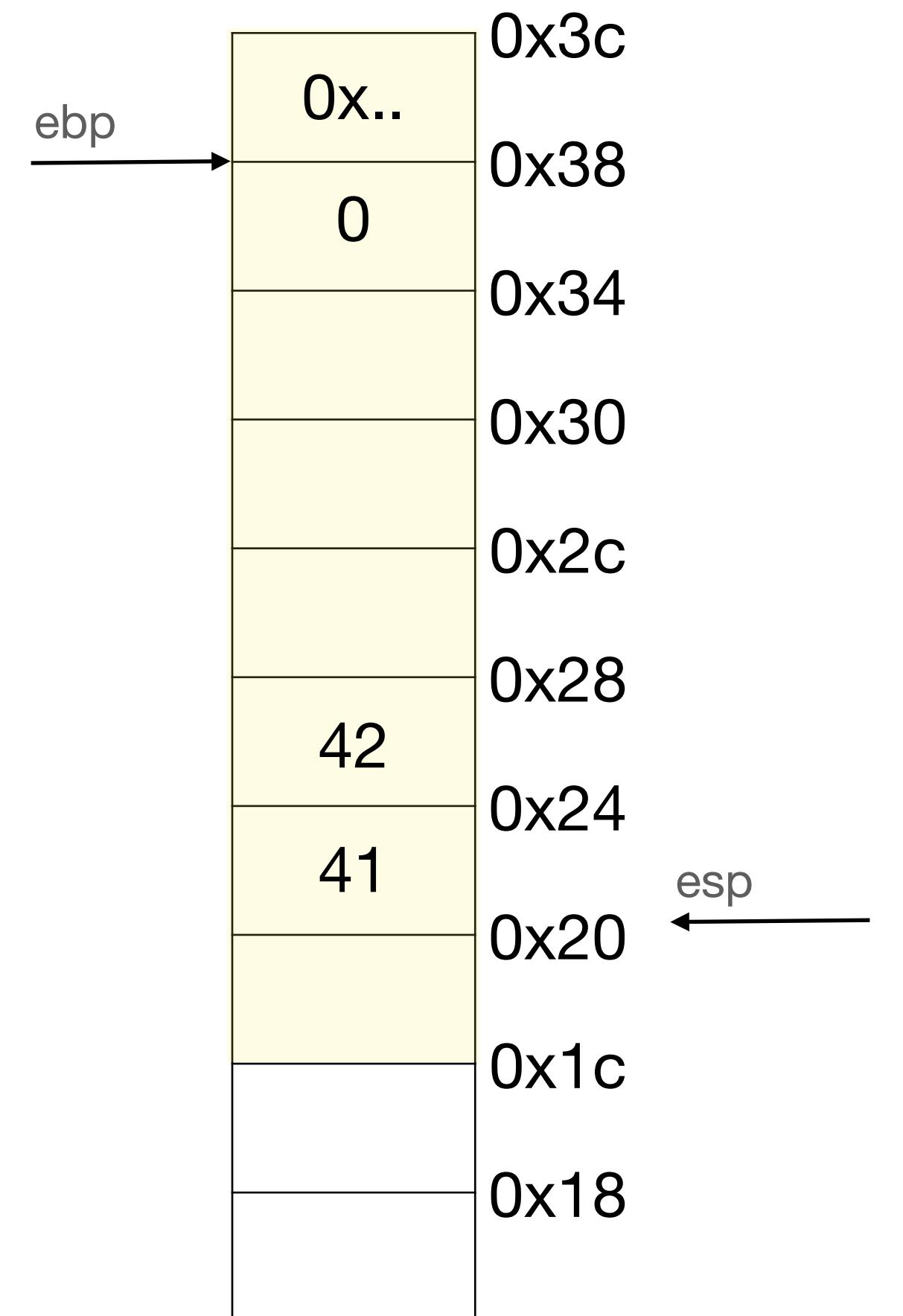
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

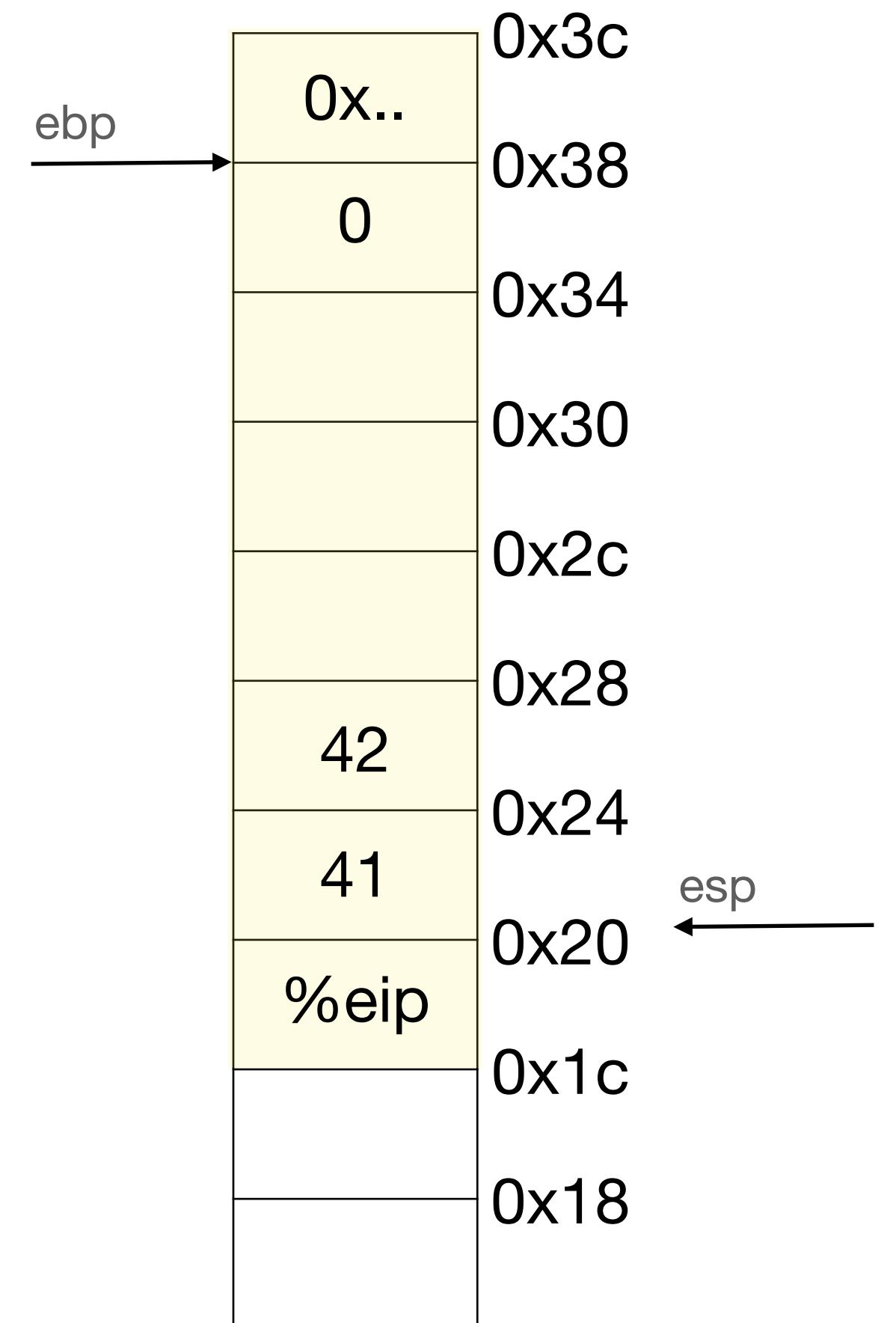
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

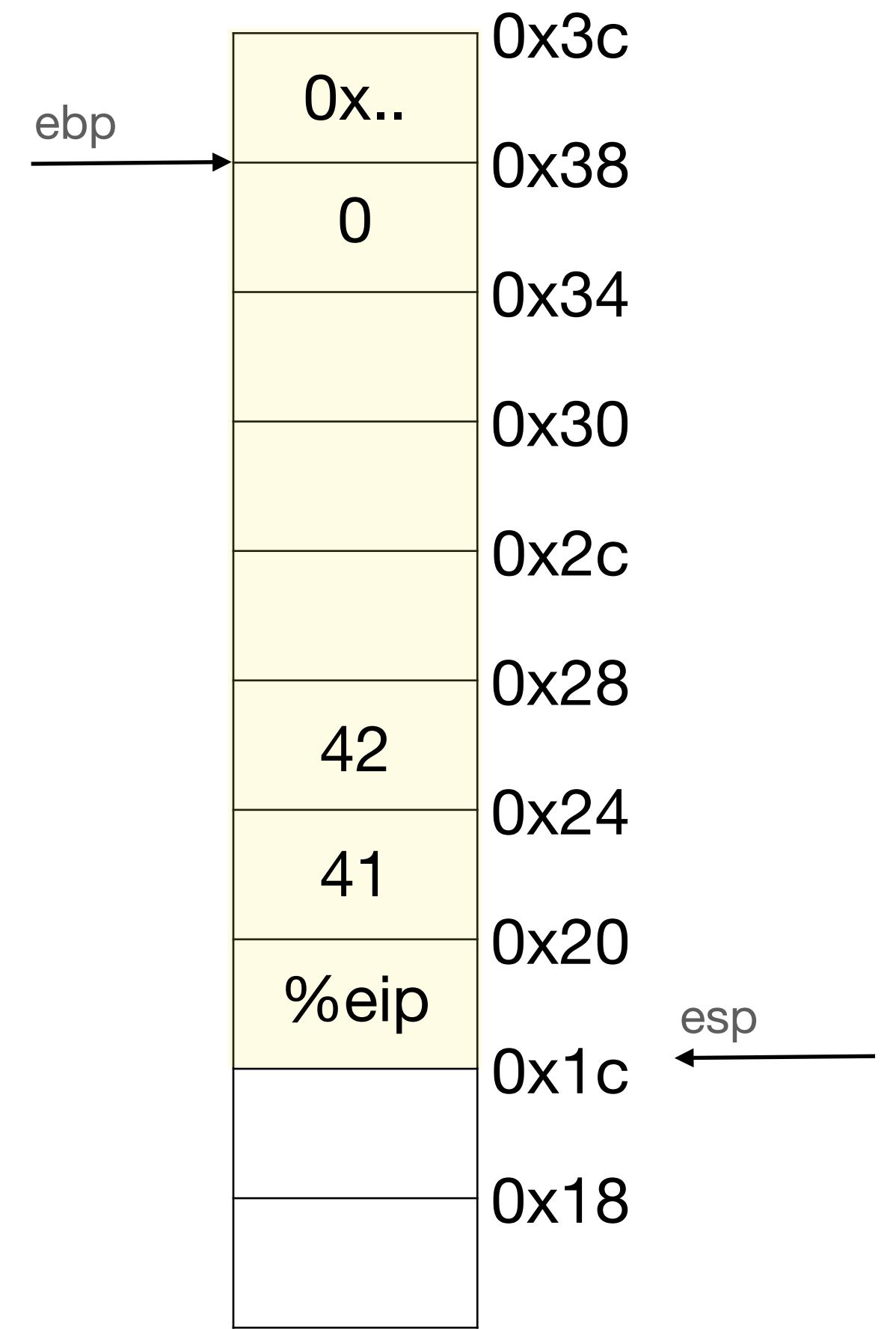
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

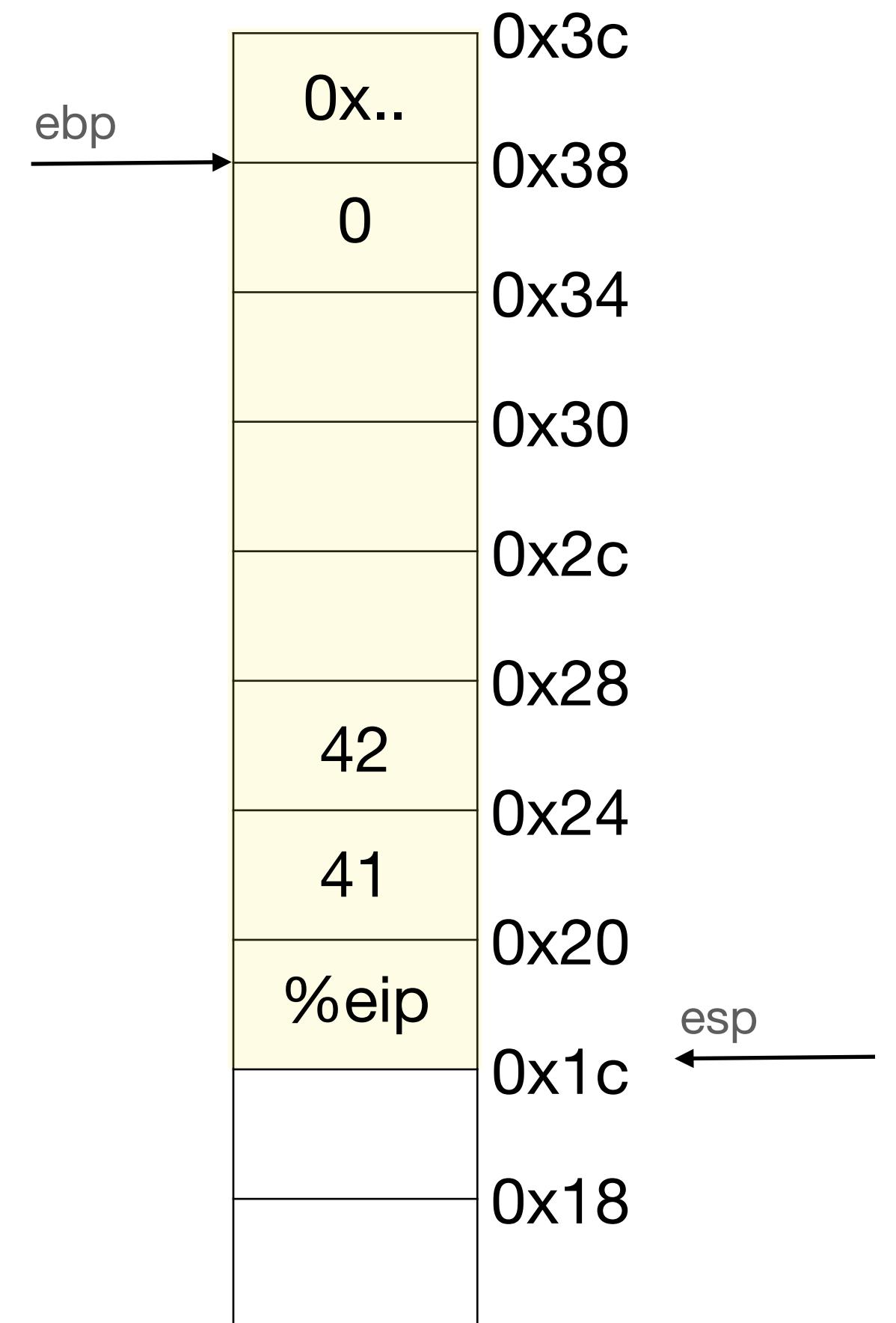
_eip →
_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

.globl _main
.p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

## -- Begin function main

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp
```



Function calling in action

Stack

```
02.s

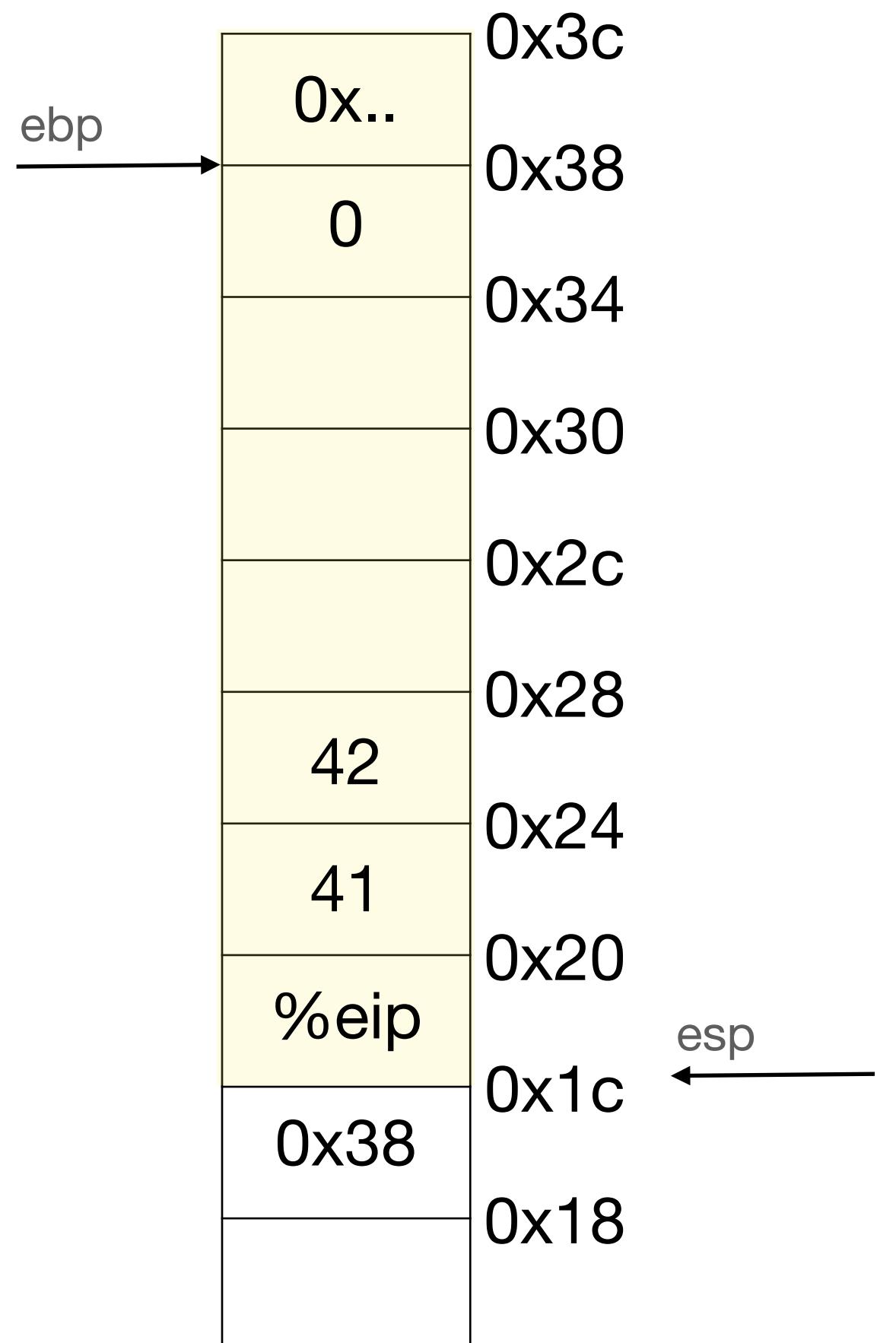
_eip →
_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

.globl _main
.p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

## -- Begin function main

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp
```



Function calling in action

Stack

```
02.s

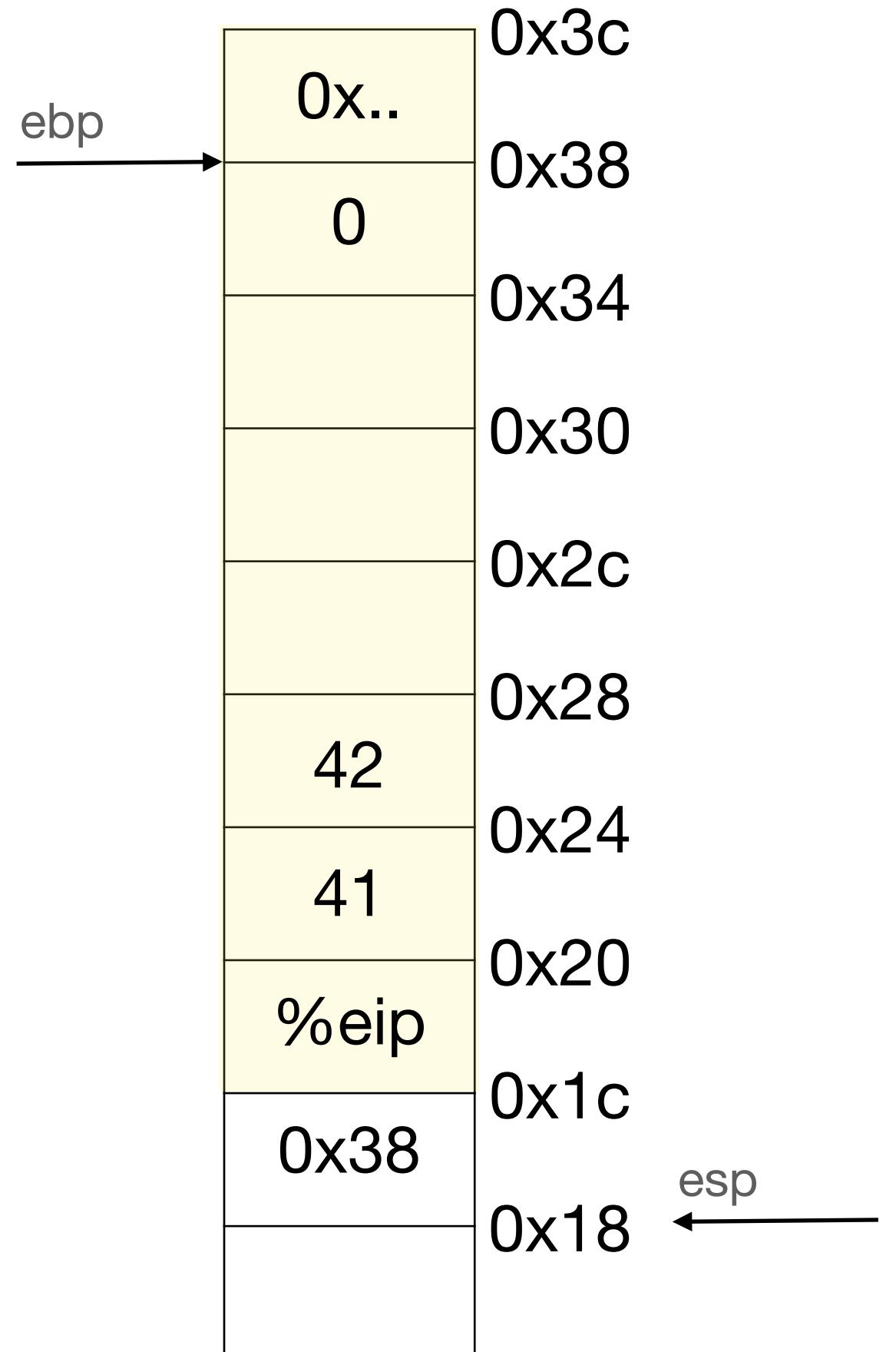
_eip →
_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

.globl _main
.p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

## -- Begin function main

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp
```



Function calling in action

Stack

```
02.s

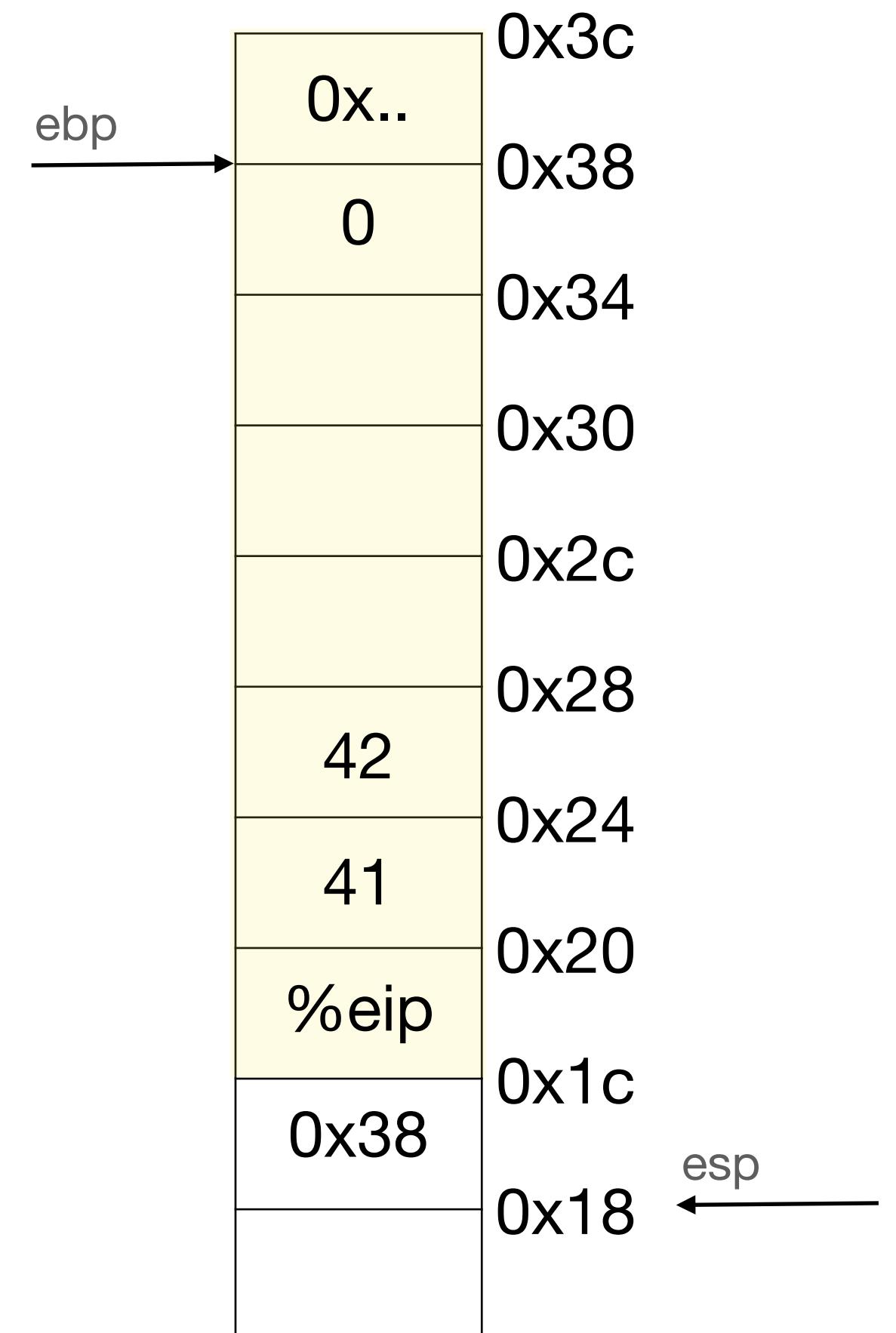
_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

## -- Begin function main
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

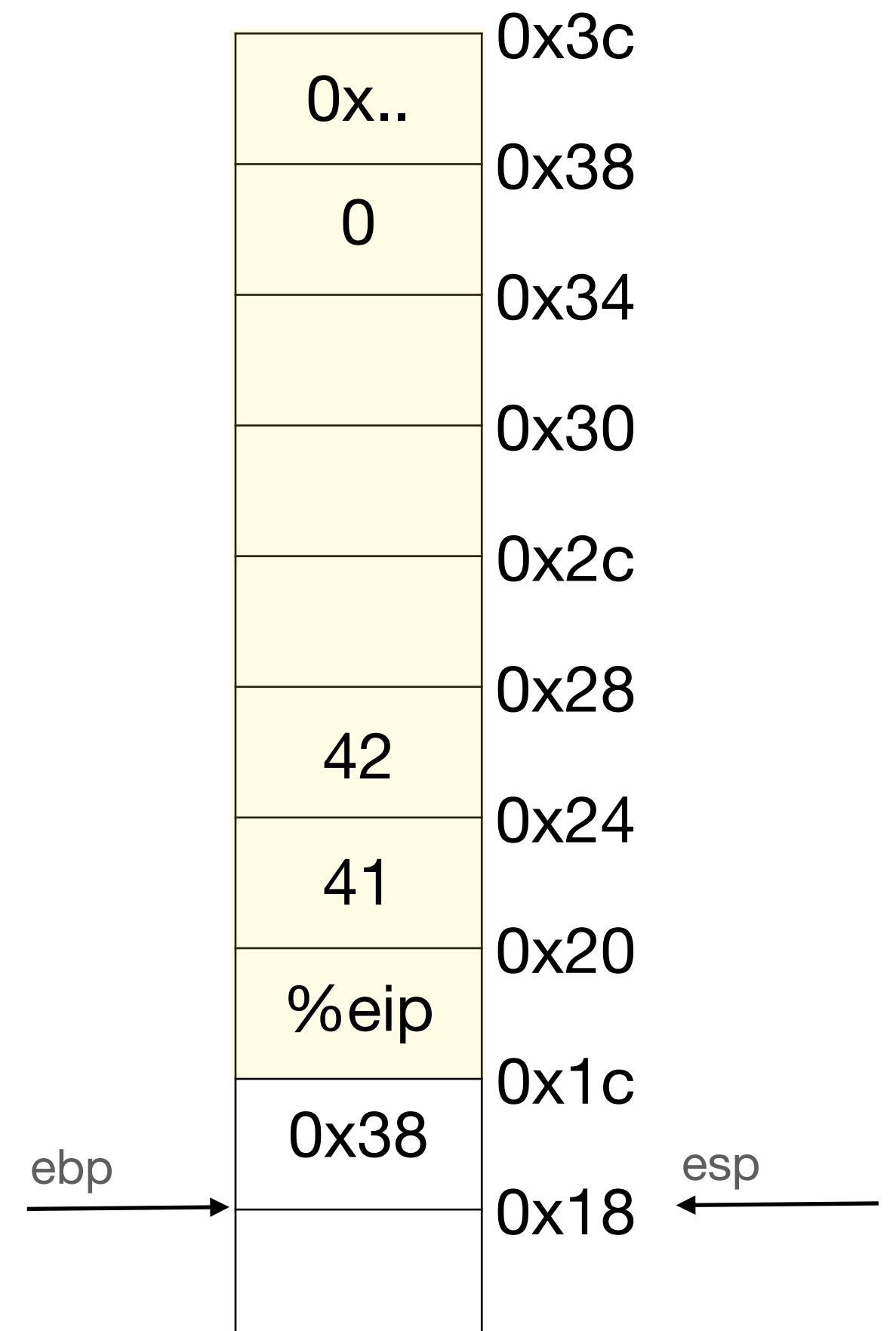
Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

## -- Begin function main
```



Function calling in action

Stack

```
02.s

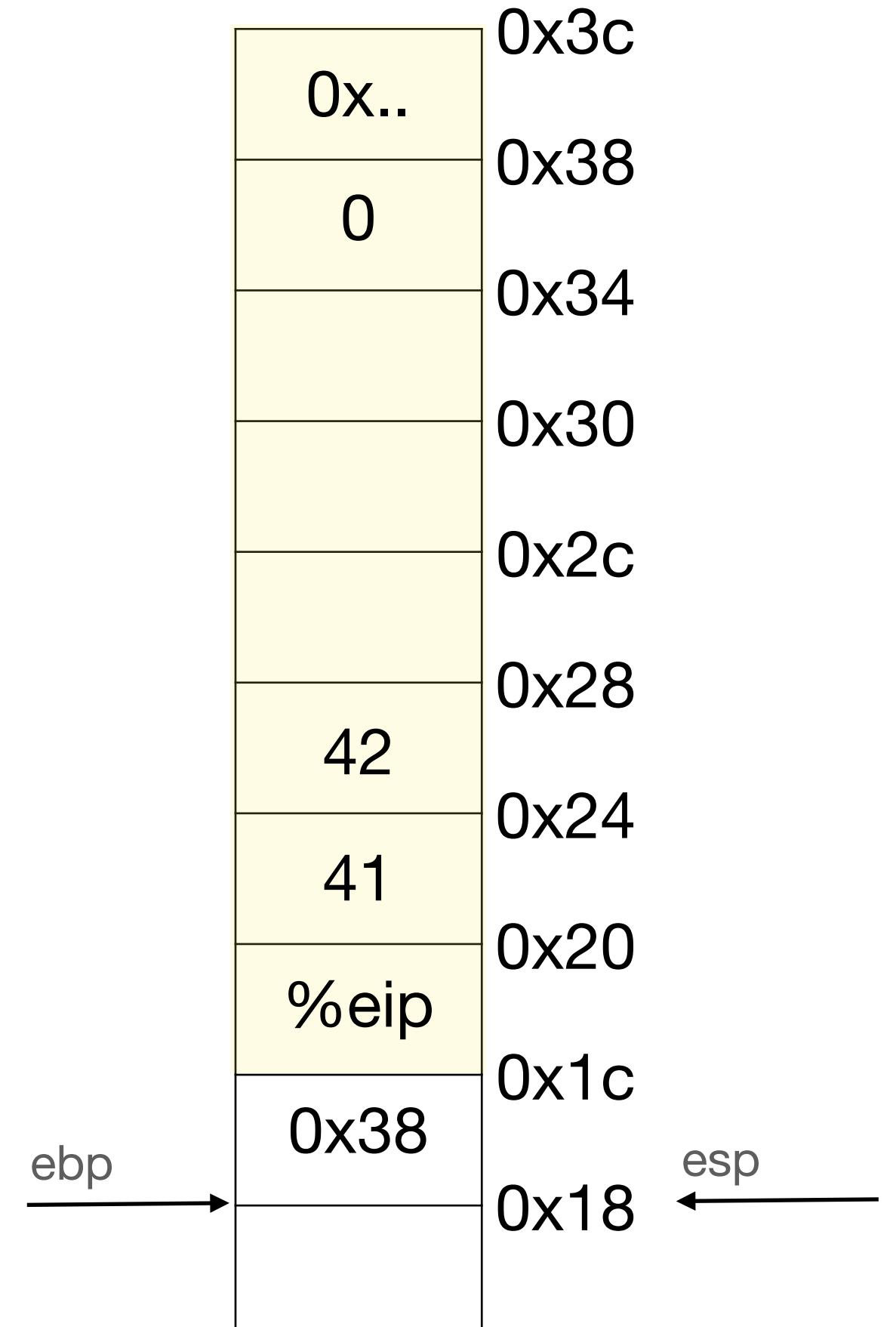
_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

## -- Begin function main
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

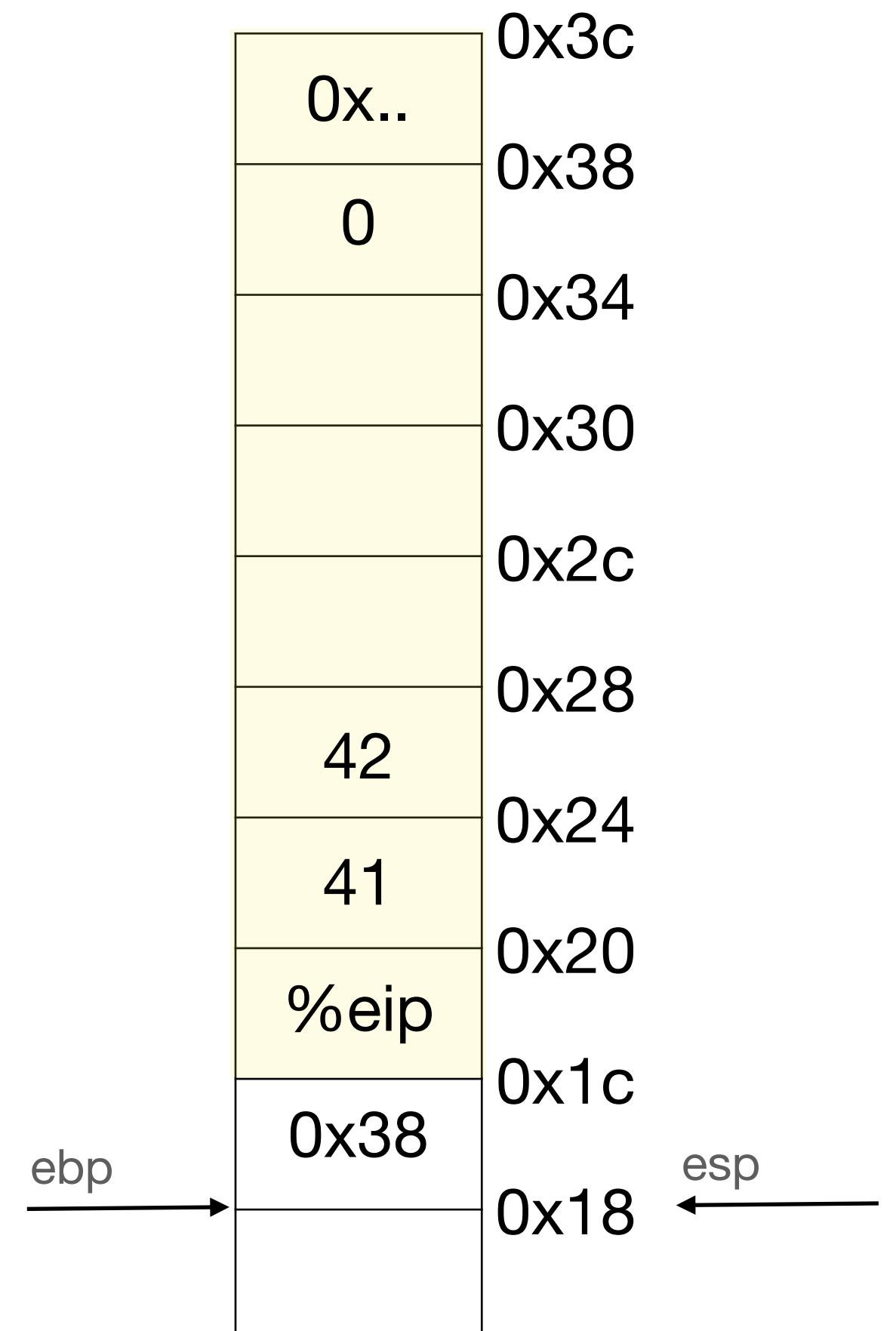
_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

## -- Begin function main
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

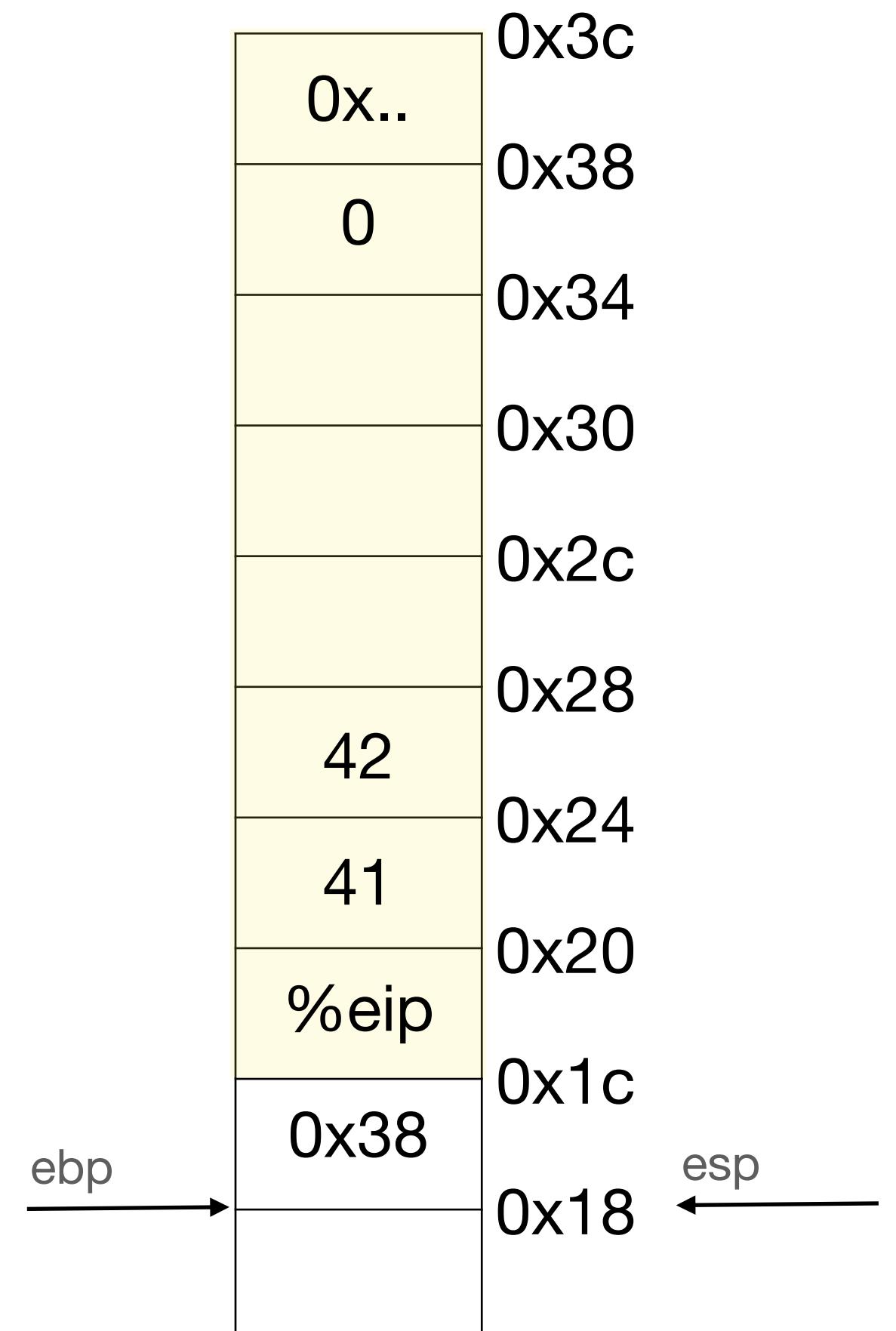
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

## -- Begin function main
```

eip →

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

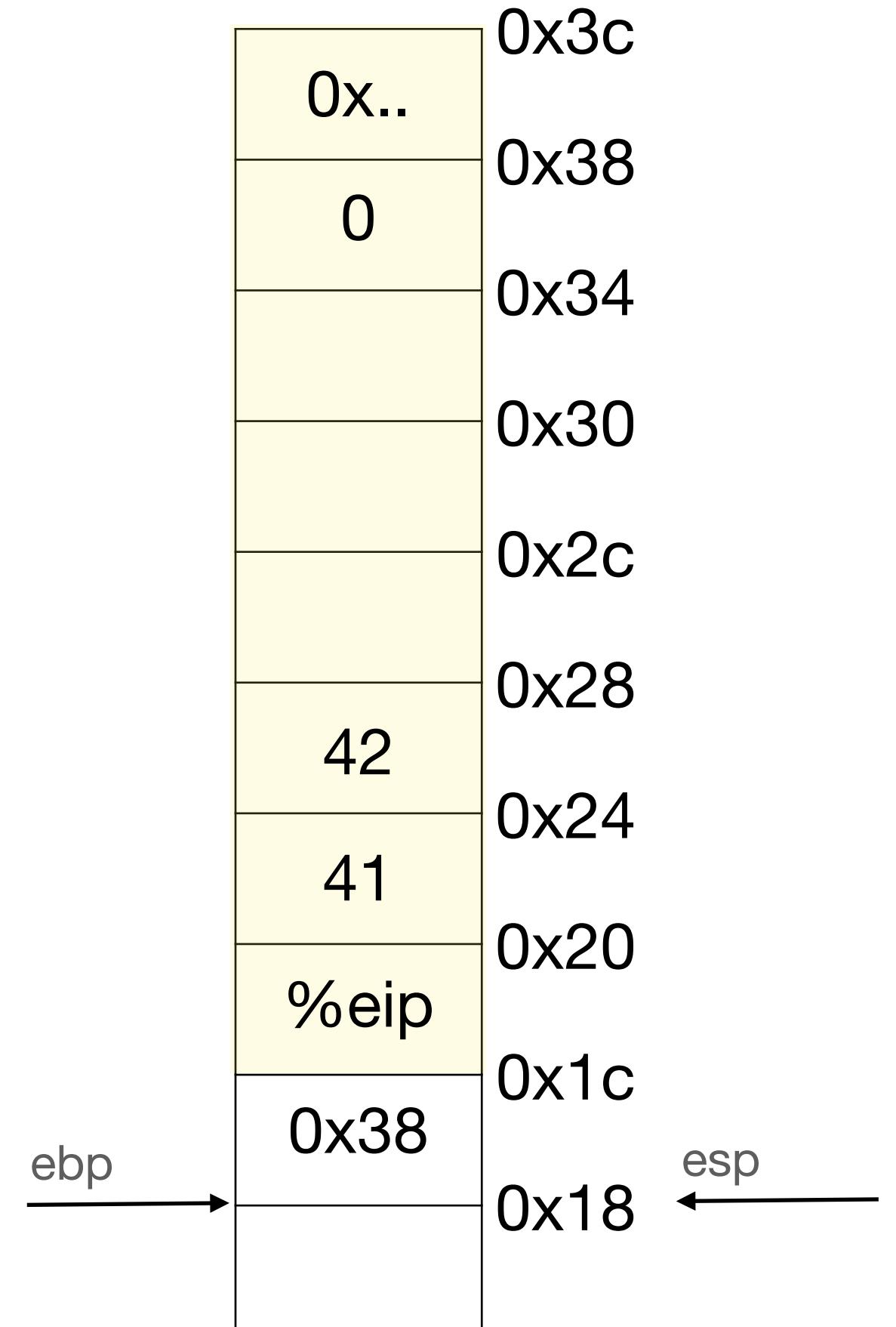
_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

## -- Begin function main
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

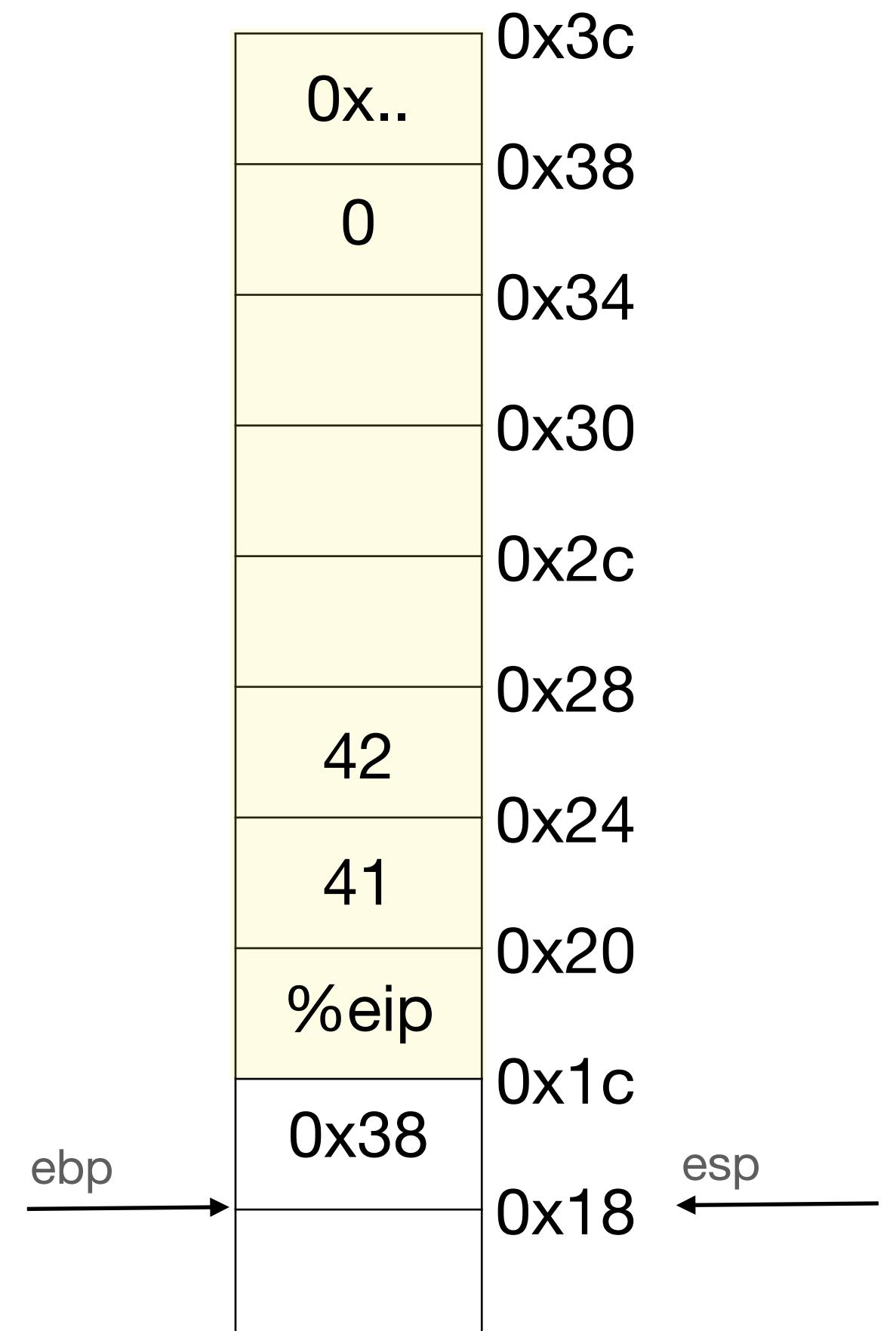
Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

## -- Begin function main
```



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

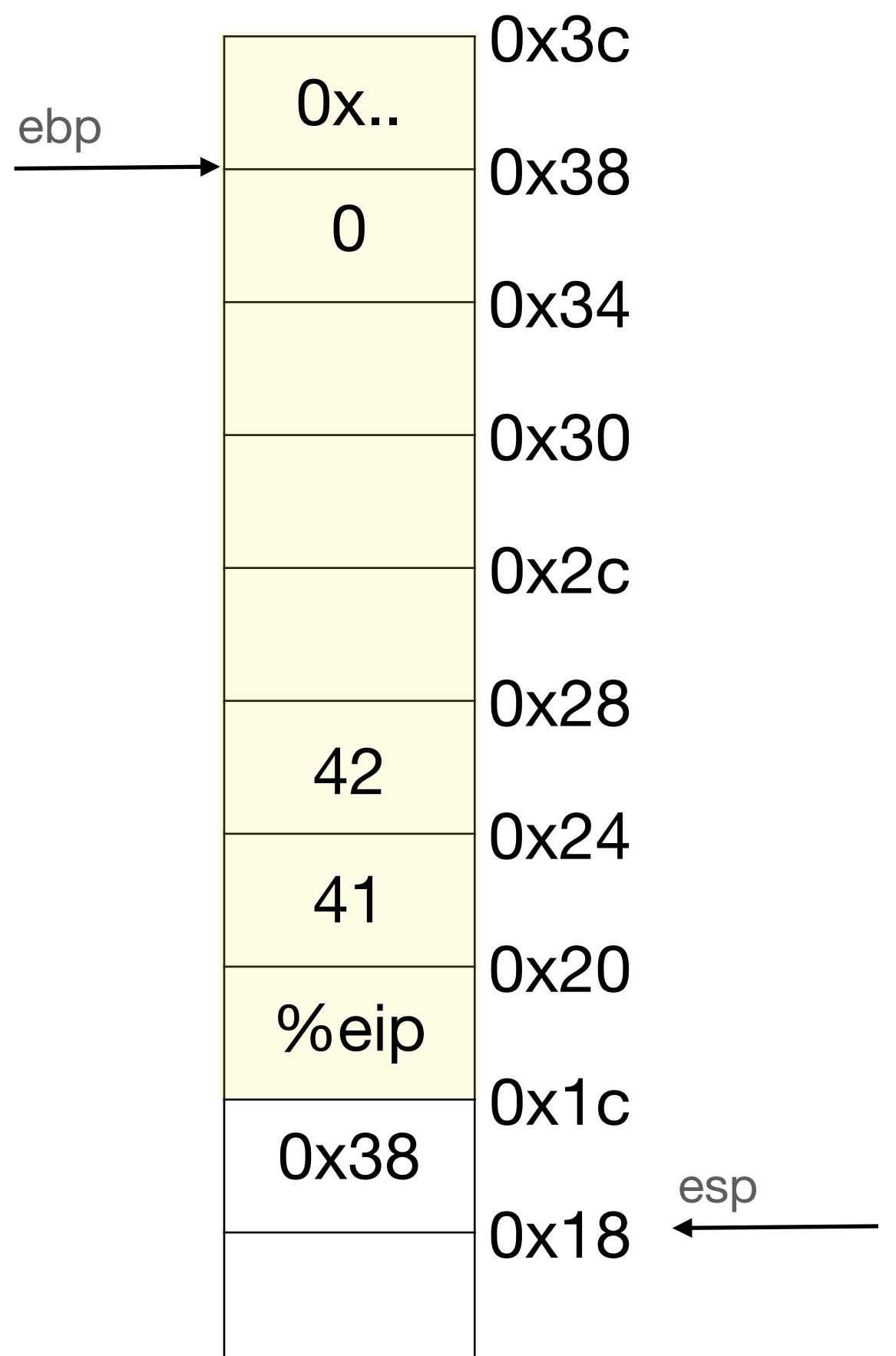
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

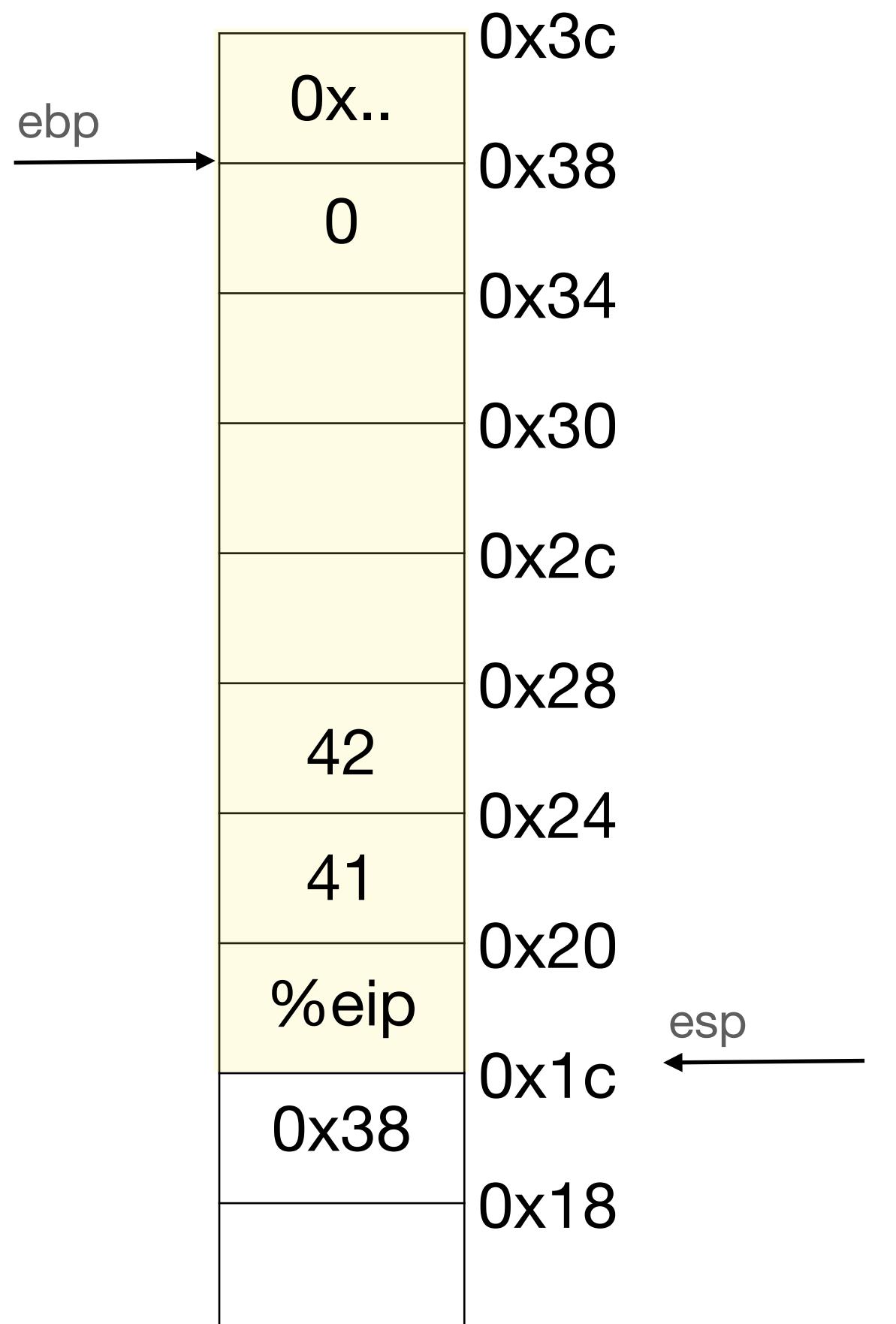
## -- Begin function main
```

Annotations for _foo:

- pushl %ebp: Save caller's base pointer
- movl %esp, %ebp: ebp = esp
- movl 8(%ebp), %eax: eax = *(ebp + 8)
- addl 12(%ebp), %eax: eax = eax + *(ebp + 12)
- popl %ebp: Restore caller's base pointer
- retl: change eip to return address

Annotations for _main:

- pushl %ebp: Save caller's base pointer
- movl %esp, %ebp: ebp = esp
- subl \$24, %esp: esp = esp - 0x18
- movl \$0, -4(%ebp): *(ebp-4)=0
- movl \$41, (%esp): *(esp) = 41
- movl \$42, 4(%esp): *(esp+4) = 42
- calll _foo: Push current eip on to stack, jump to foo
- addl \$24, %esp: esp = esp + 24 (Restore caller's esp)
- popl %ebp: Restore caller's ebp
- retl:



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

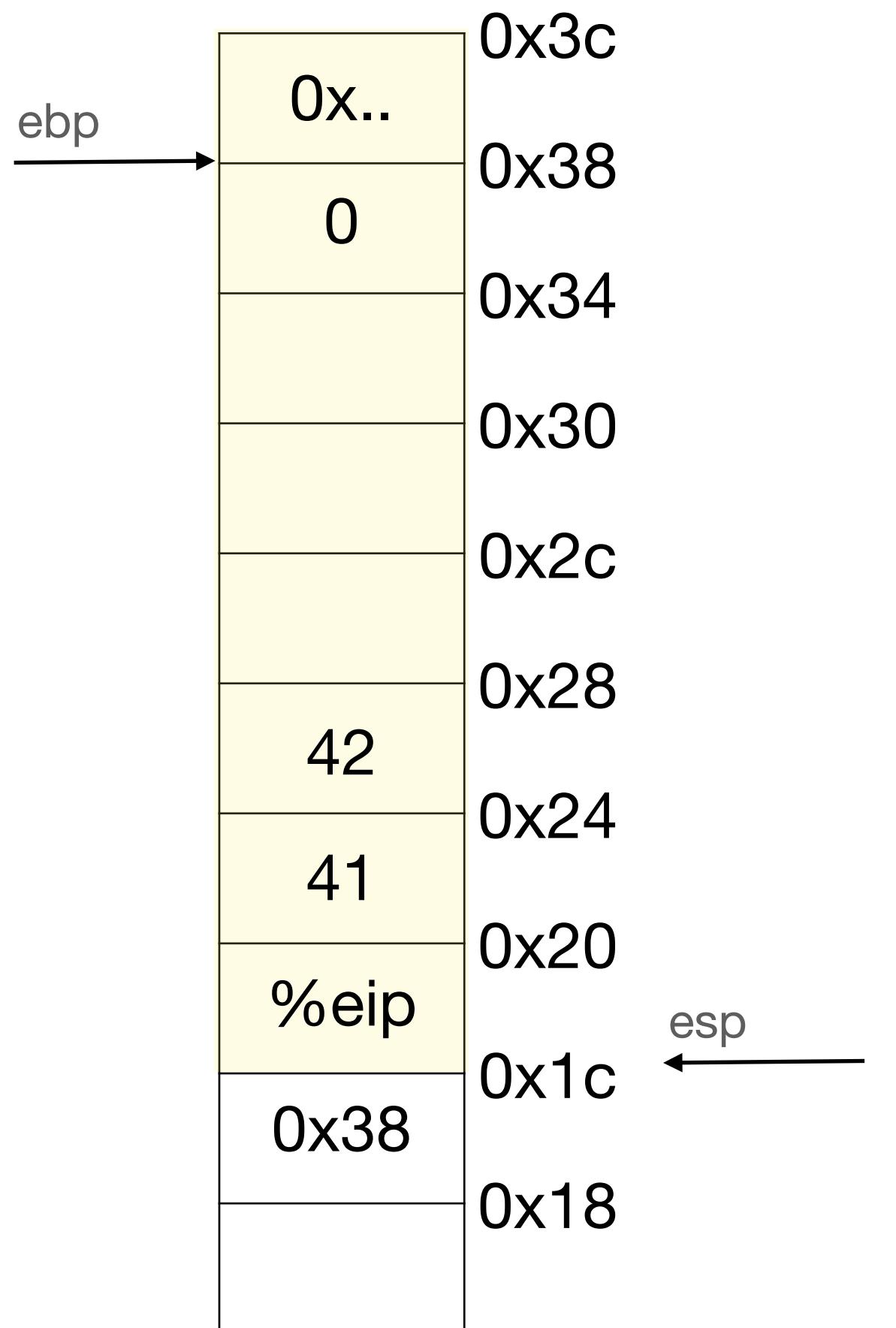
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip → retl
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

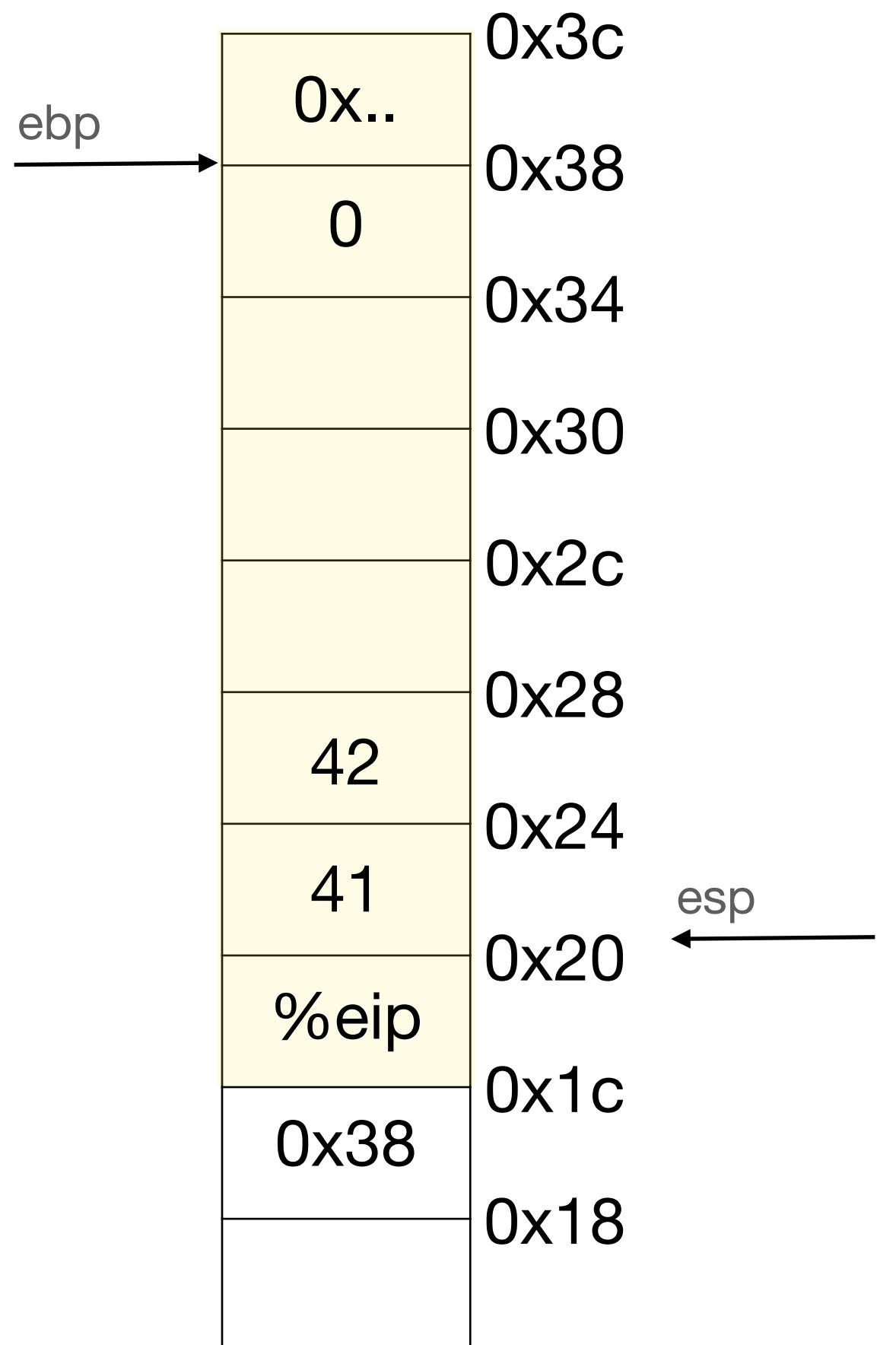
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip → retl
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

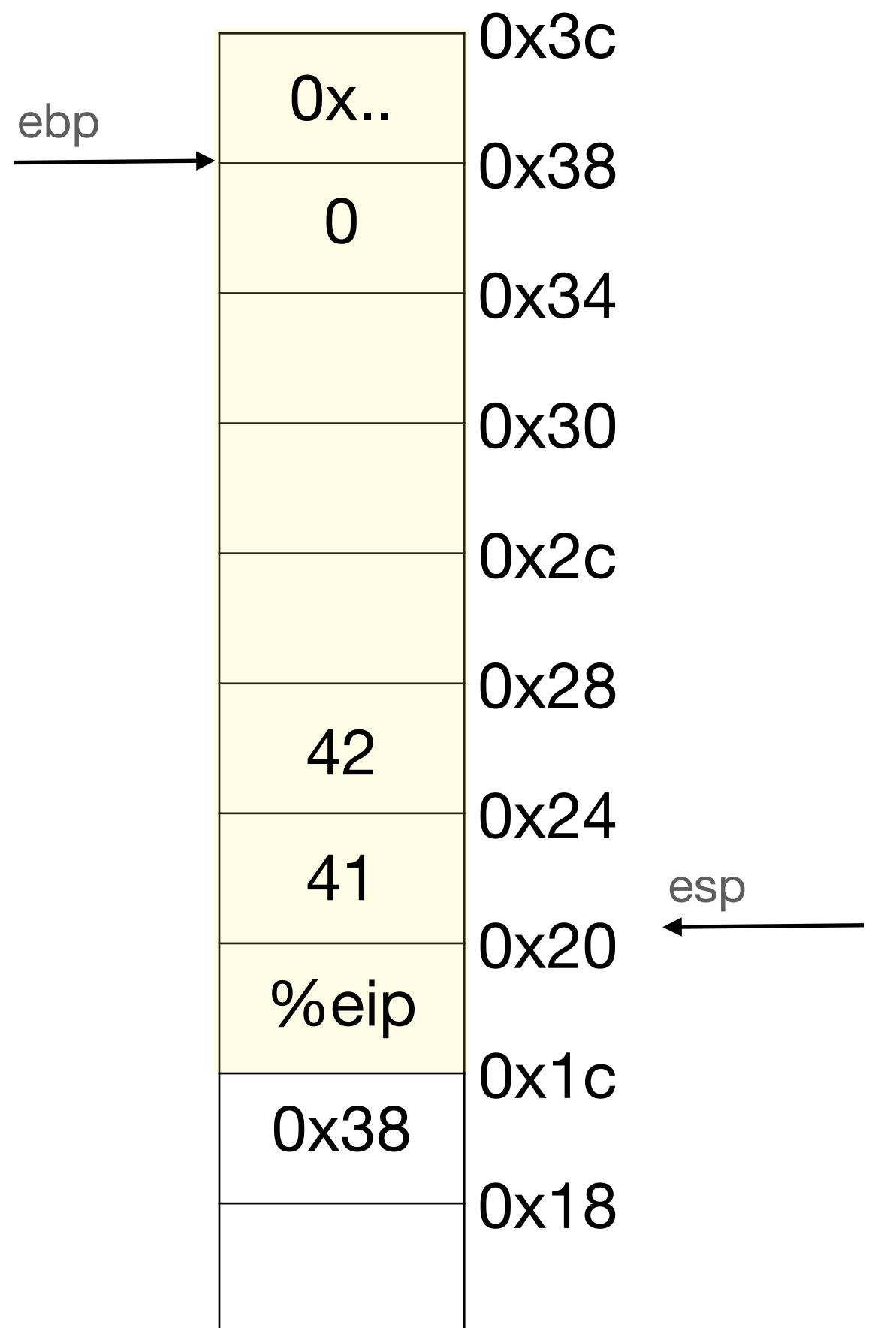
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

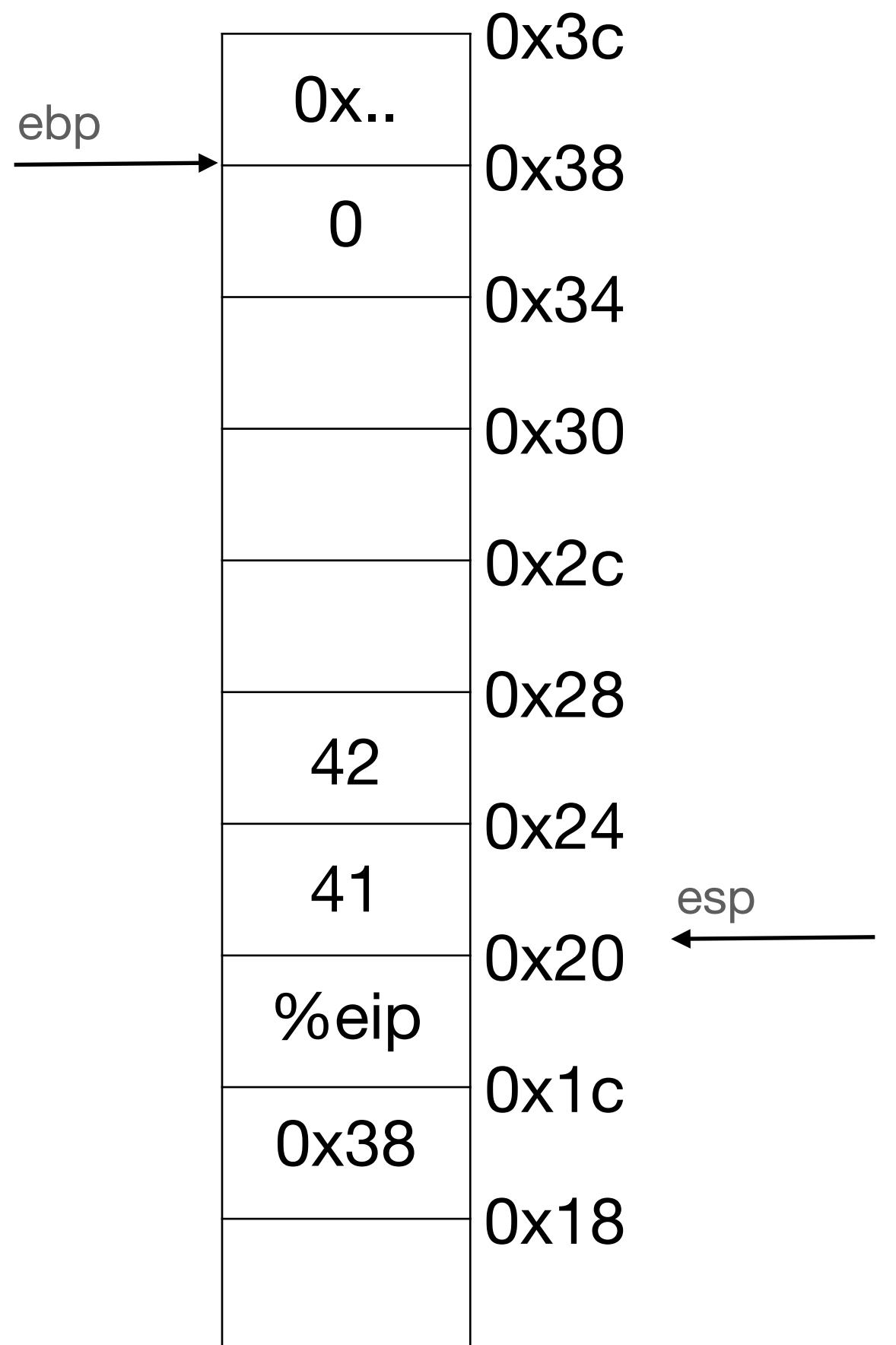
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

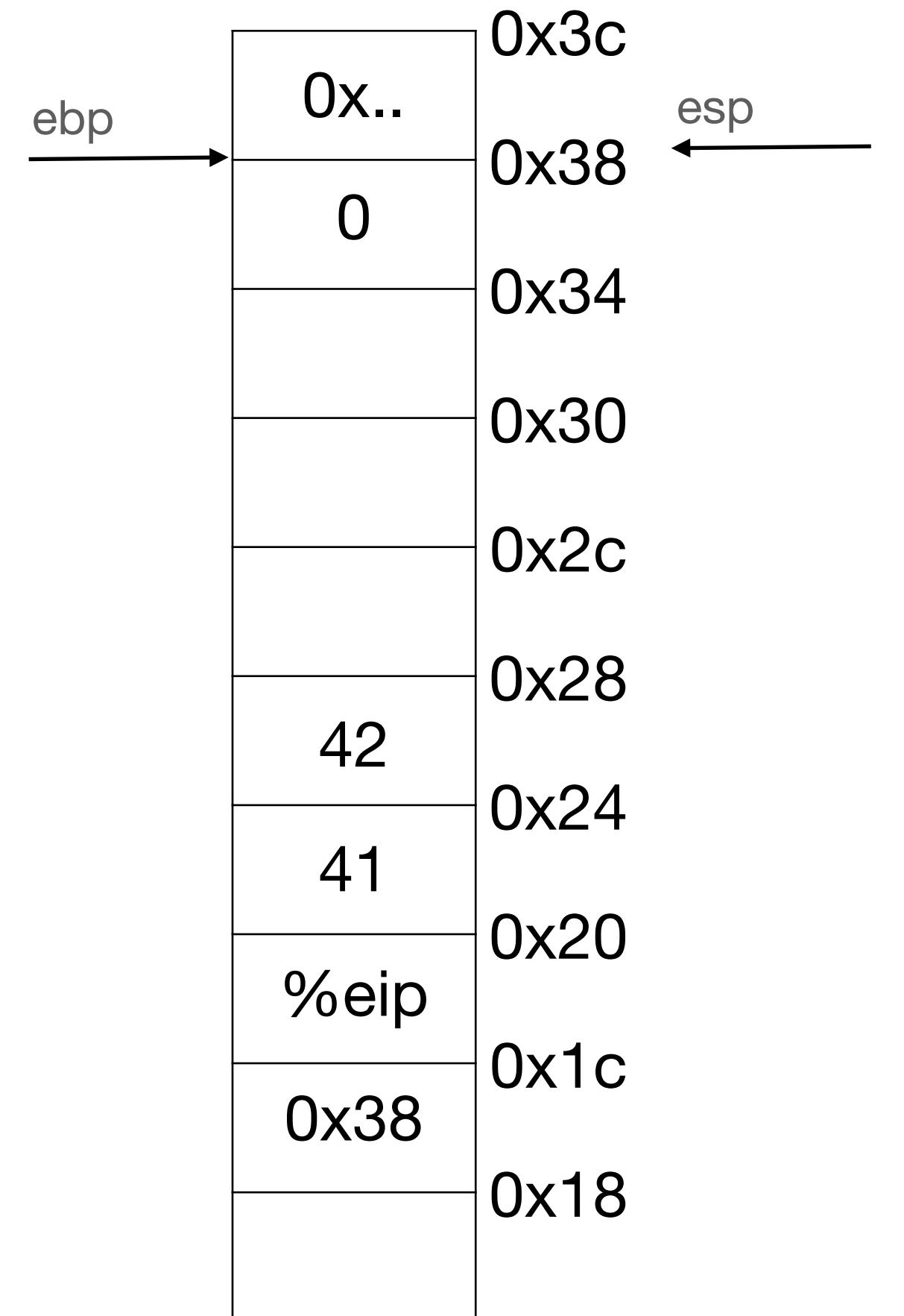
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

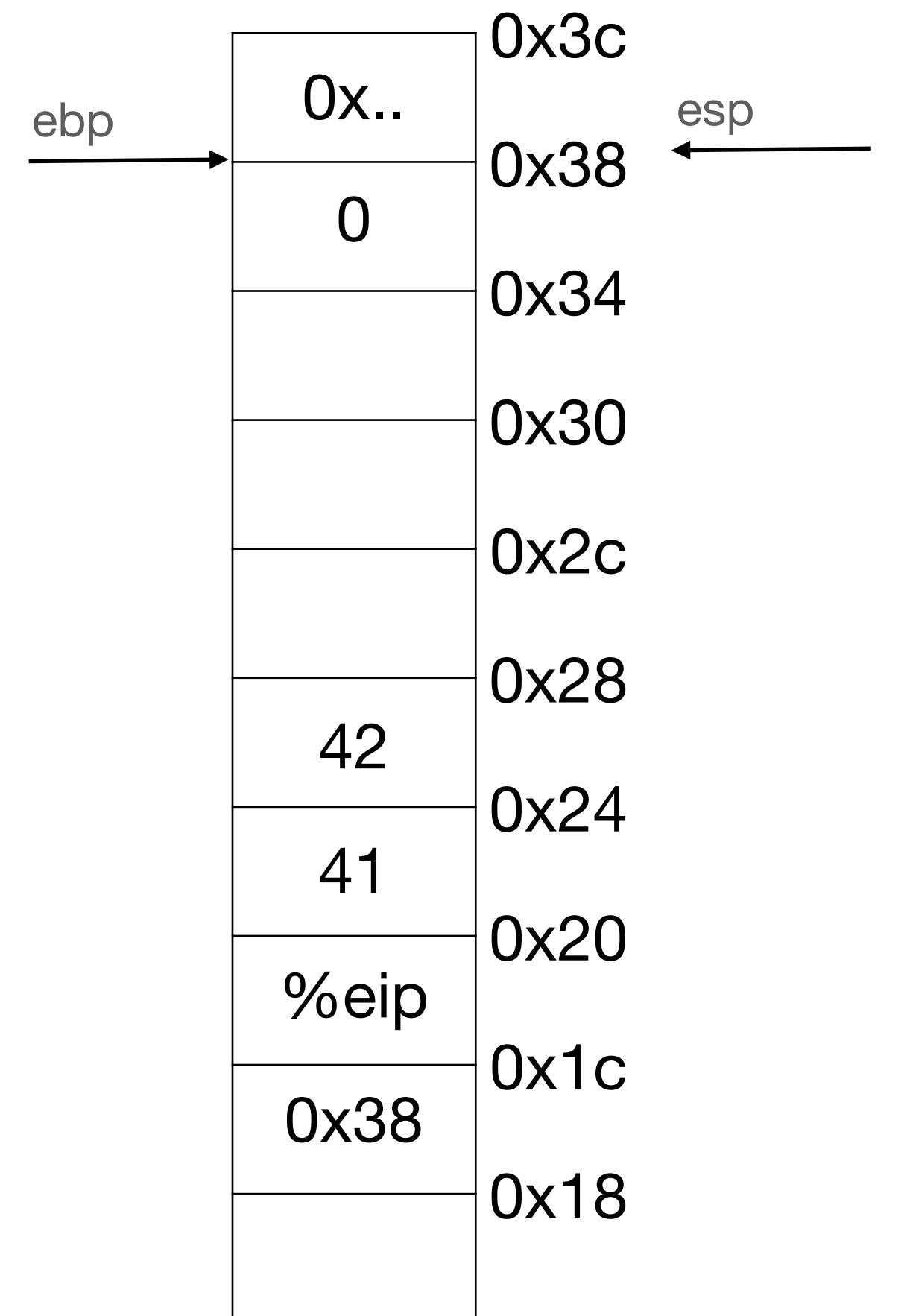
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

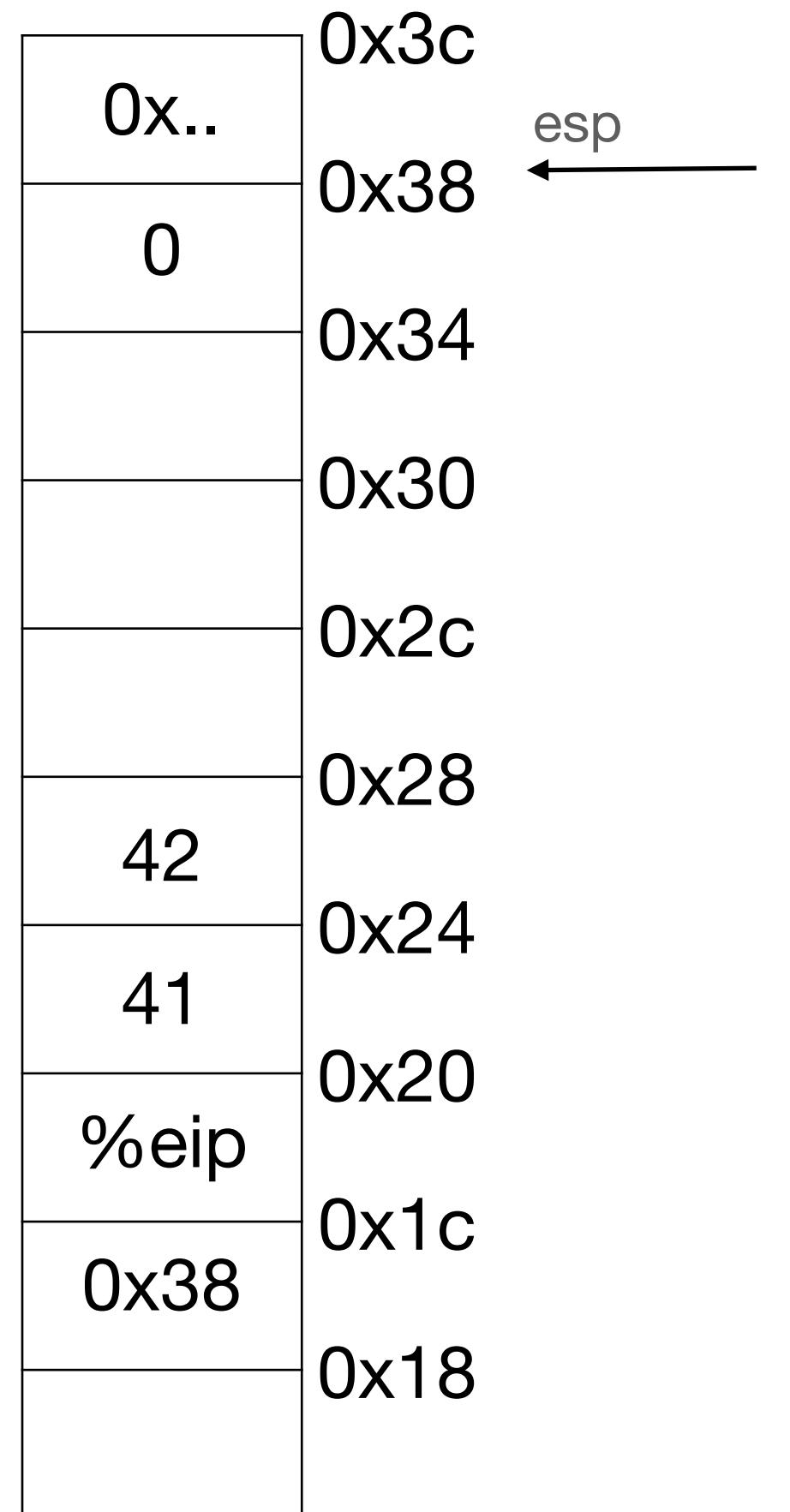
eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp

ebp →



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

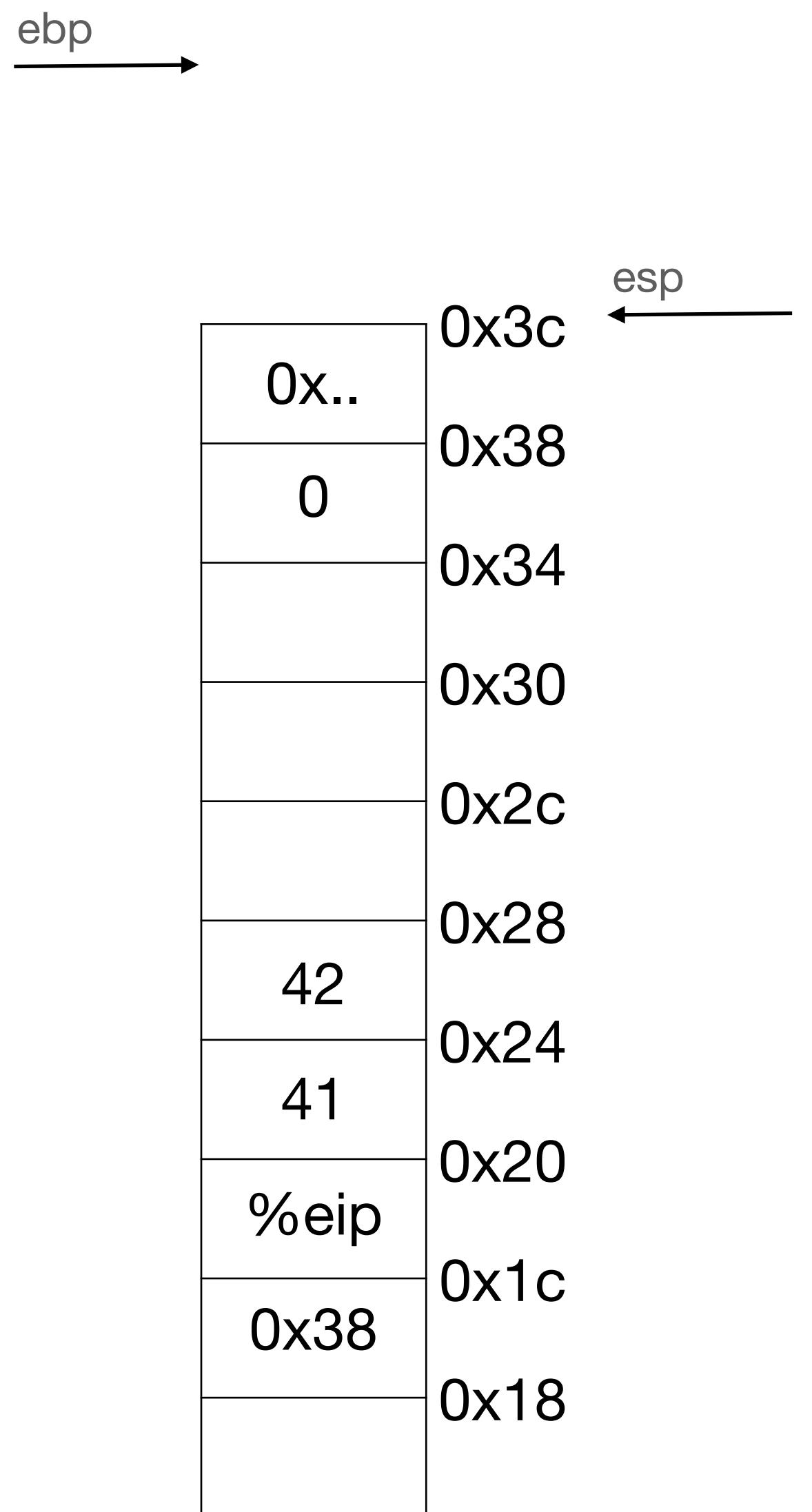
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip →
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp



Function calling in action

Stack

```
02.s

_foo:
    pushl %ebp
    movl %esp, %ebp
    movl 8(%ebp), %eax
    addl 12(%ebp), %eax
    popl %ebp
    retl

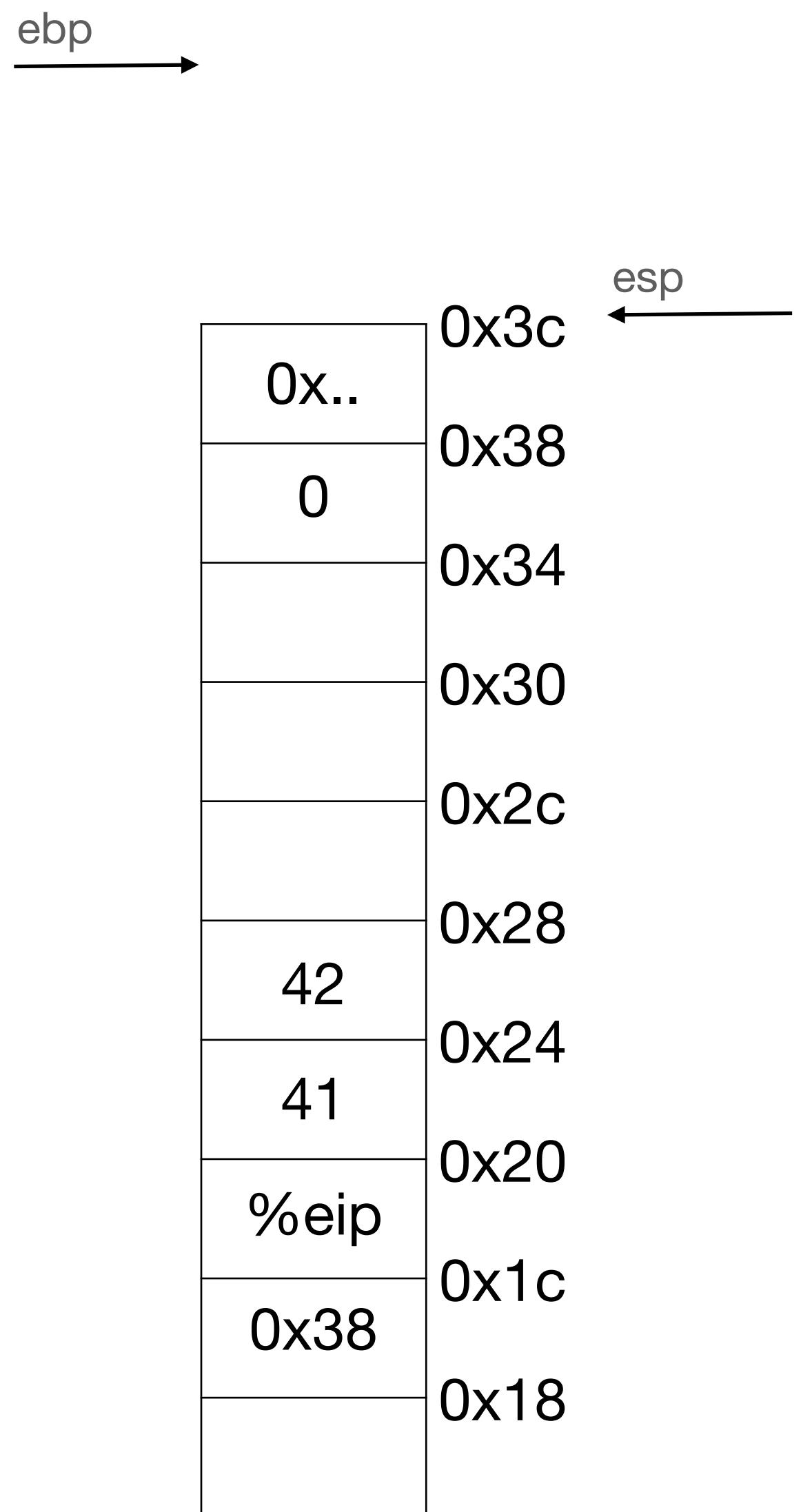
    .globl _main
    .p2align 4, 0x90
_main:
    pushl %ebp
    movl %esp, %ebp
    subl $24, %esp
    movl $0, -4(%ebp)
    movl $41, (%esp)
    movl $42, 4(%esp)
    calll _foo
    addl $24, %esp
    popl %ebp
    retl

eip
```

Save caller's base pointer
ebp = esp
eax = *(ebp + 8)
eax = eax + *(ebp + 12)
Restore caller's base pointer
change eip to return address

-- Begin function main

Save caller's base pointer
ebp = esp
esp = esp - 0x18
*(ebp-4)=0
*(esp) = 41
*(esp+4) = 42
Push current eip on to stack, jump to foo
esp = esp + 24 (Restore caller's esp)
Restore caller's ebp

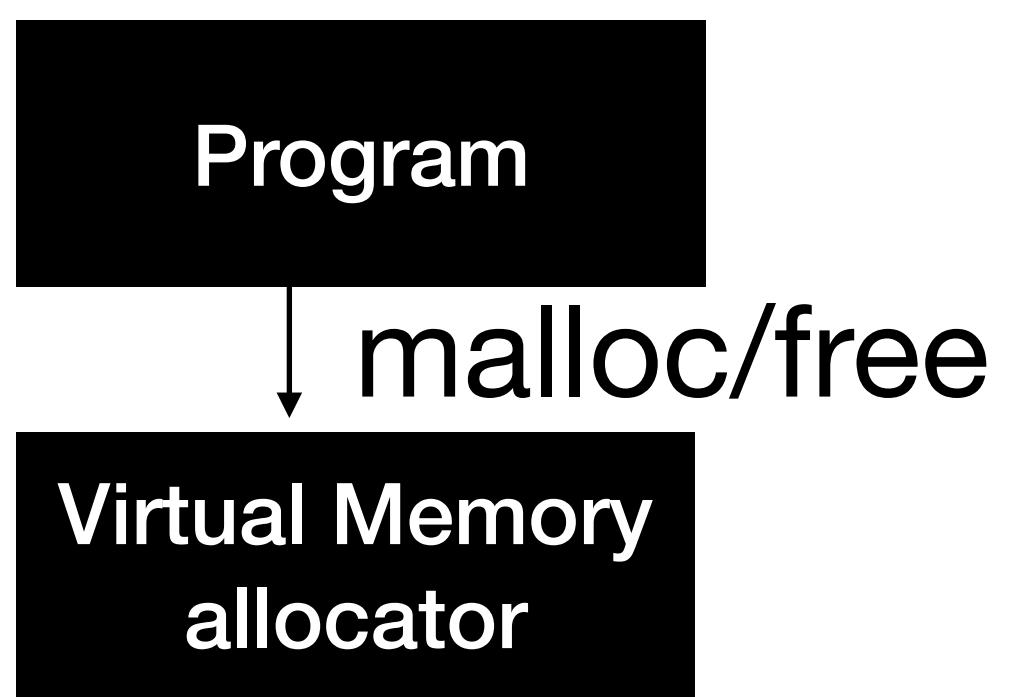


Memory APIs and bugs

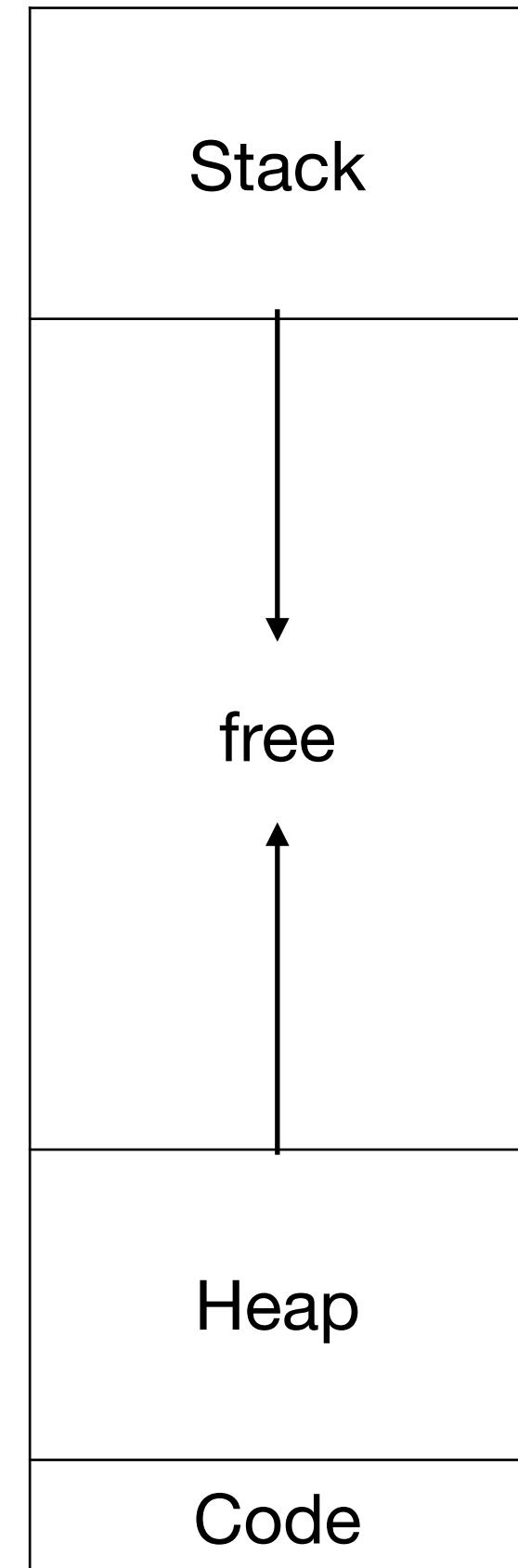
- malloc, free, va.c.
 - malloc is for dynamic allocation. Size is not known at compile time. Slower than stack allocations. Need to find free space.
- Null pointer dereference. null.c
- Memory leak. leak.c
- Buffer overflow. overflow.c
- Use after free. useafterfree.c
- Invalid free. invalidfree.c
- Double free. doublefree.c
- Uninitialised read. uninitread.c

Memory allocator

Works with virtual memory

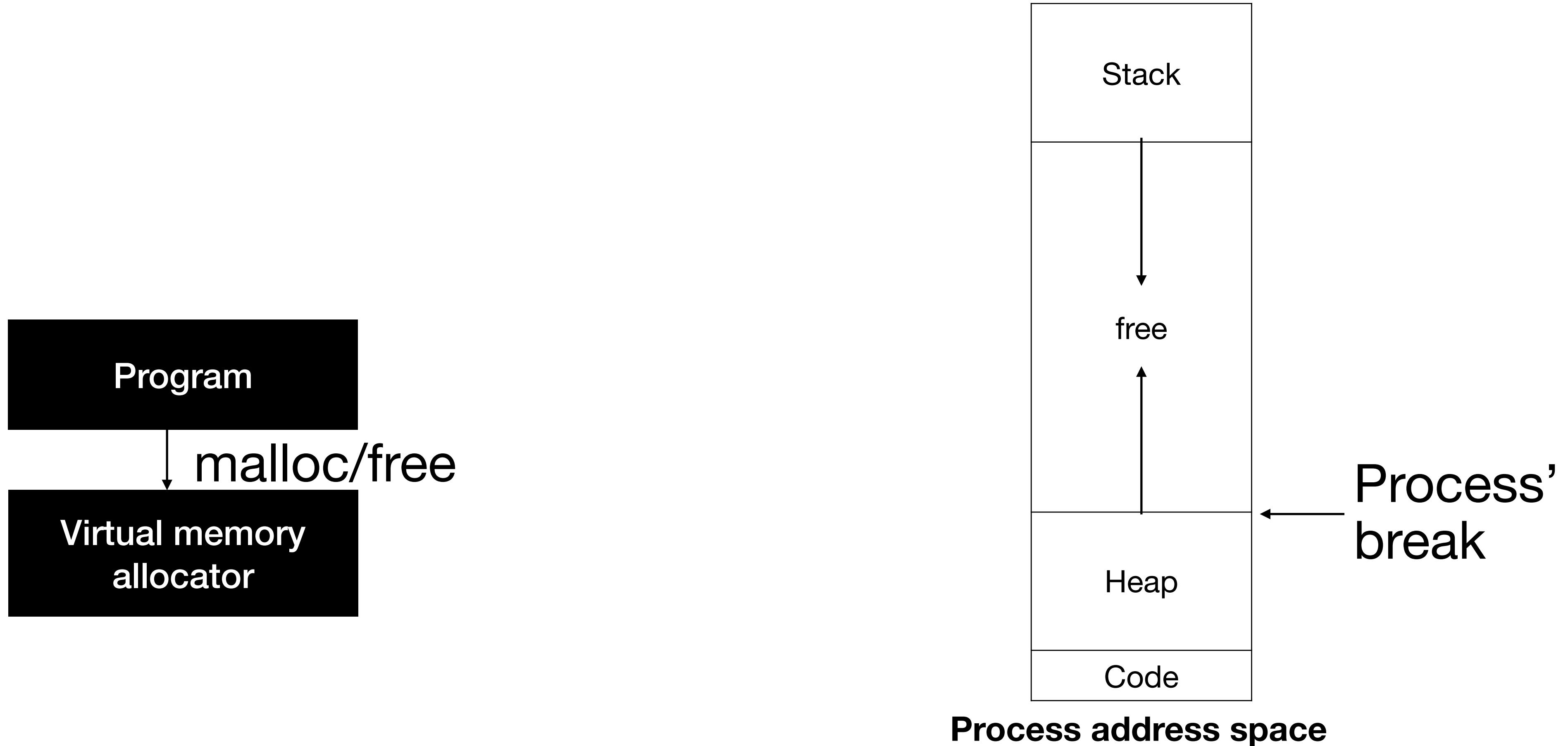


Manages heap memory



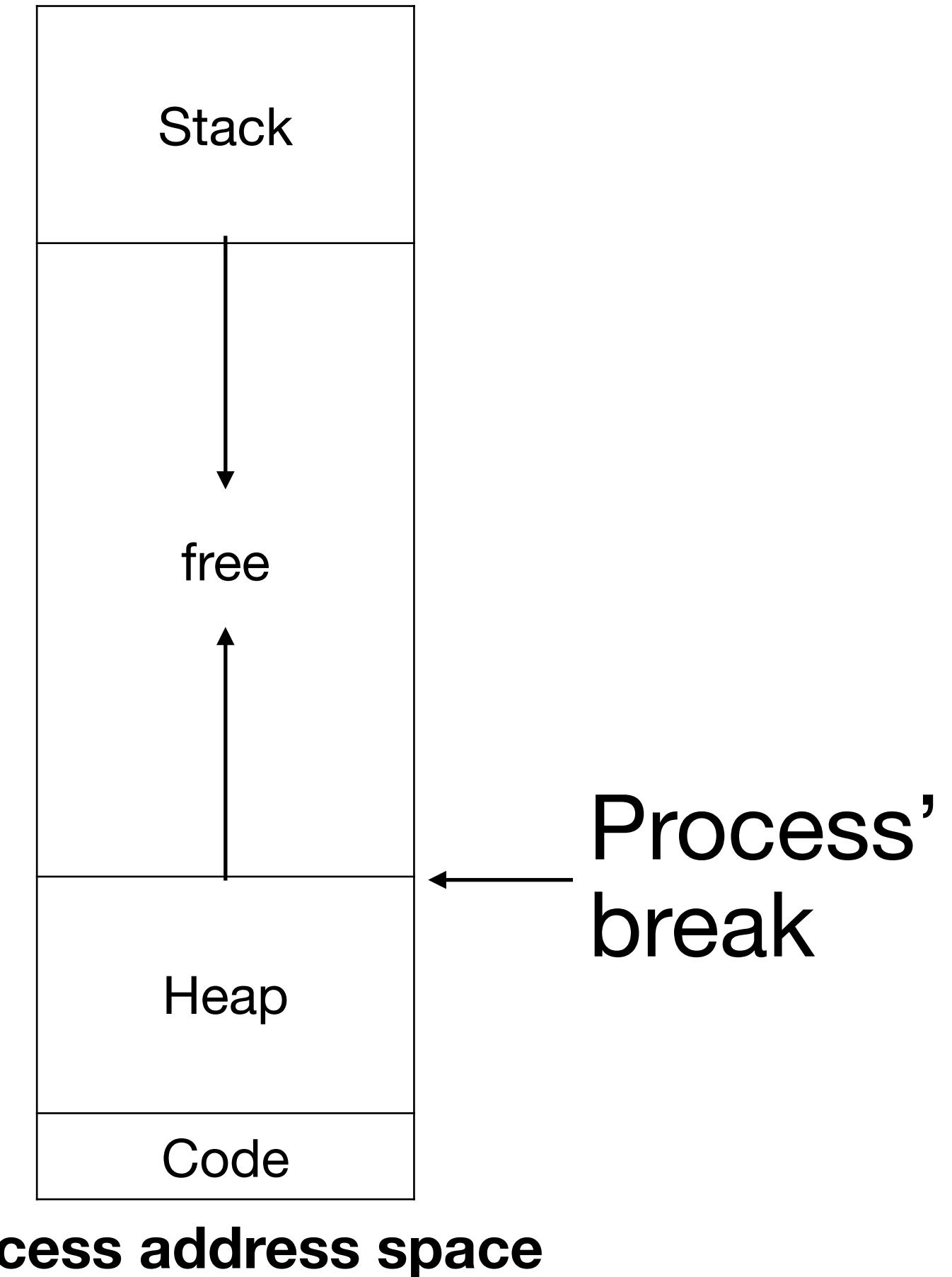
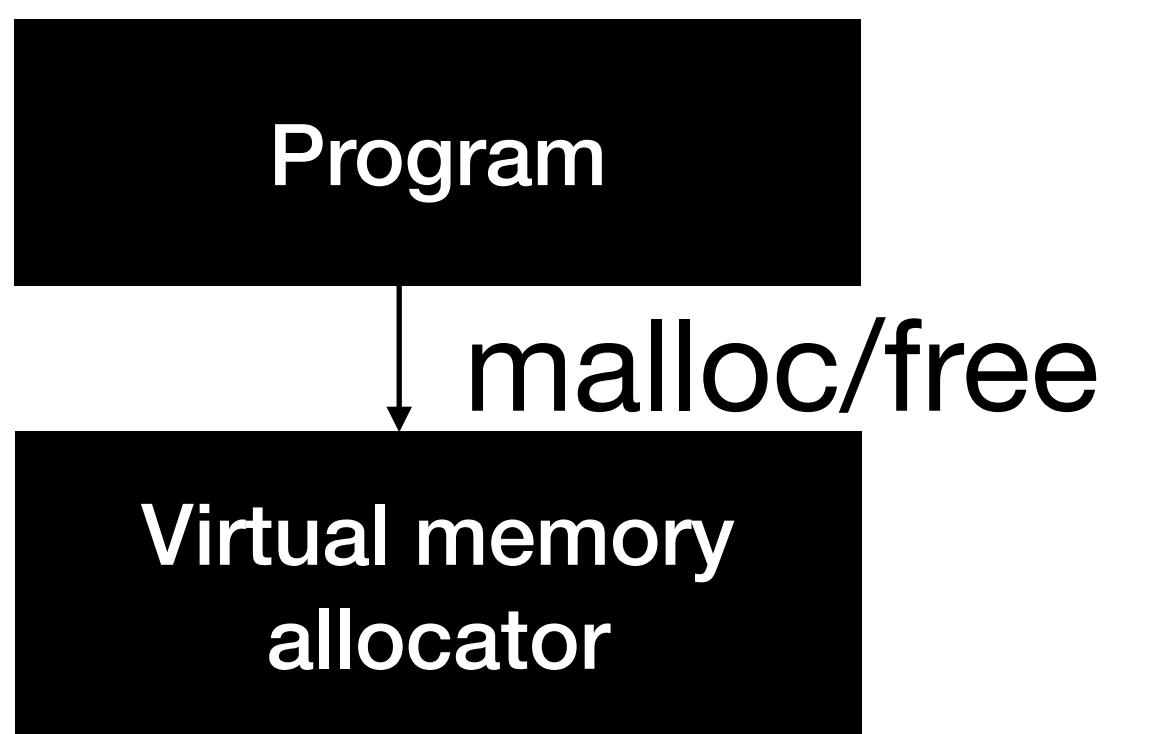
Process address space

OS memory allocator



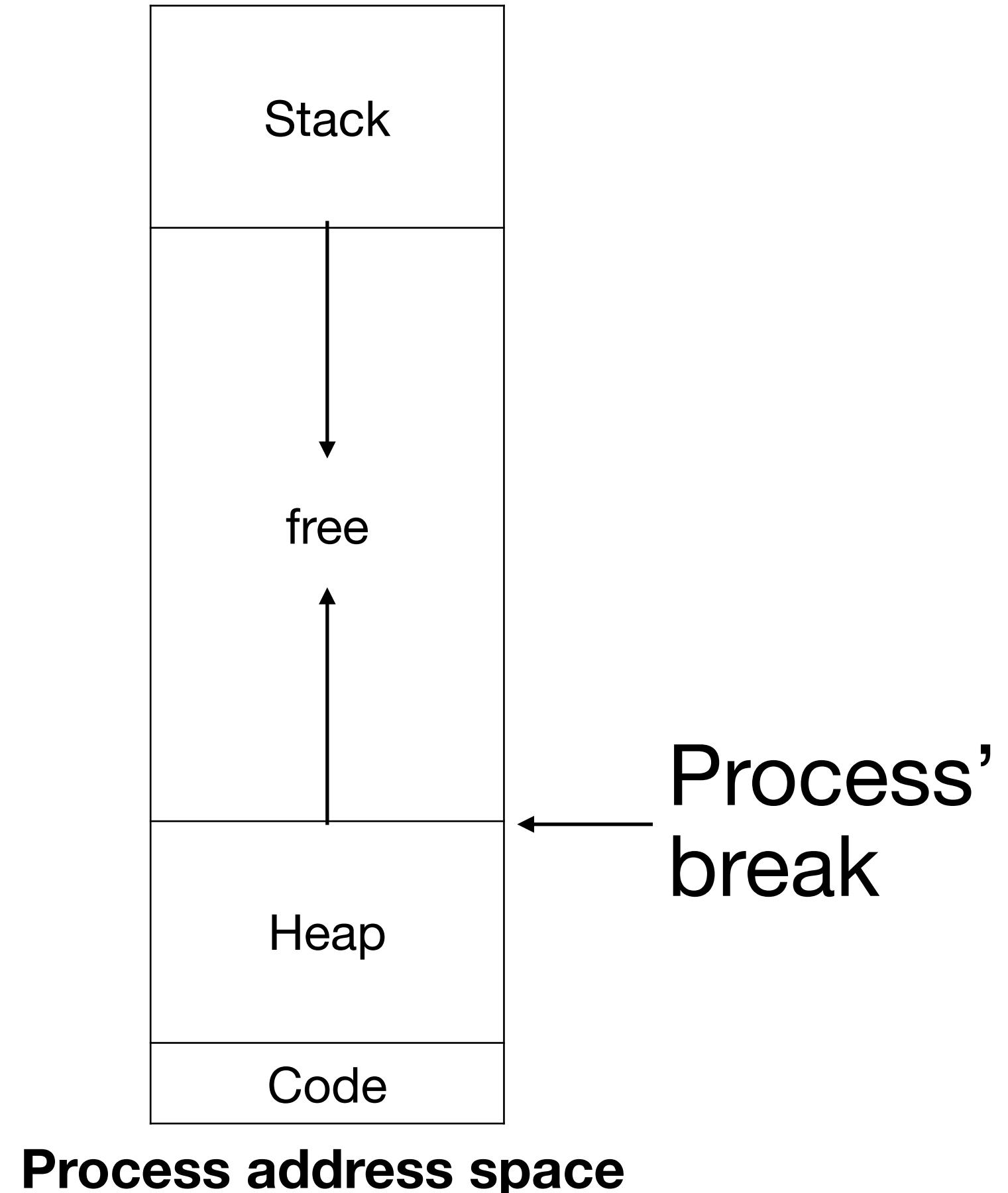
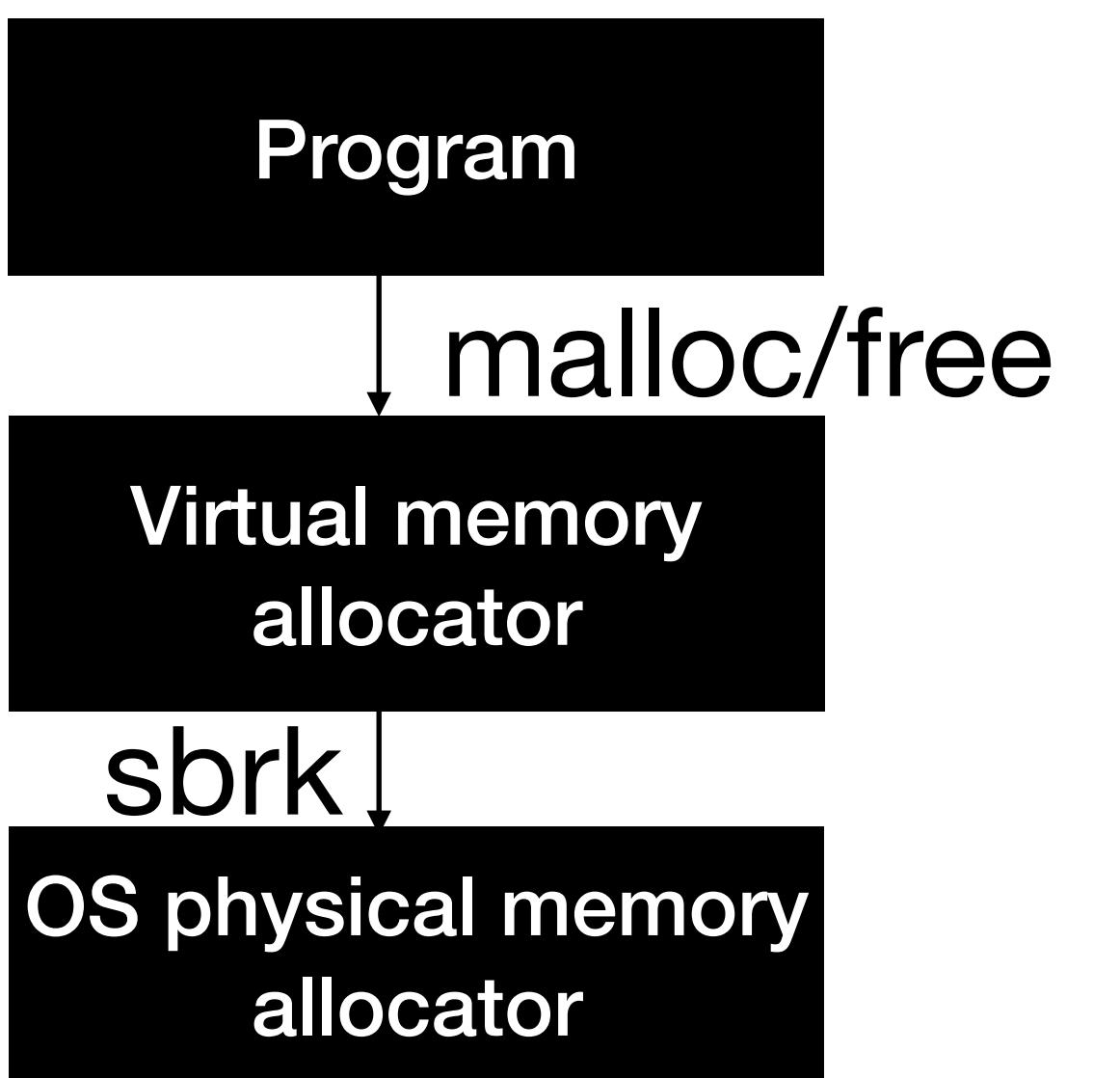
OS memory allocator

- `sbrk(int increment)` increments process' break. *increment* can be negative.



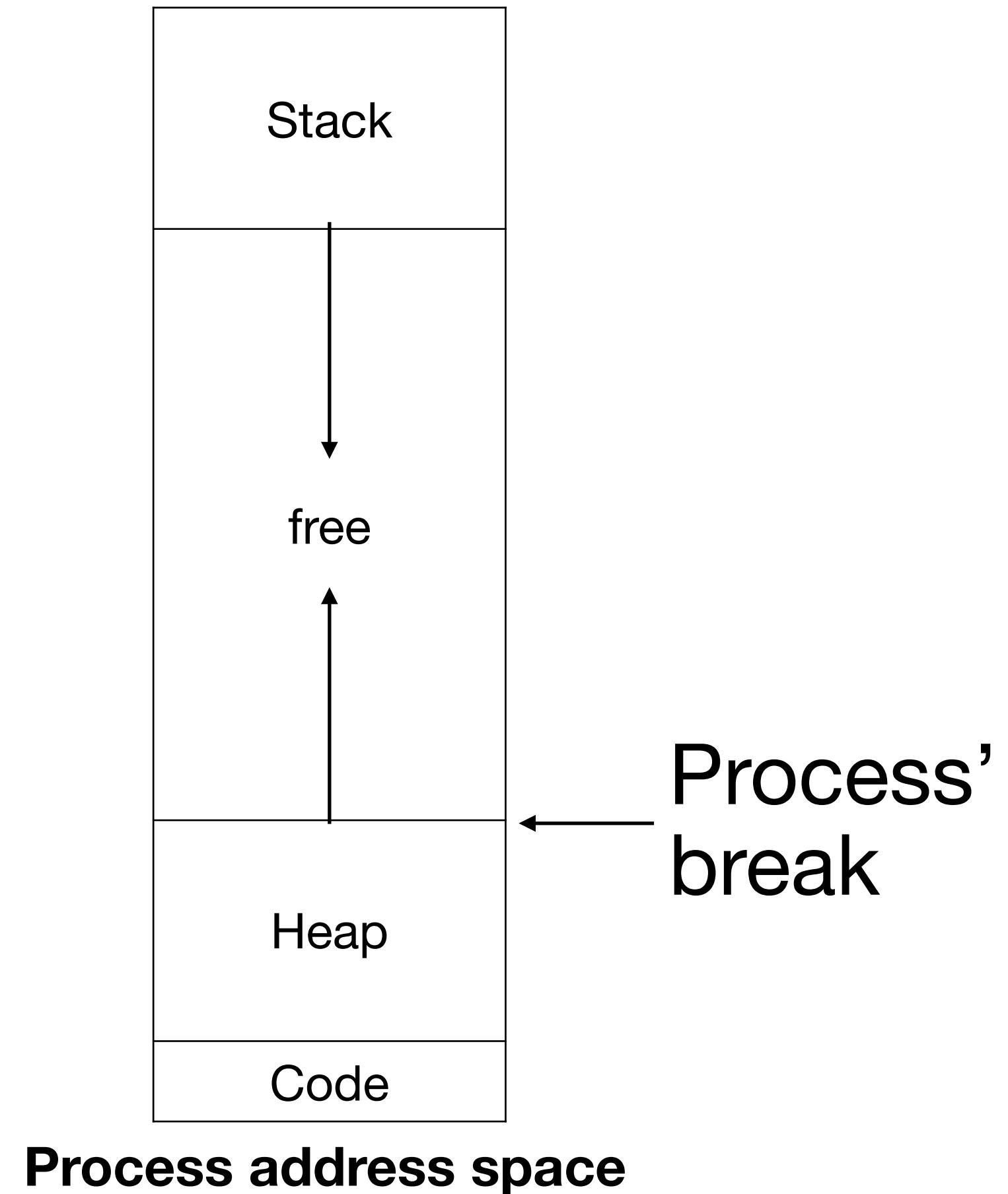
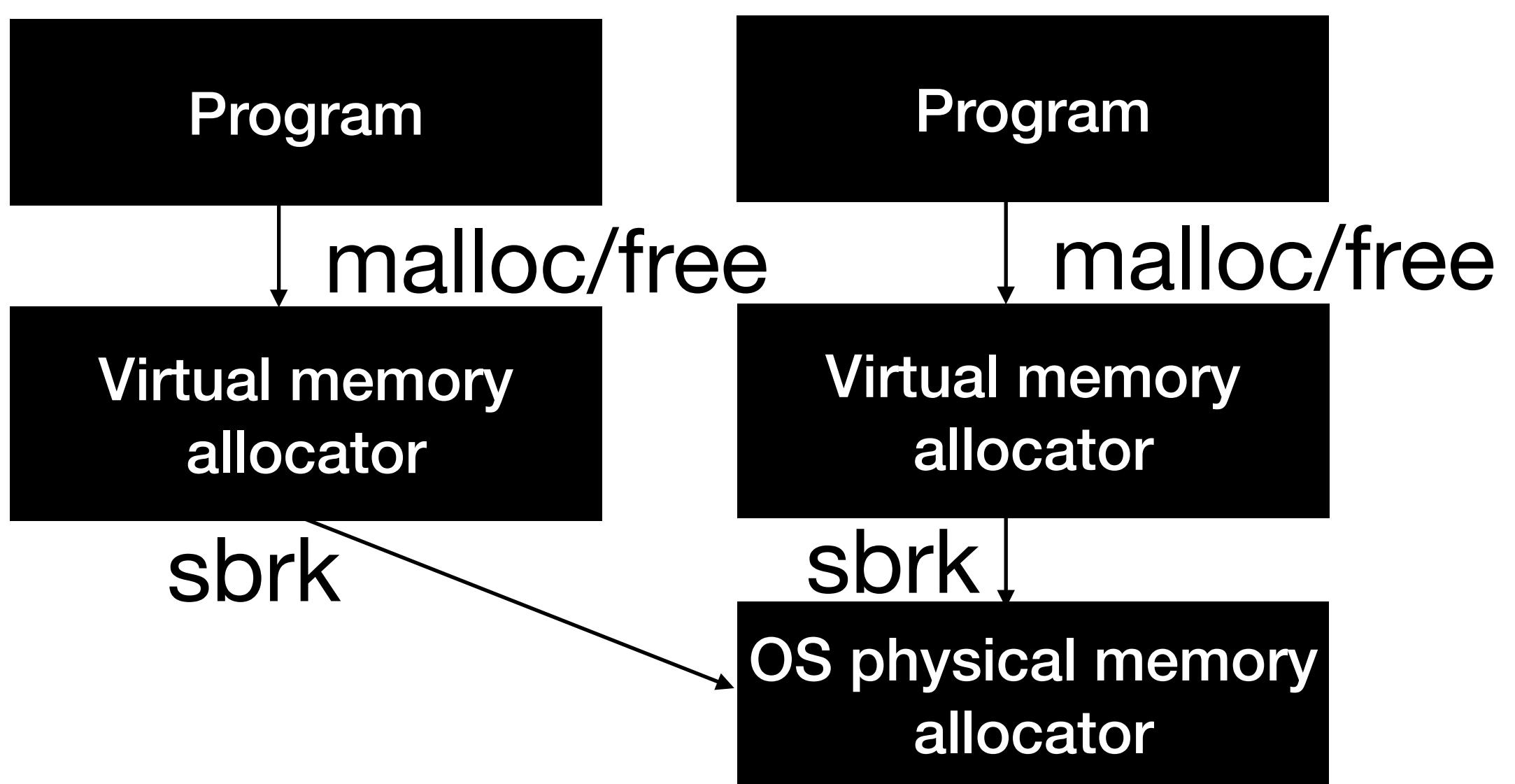
OS memory allocator

- `sbrk(int increment)` increments process' break. *increment* can be negative.



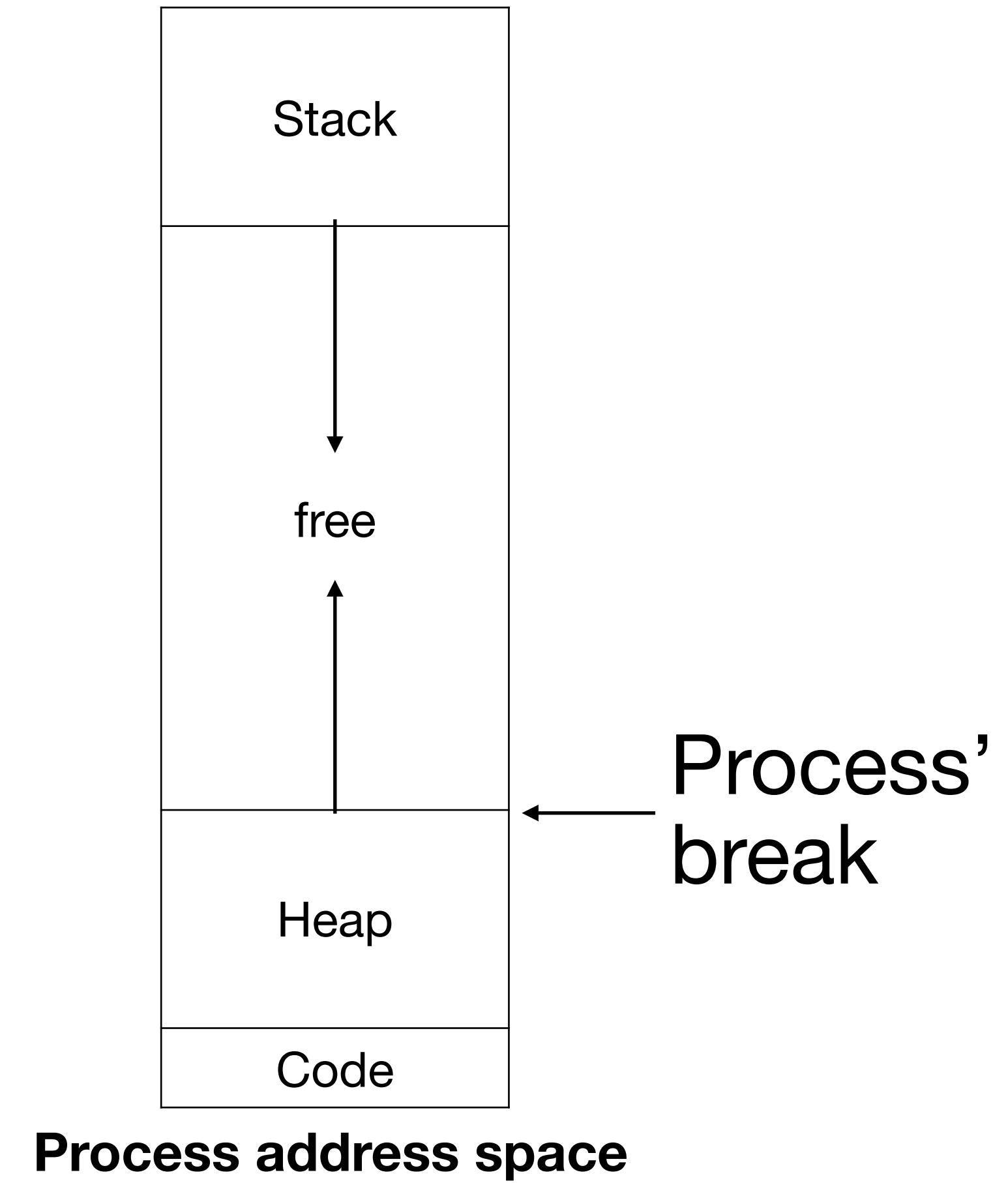
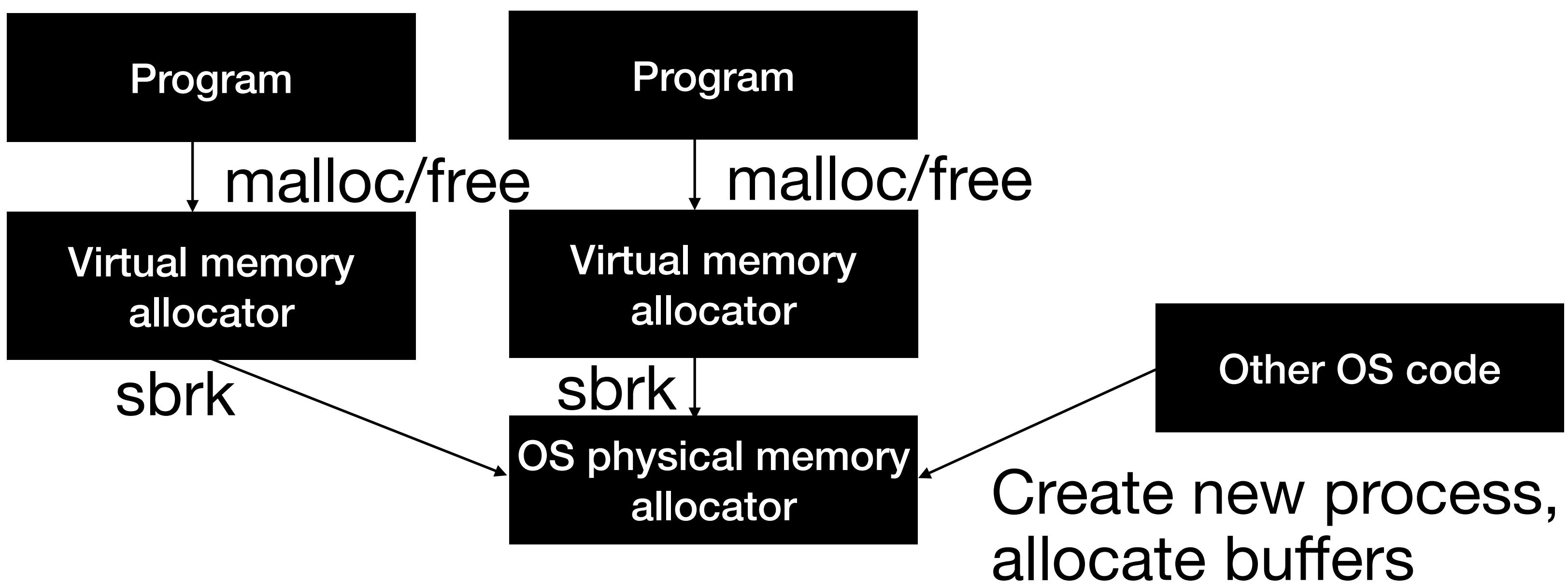
OS memory allocator

- `sbrk(int increment)` increments process' break. *increment* can be negative.



OS memory allocator

- `sbrk(int increment)` increments process' break. *increment* can be negative.



Memory allocation

```
ptr=malloc(size_t size);
```

```
free(ptr);
```

Memory allocation

```
ptr=malloc(size_t size);  
  
free(ptr);
```

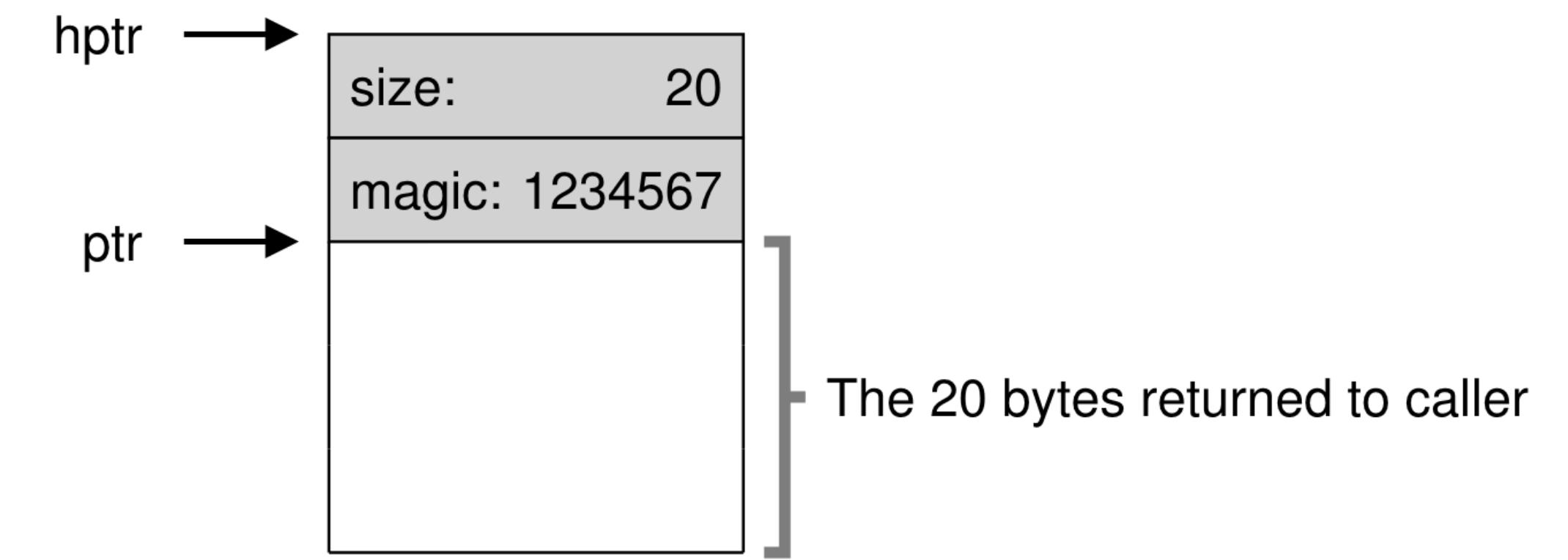


Figure 17.2: Specific Contents Of The Header

Memory allocation

```
ptr=malloc(size_t size);  
free(ptr);
```

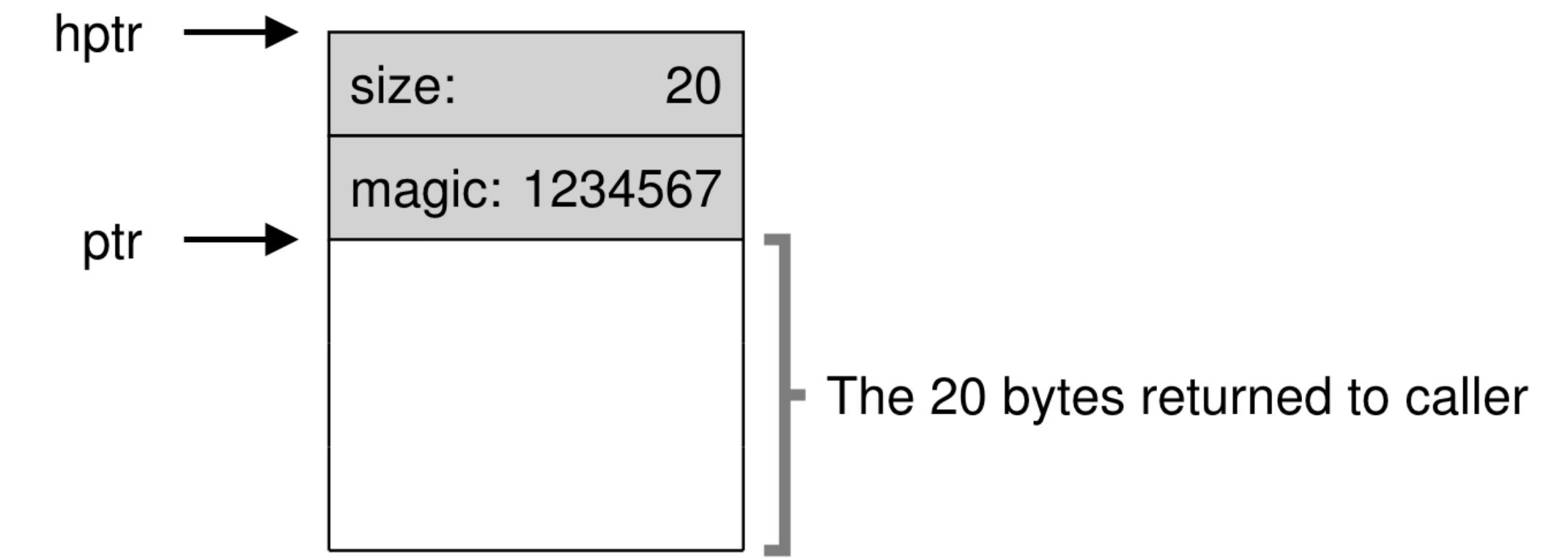
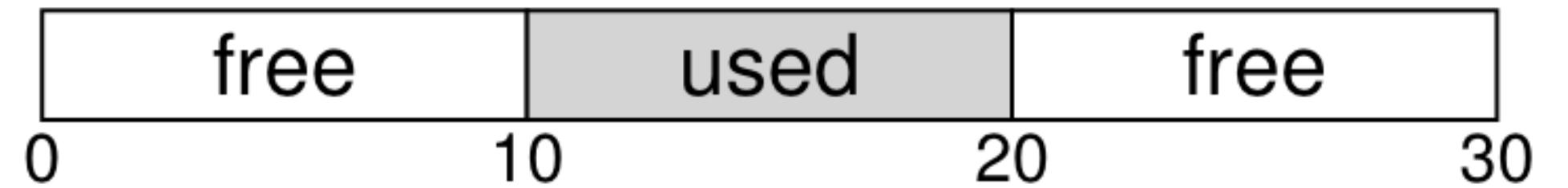


Figure 17.2: Specific Contents Of The Header

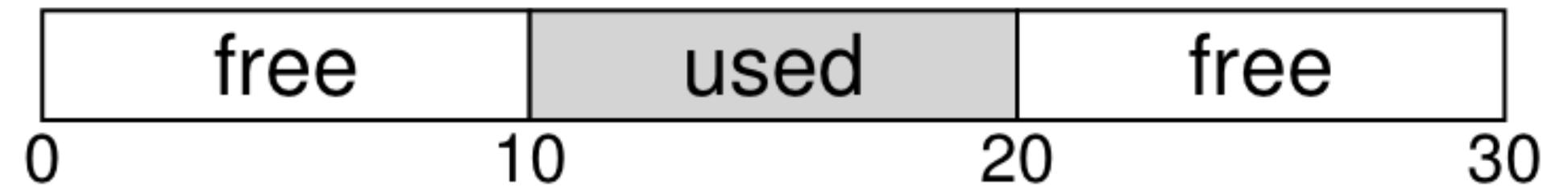
Stack allocations are faster than heap allocations. No need to find space.

Memory allocator



Fragmented heap over time

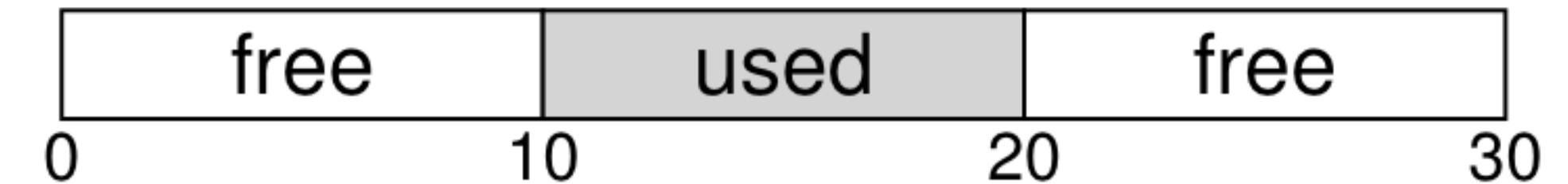
Memory allocator



Fragmented heap over time

- Assumptions
 - Do not apriori know allocation size and order
 - Cannot move memory once it is allocated. Program might have the pointer to it.

Memory allocator



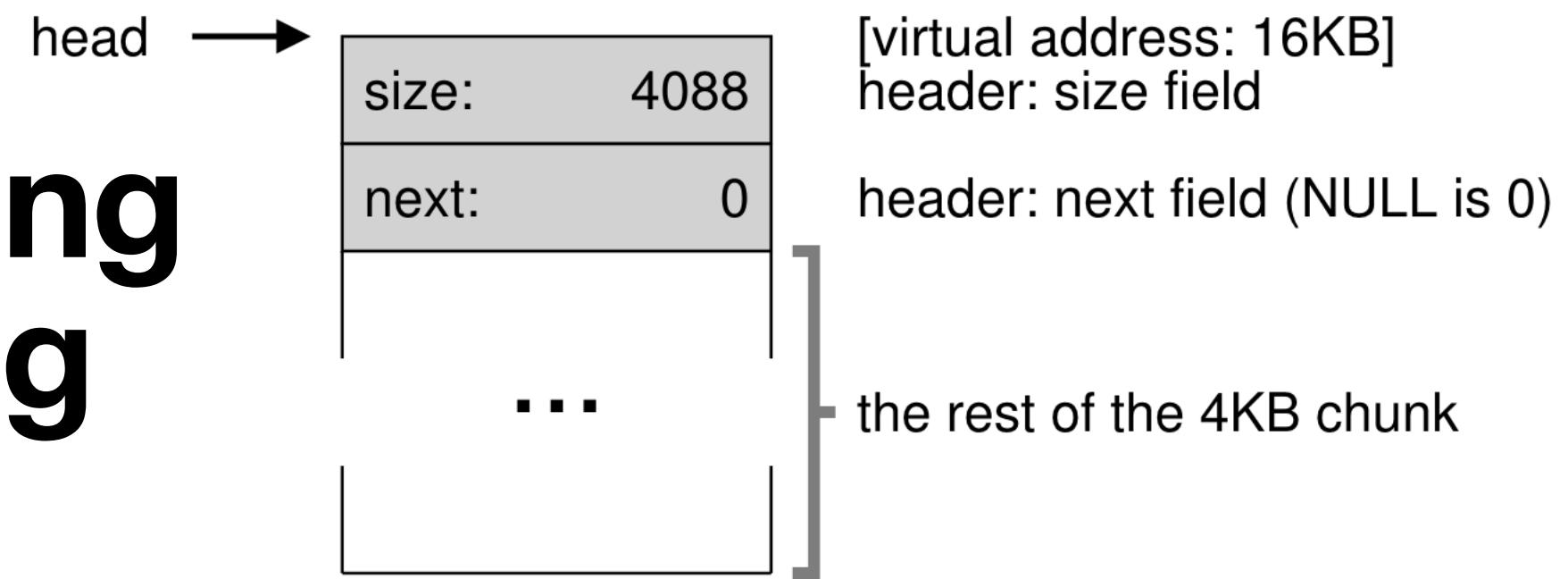
Fragmented heap over time

- Assumptions
 - Do not apriori know allocation size and order
 - Cannot move memory once it is allocated. Program might have the pointer to it.
- Goals
 - Quickly satisfy variable-sized memory allocation requests. How to track free memory?
 - Minimize fragmentation

Memory (de)allocation patterns

- Small mallocs can be frequent. Large mallocs are usually infrequent.
 - After malloc, program will initialise the memory area.
- “Clustered deaths”: Objects allocated together die together.

Free list splitting and coalescing



```
ptr = malloc(100)
```

Figure 17.3: A Heap With One Free Chunk

```
sptr = malloc(100)
```

```
optr = malloc(100)
```

```
free(sptr)
```

```
free(ptr)
```

```
free(optr)
```

Free list splitting and coalescing

```
ptr = malloc(100)
```

```
sptr = malloc(100)
```

```
optr = malloc(100)
```

```
free(sptr)
```

```
free(ptr)
```

```
free(optr)
```

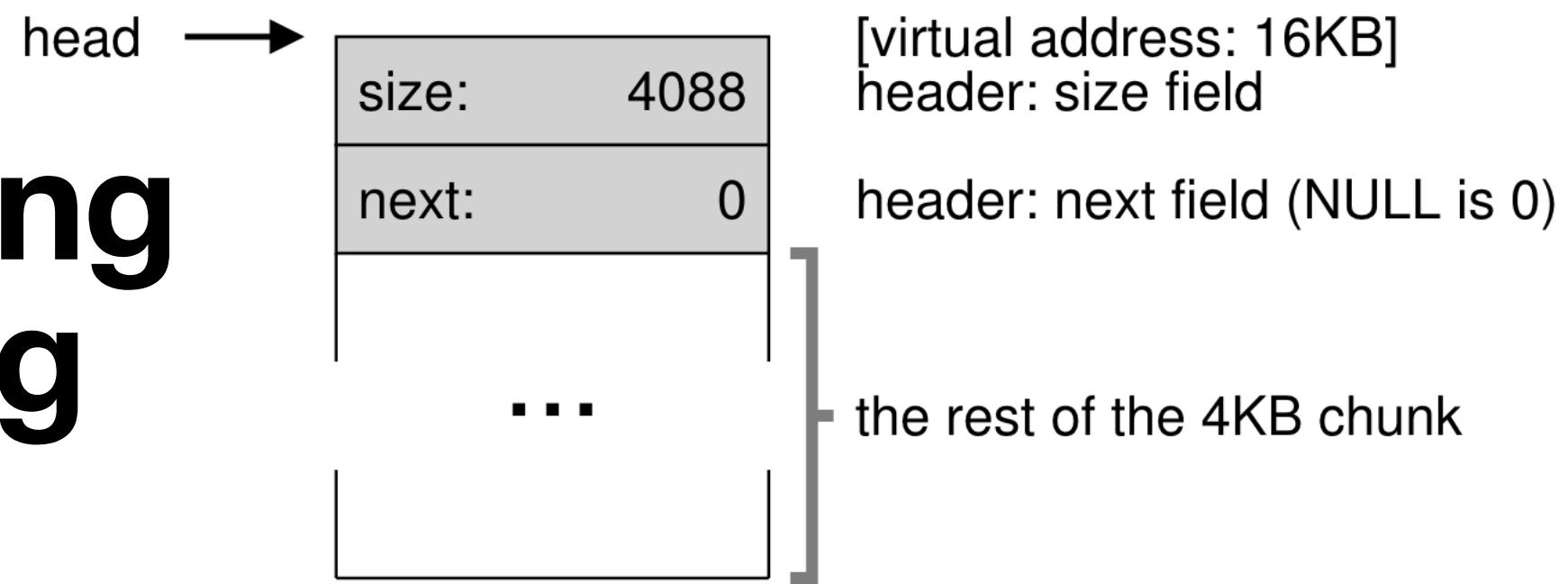


Figure 17.3: A Heap With One Free Chunk

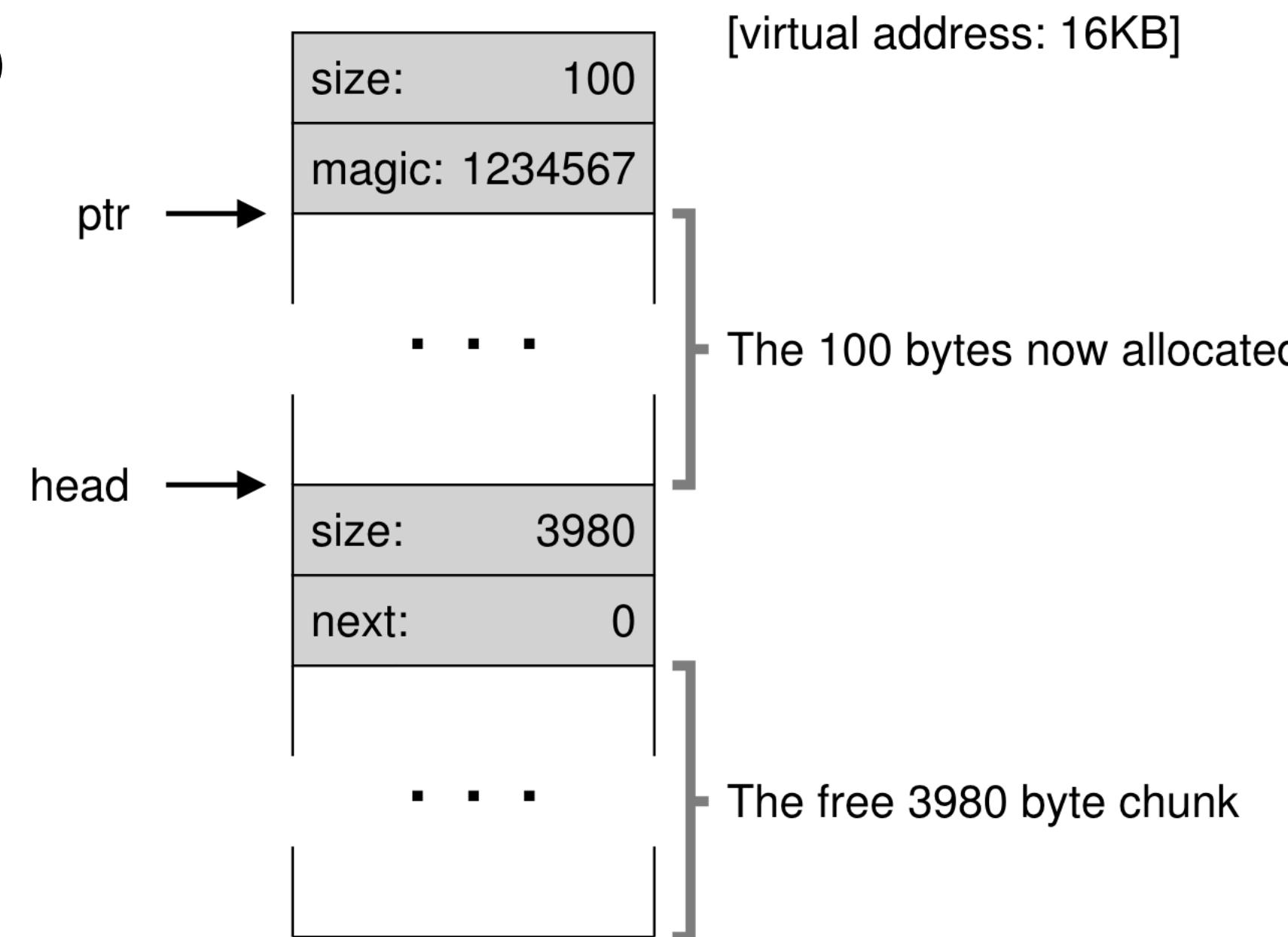


Figure 17.4: A Heap: After One Allocation

Free list splitting and coalescing

```
ptr = malloc(100)
```

```
sptr = malloc(100)
```

```
optr = malloc(100)
```

```
free(sptr)
```

```
free(ptr)
```

```
free(optr)
```

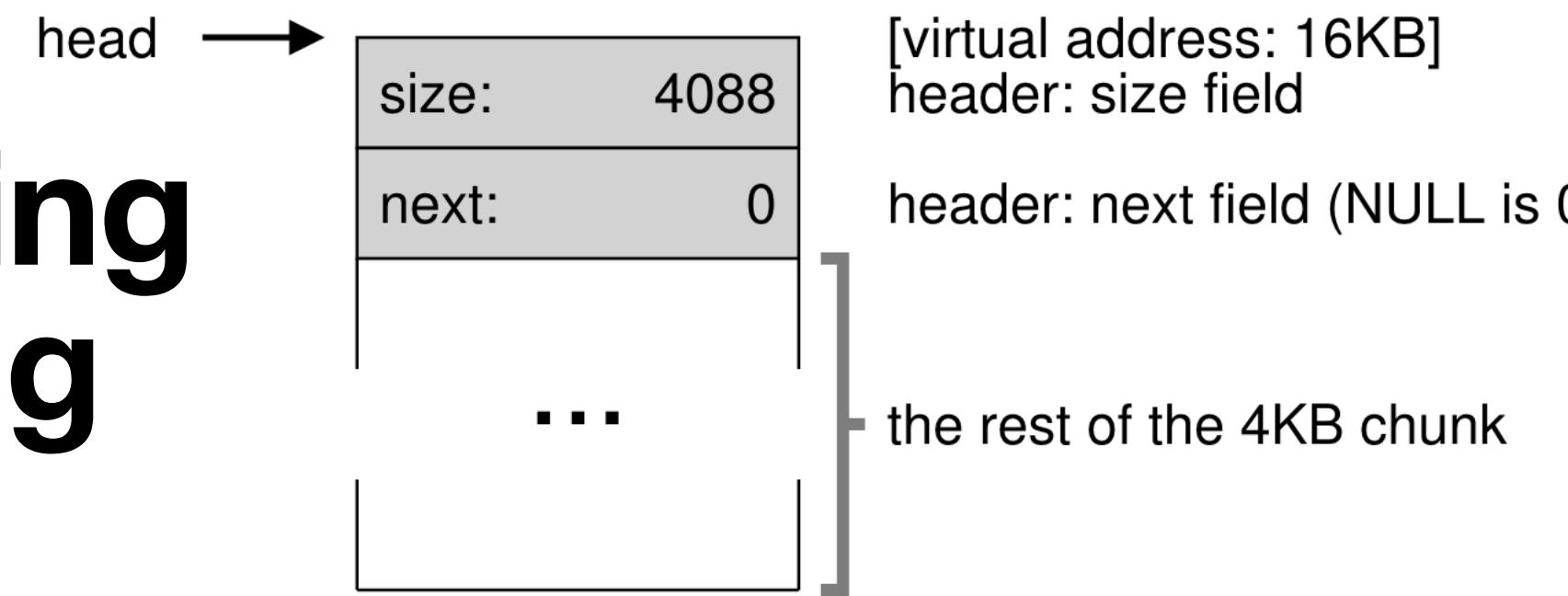


Figure 17.3: A Heap With One Free Chunk

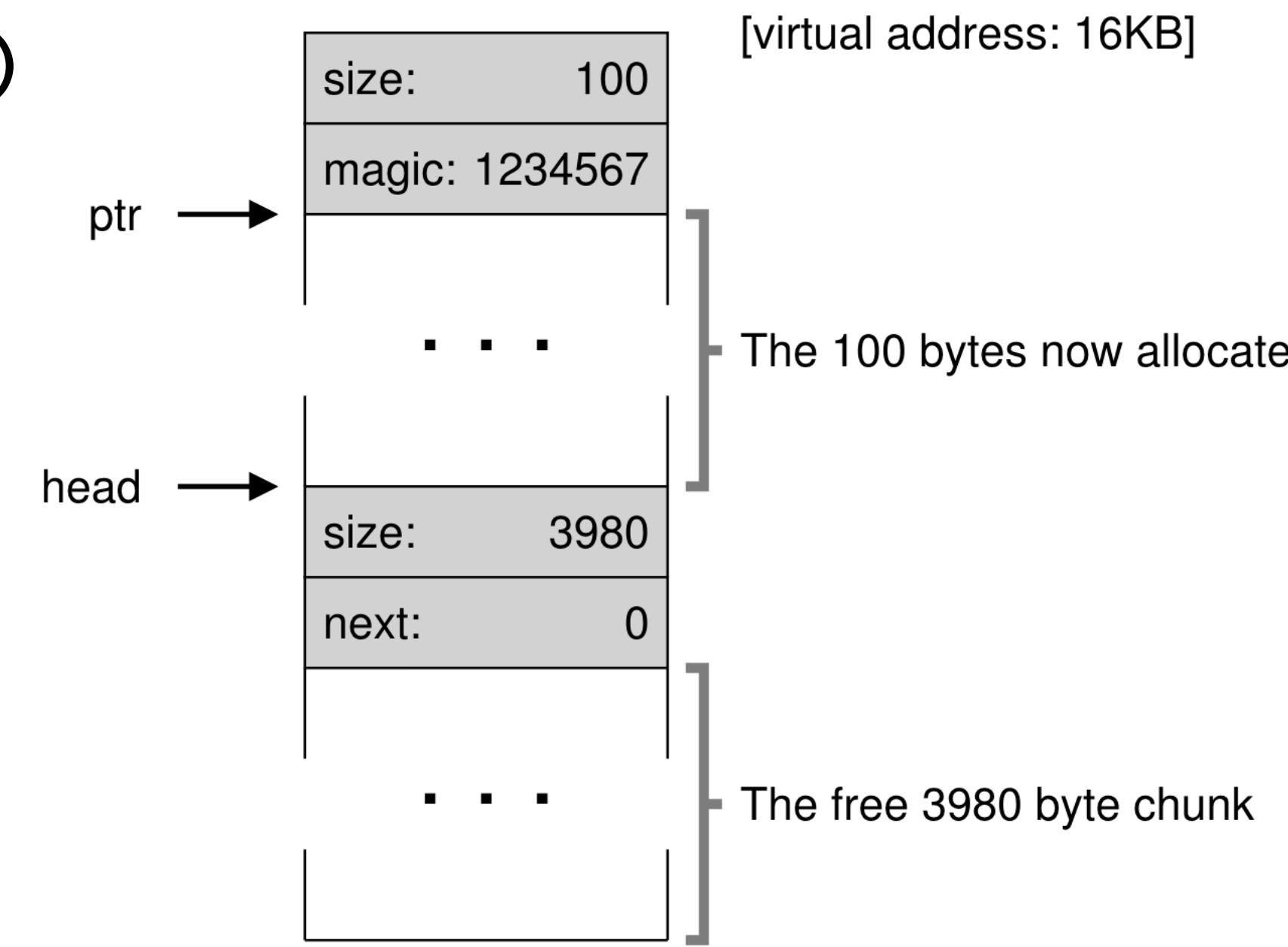


Figure 17.4: A Heap: After One Allocation

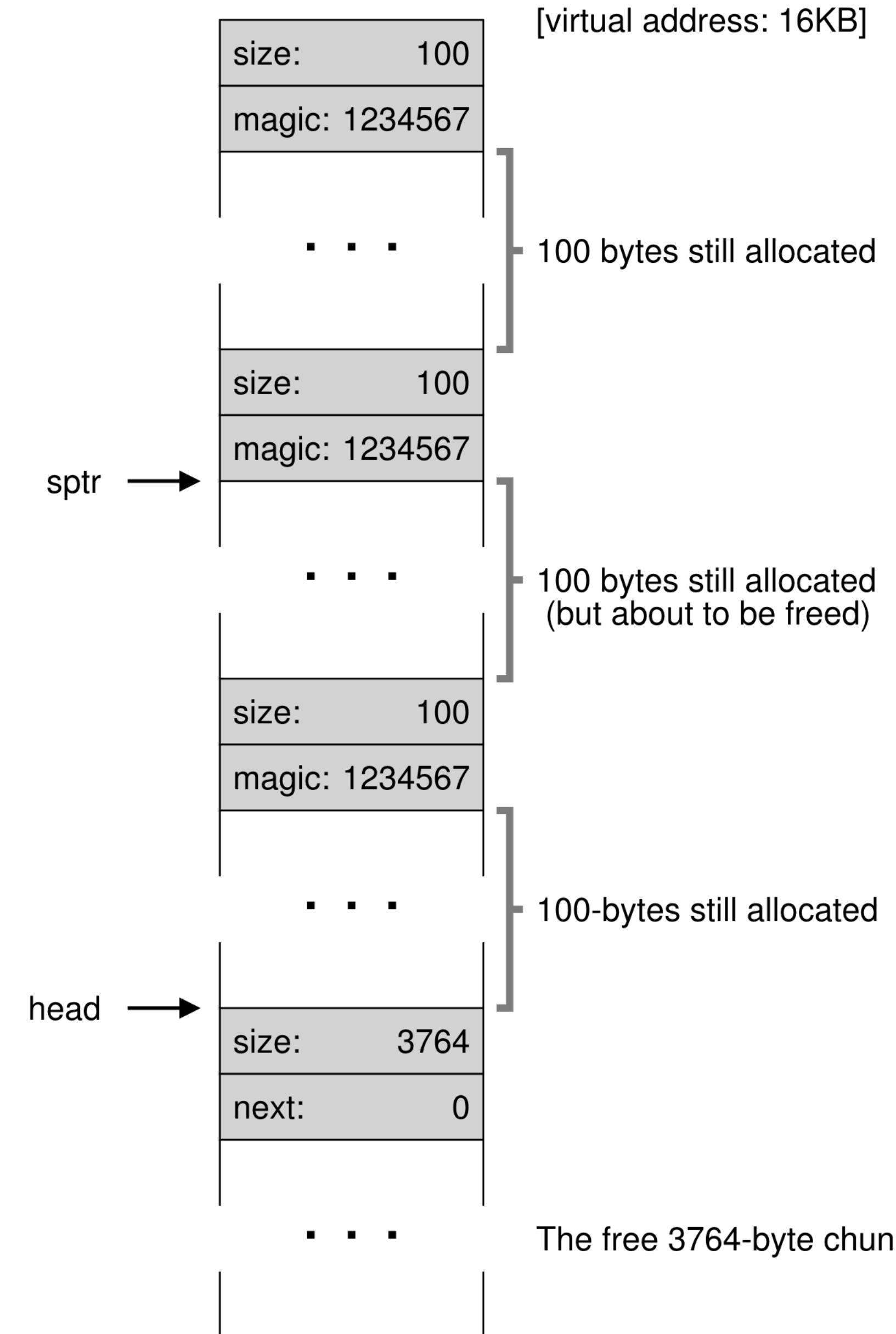


Figure 17.5: Free Space With Three Chunks Allocated

Free list splitting and coalescing

```
ptr = malloc(100)
```

```
sptr = malloc(100)
```

```
optr = malloc(100)
```

```
free(sptr)
```

```
free(ptr)
```

```
free(optr)
```

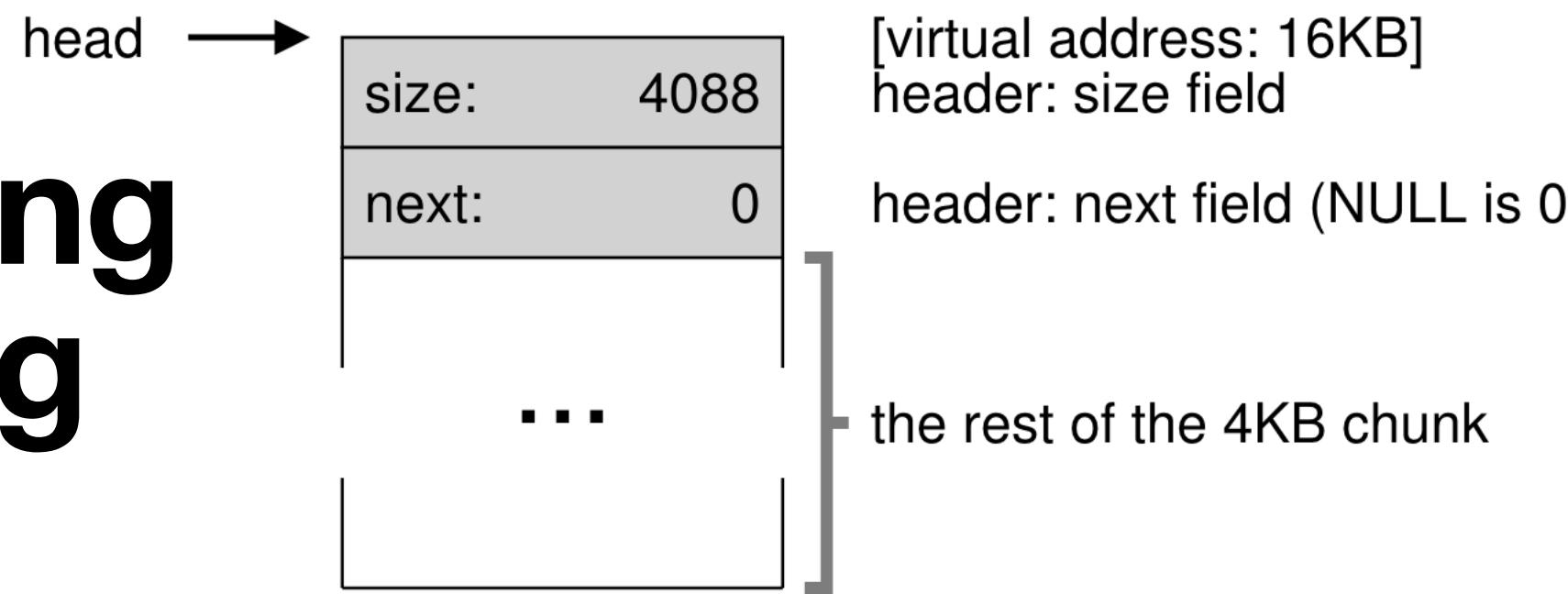


Figure 17.3: A Heap With One Free Chunk

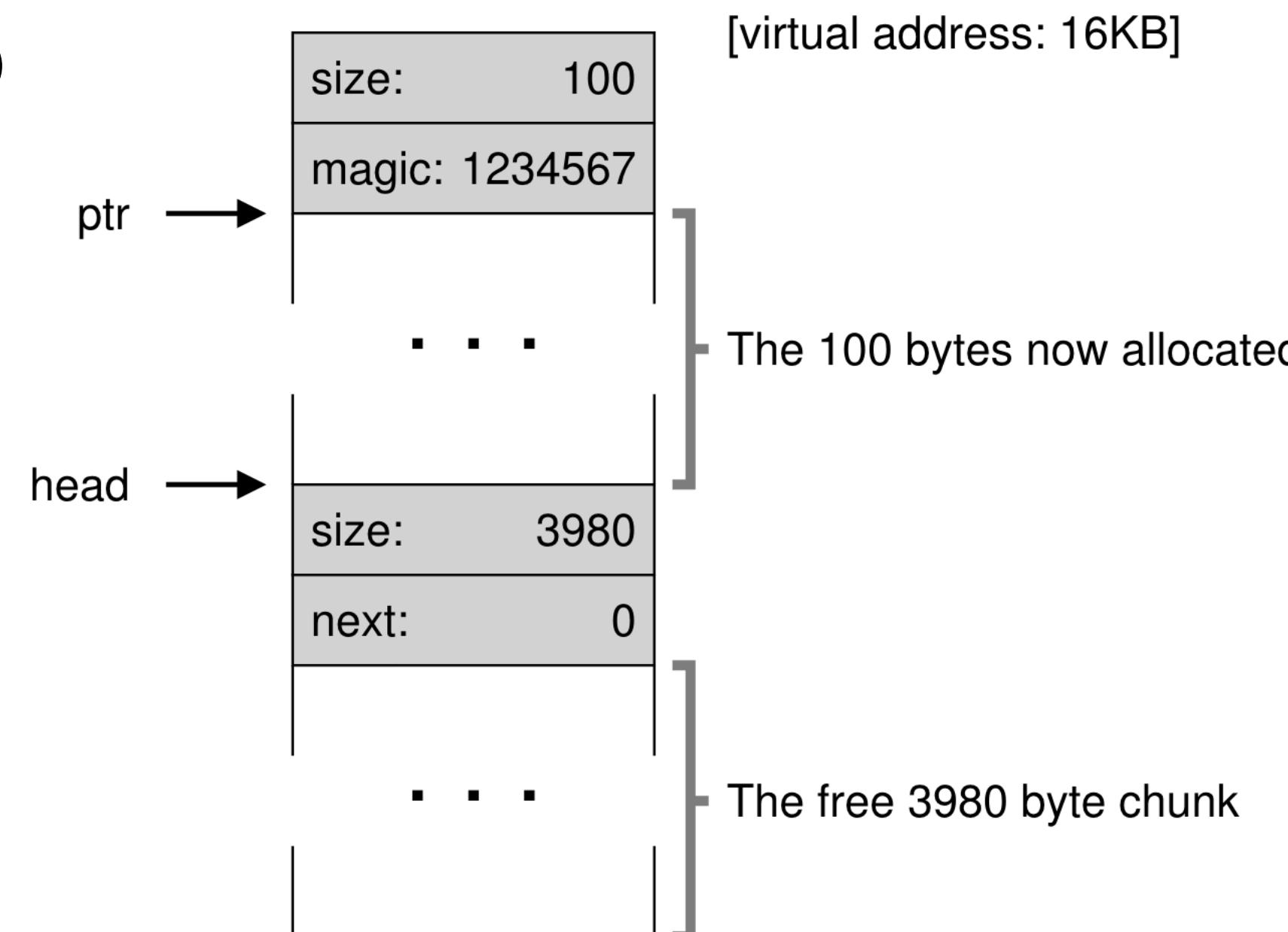


Figure 17.4: A Heap: After One Allocation

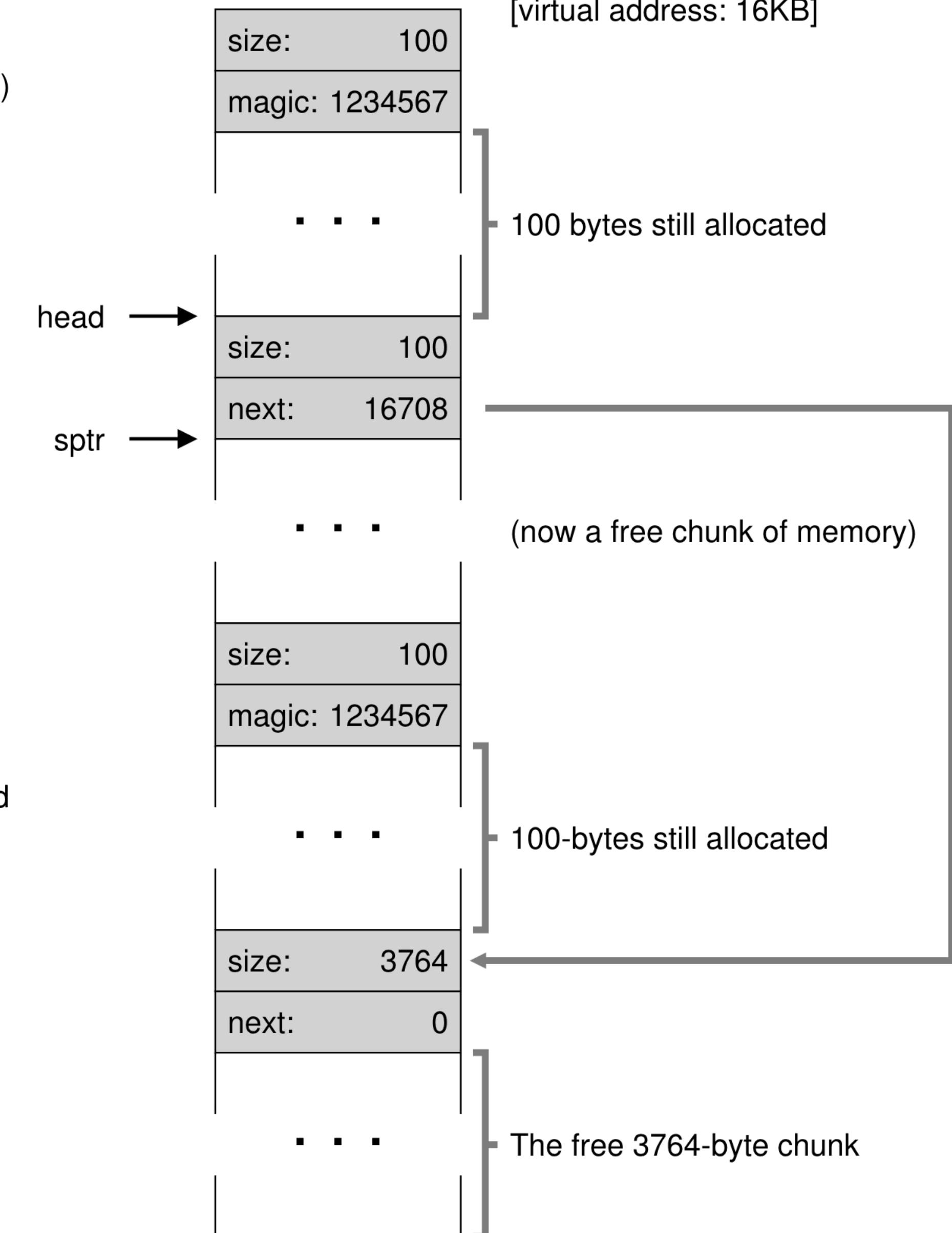


Figure 17.6: Free Space With Two Chunks Allocated

Free list splitting and coalescing

```
ptr = malloc(100)
```

```
sptr = malloc(100)
```

```
optr = malloc(100)
```

```
free(sptr)
```

```
free(ptr)
```

```
free(optr)
```

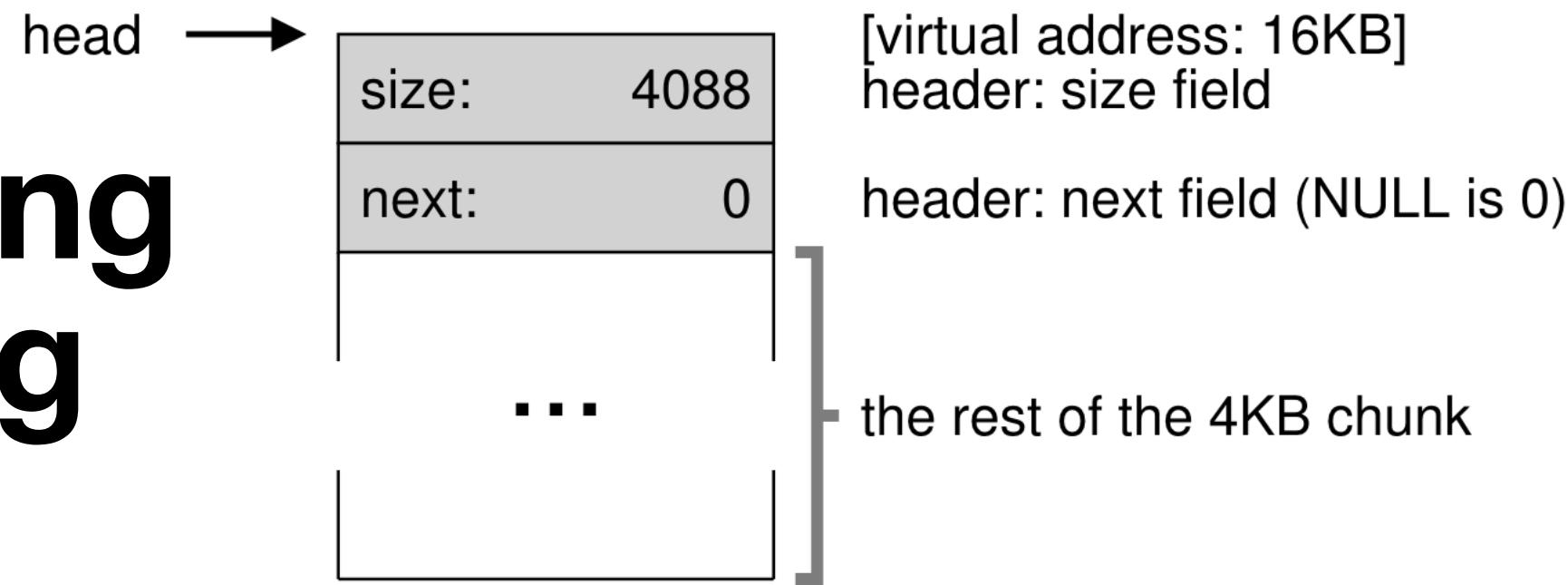


Figure 17.3: A Heap With One Free Chunk

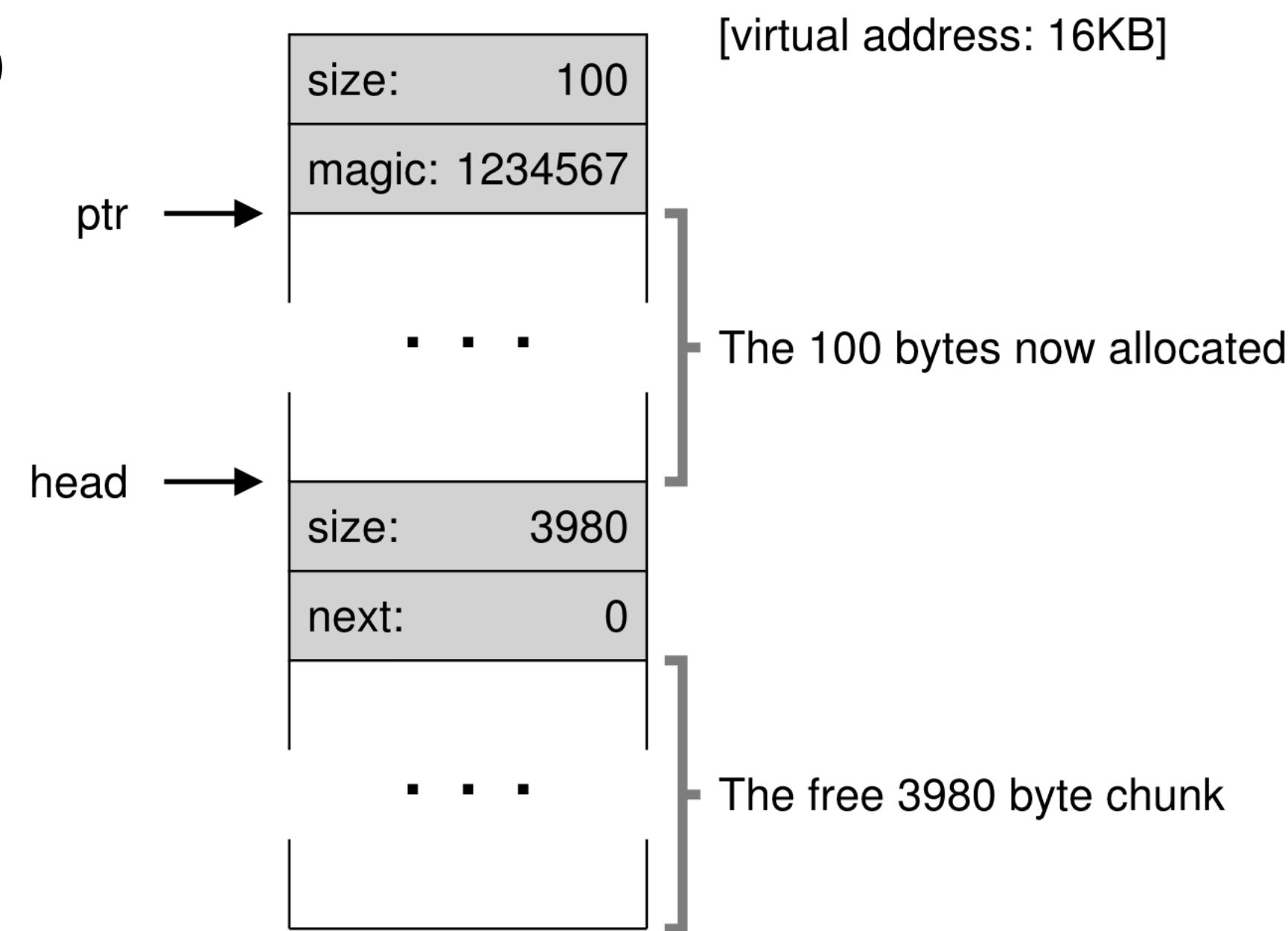


Figure 17.4: A Heap: After One Allocation

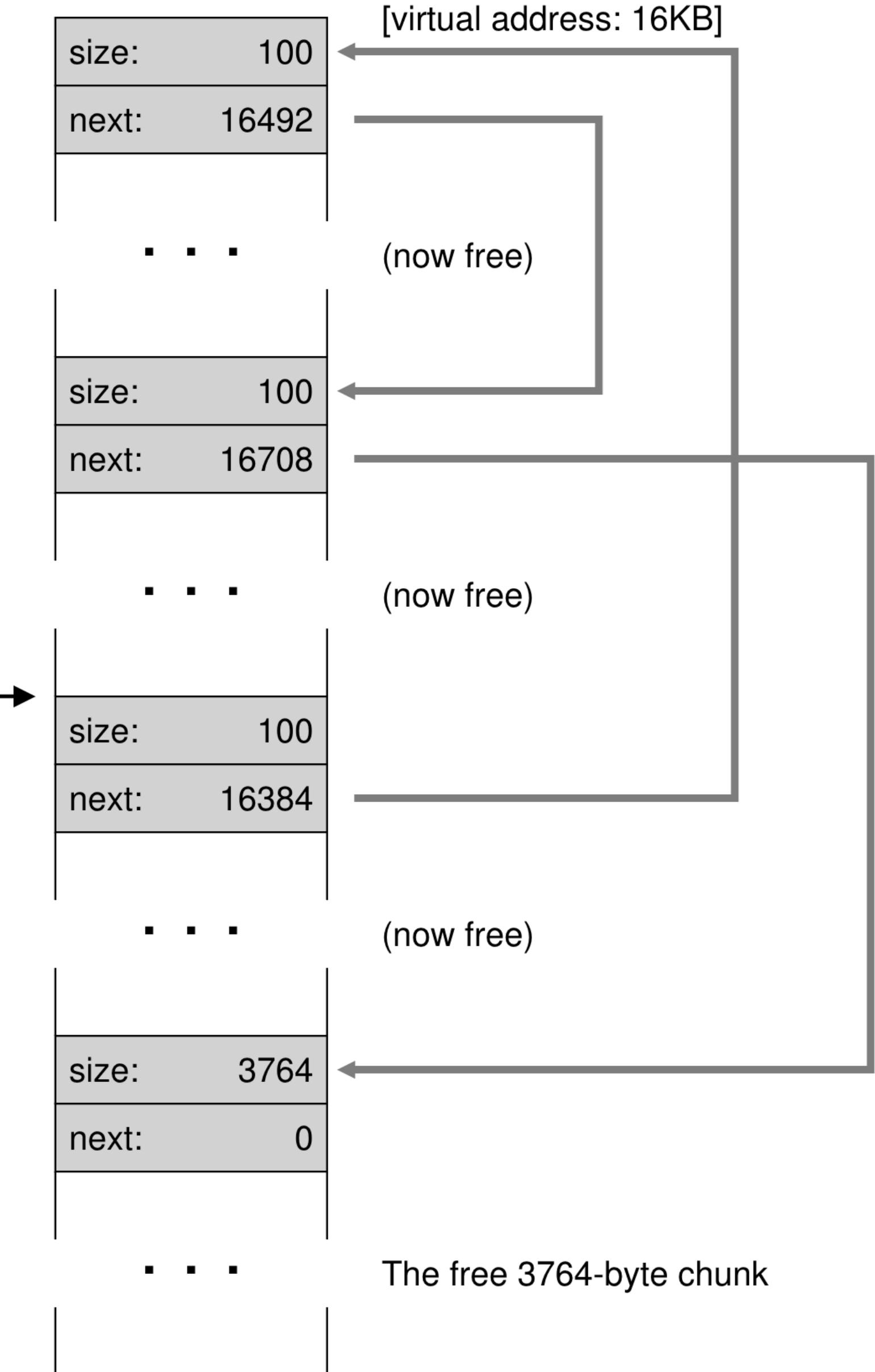


Figure 17.7: A Non-Coalesced Free List

Free list splitting and coalescing

```
ptr = malloc(100)
```

```
sptr = malloc(100)
```

```
optr = malloc(100)
```

```
free(sptr)
```

```
free(ptr)
```

```
free(optr)
```

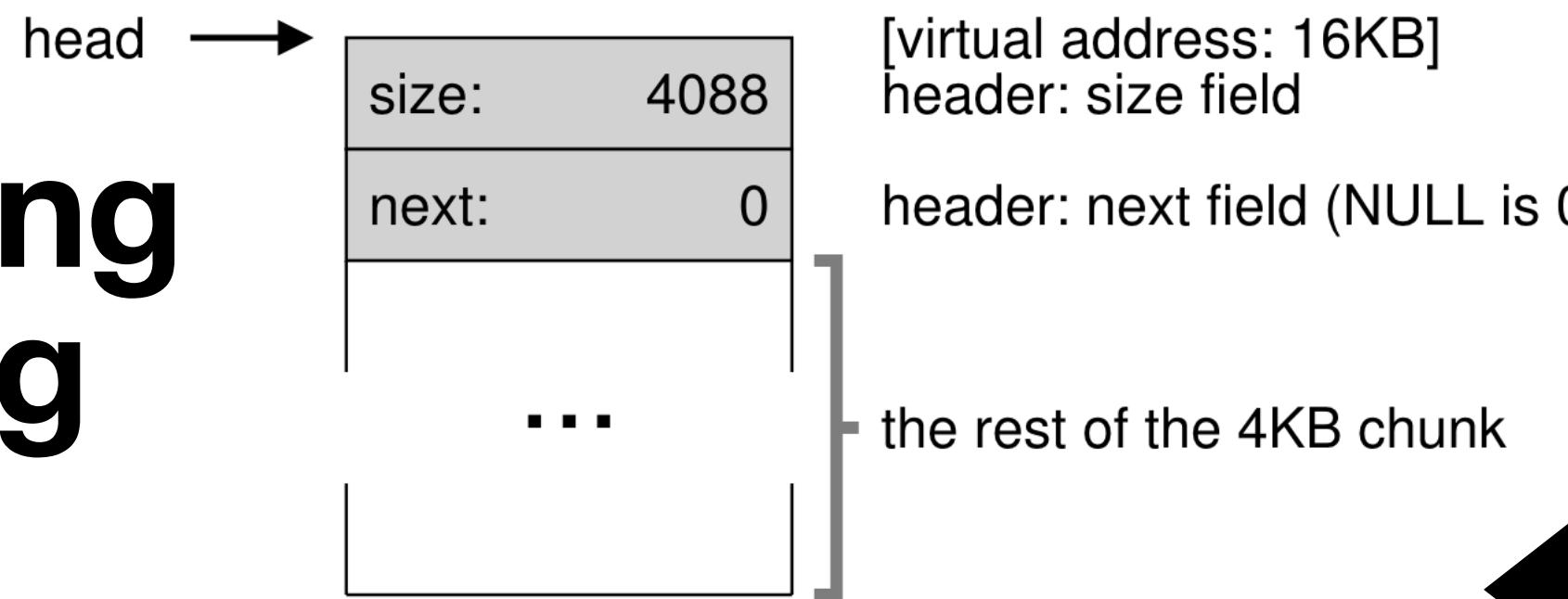


Figure 17.3: A Heap With One Free Chunk

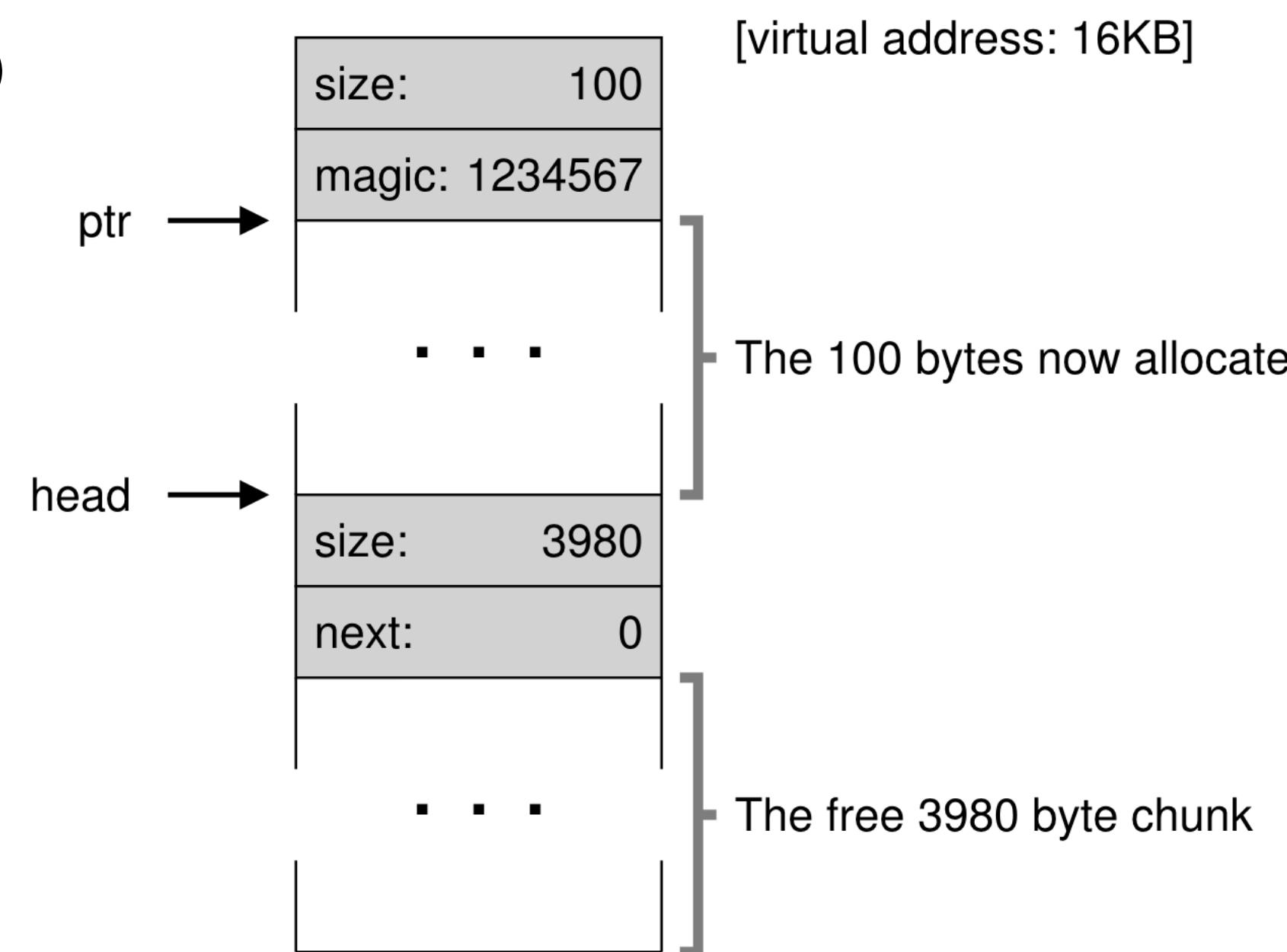


Figure 17.4: A Heap: After One Allocation

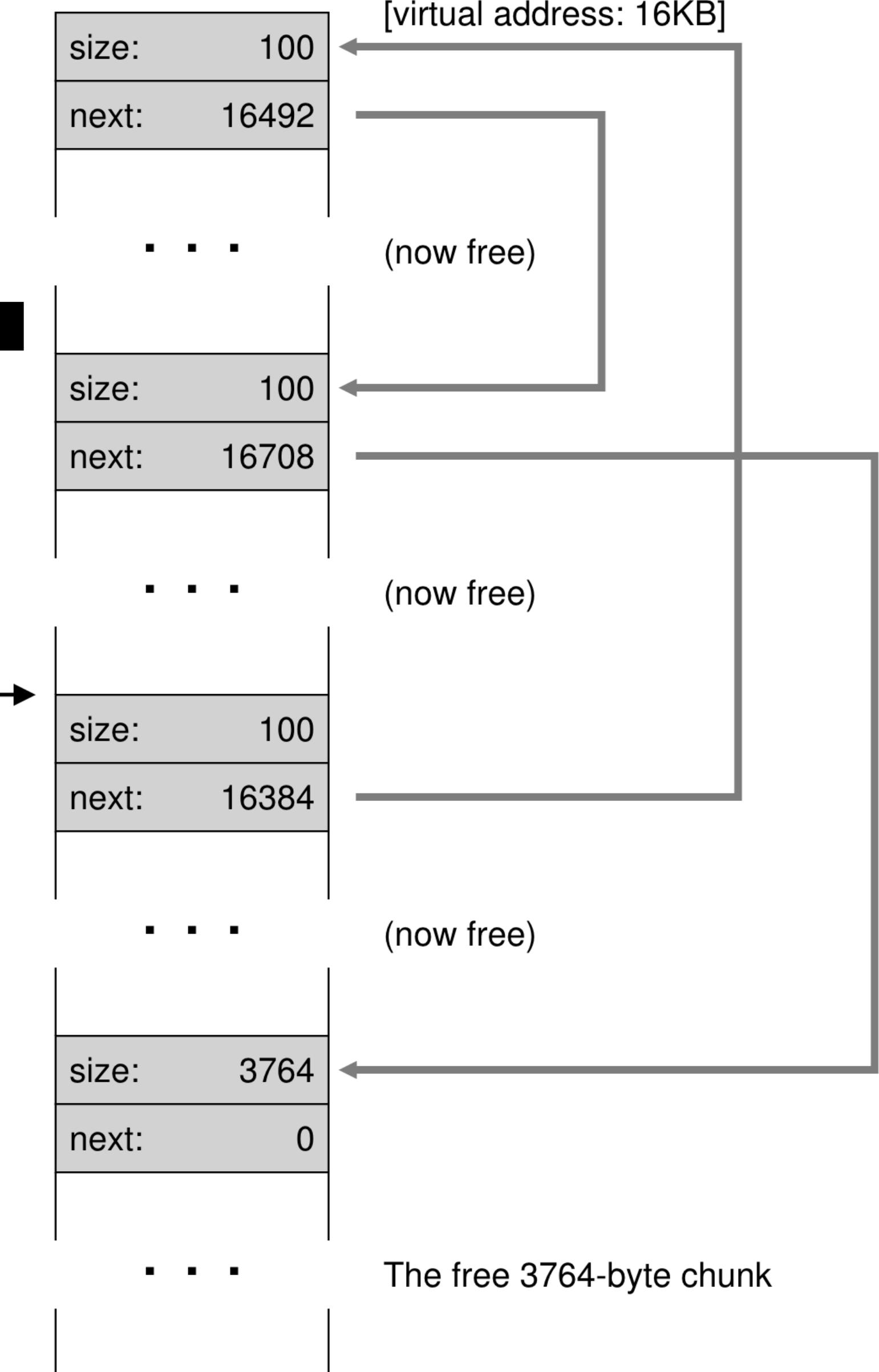


Figure 17.7: A Non-Coalesced Free List

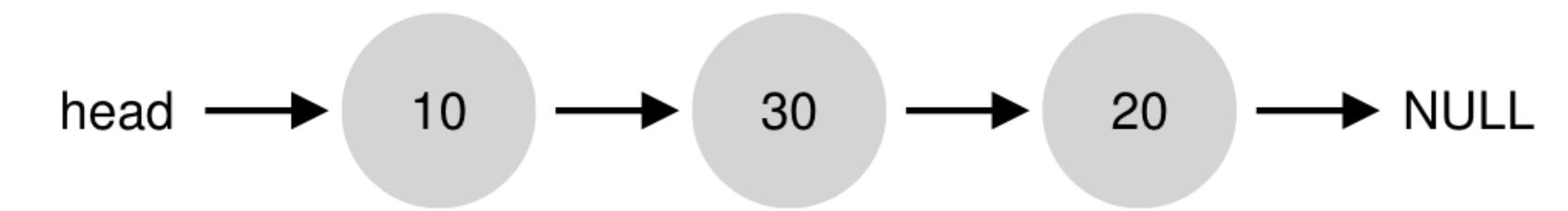
Which block to allocate?

Example: malloc(15)



Which block to allocate?

Example: malloc(15)



- Best fit
 - Slow. need to search the whole list

Which block to allocate?

Example: malloc(15)

- Best fit
 - Slow. need to search the whole list



Which block to allocate?

Example: malloc(15)

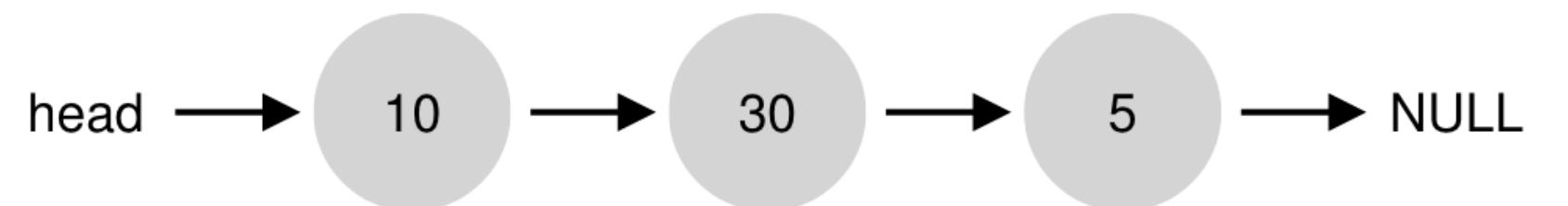
- Best fit
 - Slow. need to search the whole list
- First fit
 - Faster. (xv6: umalloc.c)



Which block to allocate?

Example: malloc(15)

- Best fit
 - Slow. need to search the whole list
- First fit
 - Faster. (xv6: umalloc.c)



Which block to allocate?

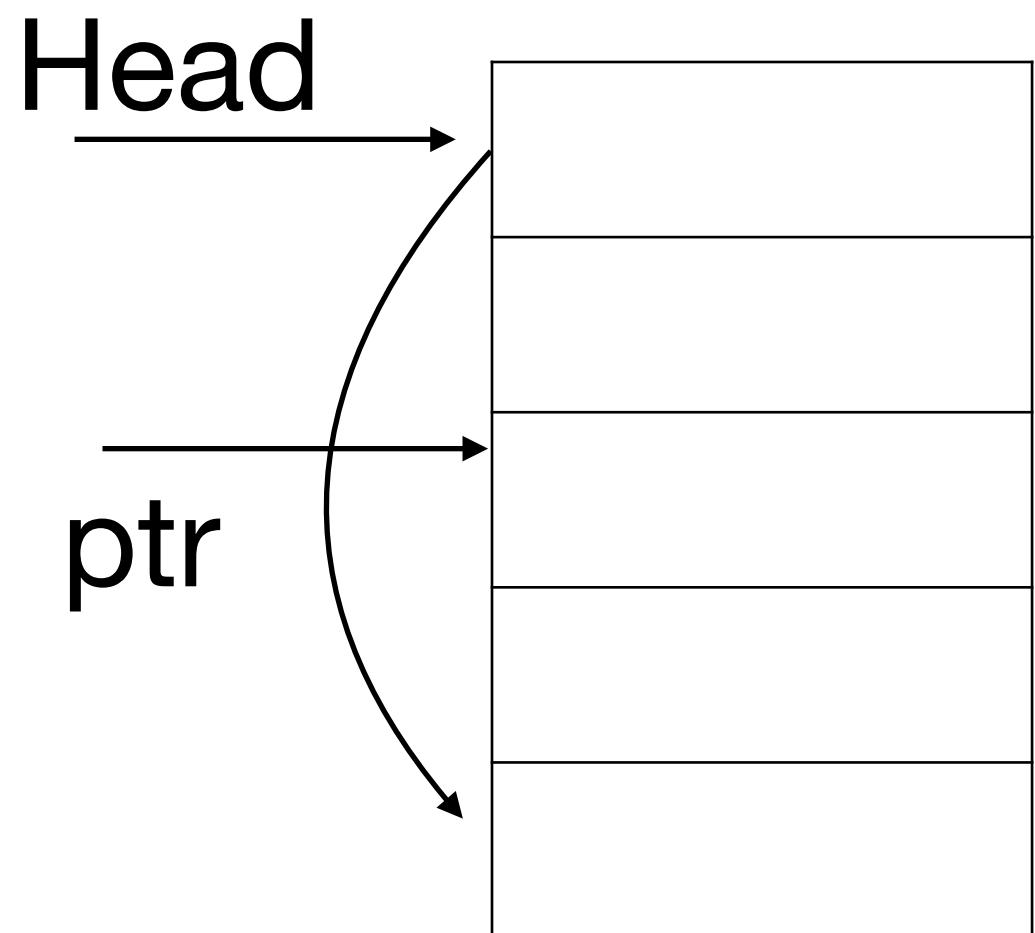
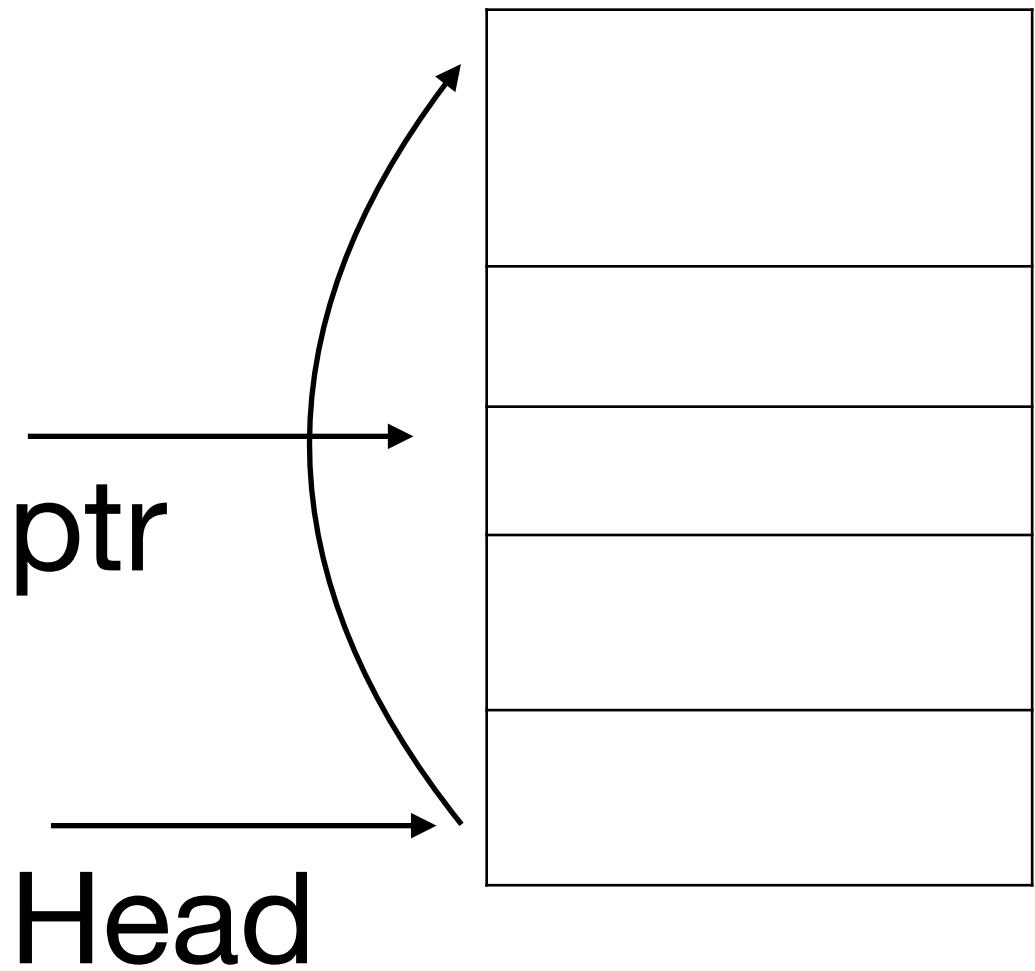
Example: malloc(15)

- Best fit
 - Slow. need to search the whole list
- First fit
 - Faster. (xv6: umalloc.c)
- Fragmentation
 - Example: malloc(25)



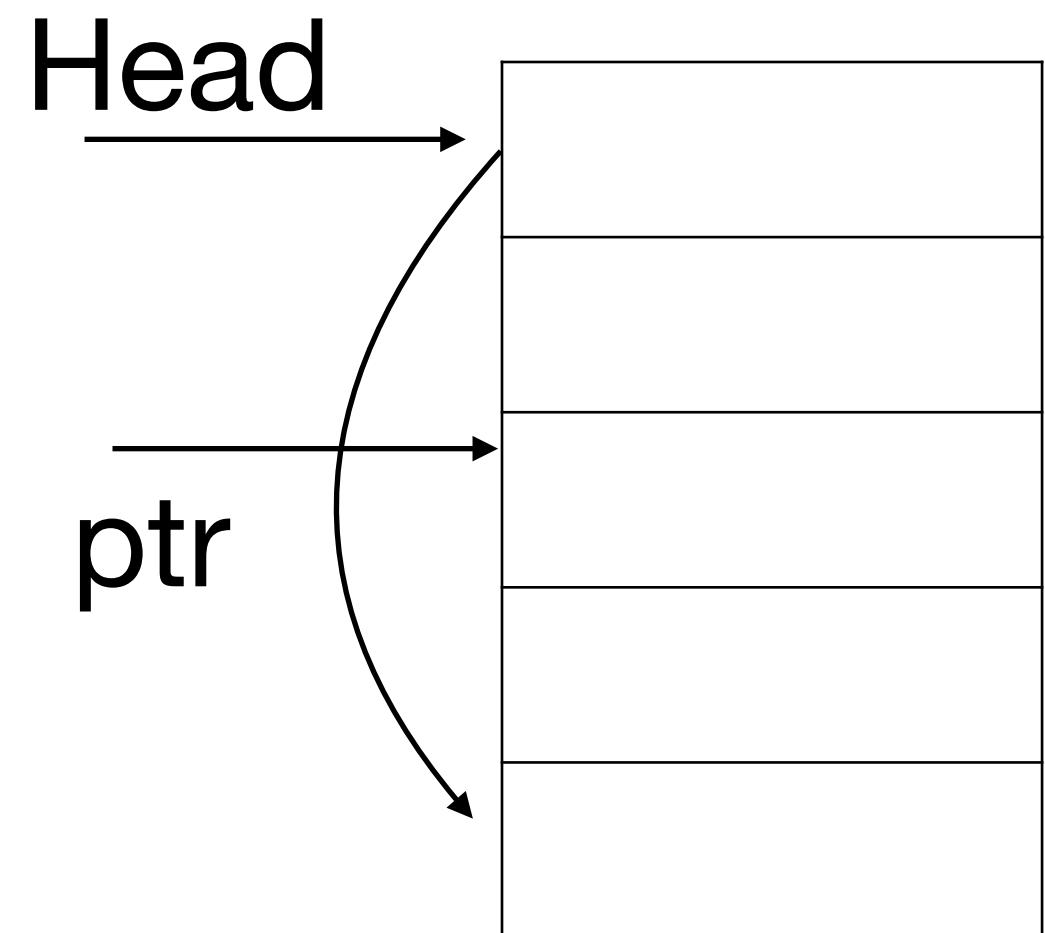
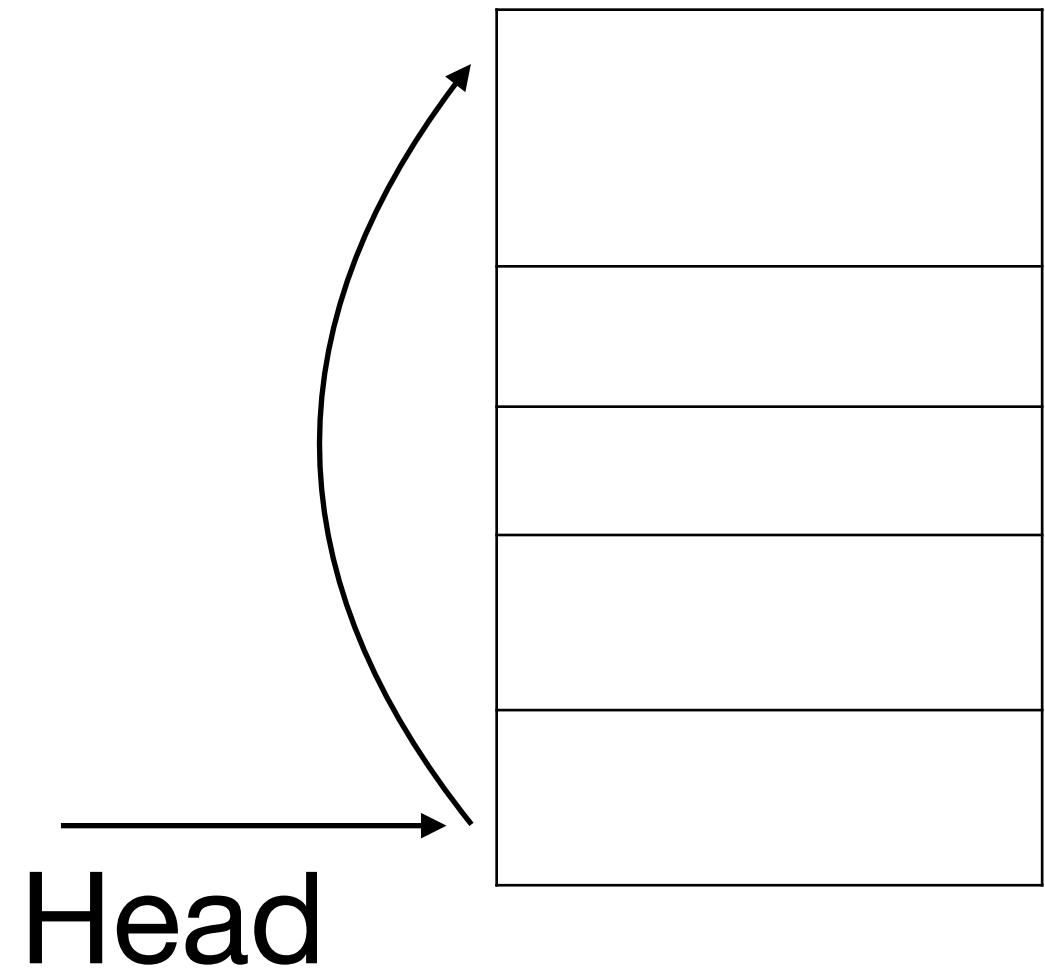
In which order to maintain lists?

- (De)allocation order



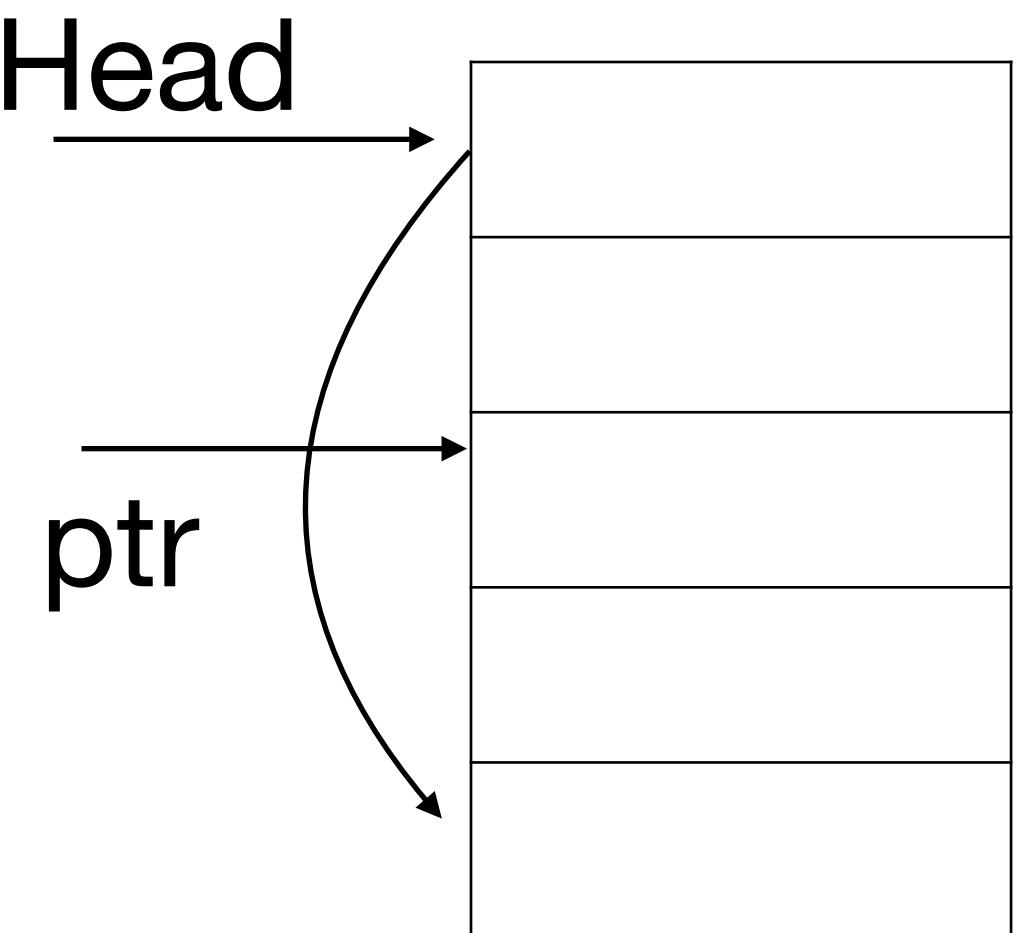
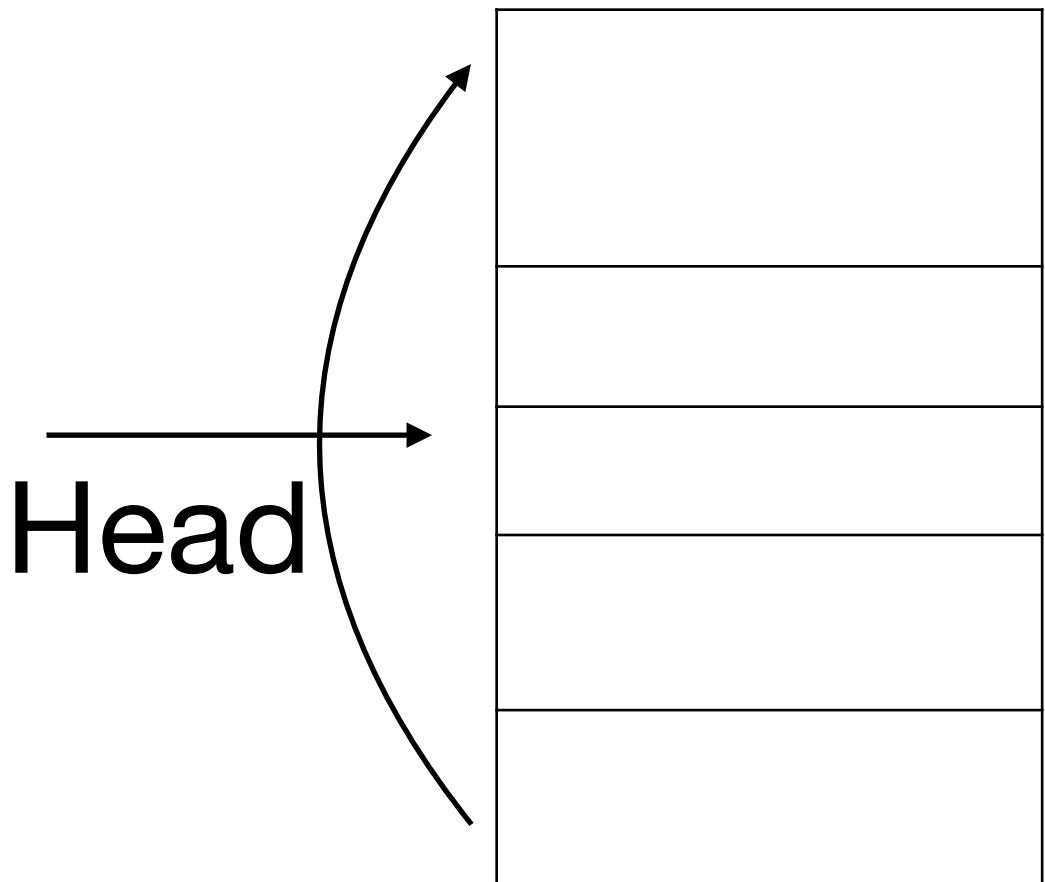
In which order to maintain lists?

- (De)allocation order



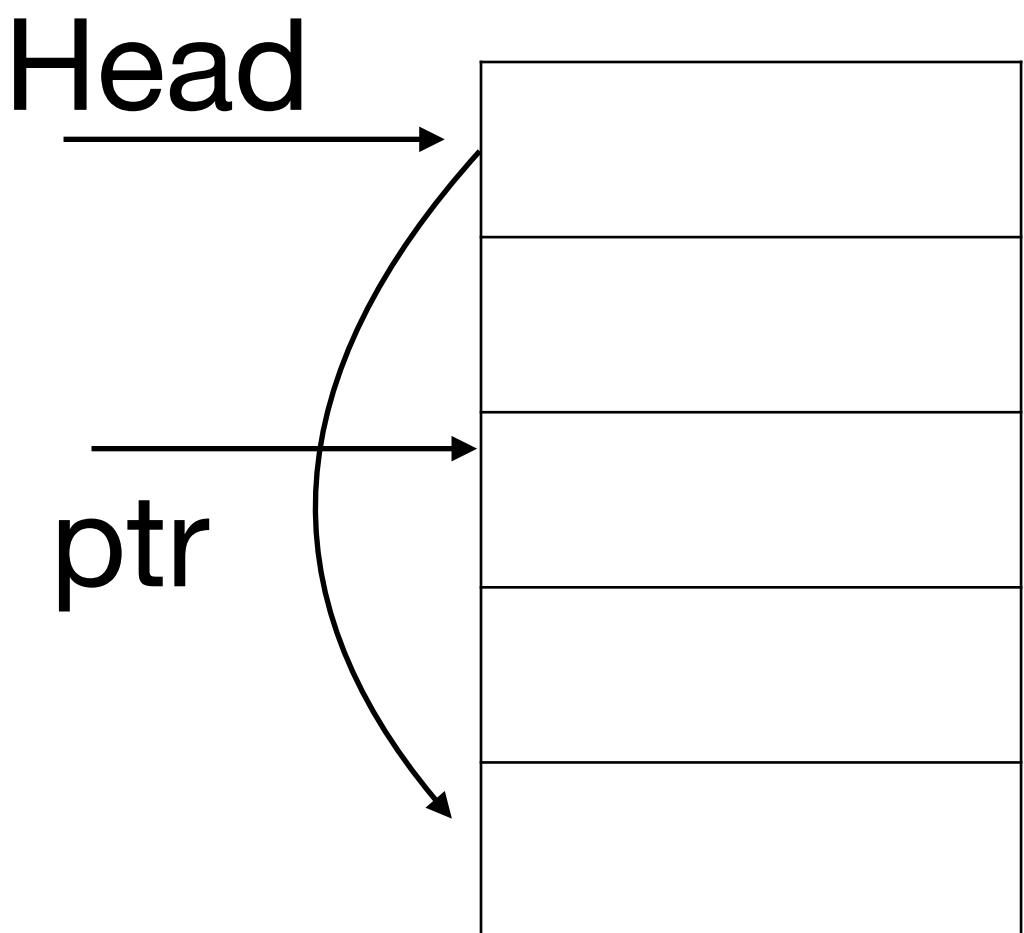
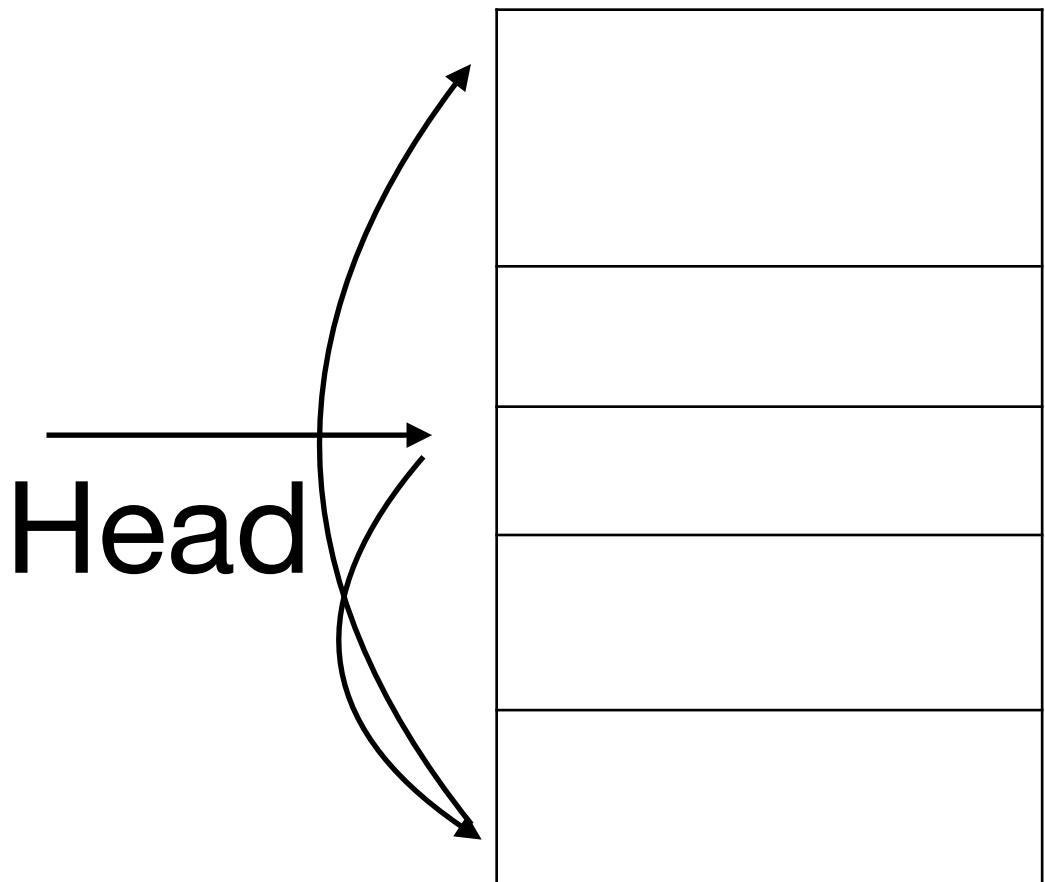
In which order to maintain lists?

- (De)allocation order



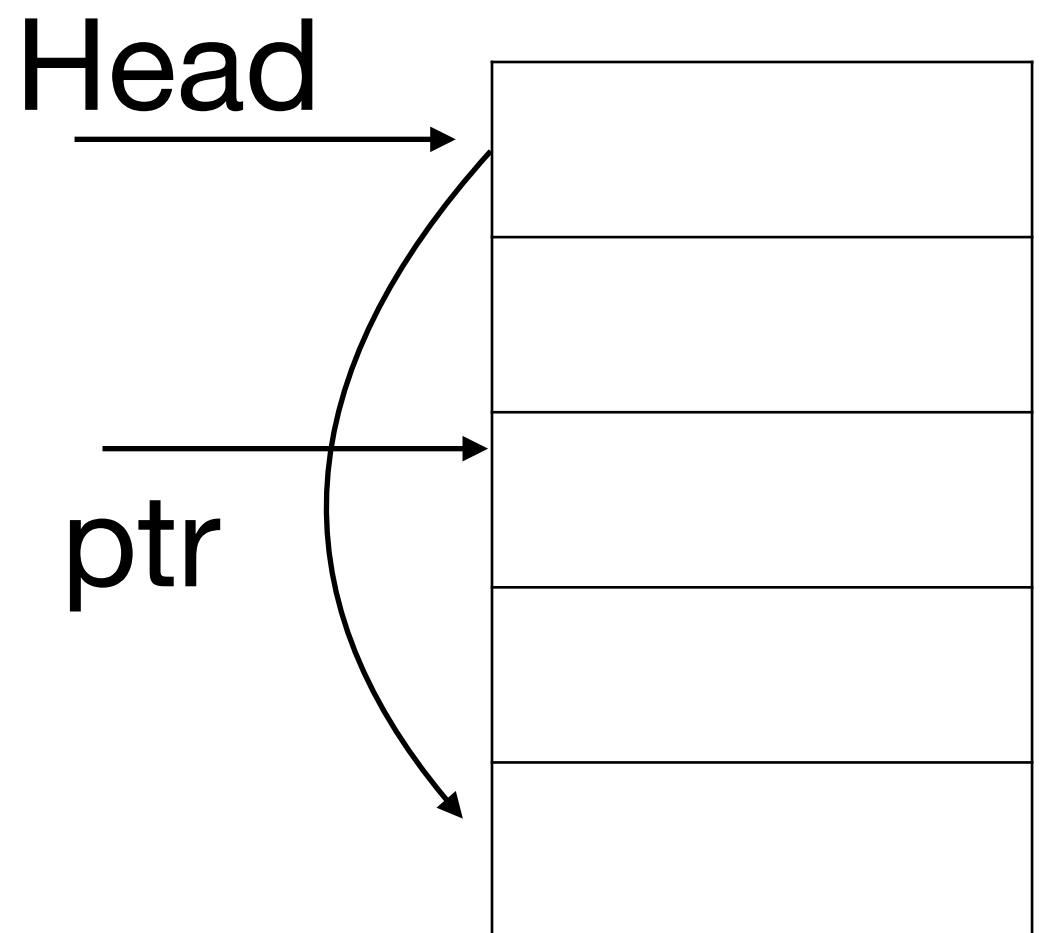
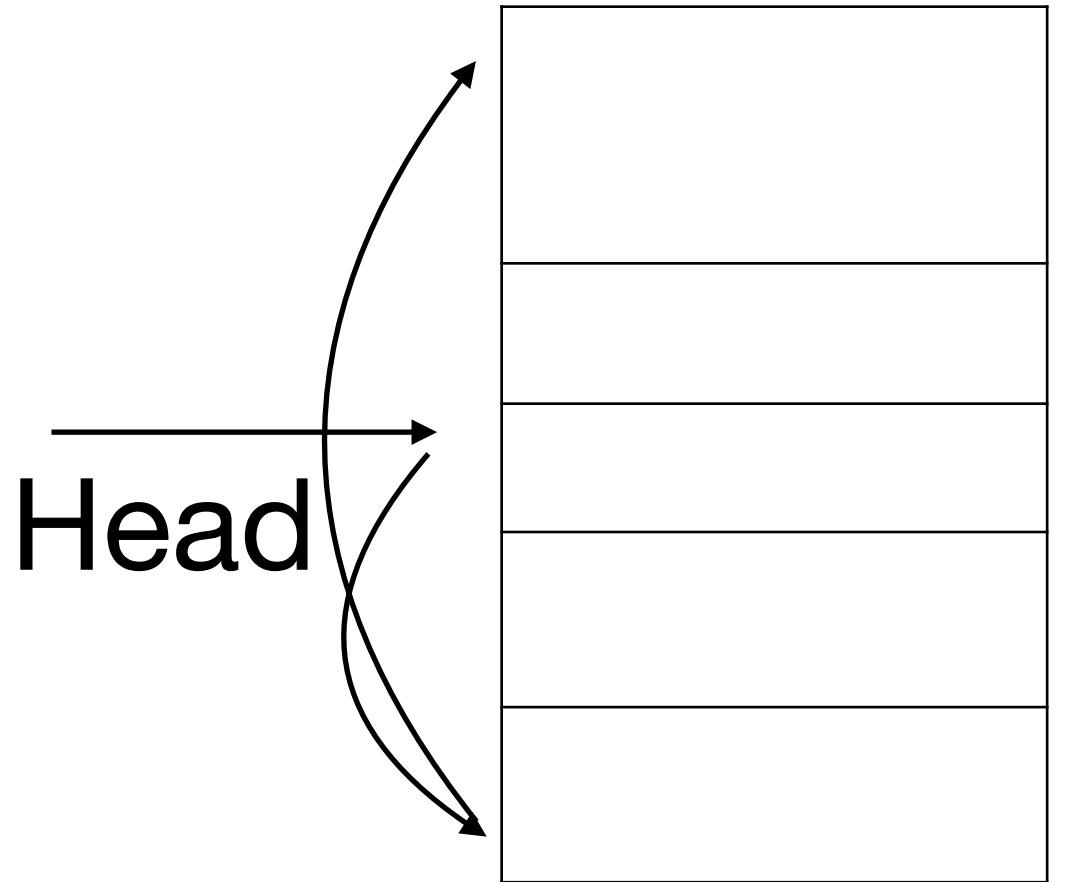
In which order to maintain lists?

- (De)allocation order



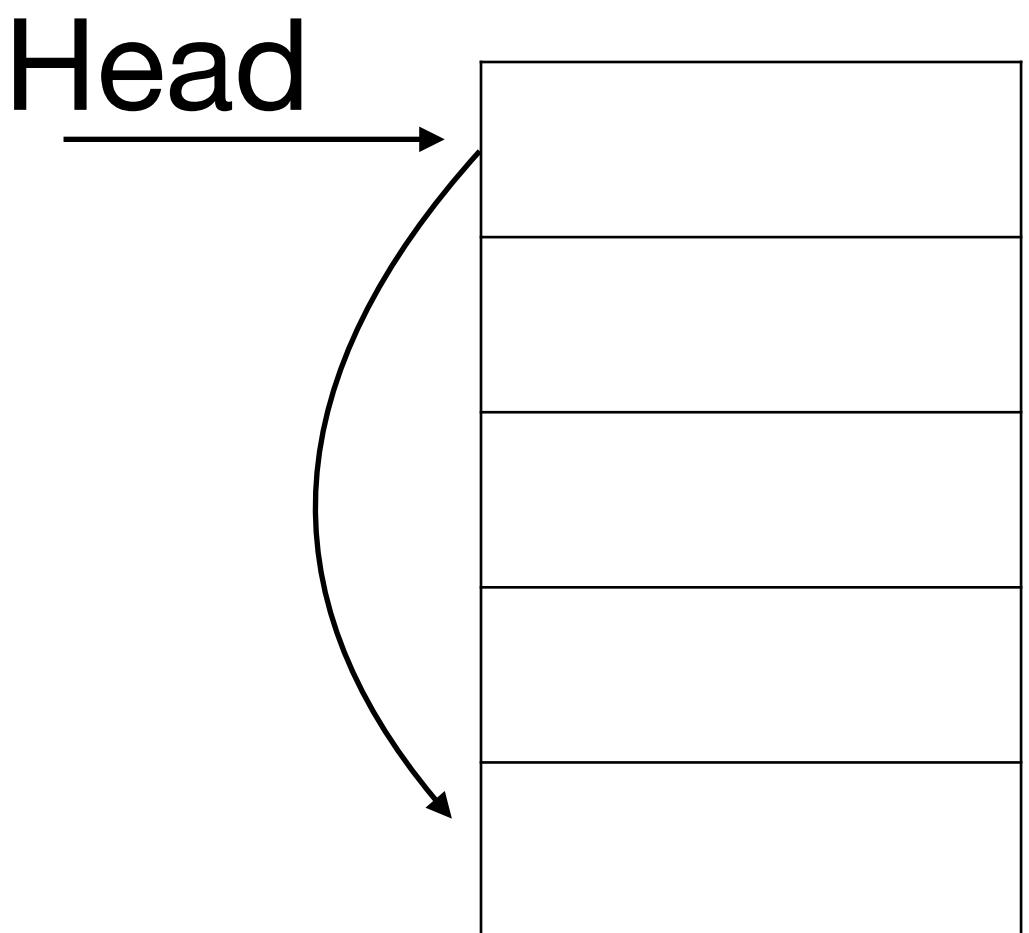
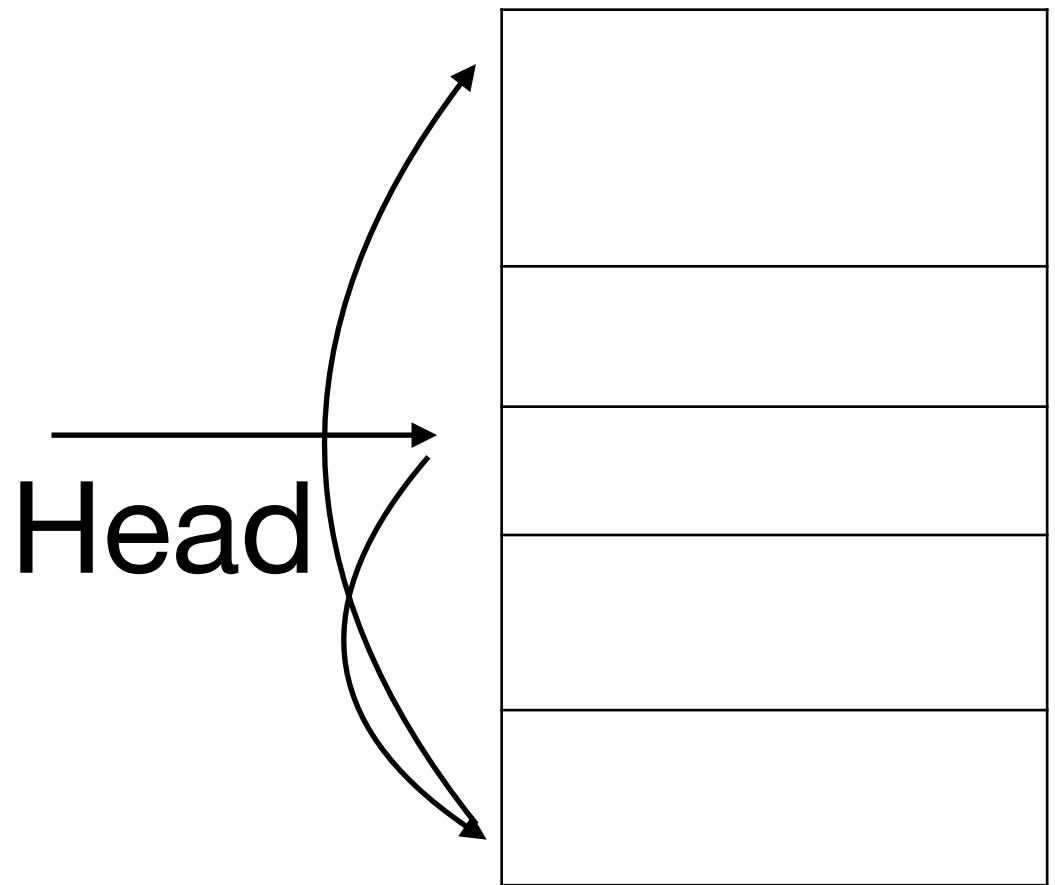
In which order to maintain lists?

- (De)allocation order
- Address order



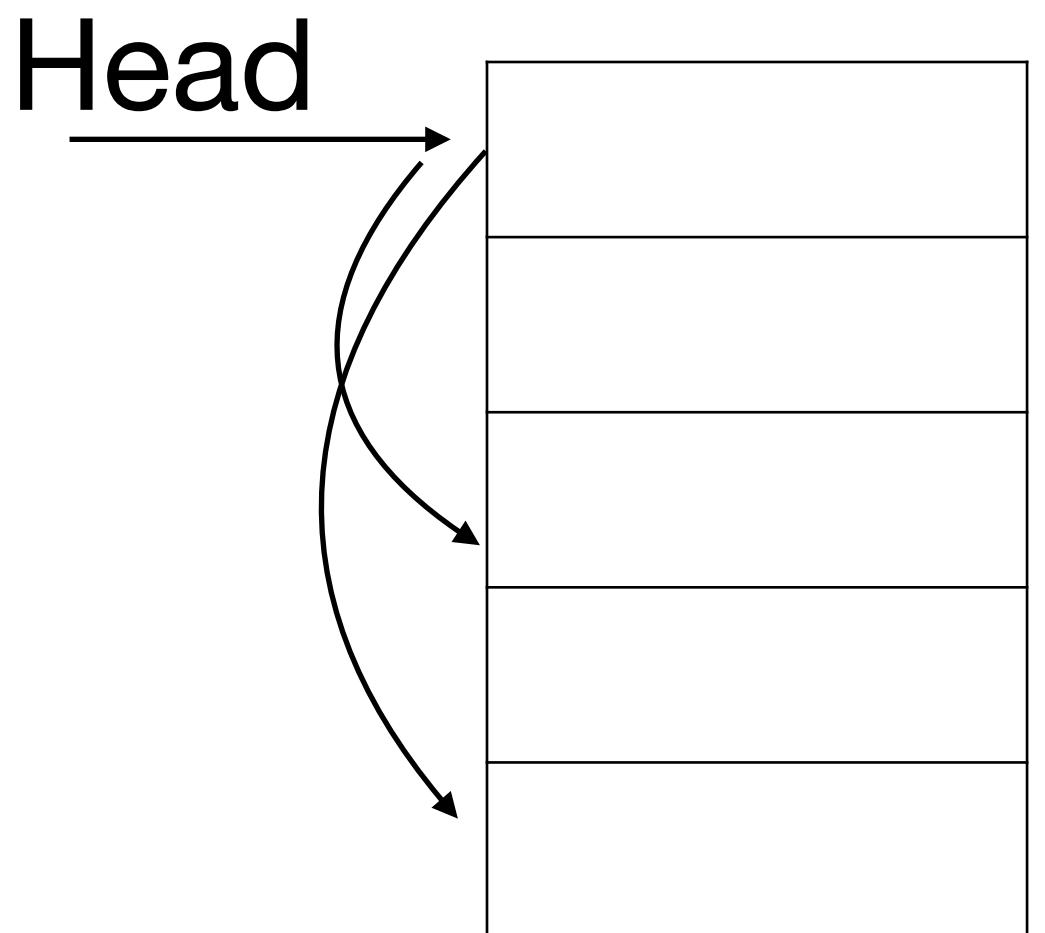
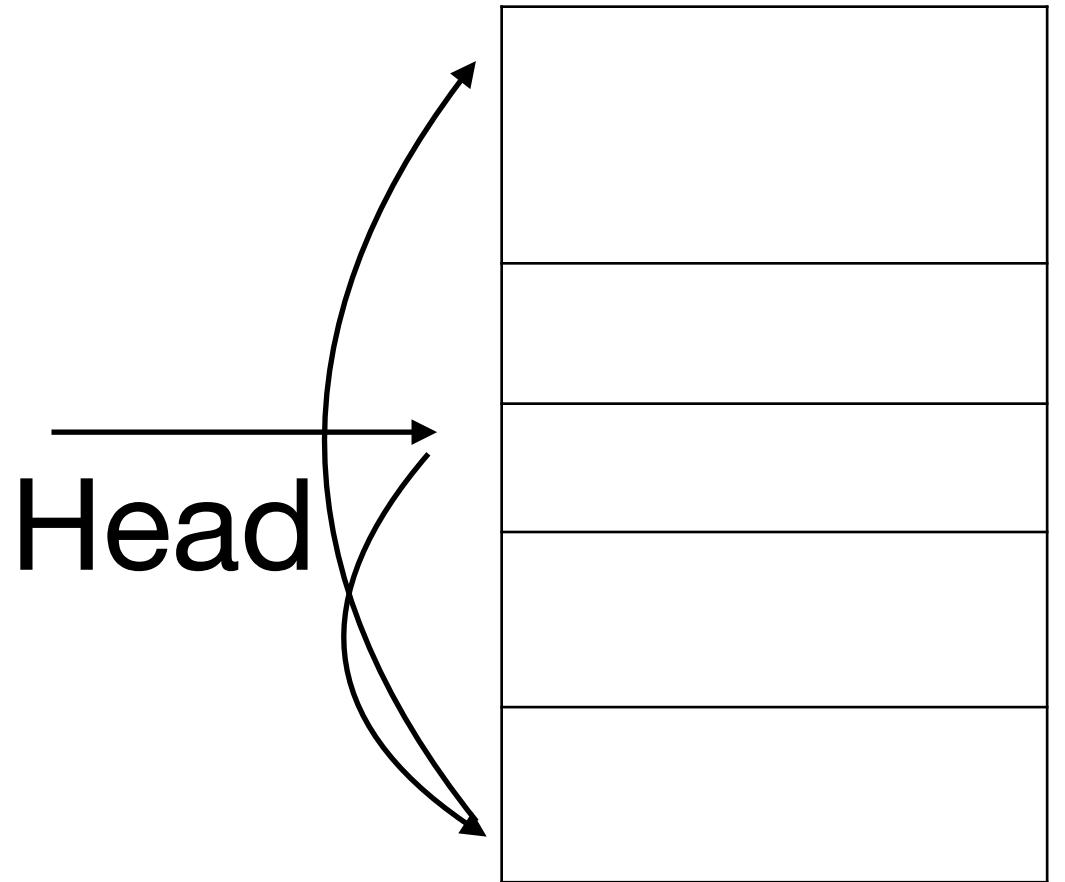
In which order to maintain lists?

- (De)allocation order
- Address order



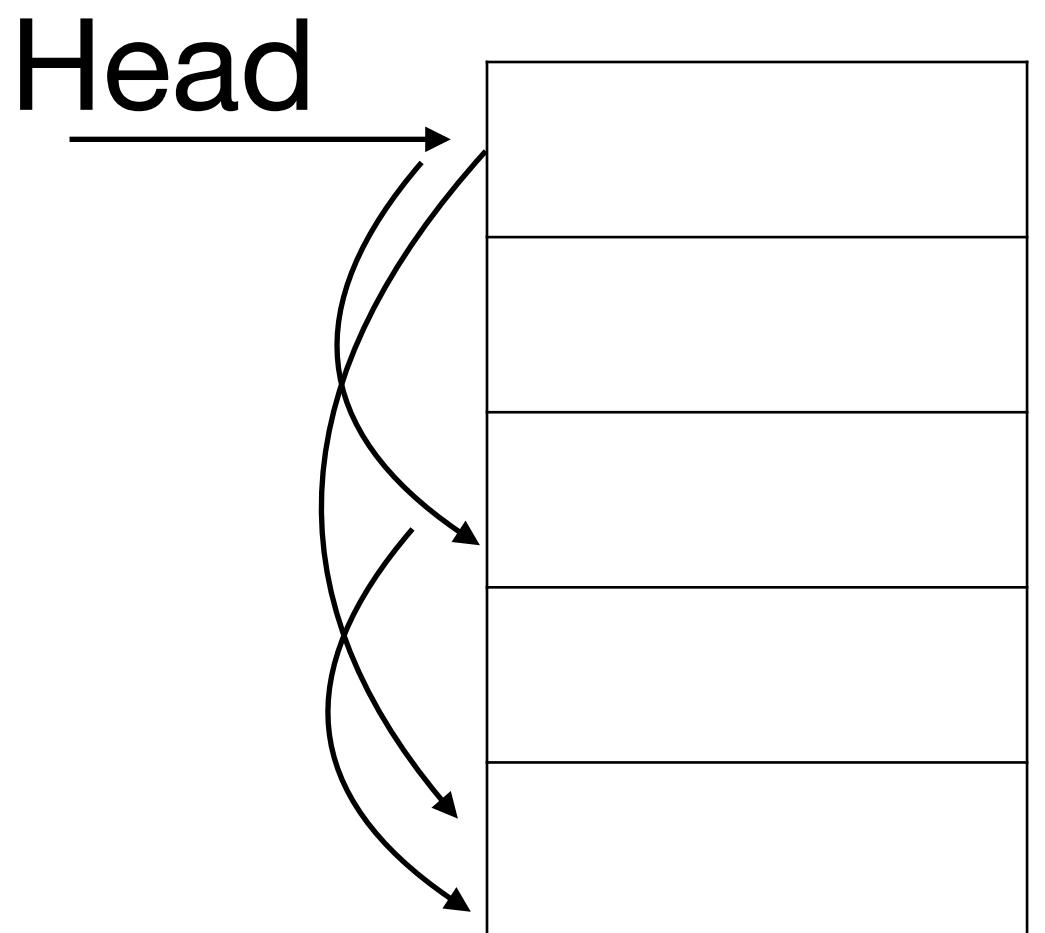
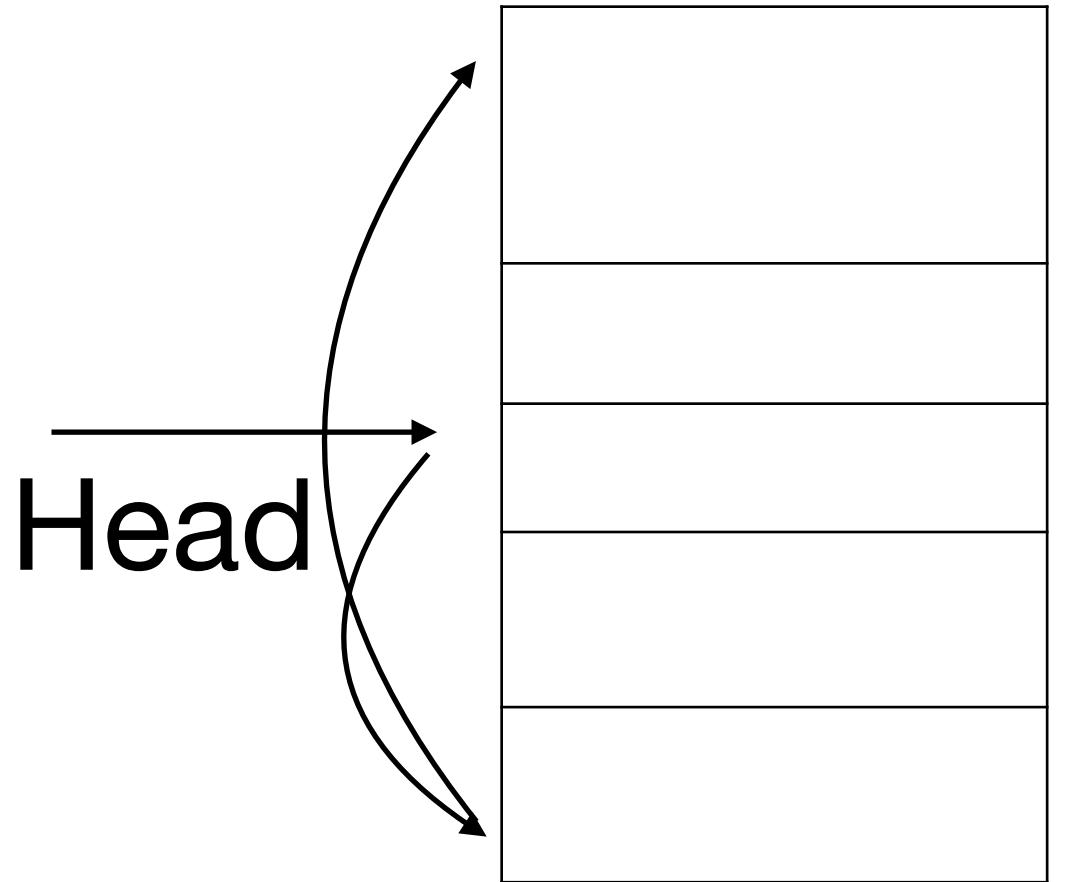
In which order to maintain lists?

- (De)allocation order
- Address order



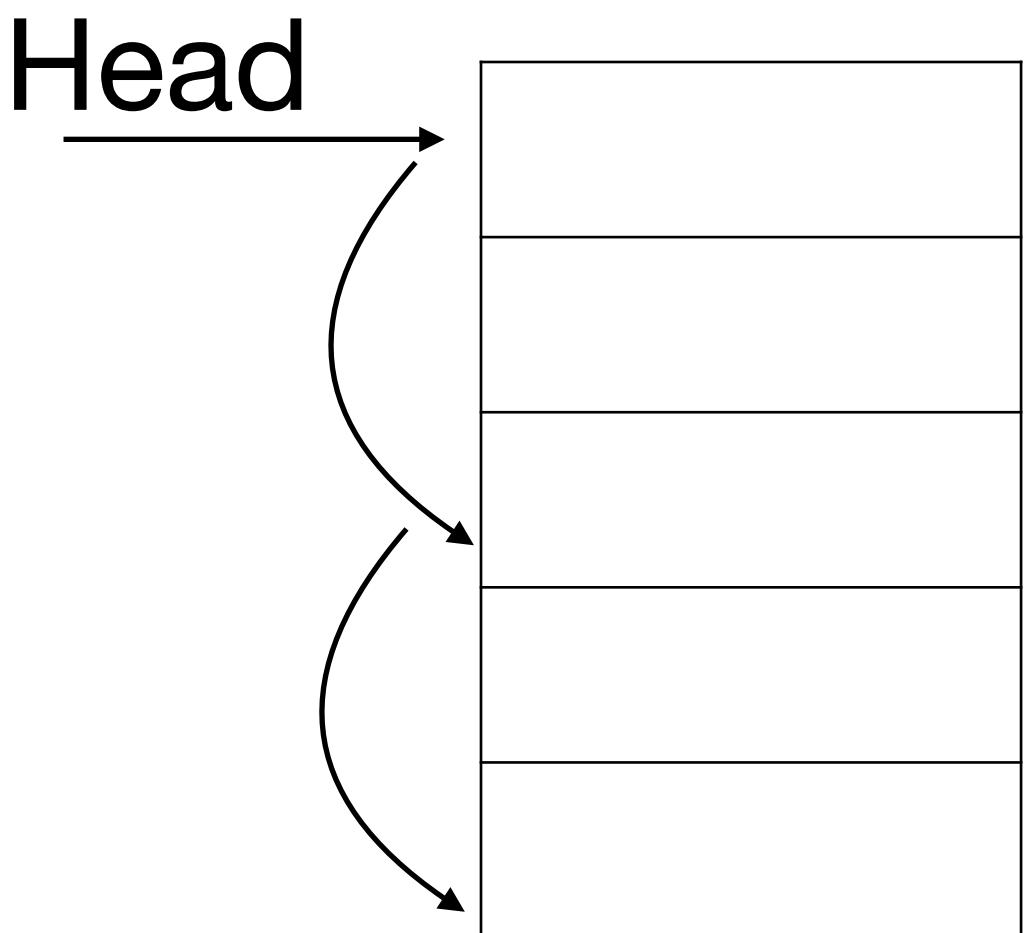
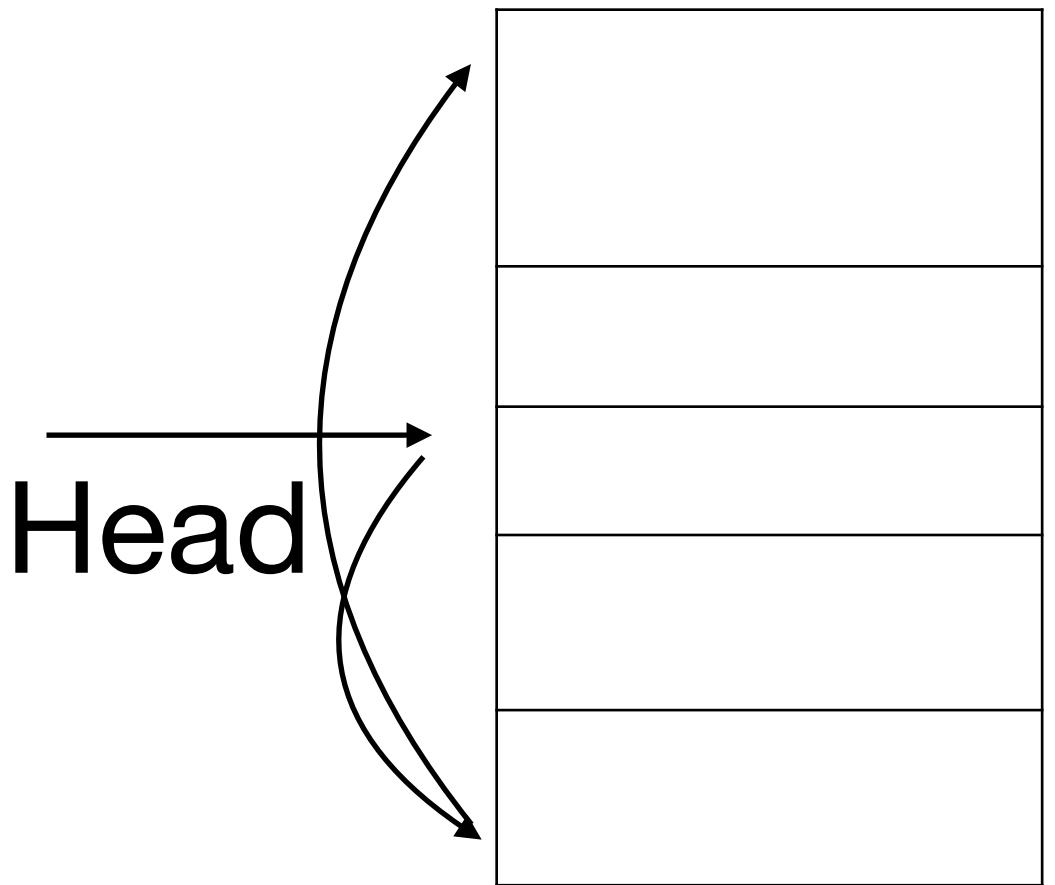
In which order to maintain lists?

- (De)allocation order
- Address order



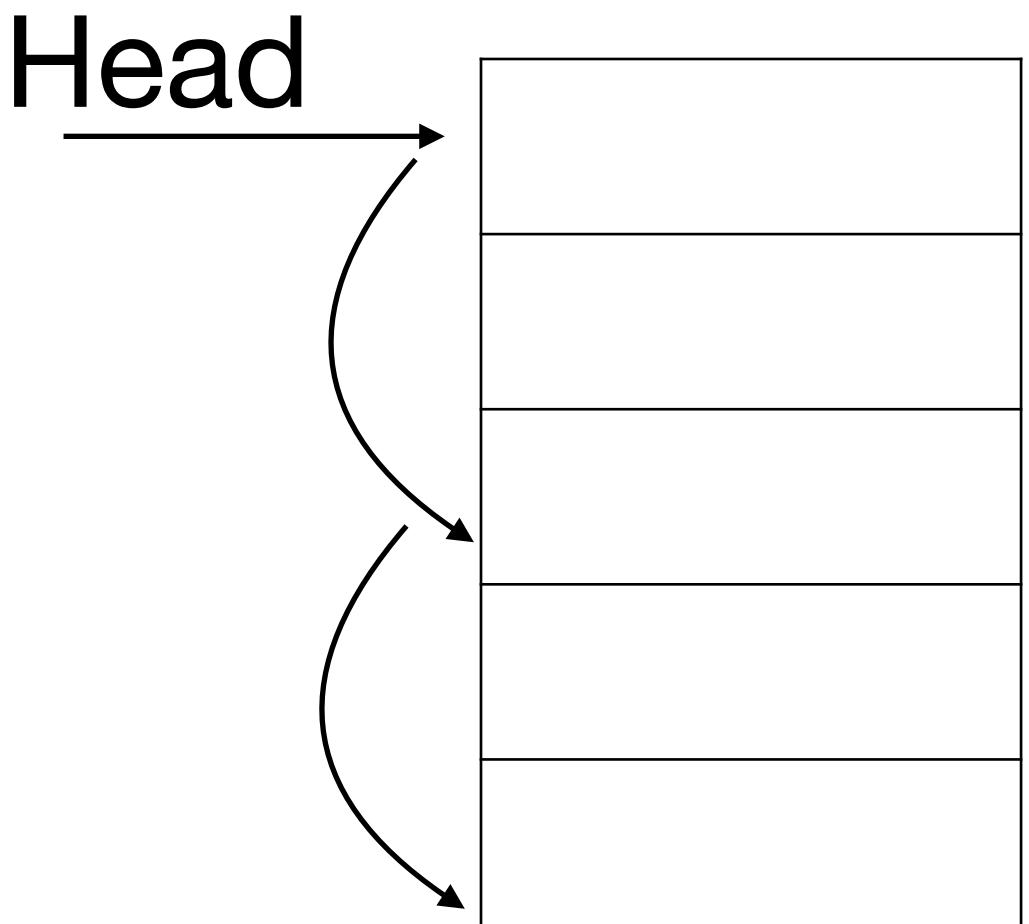
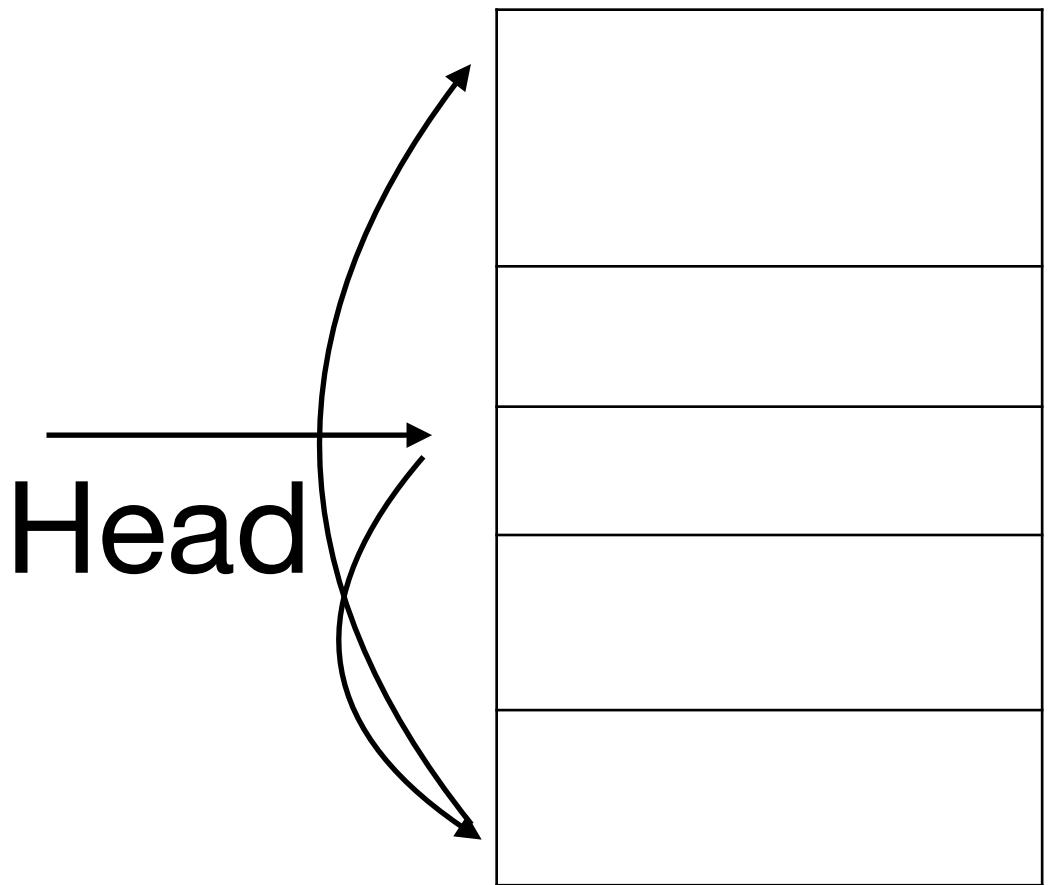
In which order to maintain lists?

- (De)allocation order
- Address order



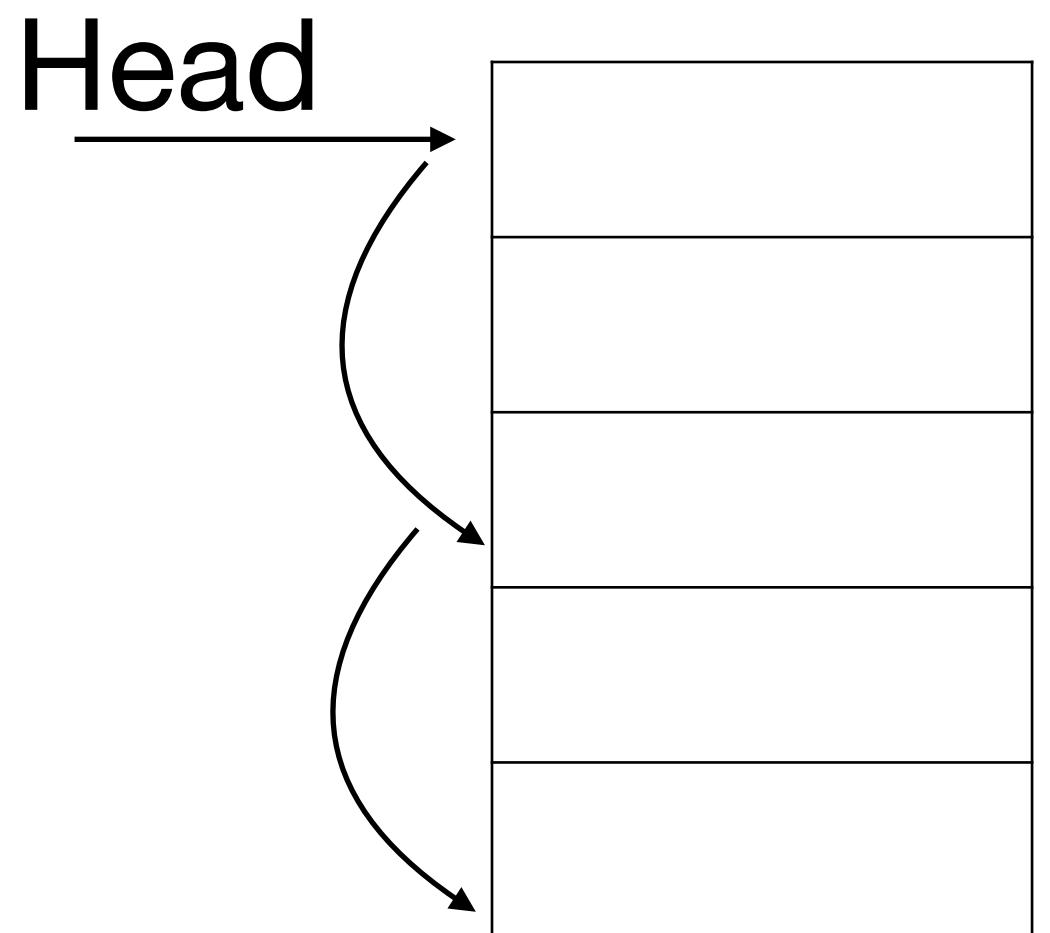
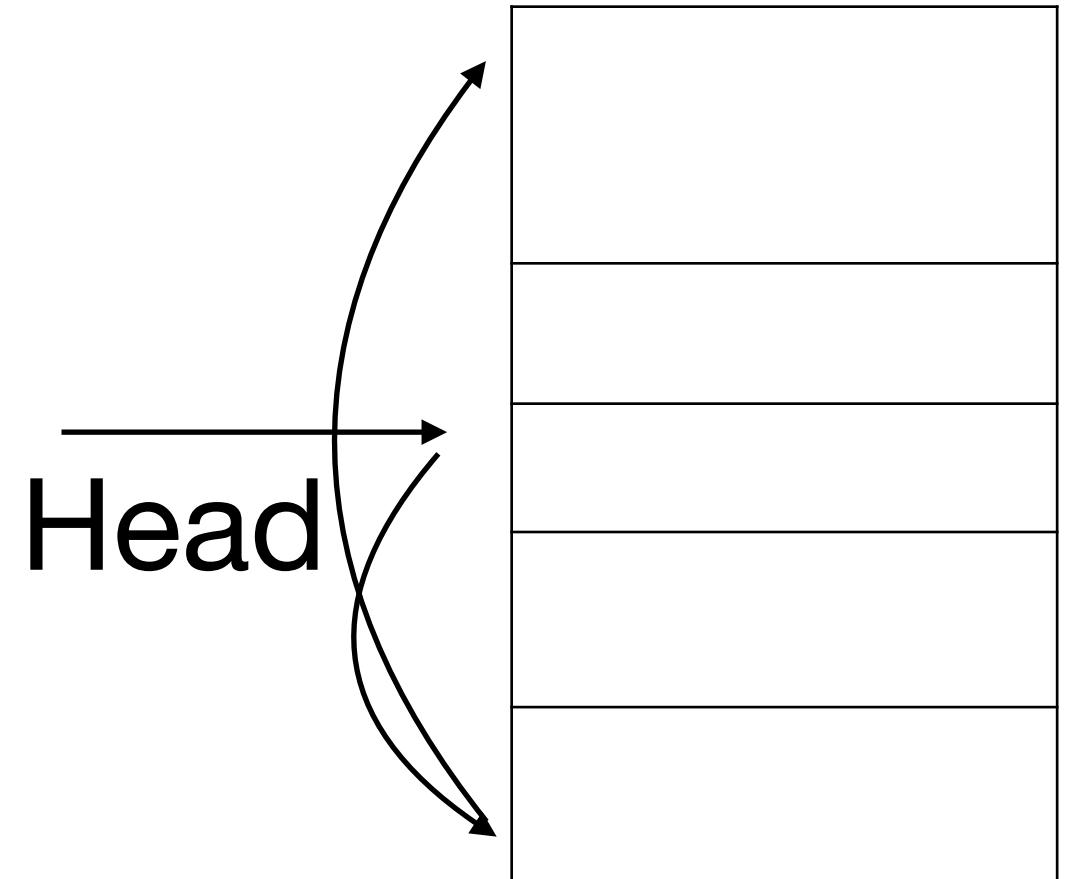
In which order to maintain lists?

- (De)allocation order
- Address order
 - Slow frees: need to traverse the free list



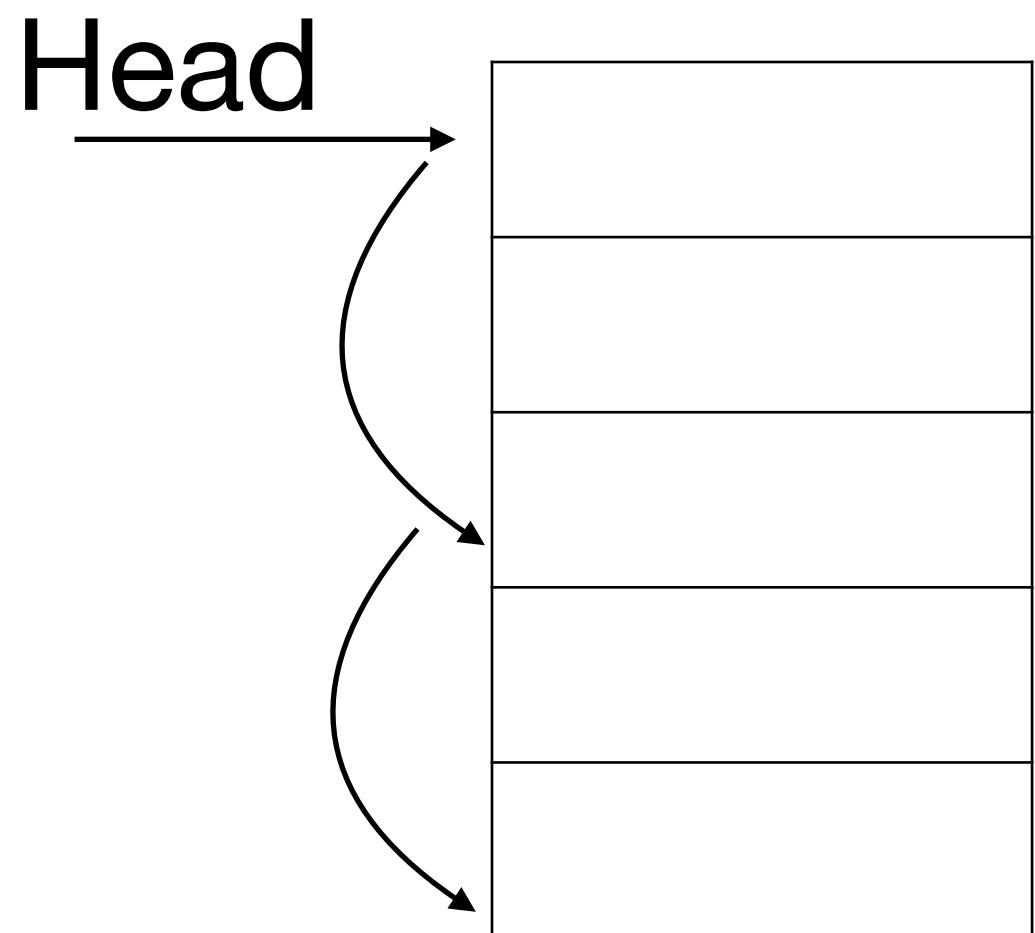
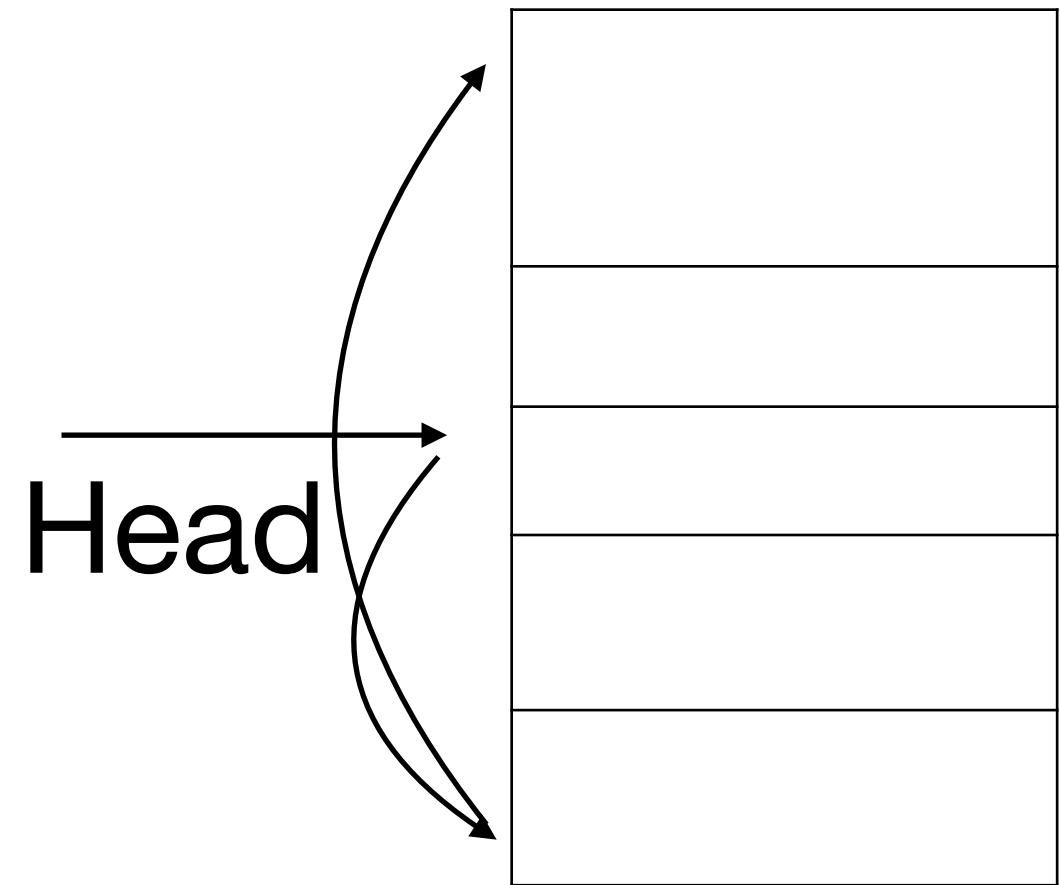
In which order to maintain lists?

- (De)allocation order
- Address order
 - Slow frees: need to traverse the free list
 - (xv6: umalloc.c)



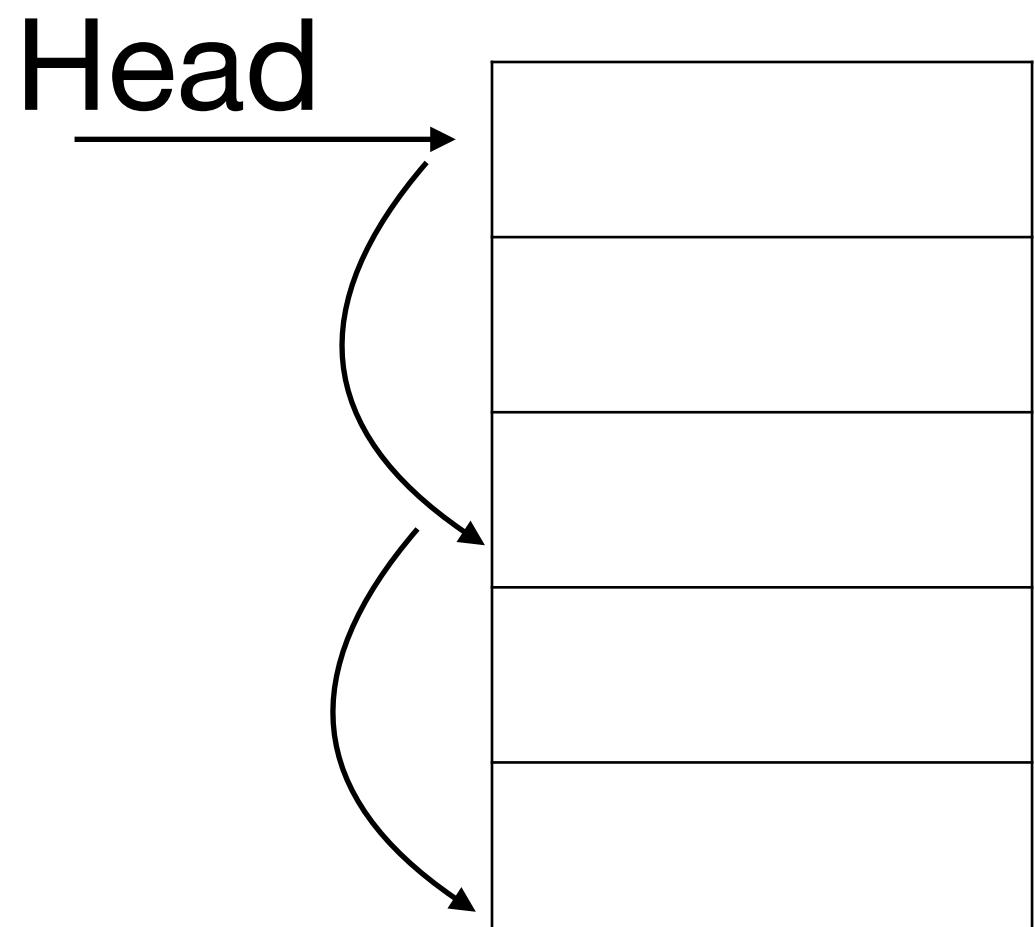
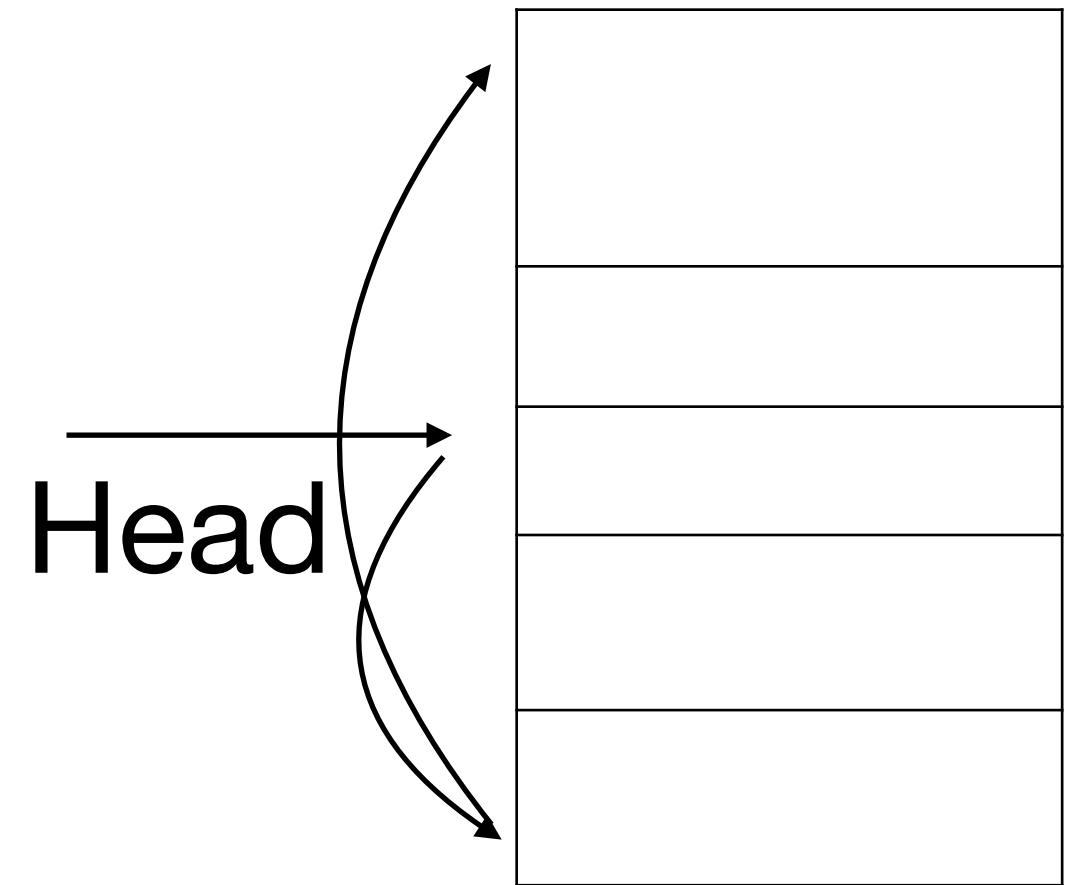
In which order to maintain lists?

- (De)allocation order
- Address order
 - Slow frees: need to traverse the free list
 - (xv6: umalloc.c)
 - Address order, first fit will allocate back-to-back allocations contiguously.



In which order to maintain lists?

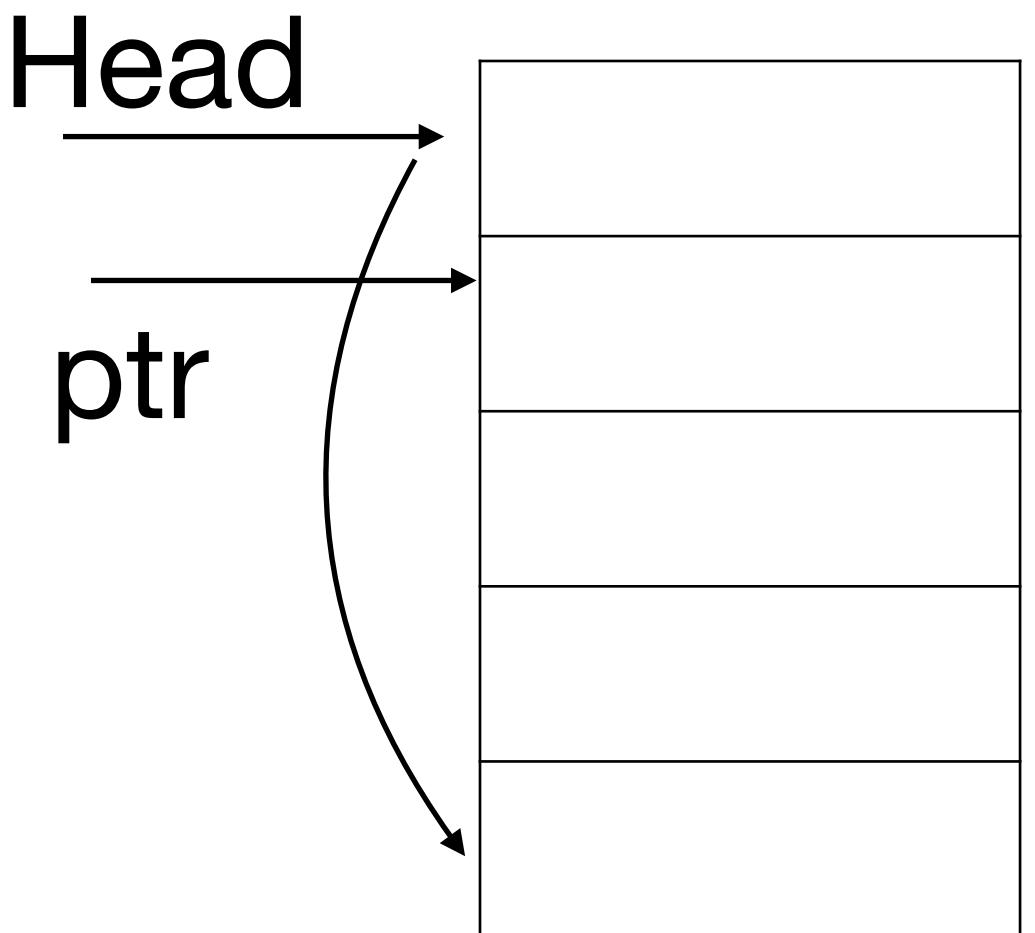
- (De)allocation order
- Address order
 - Slow frees: need to traverse the free list
 - (xv6: umalloc.c)
 - Address order, first fit will allocate back-to-back allocations contiguously.
 - Due to “clustered deaths”, we may get better chances of coalescing



How to do coalescing?

Example: free(ptr)

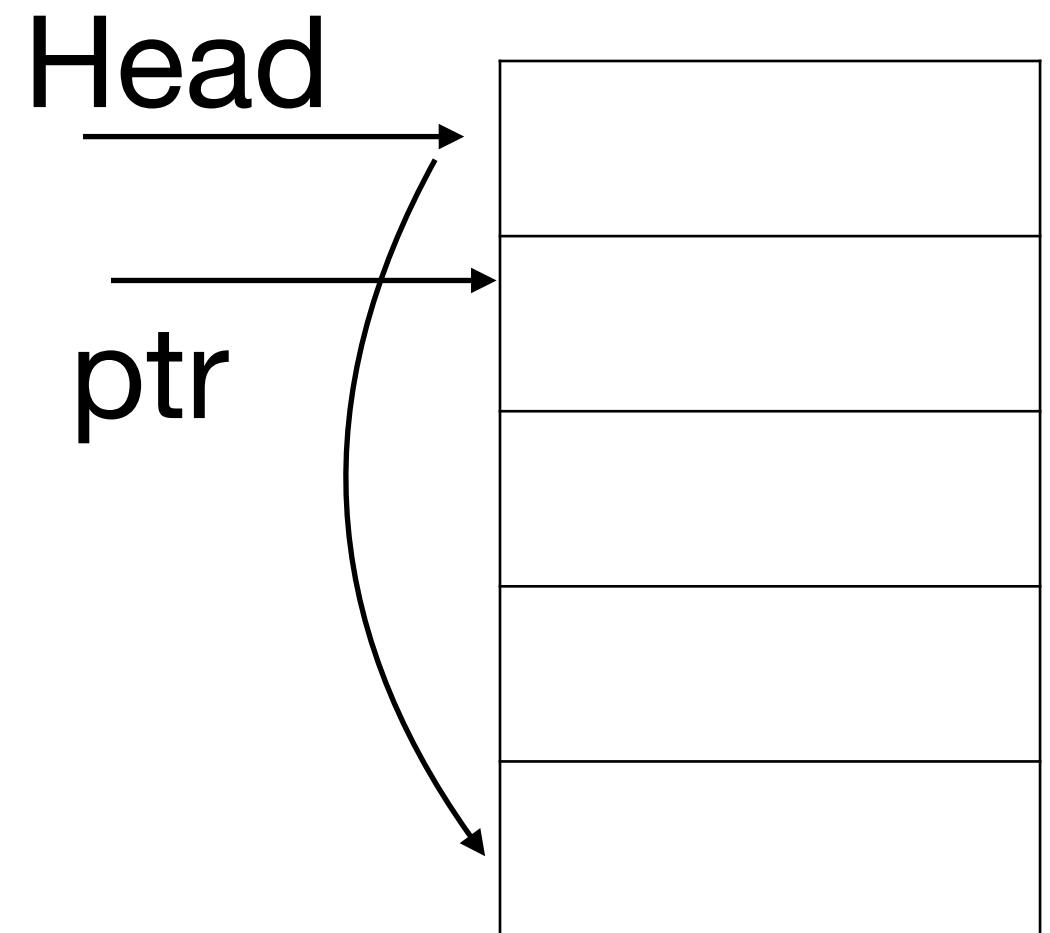
- Straightforward in address order since we are traversing the free list in address order



How to do coalescing?

Example: free(ptr)

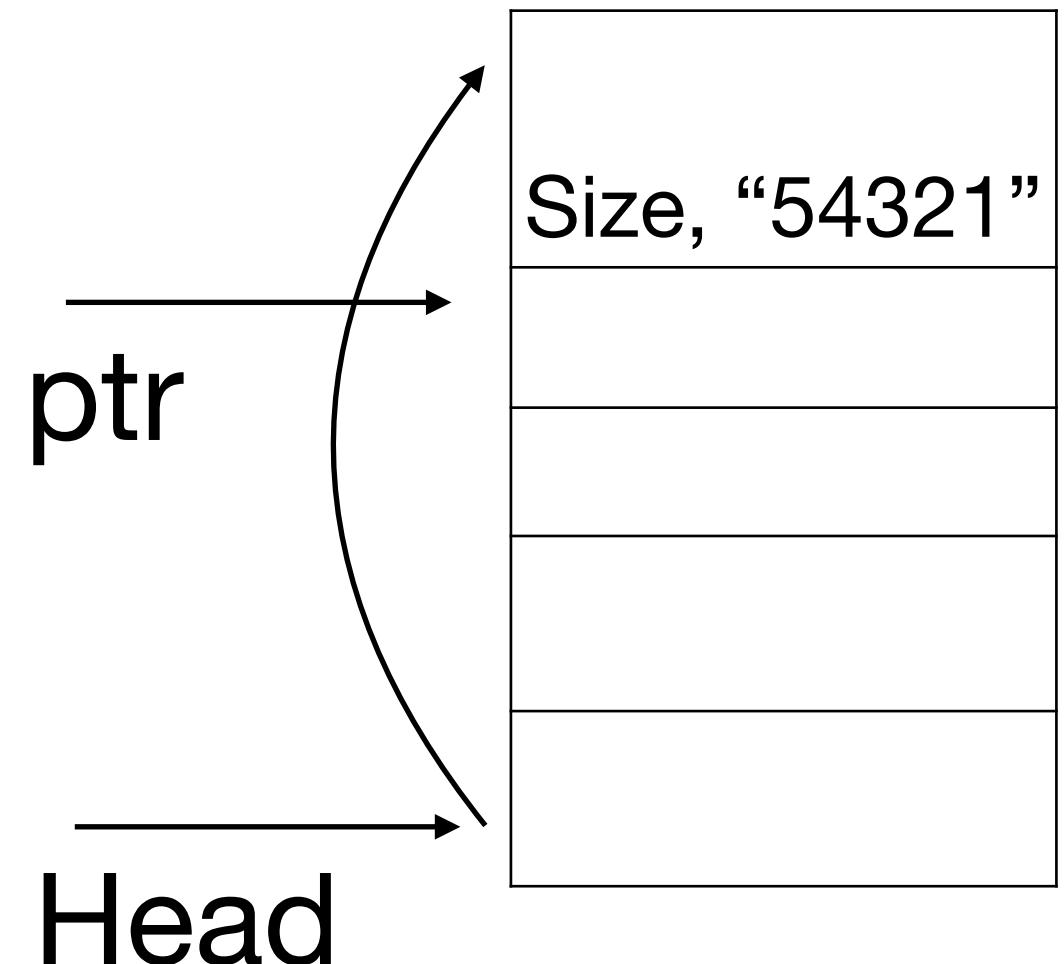
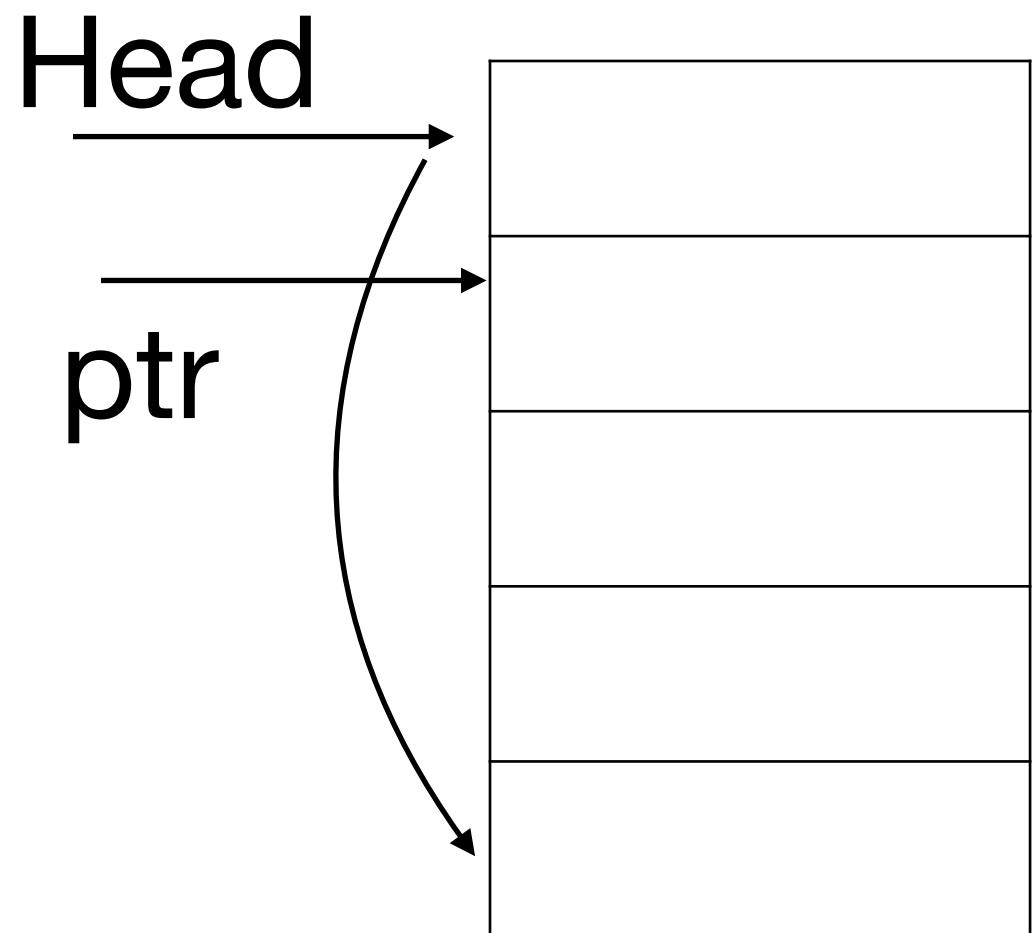
- Straightforward in address order since we are traversing the free list in address order
- In deallocation order: when an area is freed, check if the “boundary tag” is present in the footer above



How to do coalescing?

Example: free(ptr)

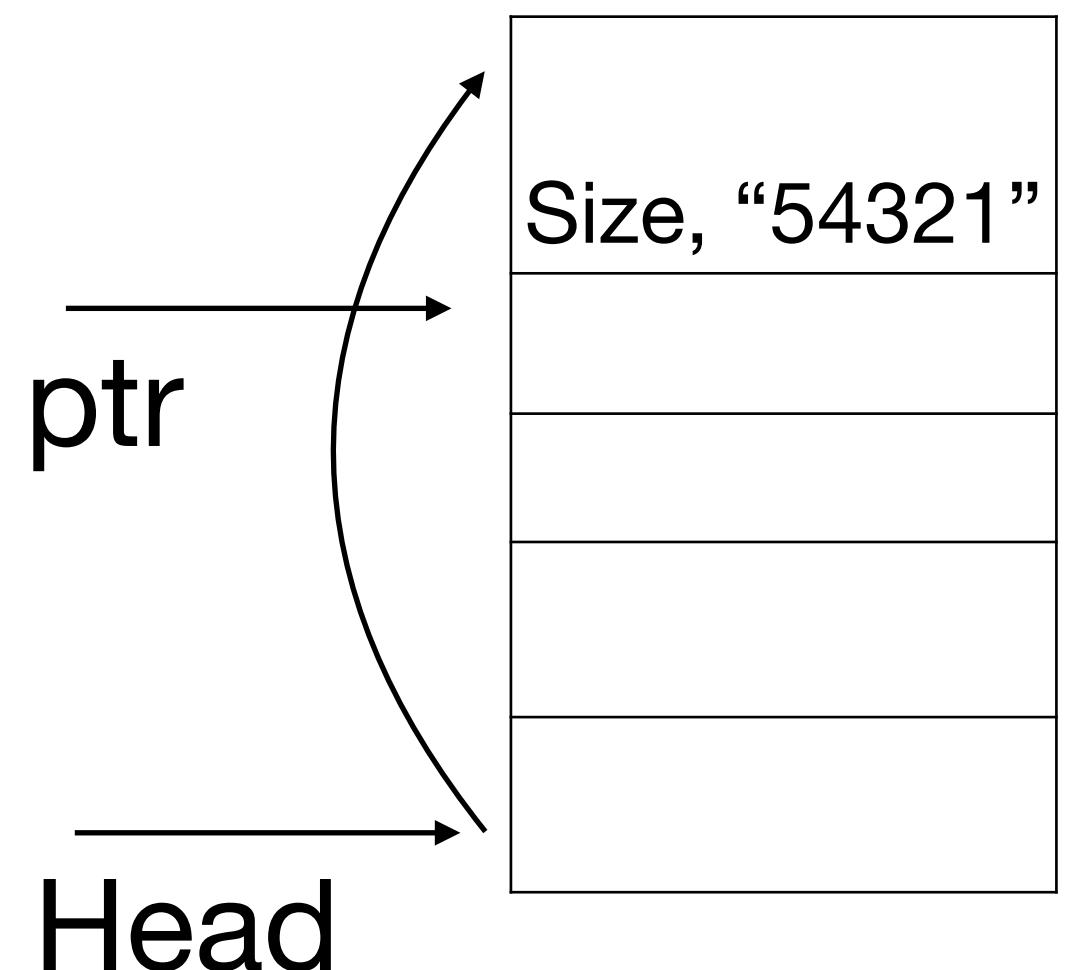
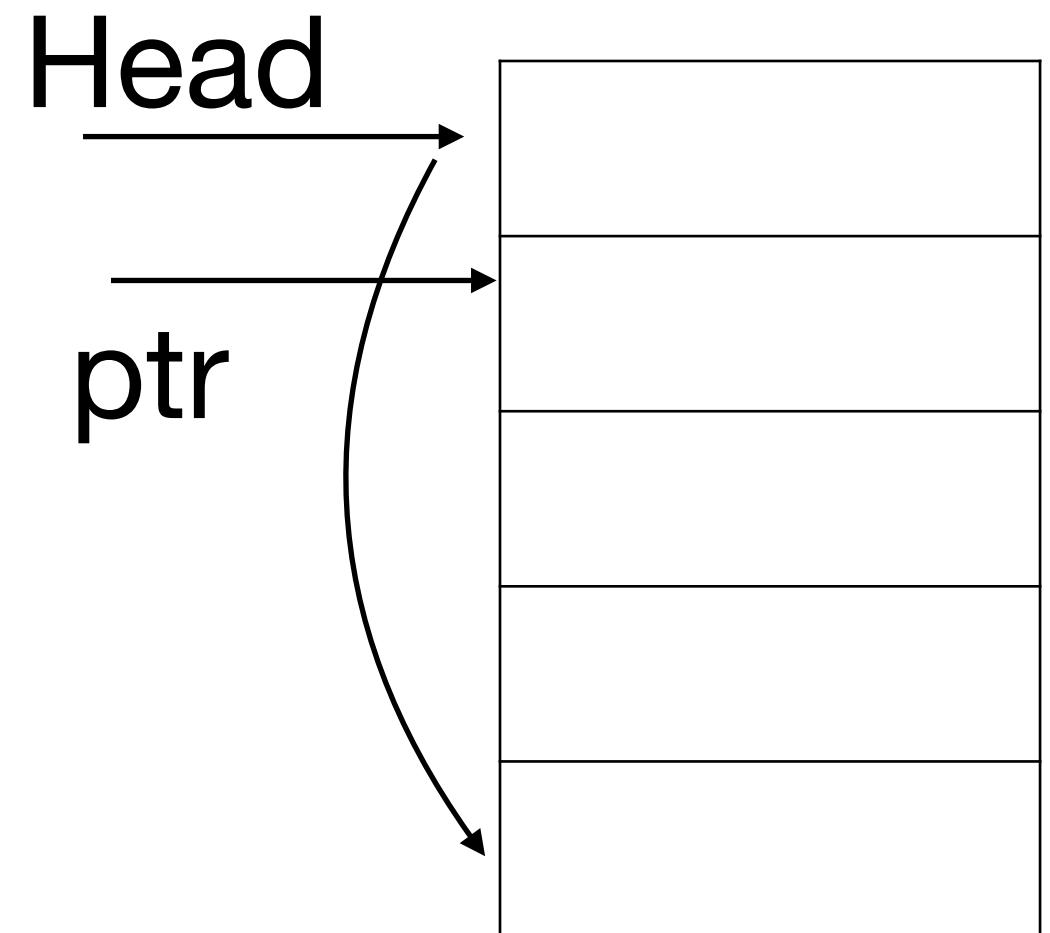
- Straightforward in address order since we are traversing the free list in address order
- In deallocation order: when an area is freed, check if the “boundary tag” is present in the footer above



How to do coalescing?

Example: free(ptr)

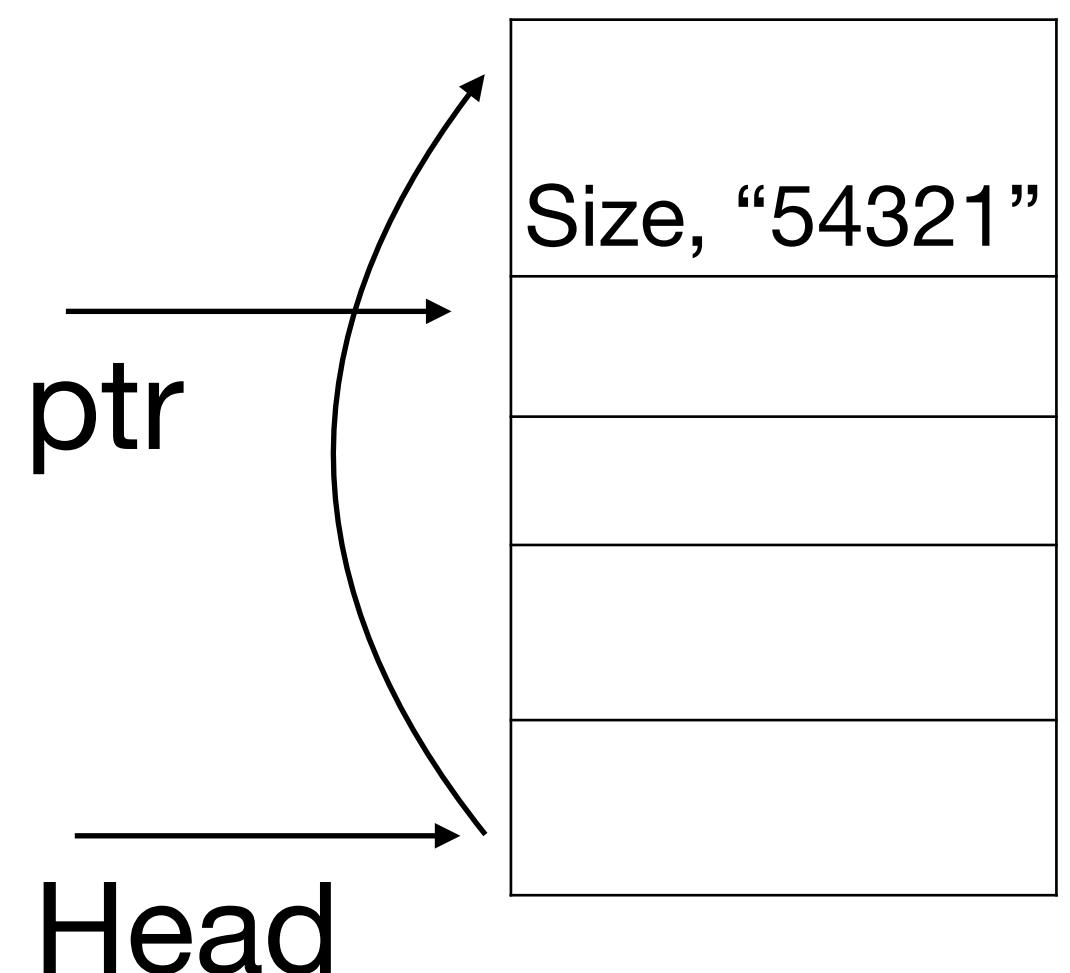
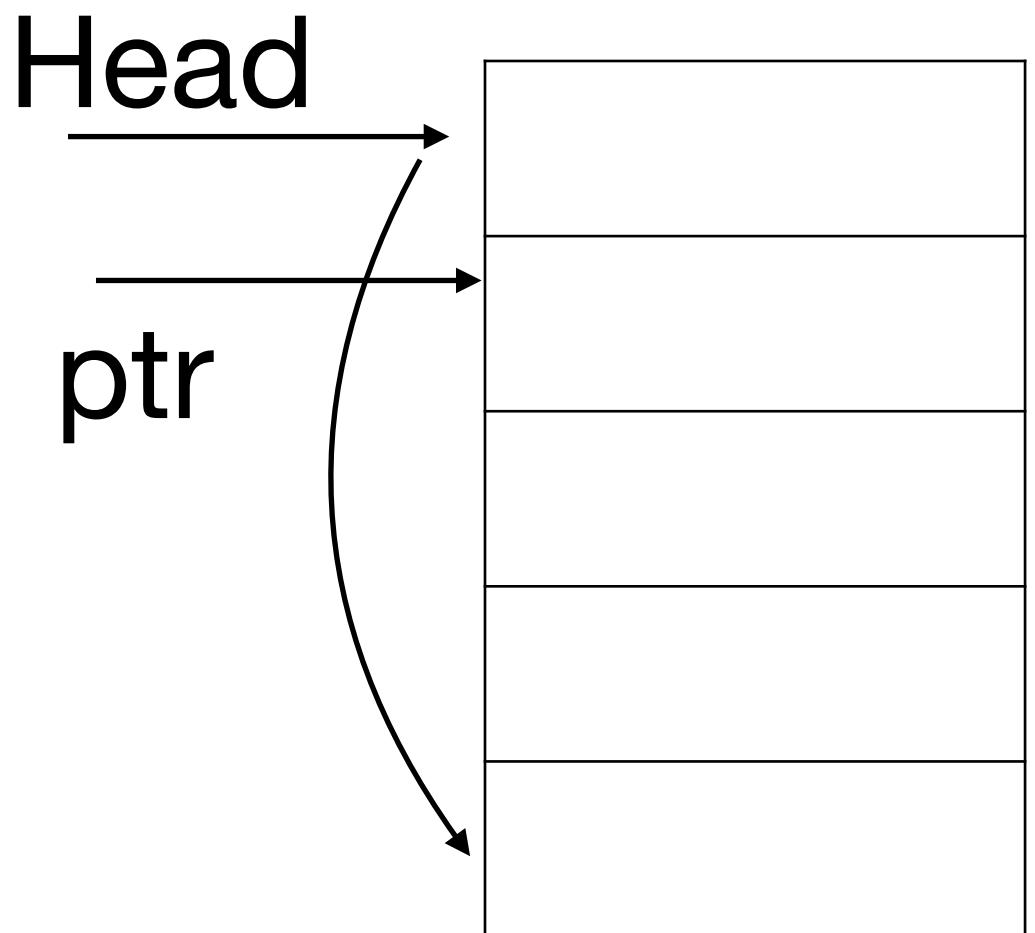
- Straightforward in address order since we are traversing the free list in address order
- In deallocation order: when an area is freed, check if the “boundary tag” is present in the footer above
- First fit and address order



How to do coalescing?

Example: free(ptr)

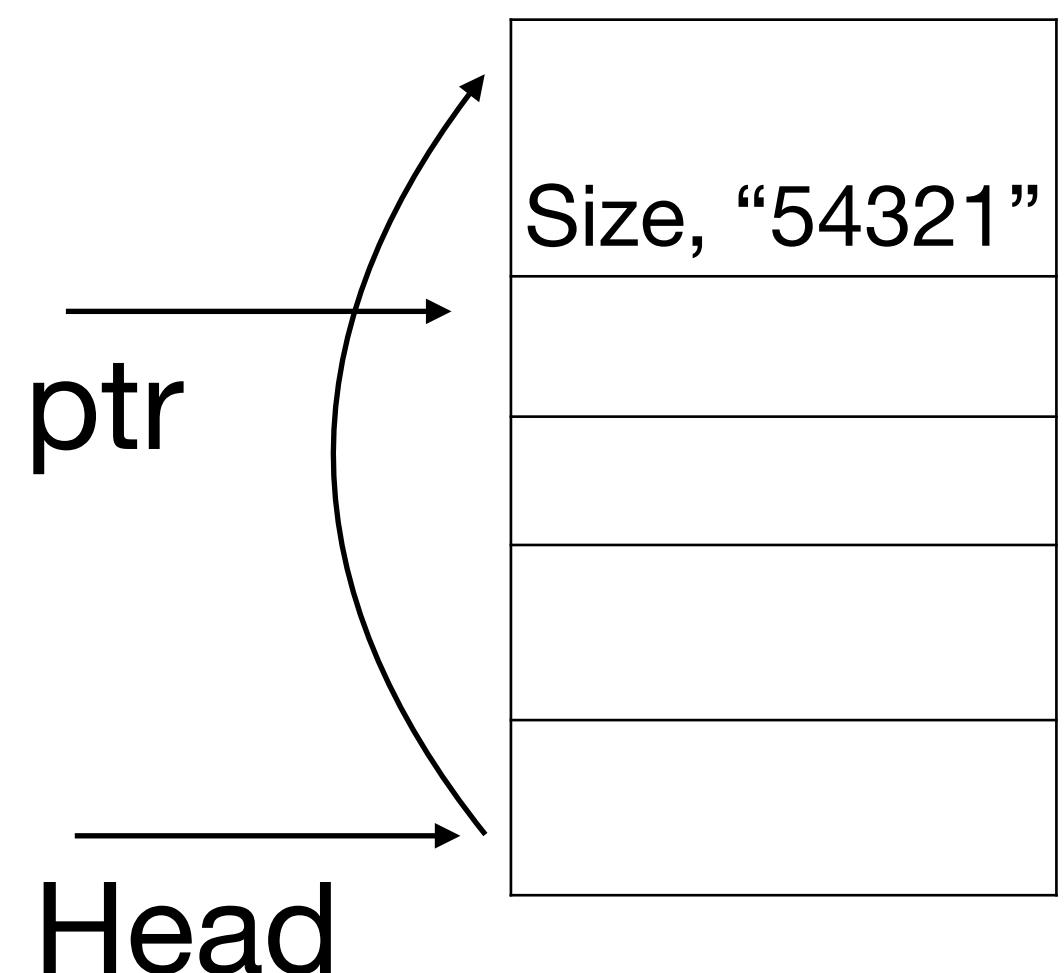
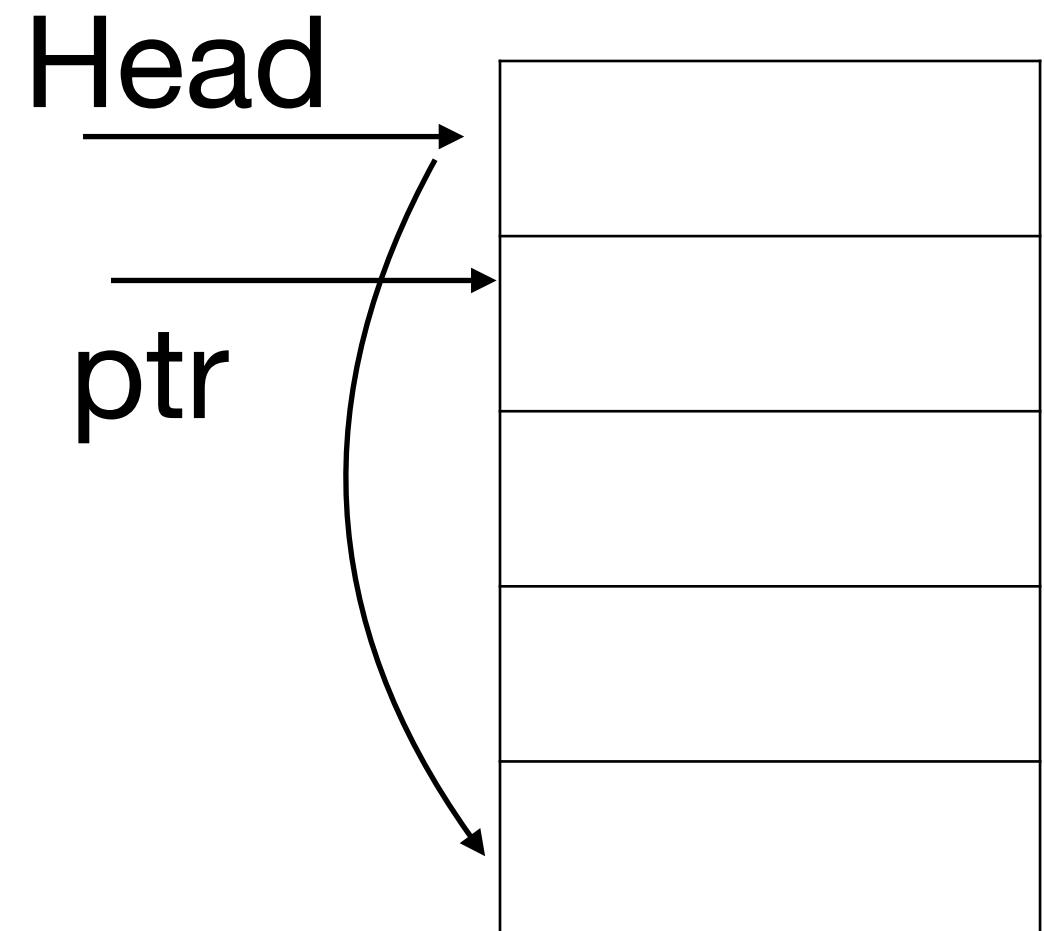
- Straightforward in address order since we are traversing the free list in address order
- In deallocation order: when an area is freed, check if the “boundary tag” is present in the footer above
- First fit and address order
 - Better chances of coalescing for clustered deaths, simpler coalescing (no boundary tag)



How to do coalescing?

Example: free(ptr)

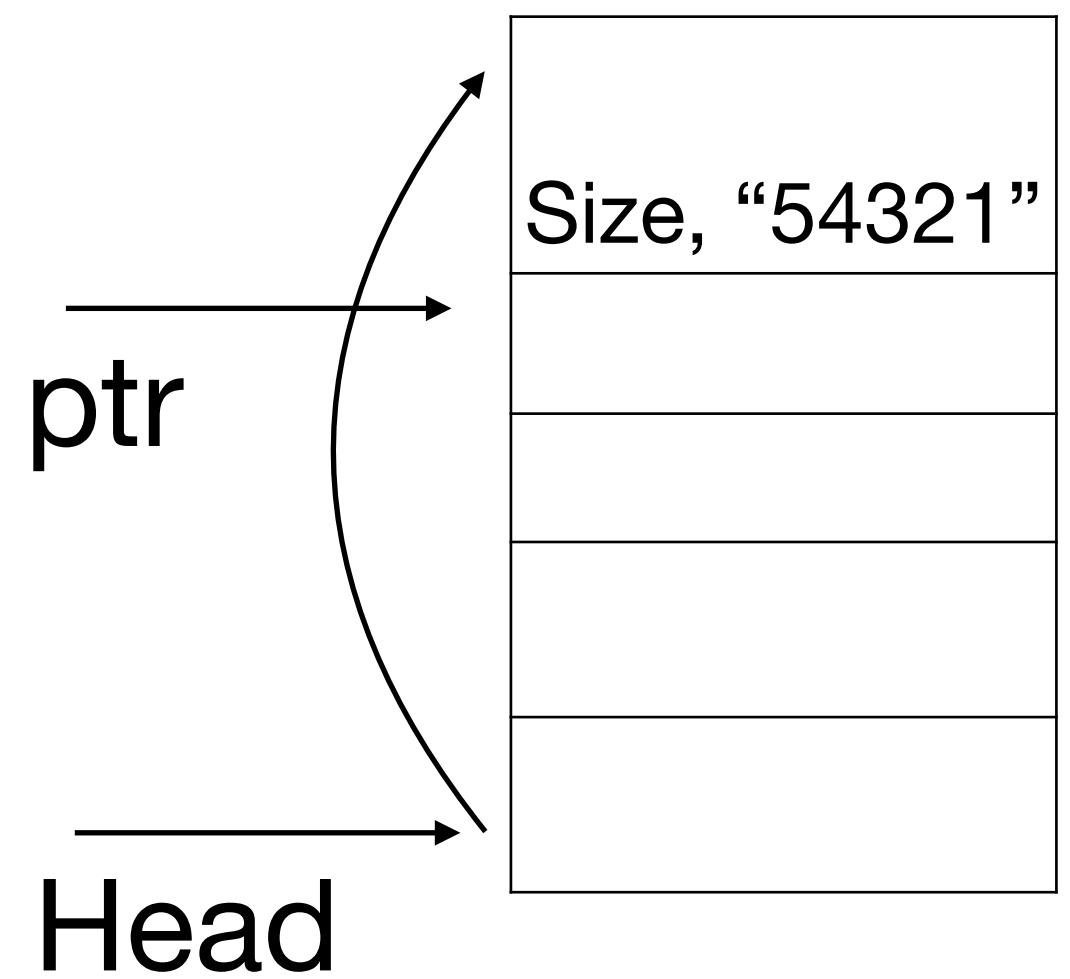
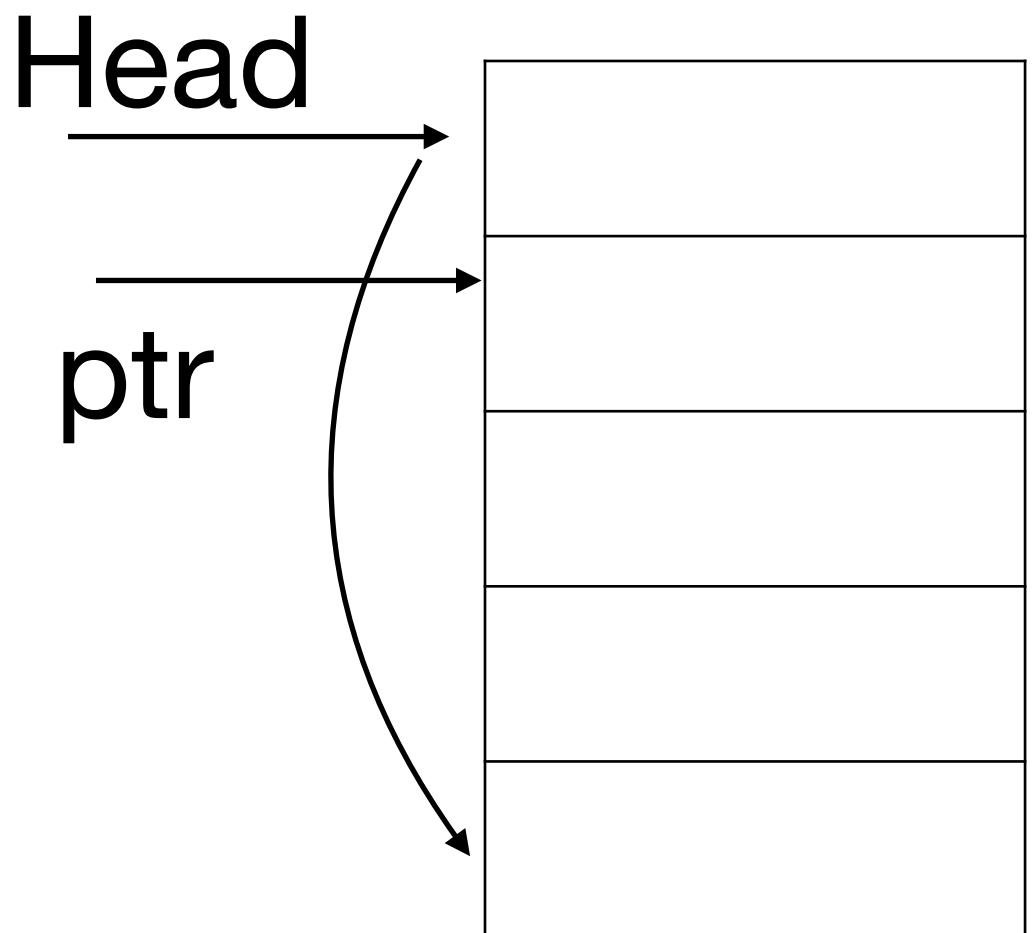
- Straightforward in address order since we are traversing the free list in address order
- In deallocation order: when an area is freed, check if the “boundary tag” is present in the footer above
- First fit and address order
 - Better chances of coalescing for clustered deaths, simpler coalescing (no boundary tag)
 - First fit causes fragmentation



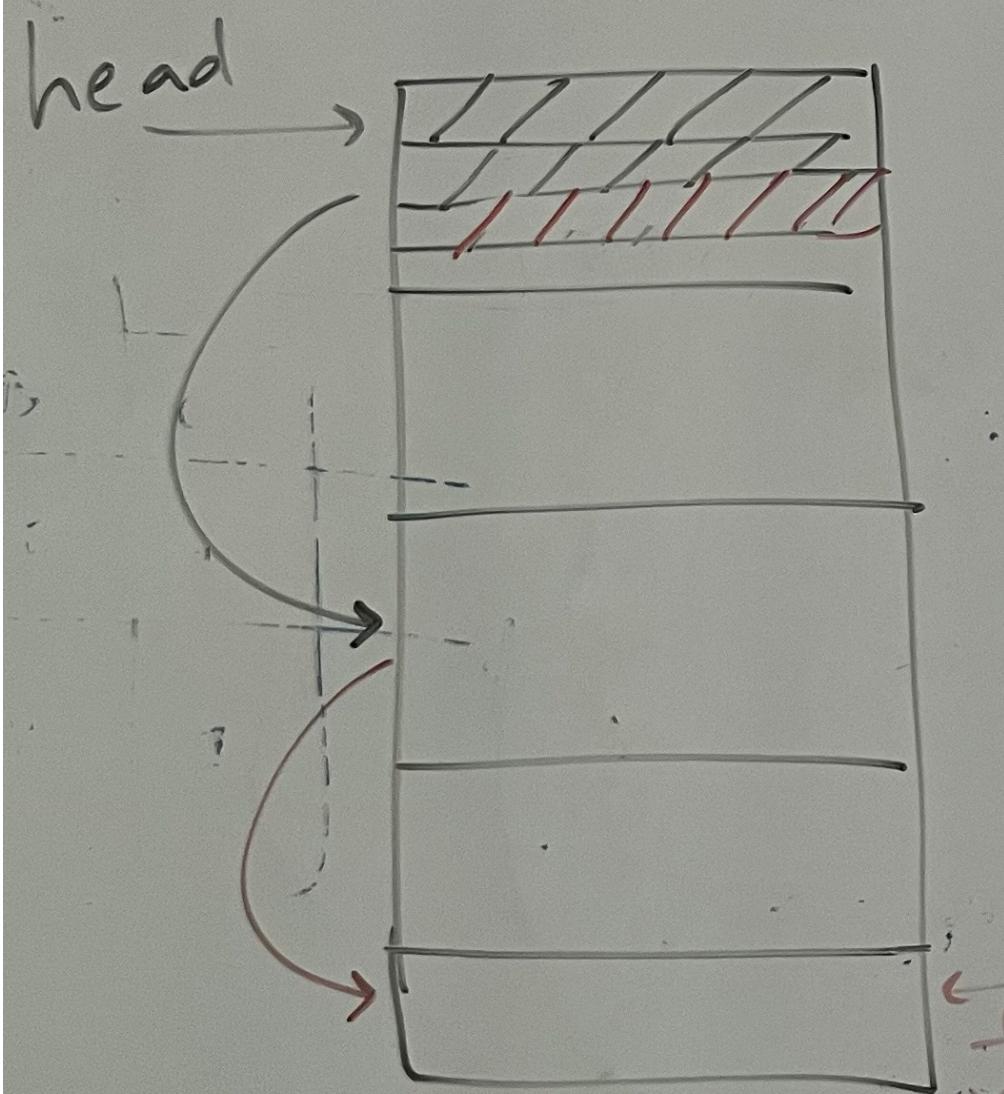
How to do coalescing?

Example: free(ptr)

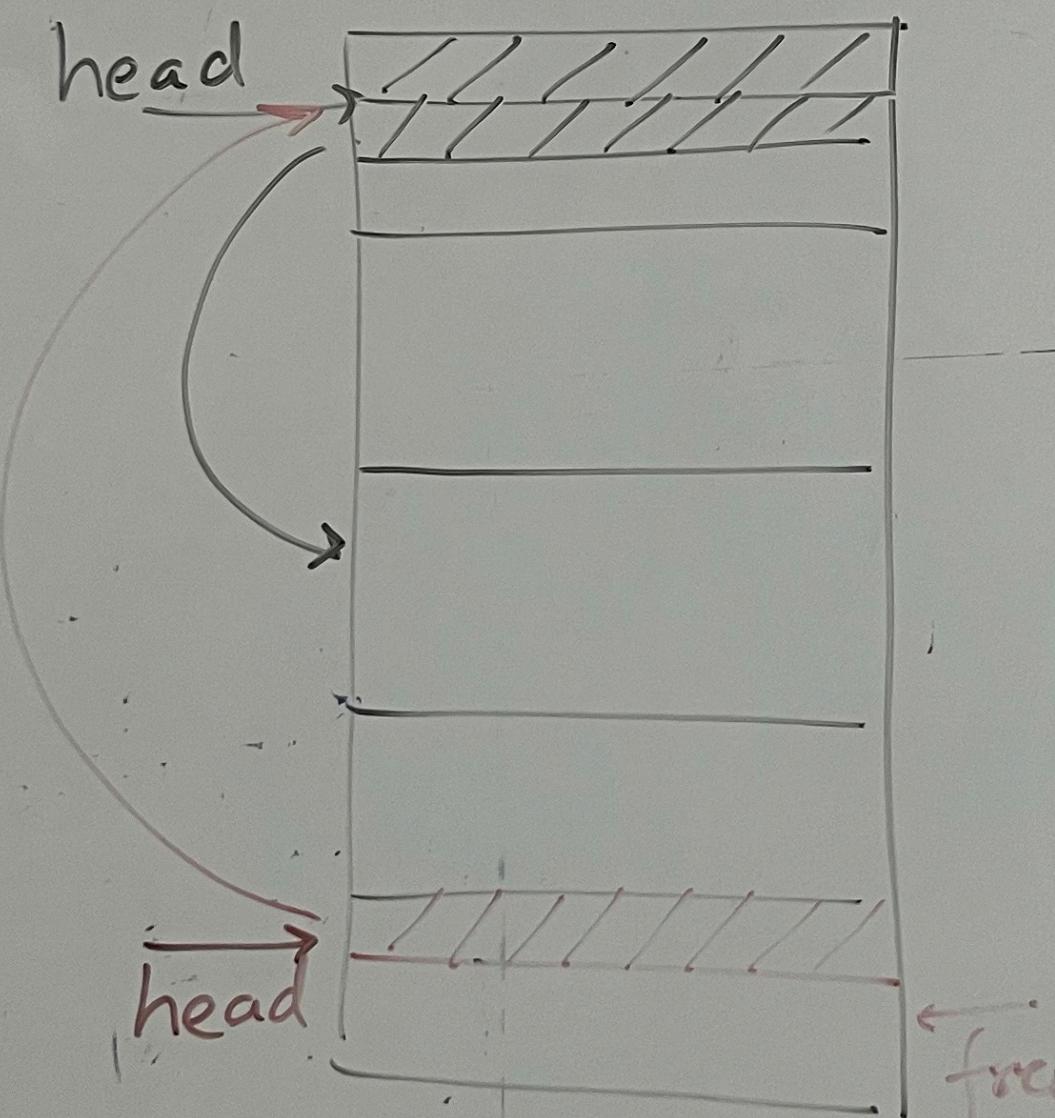
- Straightforward in address order since we are traversing the free list in address order
- In deallocation order: when an area is freed, check if the “boundary tag” is present in the footer above
- First fit and address order
 - Better chances of coalescing for clustered deaths, simpler coalescing (no boundary tag)
 - First fit causes fragmentation
 - Address order slows frees due to traversing free list



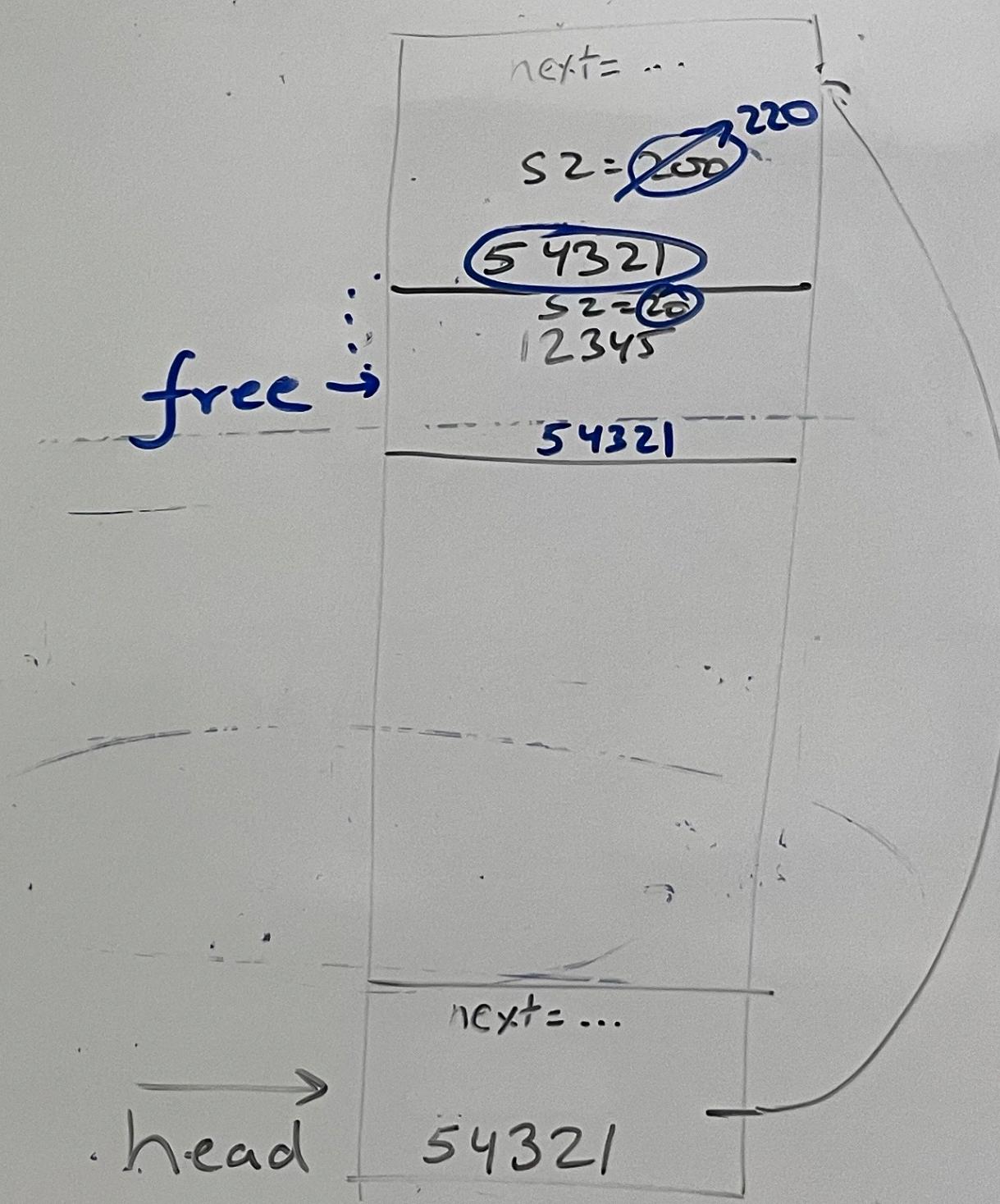
First fit
address order



First fit
(de) allocation order

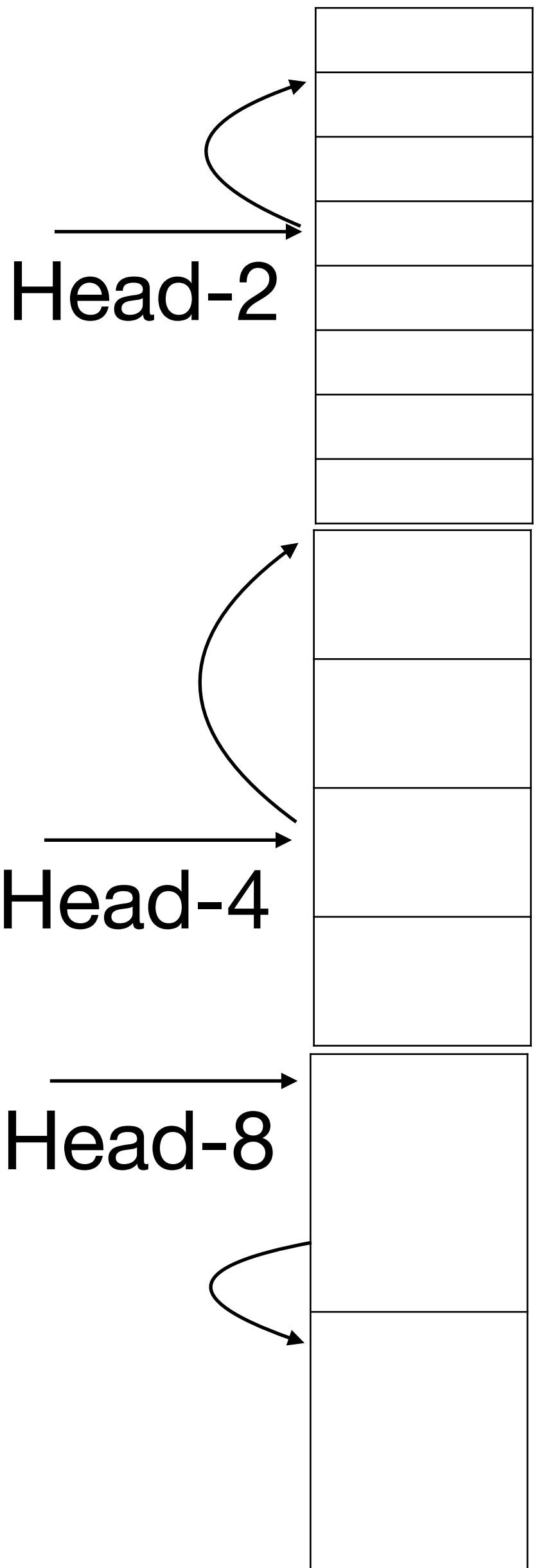


Boundary tag



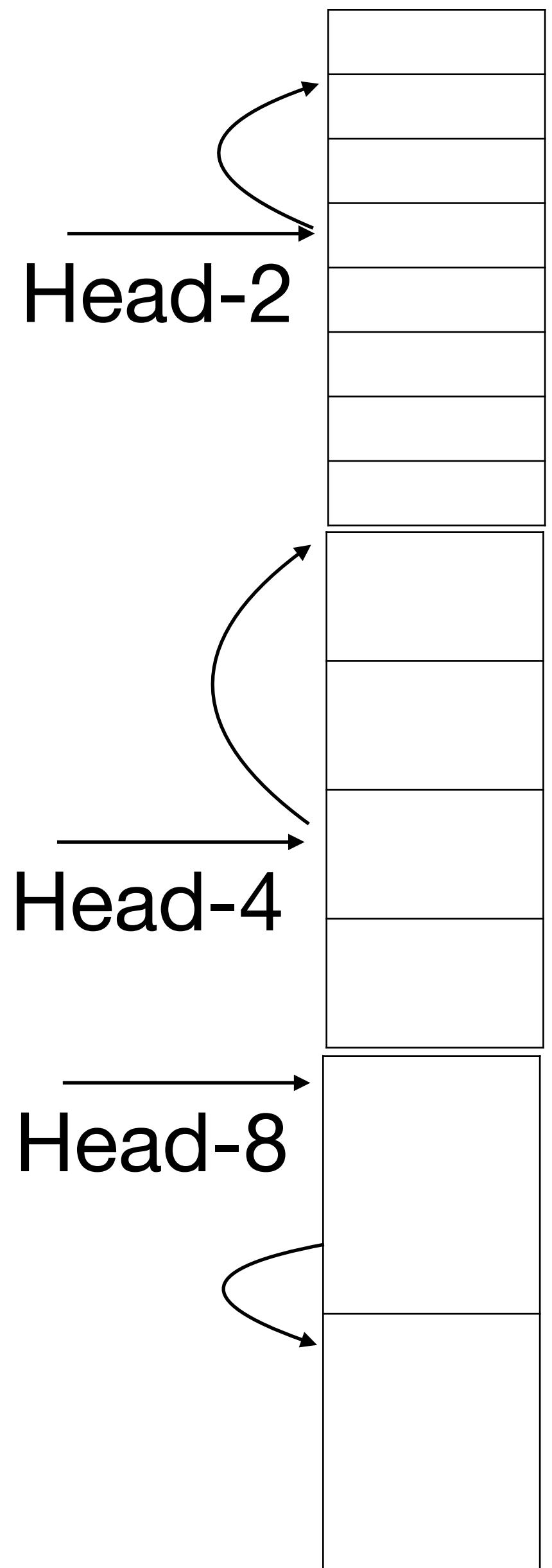
Segregated lists

- Separate lists for each size.



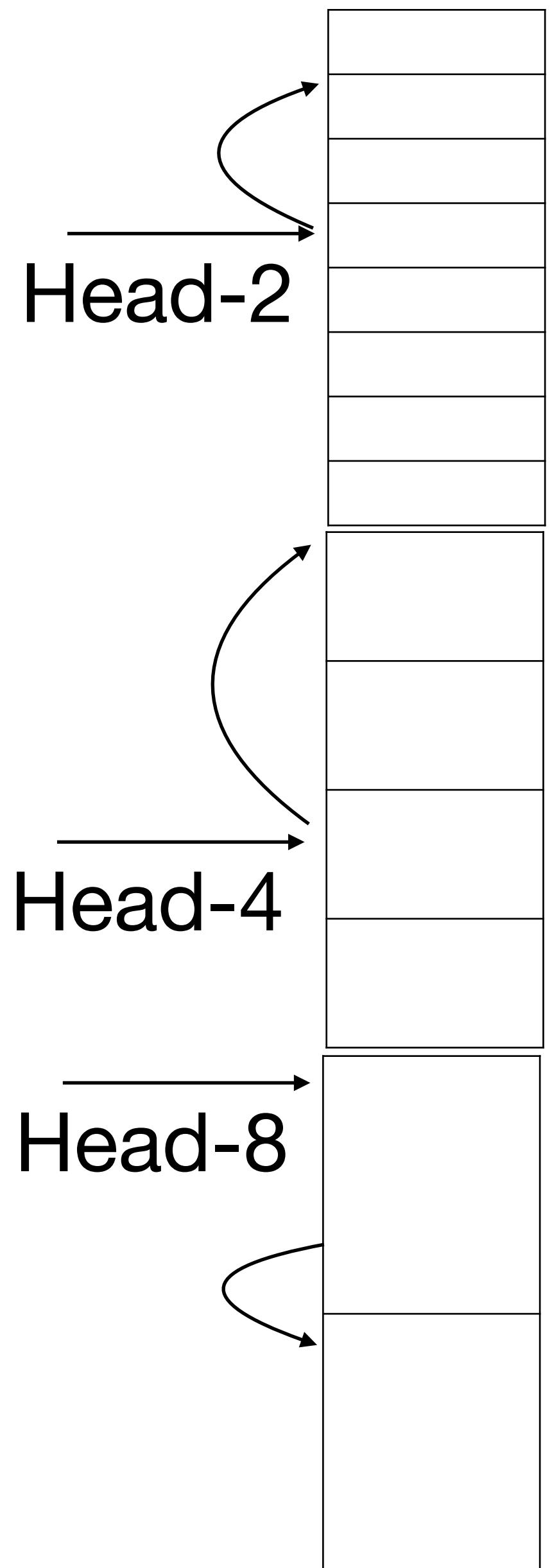
Segregated lists

- Separate lists for each size.
- “Segregated fit”: First fit in the smallest object list that can fit the object. Approximates best fit.



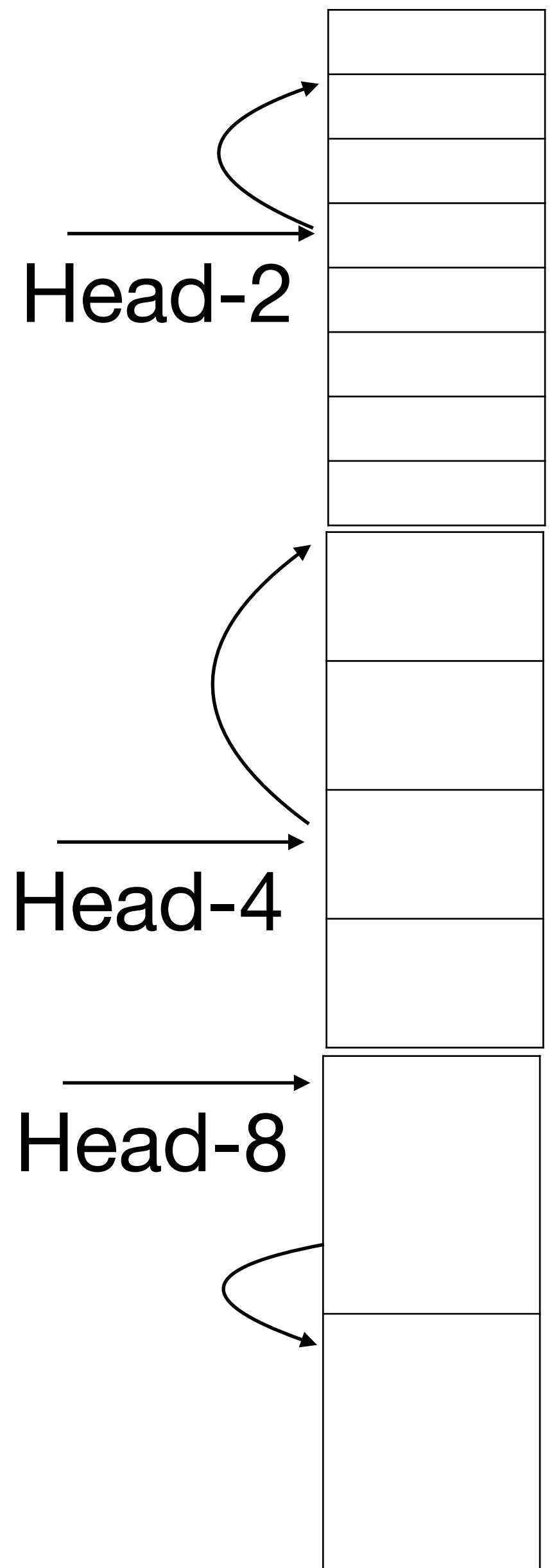
Segregated lists

- Separate lists for each size.
- “Segregated fit”: First fit in the smallest object list that can fit the object. Approximates best fit.
 - No splitting, coalescing



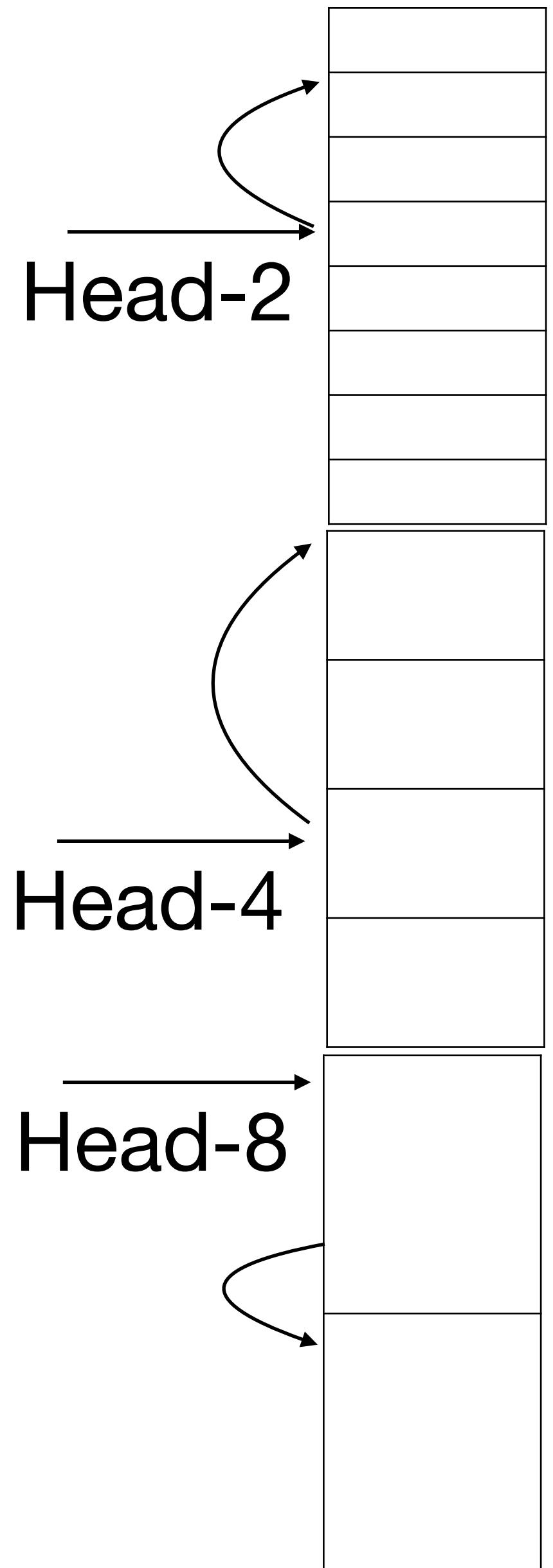
Segregated lists

- Separate lists for each size.
- “Segregated fit”: First fit in the smallest object list that can fit the object. Approximates best fit.
 - No splitting, coalescing
- Internal fragmentation: allocates more than asked



Segregated lists

- Separate lists for each size.
- “Segregated fit”: First fit in the smallest object list that can fit the object. Approximates best fit.
 - No splitting, coalescing
- Internal fragmentation: allocates more than asked
- Wastes memory: If no allocation from object size, have unnecessary reserved space for it



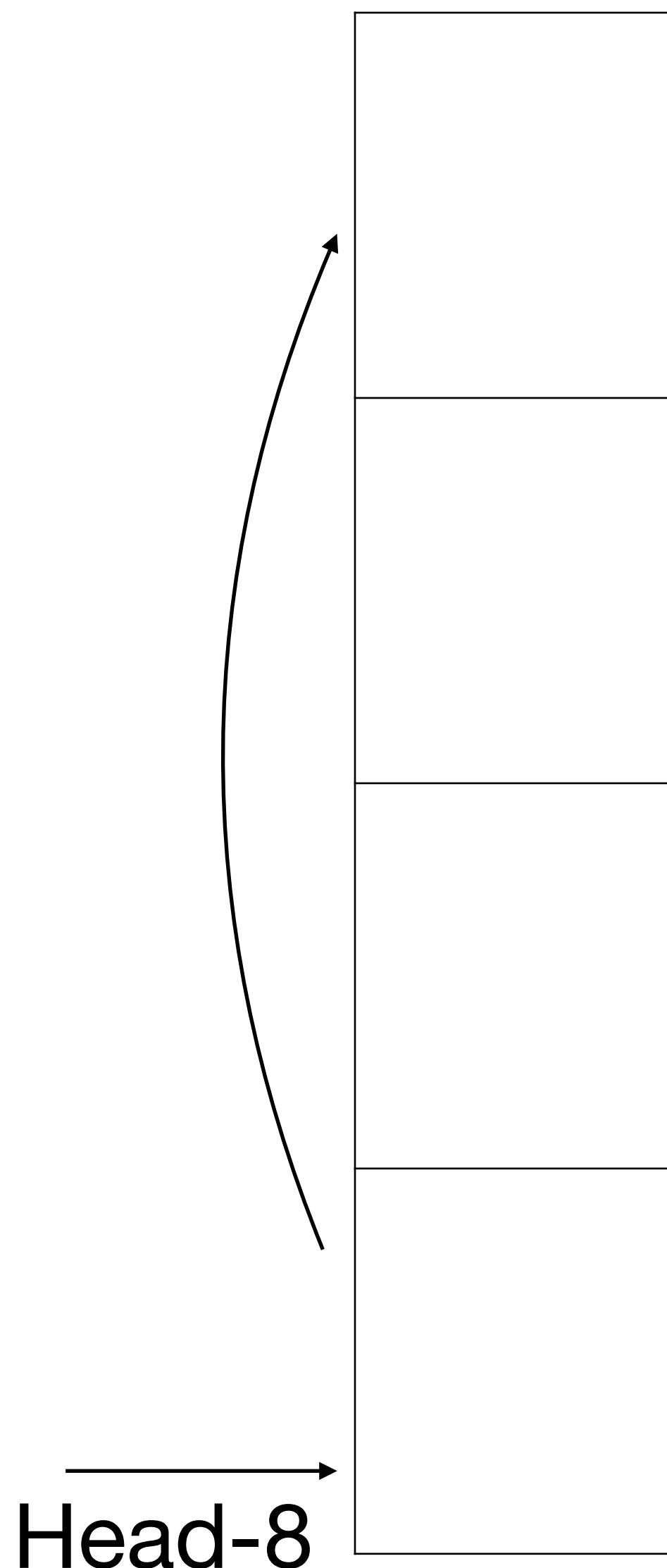
Slab allocator

Used in Linux Kernel

- Knows the size of allocations made by the Linux kernel. Keep segregated lists for each such struct.
 - No internal fragmentation: Exactly the size of the struct
 - Hierarchical allocator: Return unused lists back to global allocator

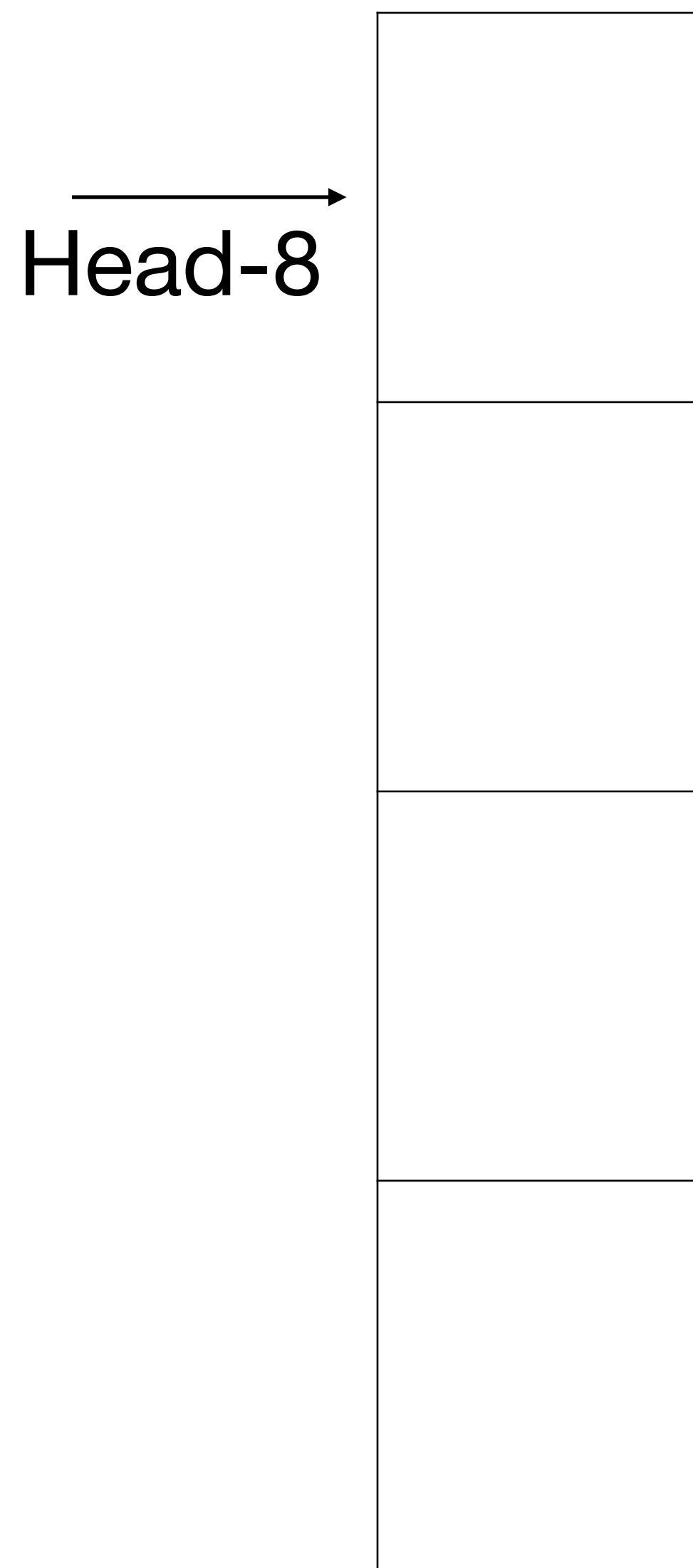
Buddy allocator

- Example:
 - $p1 = \text{malloc}(3)$
 - $p2 = \text{malloc}(2)$



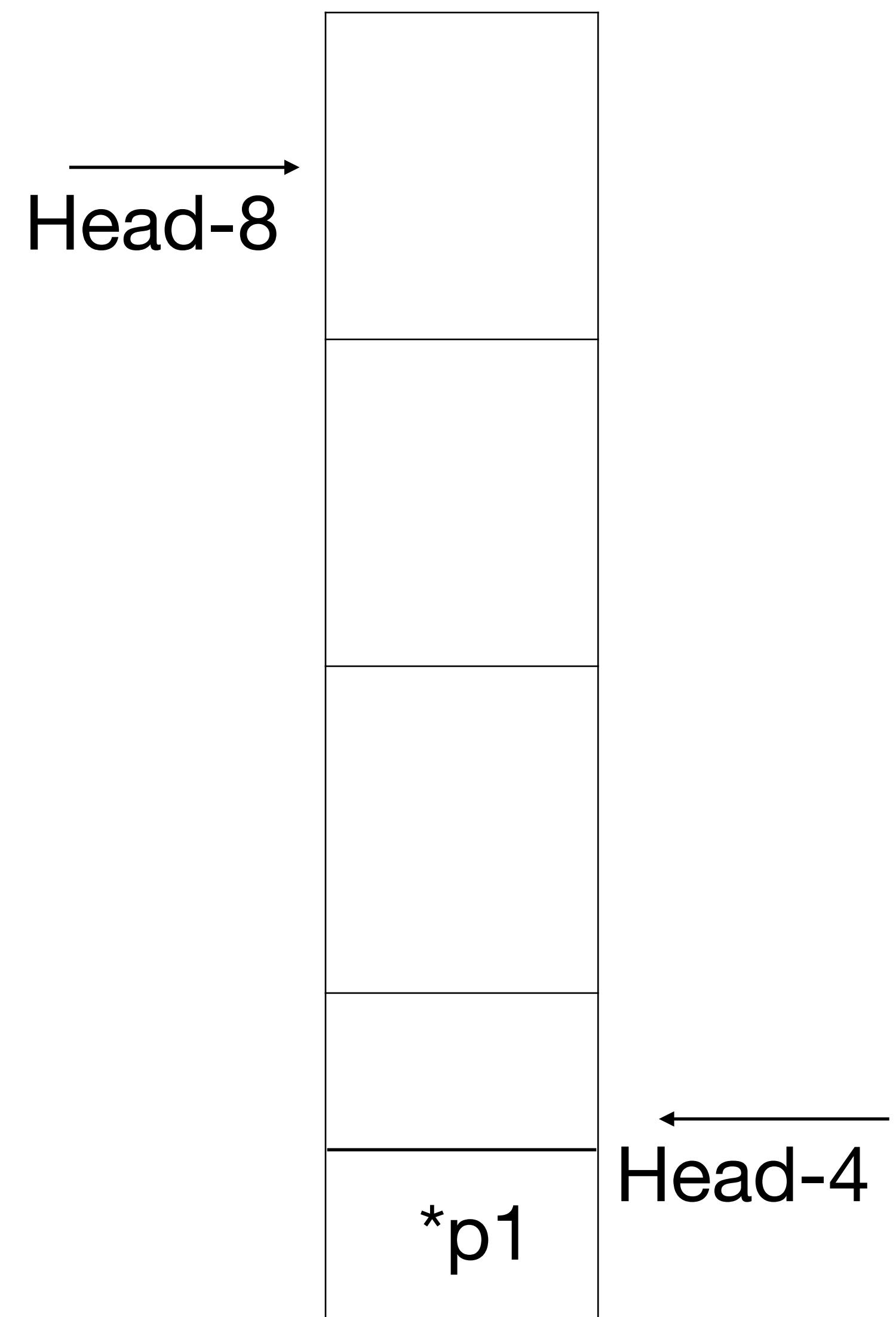
Buddy allocator

- Example:
 - $p1 = \text{malloc}(3)$
 - $p2 = \text{malloc}(2)$



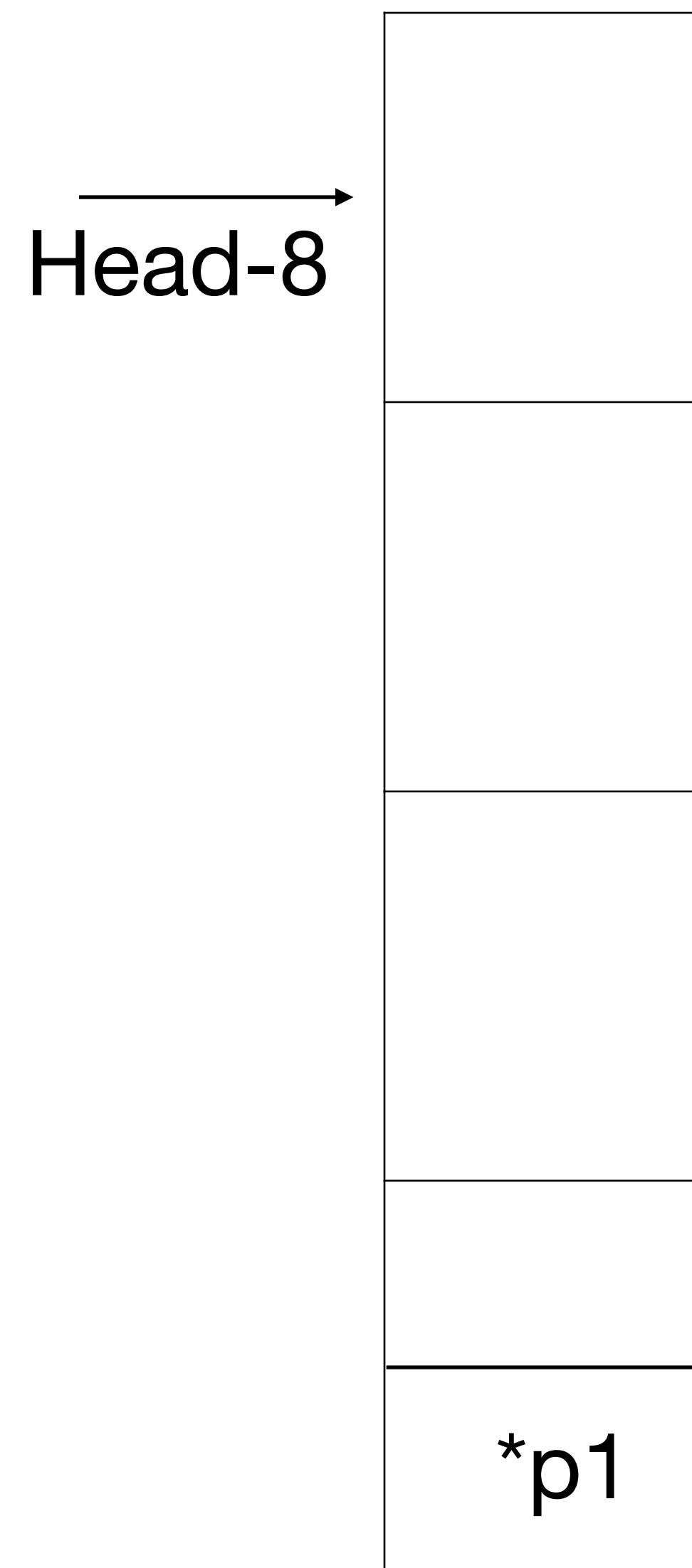
Buddy allocator

- Example:
 - $p1 = \text{malloc}(3)$
 - $p2 = \text{malloc}(2)$



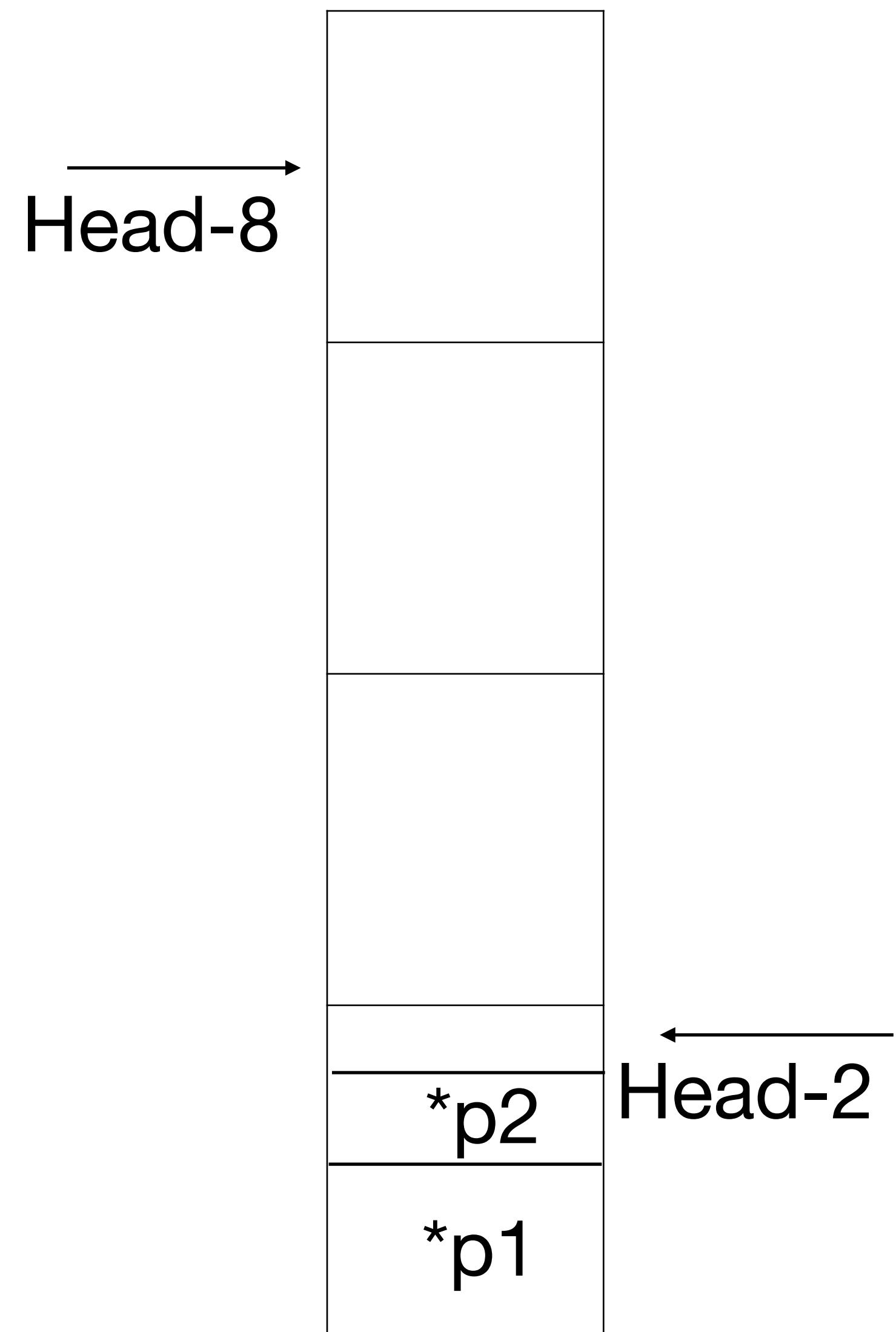
Buddy allocator

- Example:
 - $p1 = \text{malloc}(3)$
 - $p2 = \text{malloc}(2)$



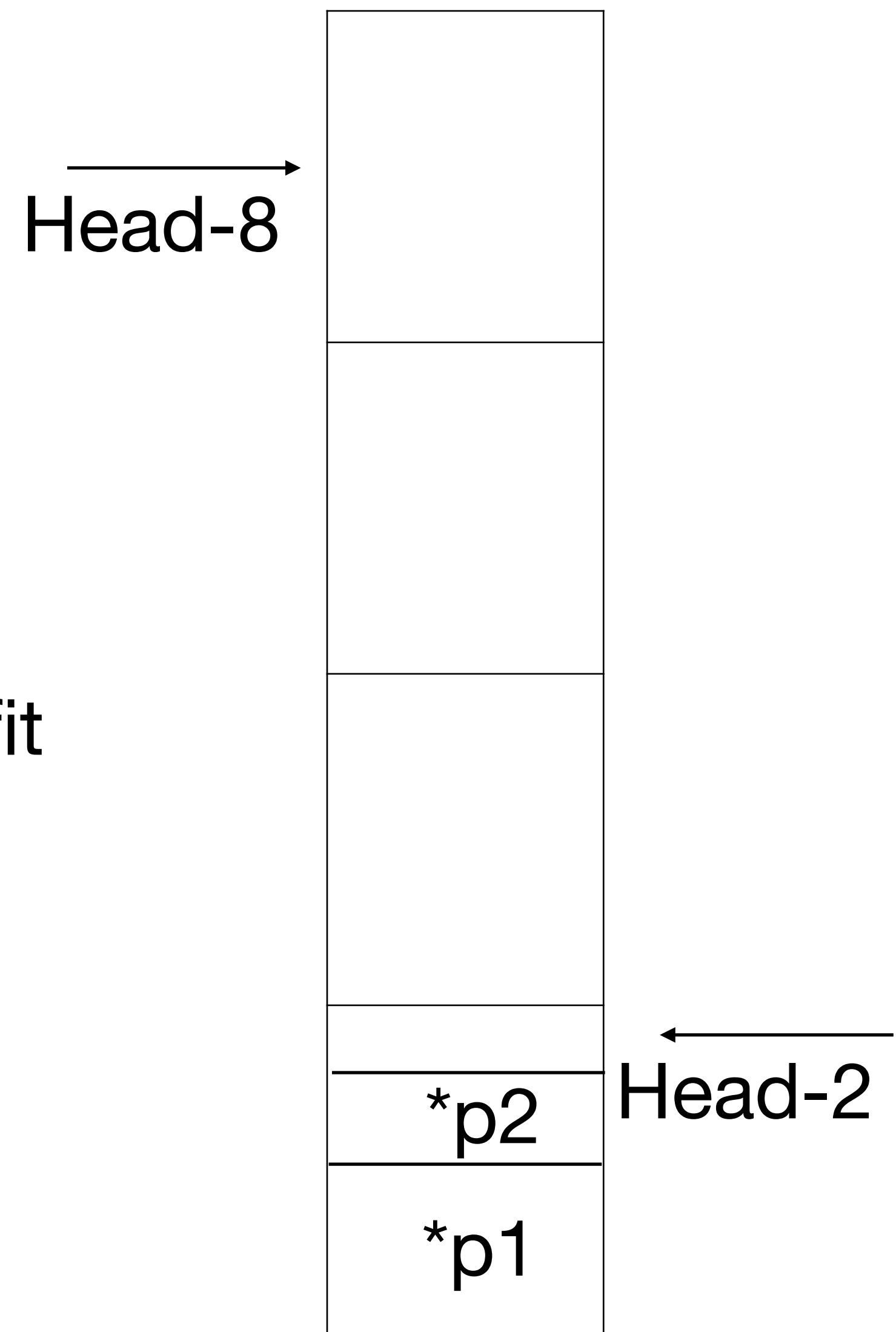
Buddy allocator

- Example:
 - $p1 = \text{malloc}(3)$
 - $p2 = \text{malloc}(2)$



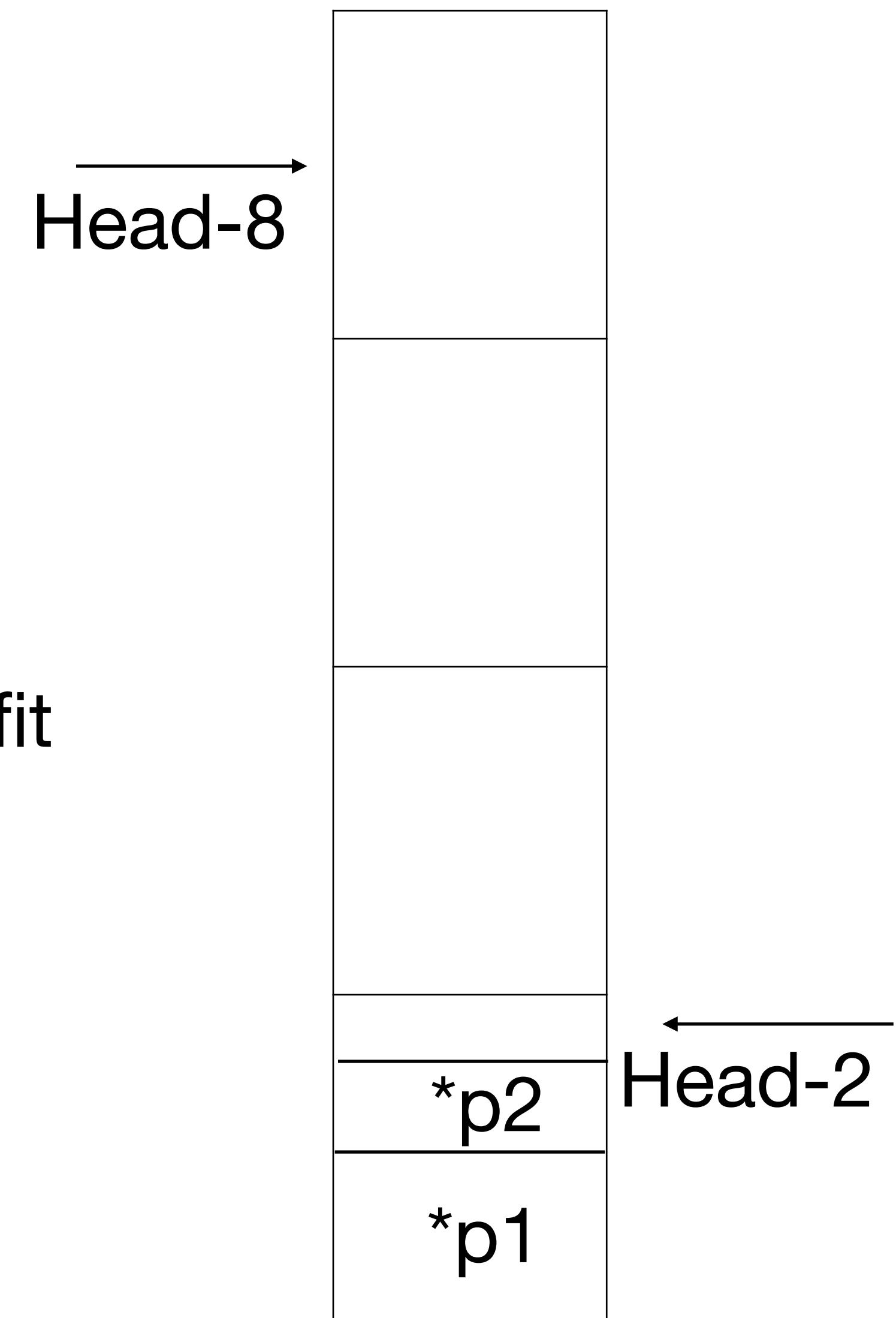
Buddy allocator

- Example:
 - $p1 = \text{malloc}(3)$
 - $p2 = \text{malloc}(2)$
- First fit with segregated lists approximates best fit



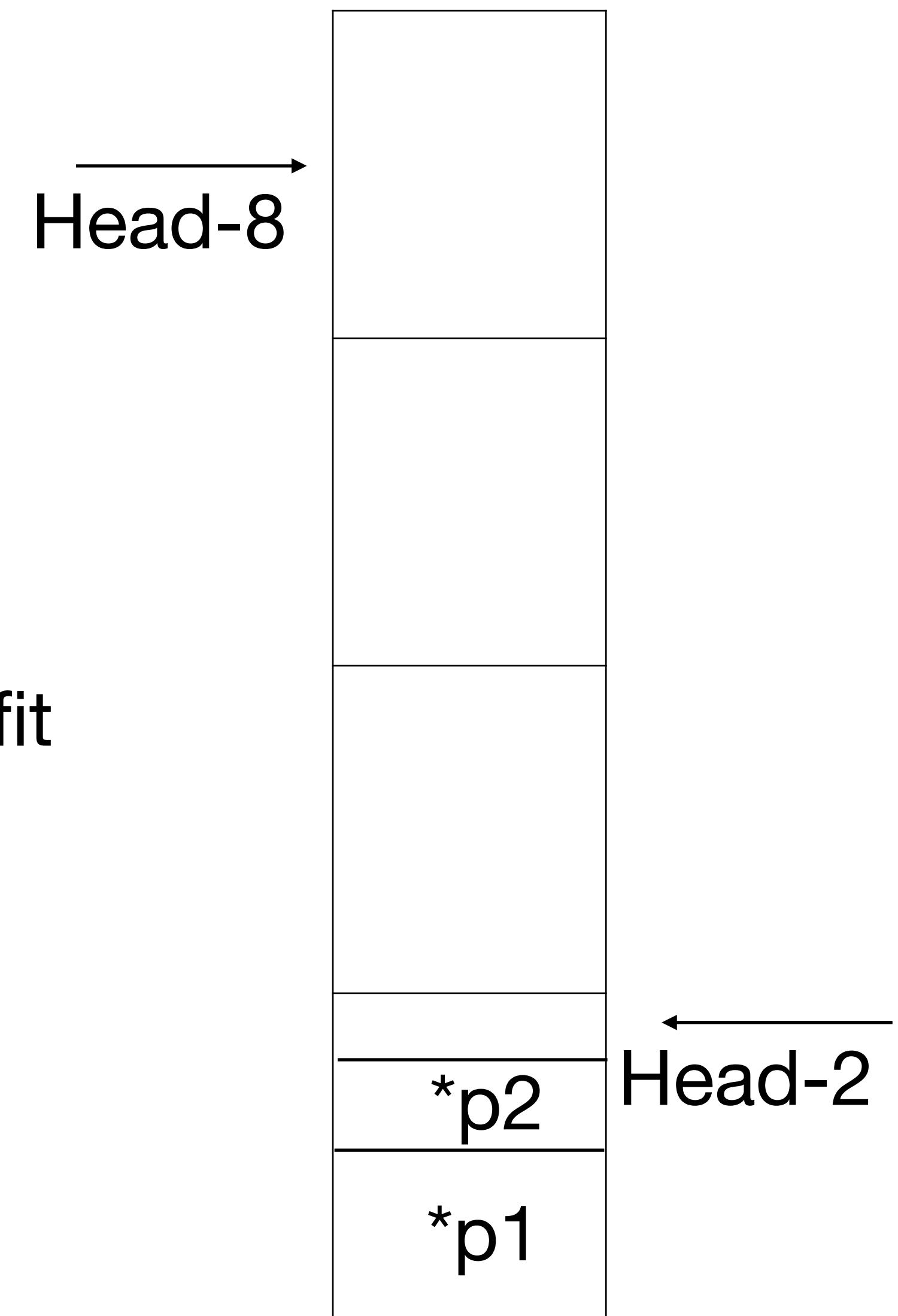
Buddy allocator

- Example:
 - $p1 = \text{malloc}(3)$
 - $p2 = \text{malloc}(2)$
- First fit with segregated lists approximates best fit
- Straightforward splitting and coalescing



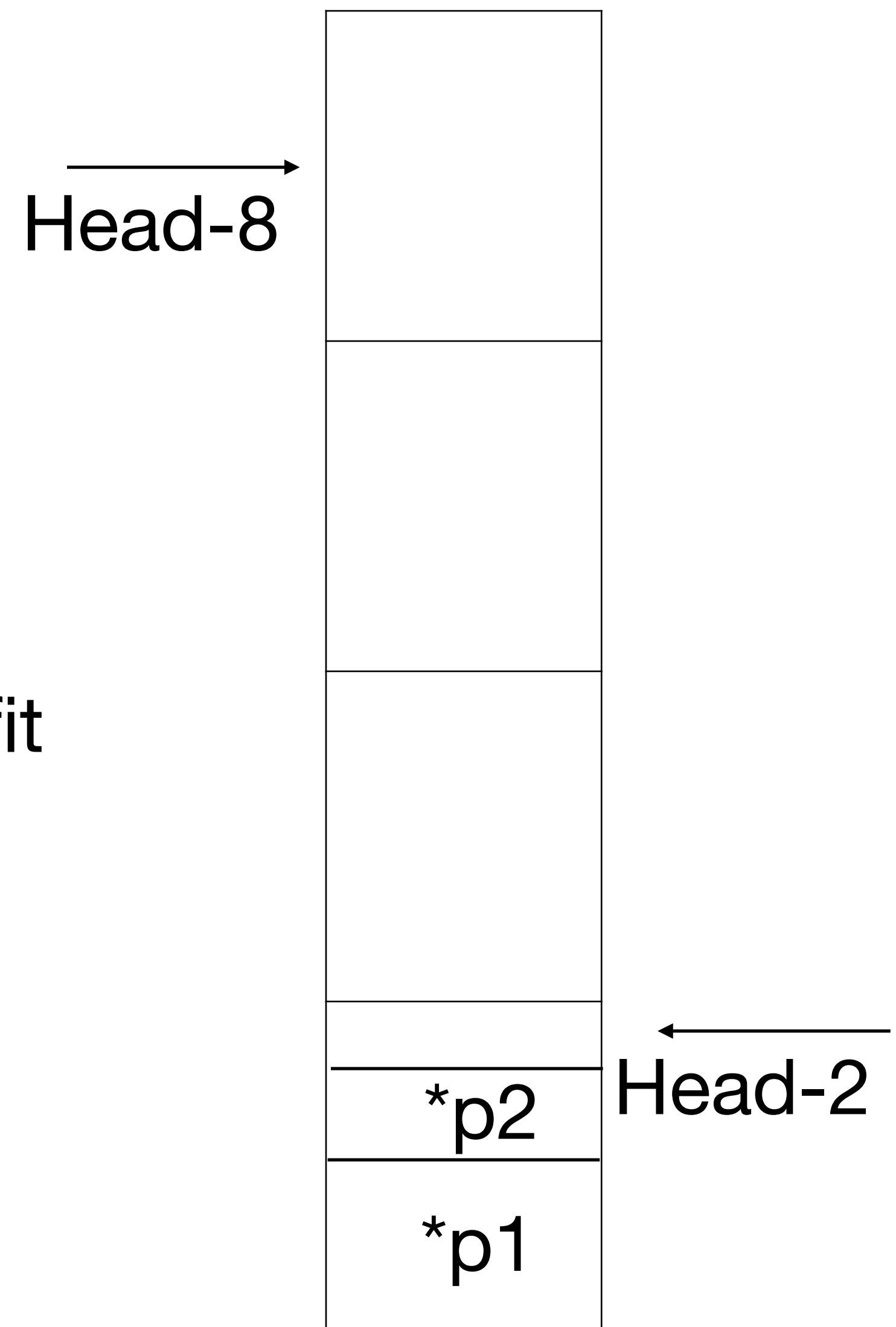
Buddy allocator

- Example:
 - $p1 = \text{malloc}(3)$
 - $p2 = \text{malloc}(2)$
- First fit with segregated lists approximates best fit
- Straightforward splitting and coalescing
- Deallocation order: fast frees



Buddy allocator

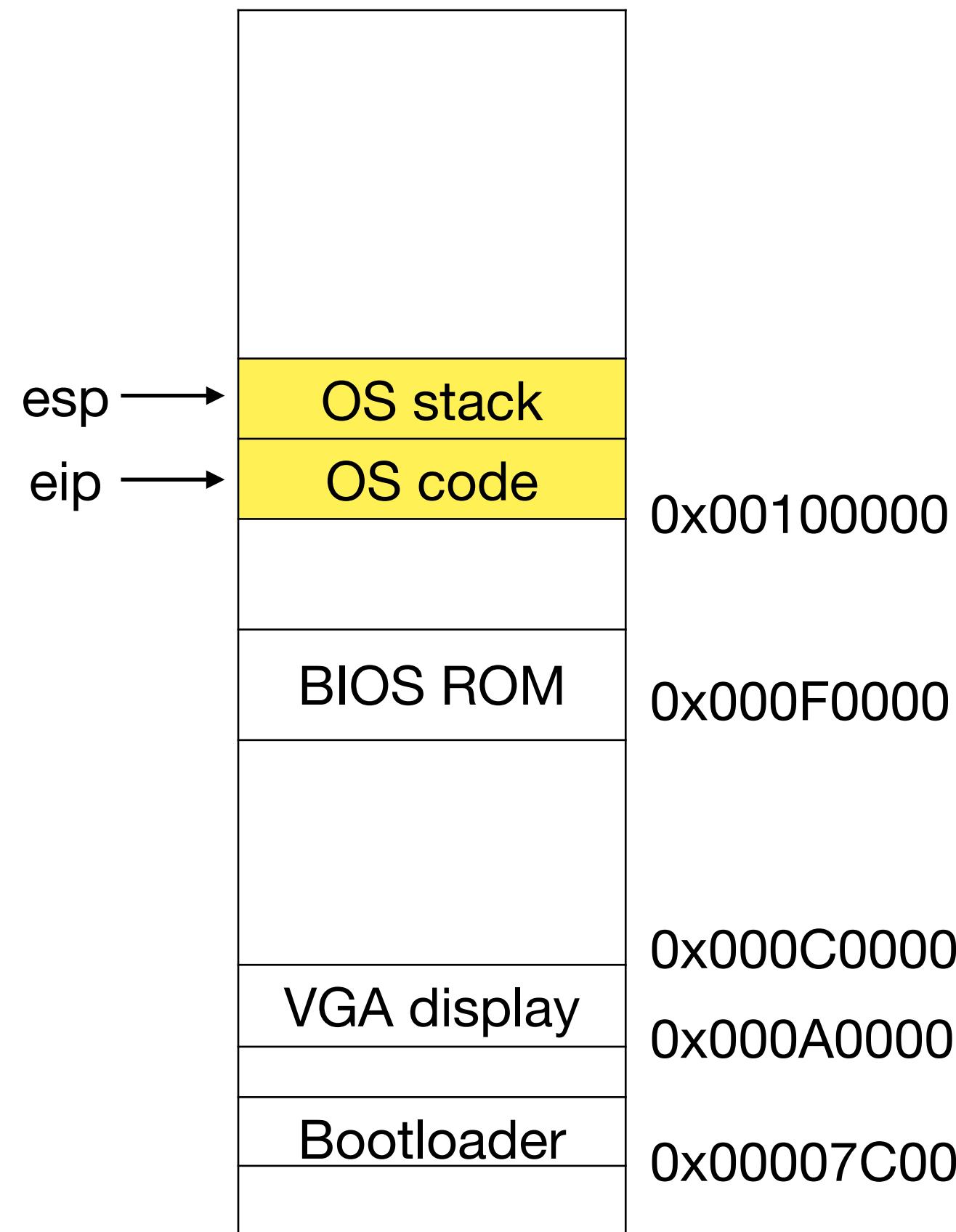
- Example:
 - $p1 = \text{malloc}(3)$
 - $p2 = \text{malloc}(2)$
- First fit with segregated lists approximates best fit
- Straightforward splitting and coalescing
- Deallocation order: fast frees
- Used in Linux kernel



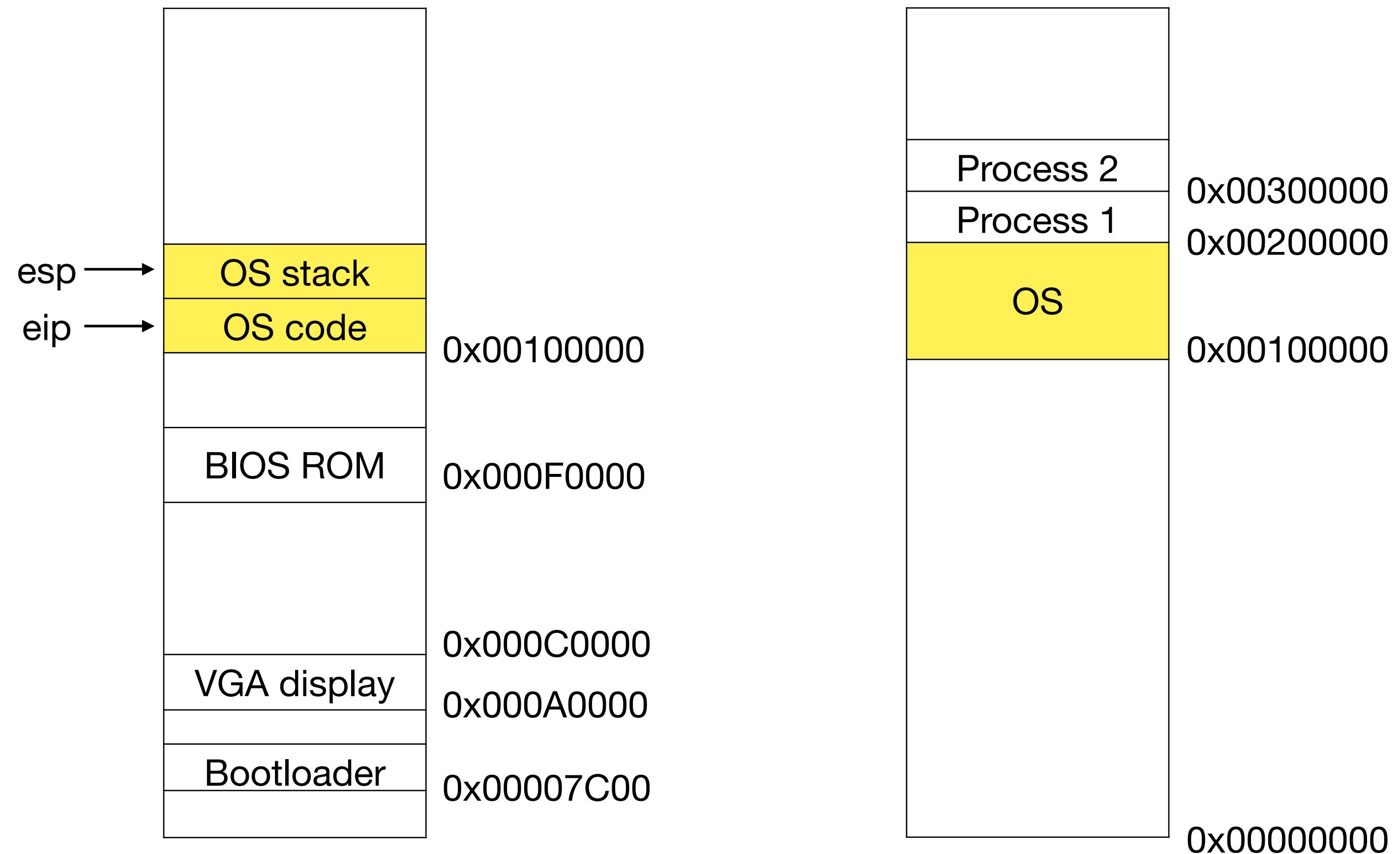
Aside: automatic memory management

- “Higher level” languages do not expose raw pointers to programmers.
Example: Java, Python, Go
 - The language runtime manages memory. Programmer need not call free.
 - Largely prevents memory leaks, dangling references, double free, null pointer dereference
 - Mark and sweep garbage collector, reference counting based garbage collector
 - Copying GC can do compaction to defragment heap. Will rewrite pointers.
 - Can incur heavy performance penalty

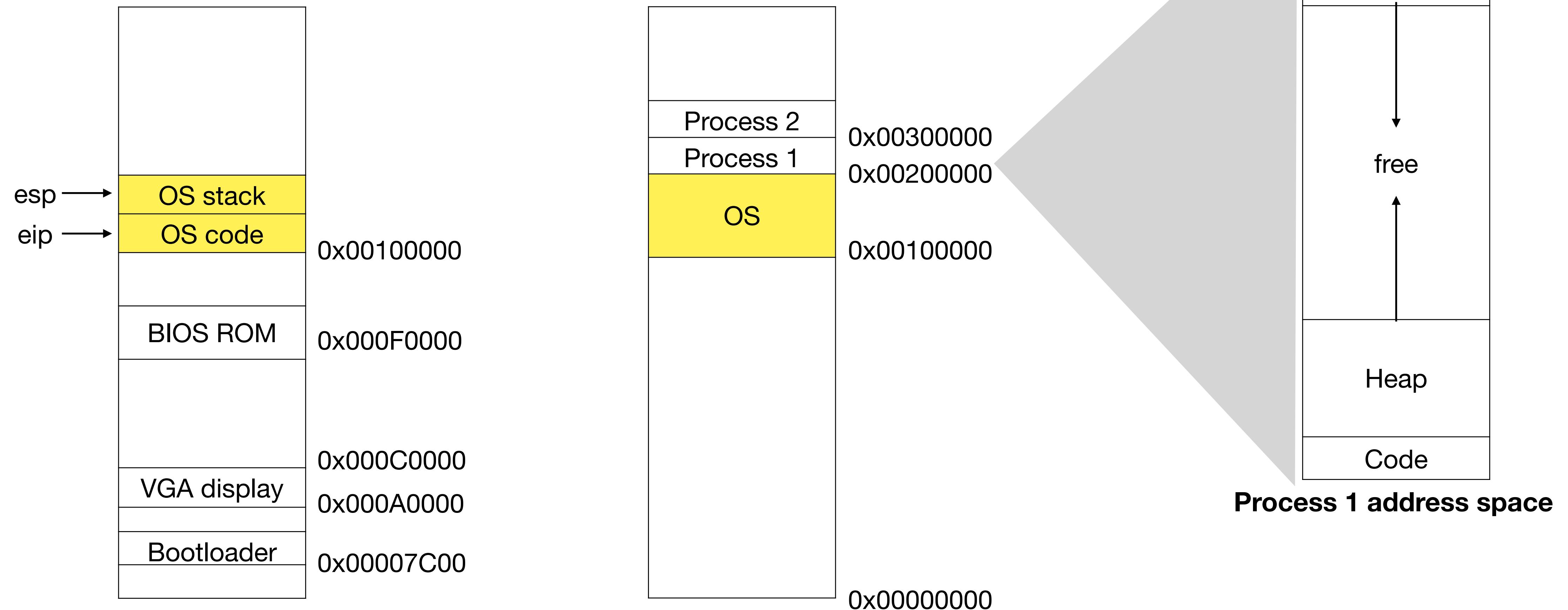
Memory isolation and address space



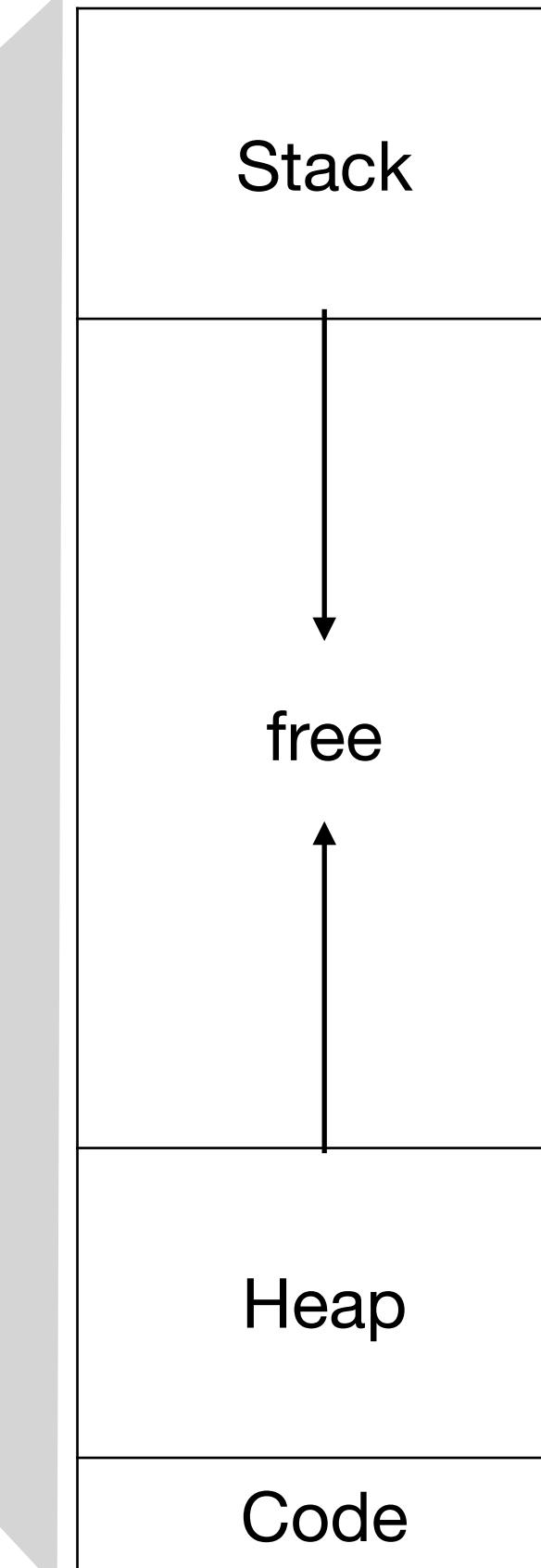
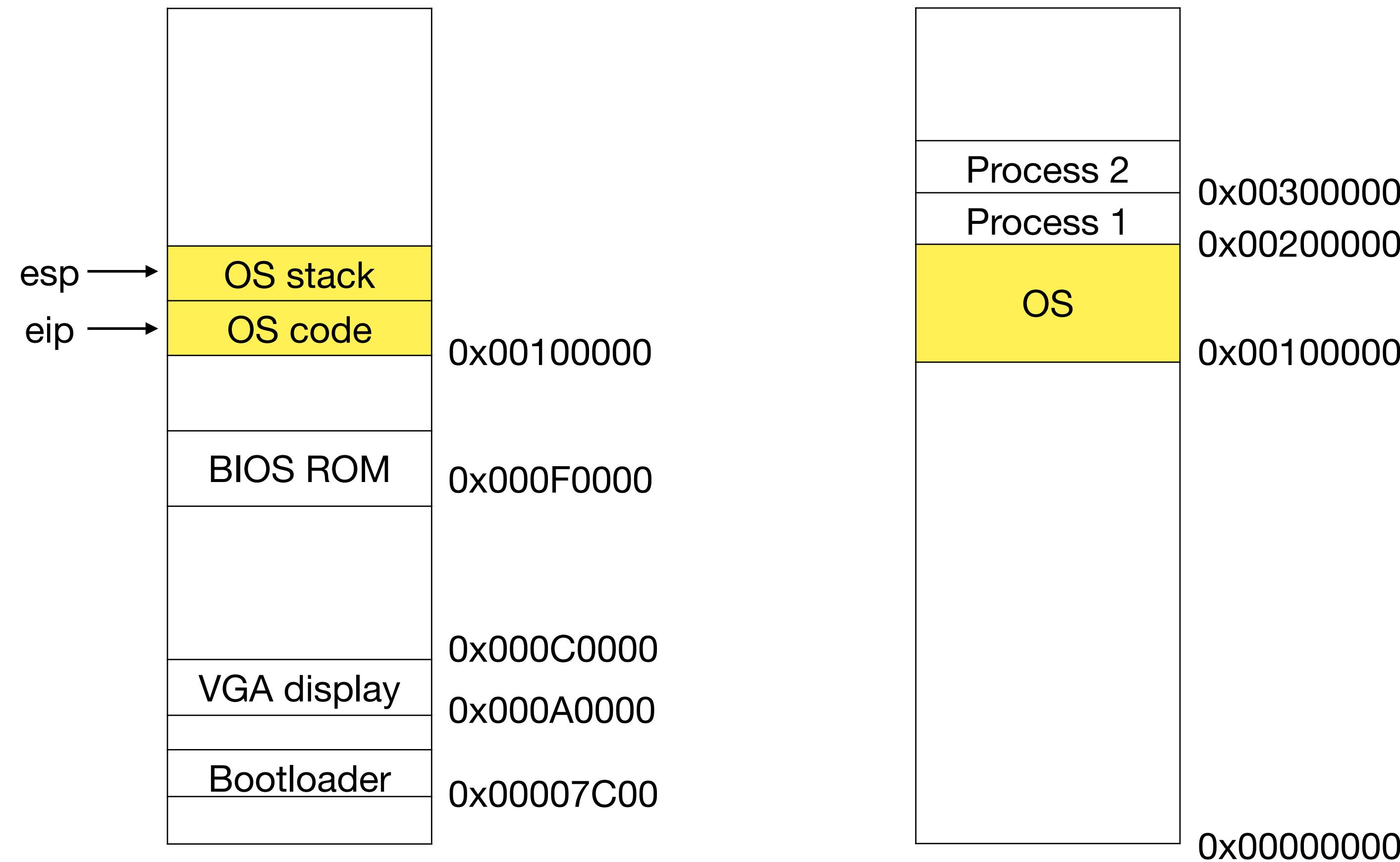
Memory isolation and address space



Memory isolation and address space



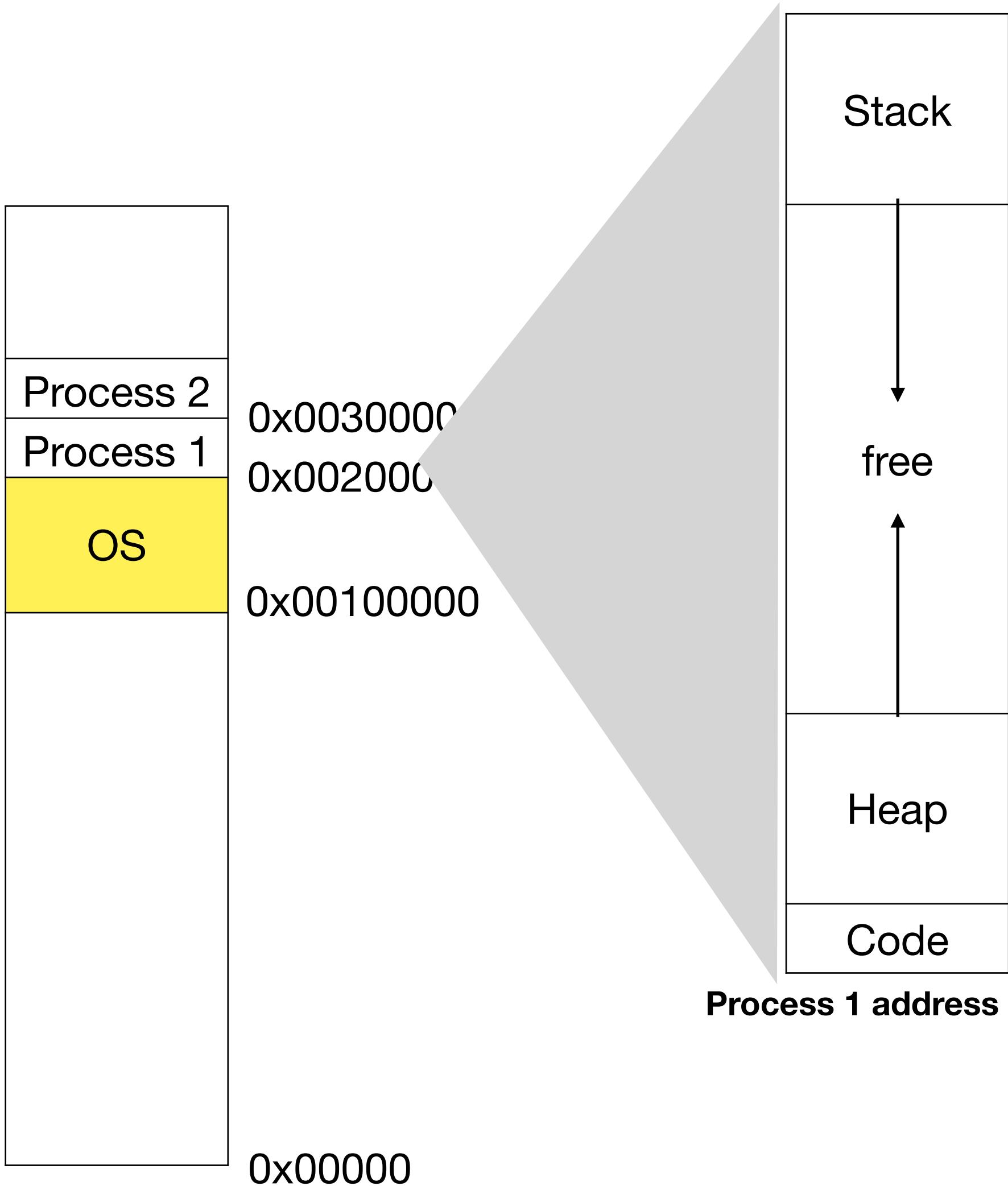
Memory isolation and address space



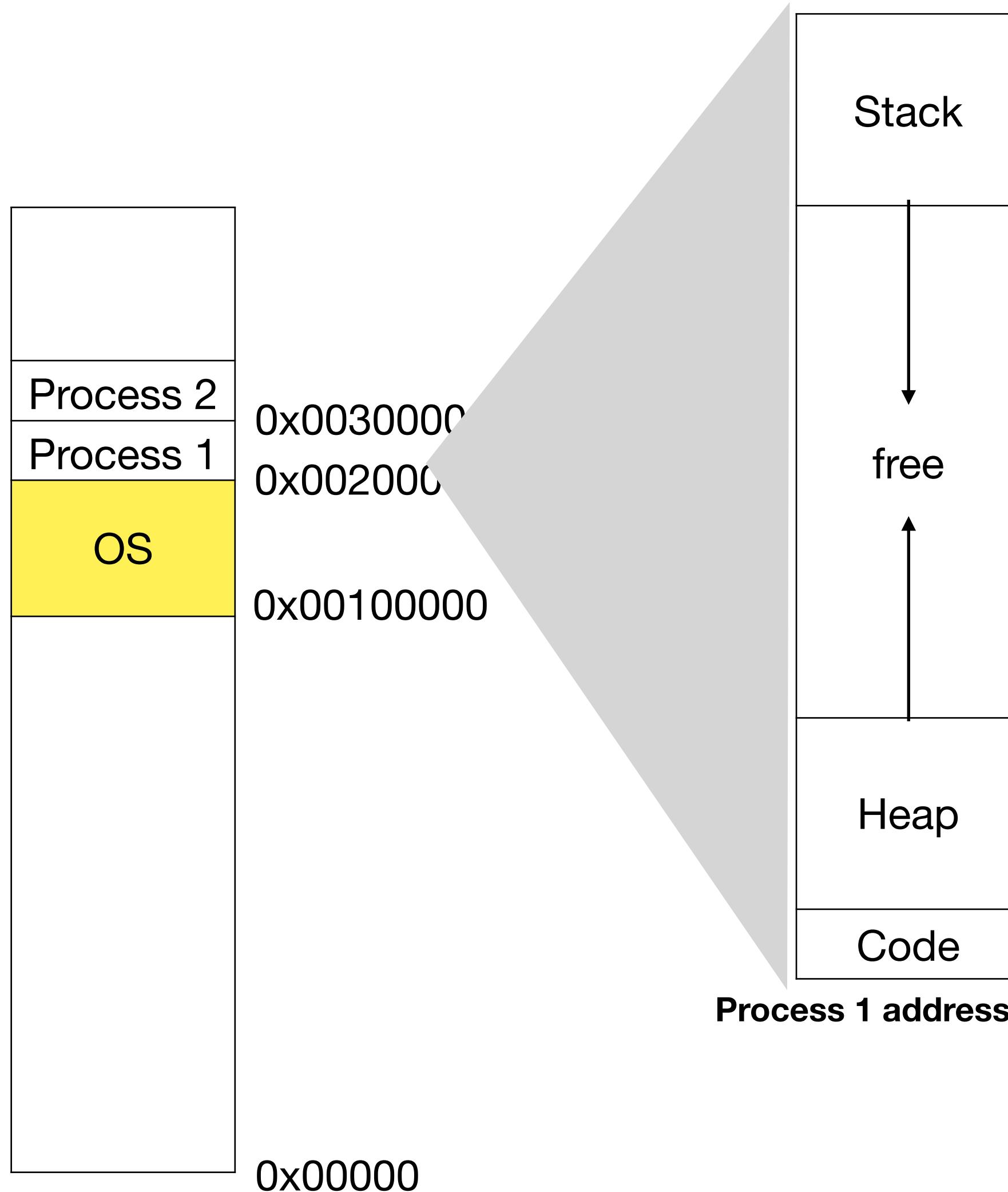
Process 1 address space

Due to address translation, compiler need not worry where the program will be loaded!

Segmentation

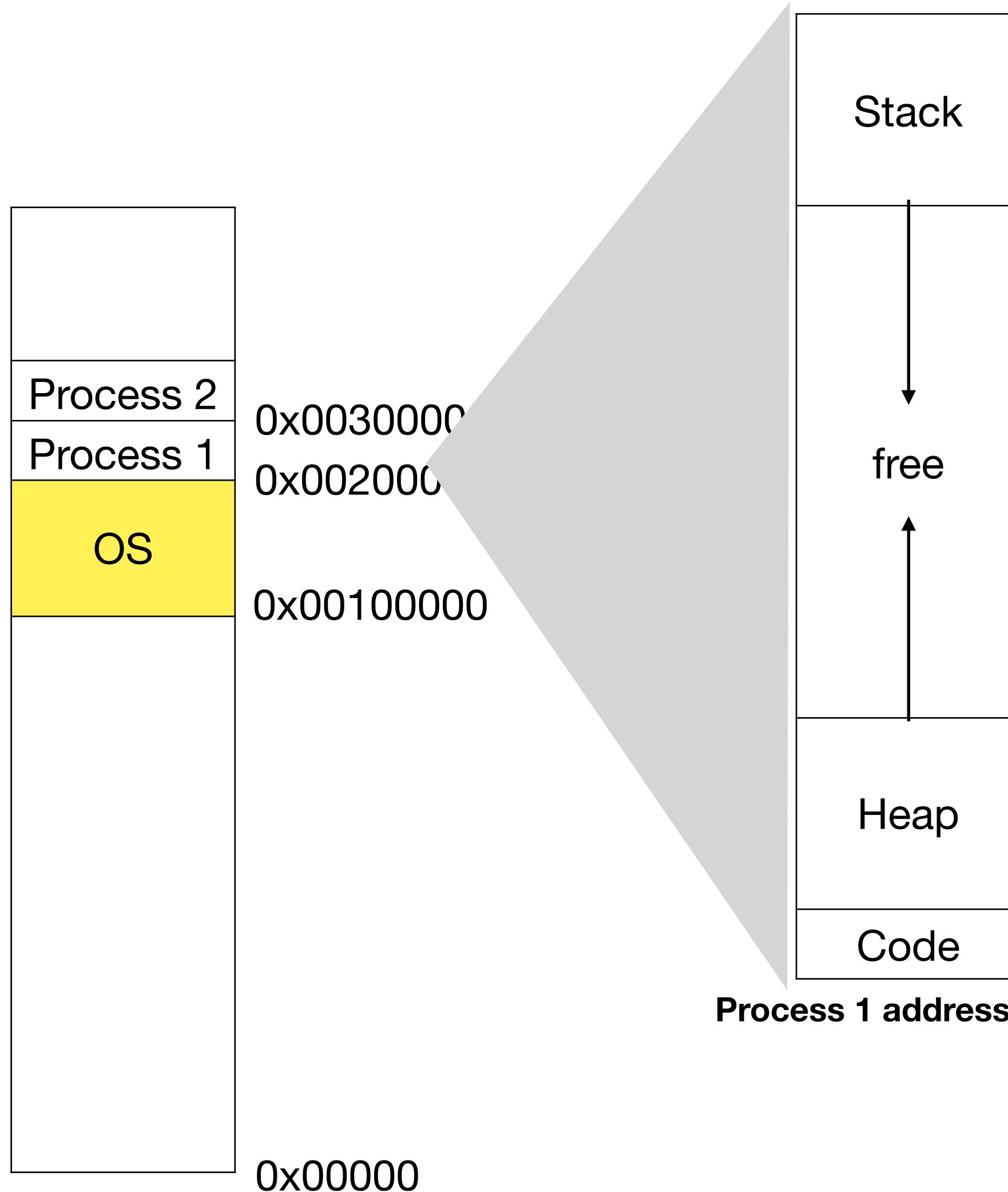


Segmentation



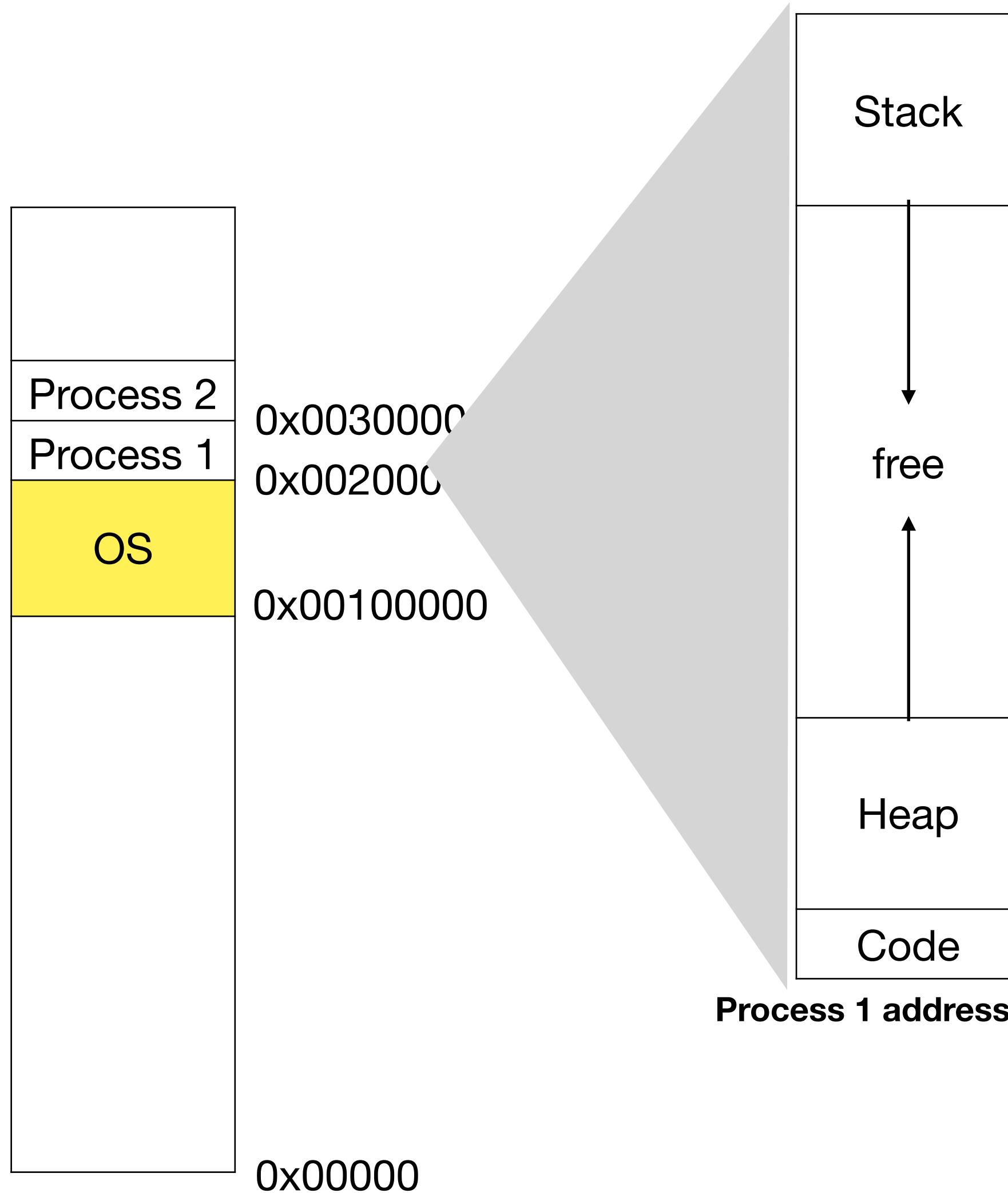
- Place each segment independently to not map free space

Segmentation

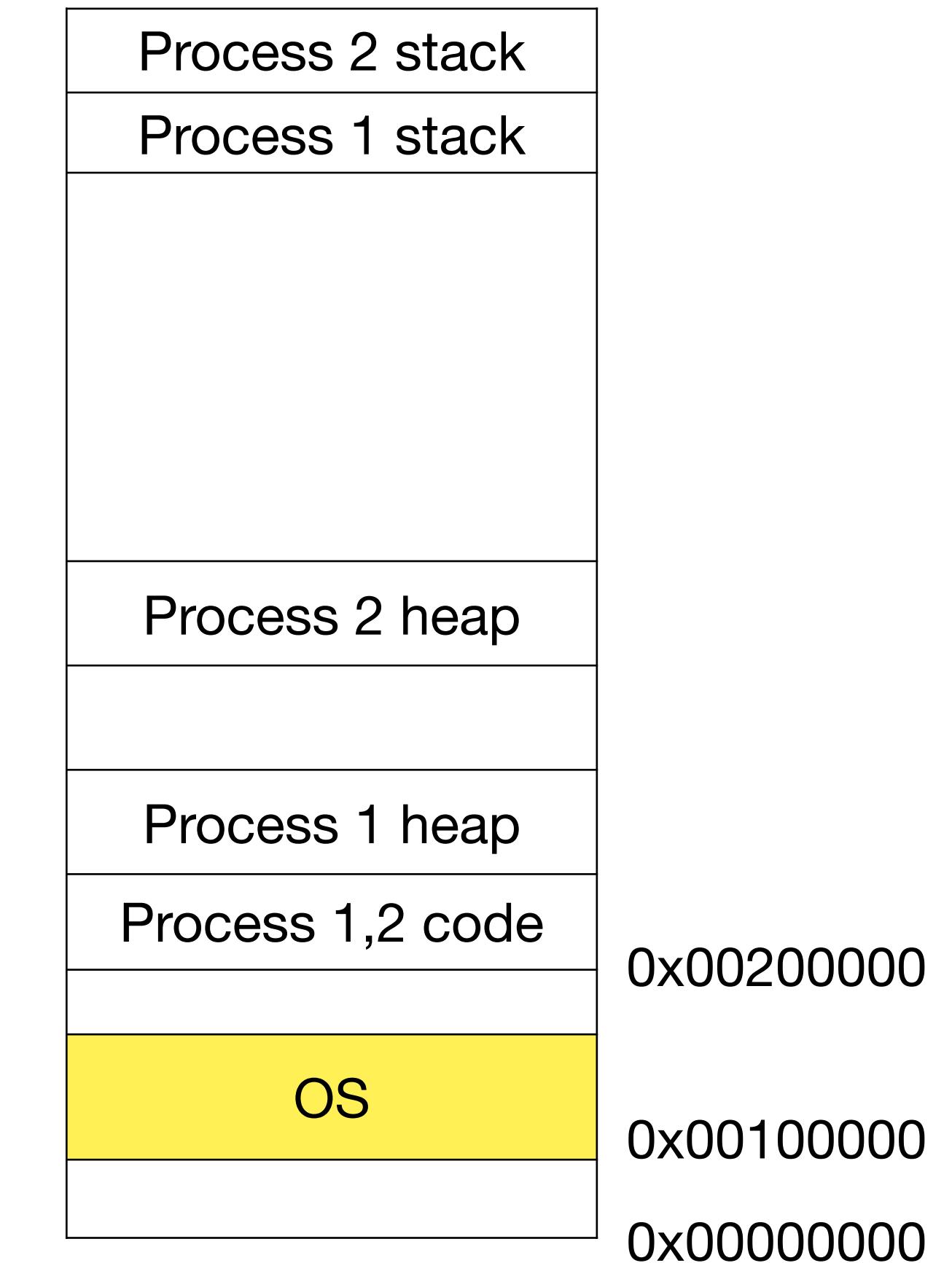


- Place each segment independently to not map free space
- Share code segments to save space

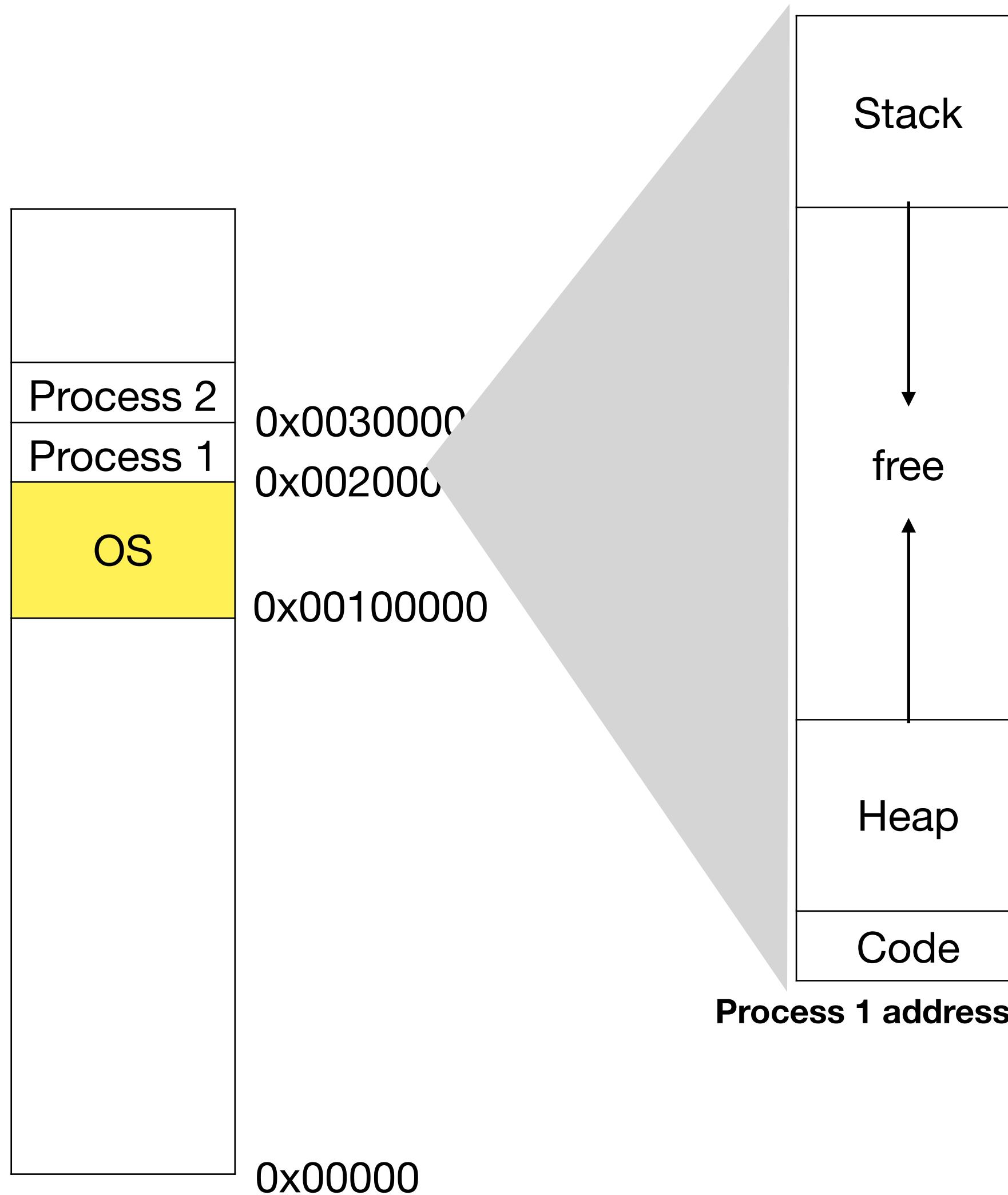
Segmentation



- Place each segment independently to not map free space
- Share code segments to save space

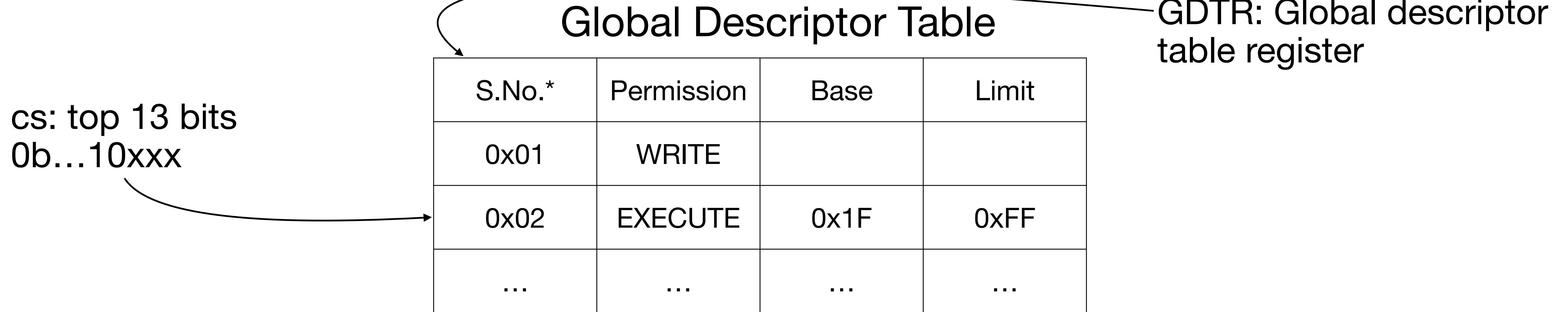


Segmentation



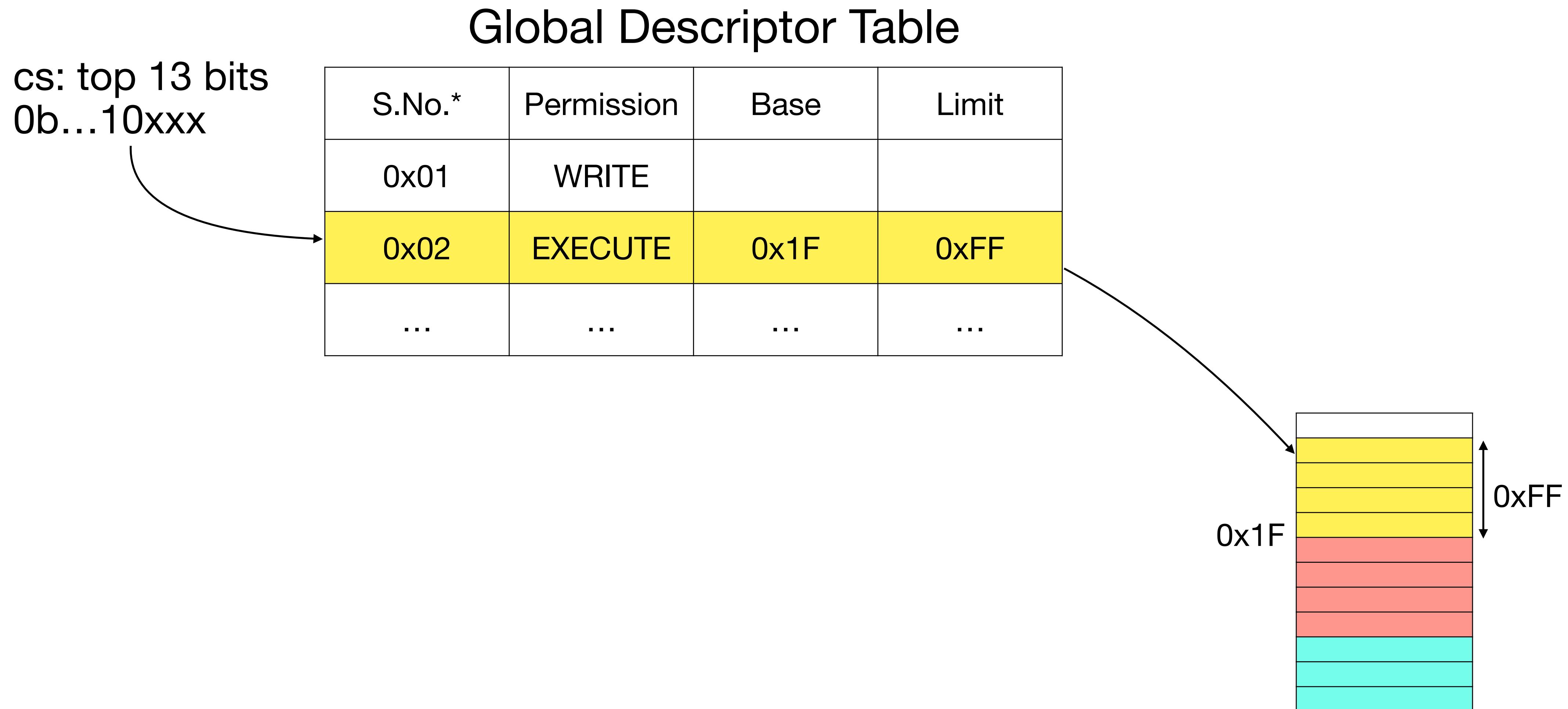
- Place each segment independently to not map free space
- Share code segments to save space
 - Mark non-writeable

Many segments can be initialised in GDT

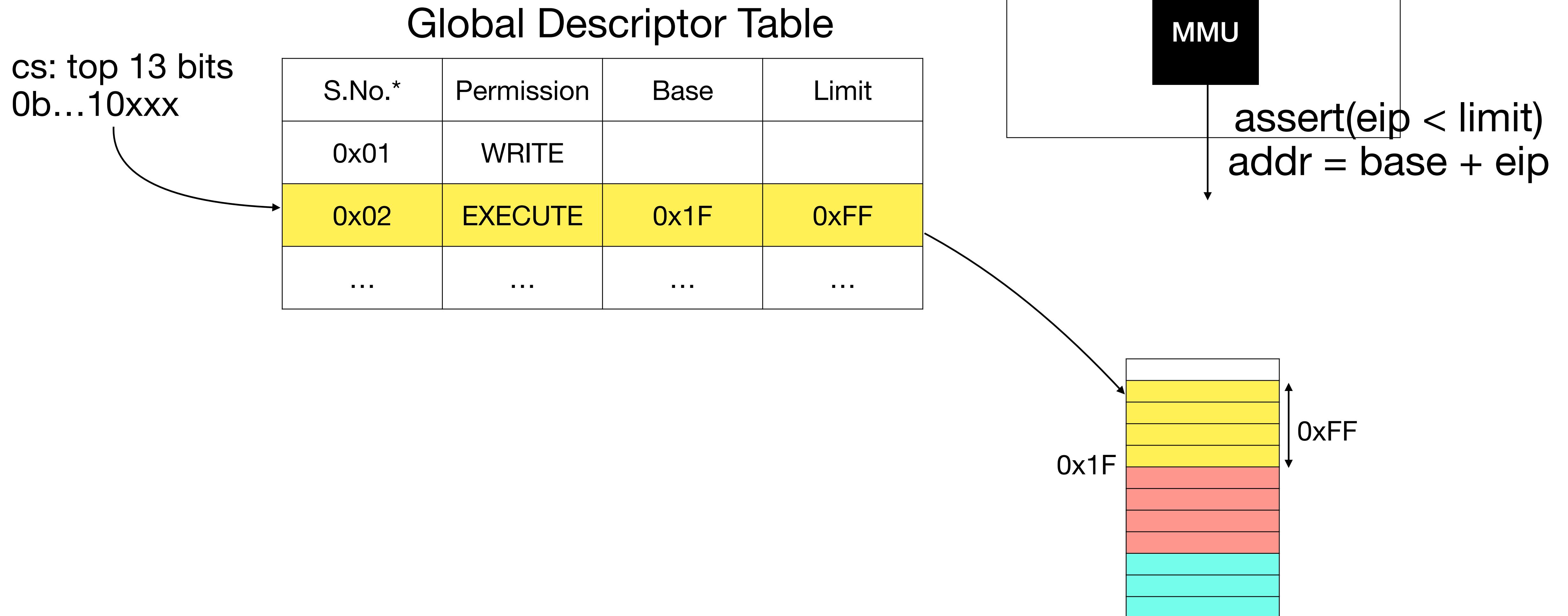


*: S.No. added only for illustration

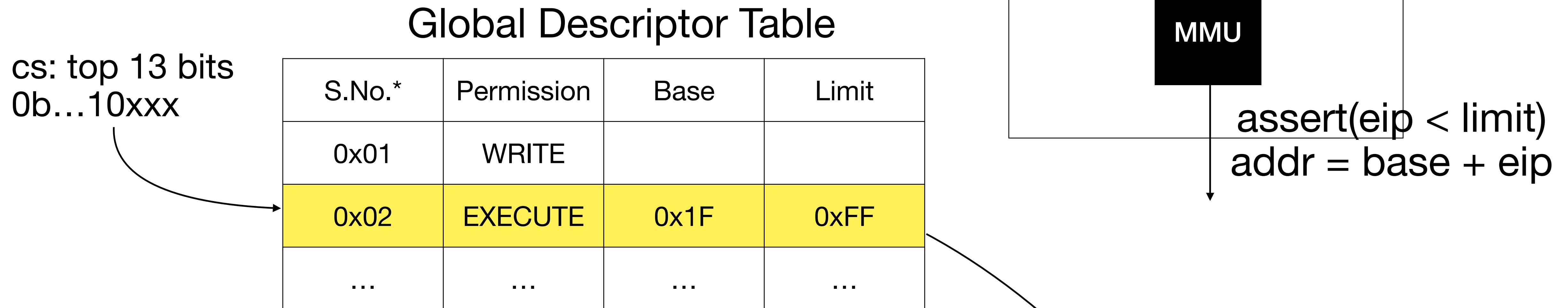
Address translation



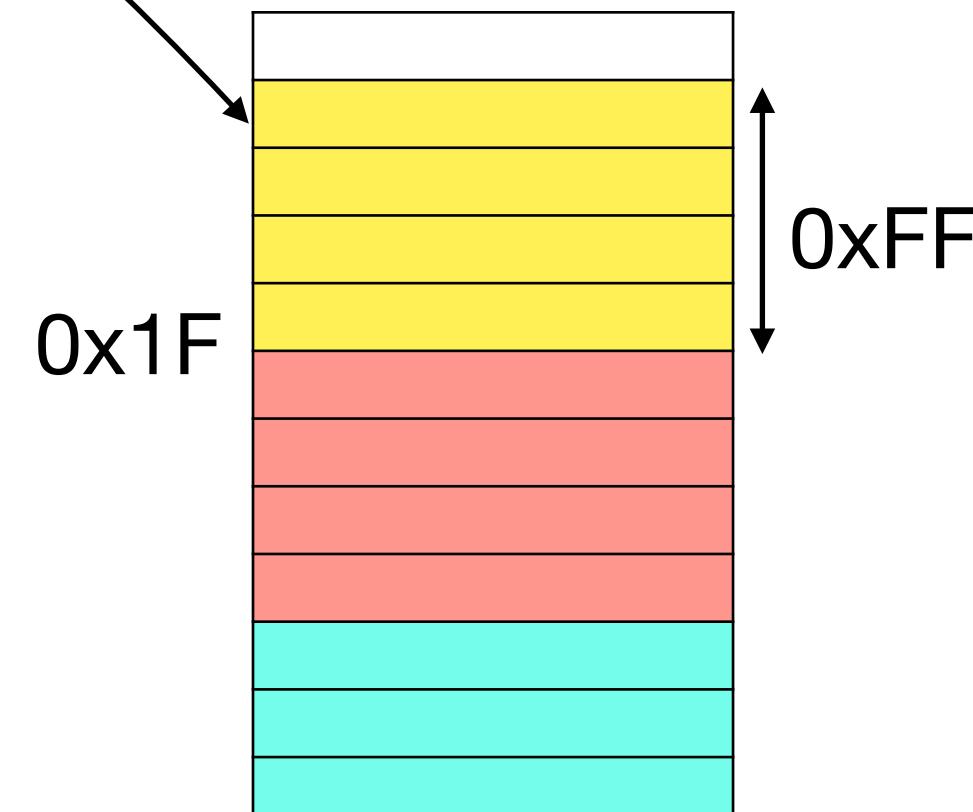
Address translation



Address translation

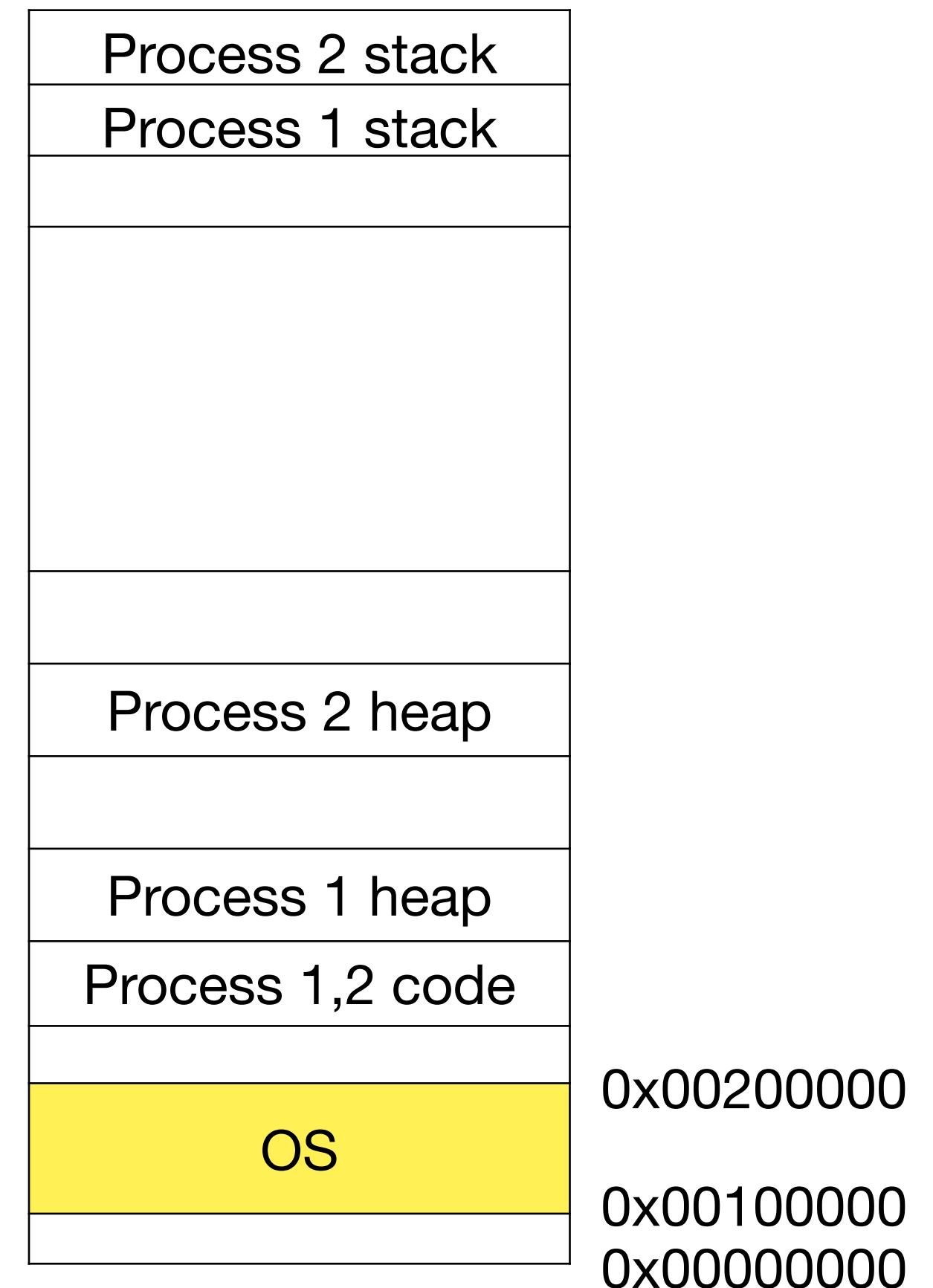


- Can “protect” different segments from each other



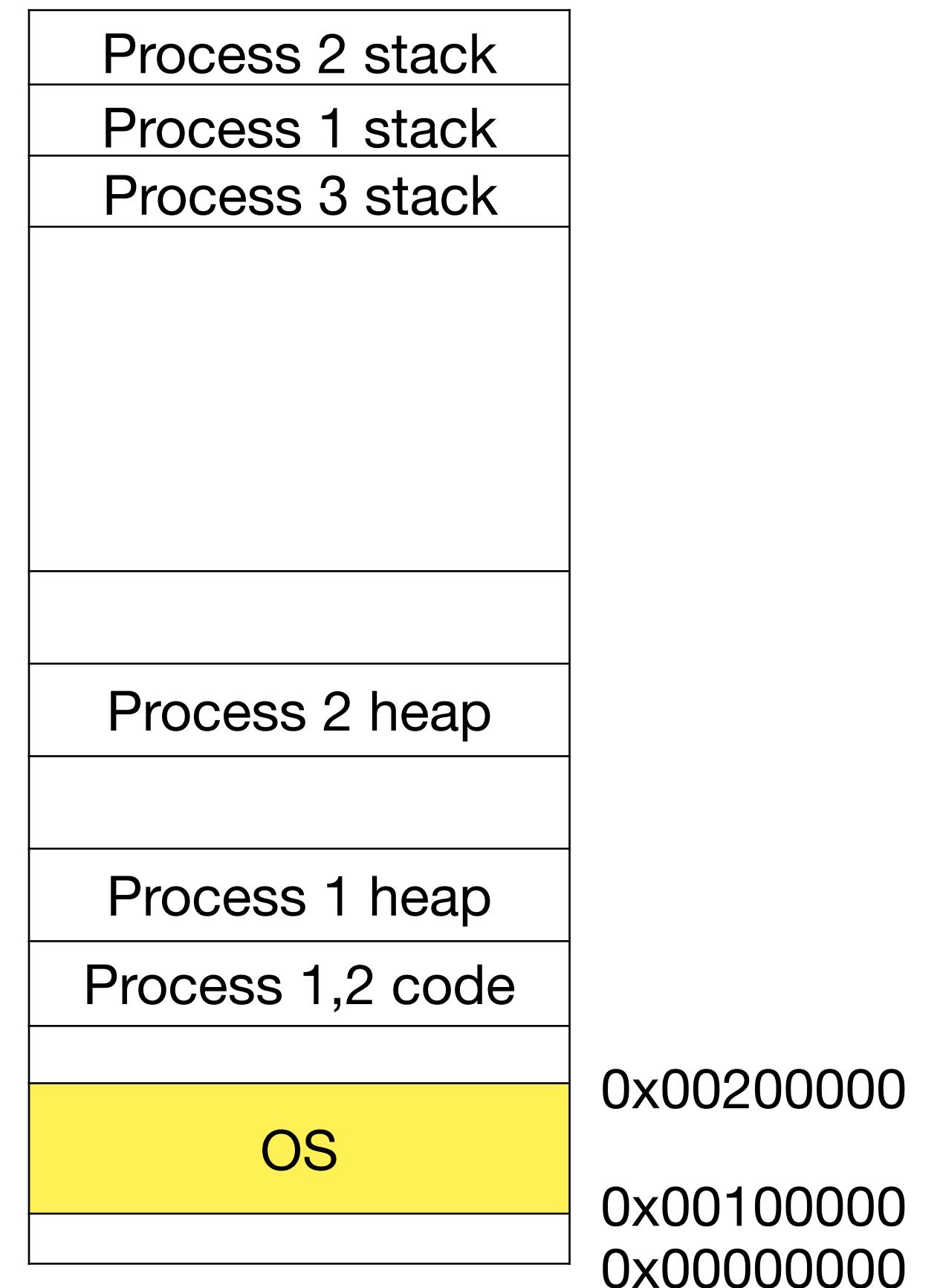
Allocating memory to a new process

- Find free spaces in physical memory.
- Create new entries in GDT for the new process.



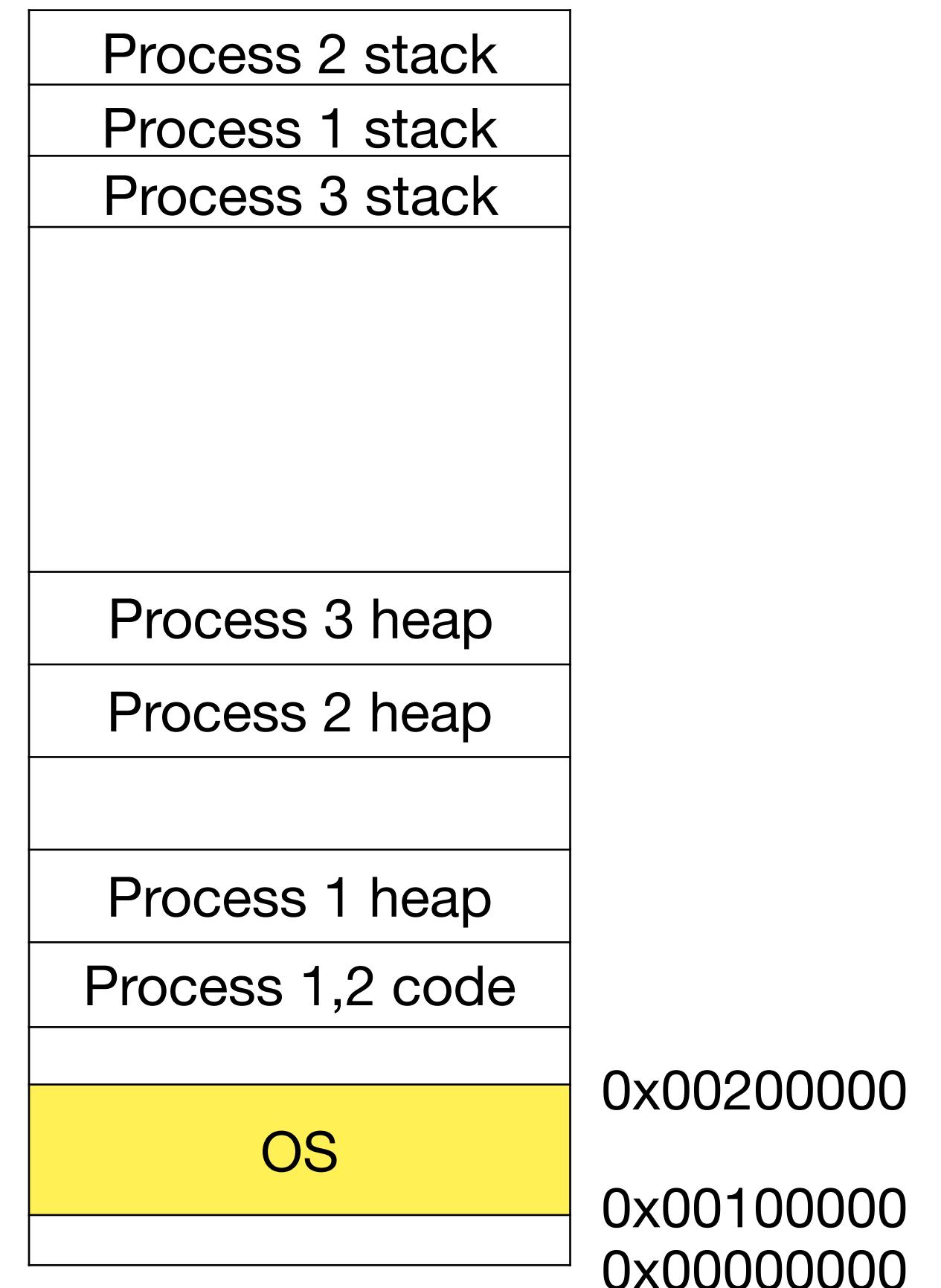
Allocating memory to a new process

- Find free spaces in physical memory.
- Create new entries in GDT for the new process.



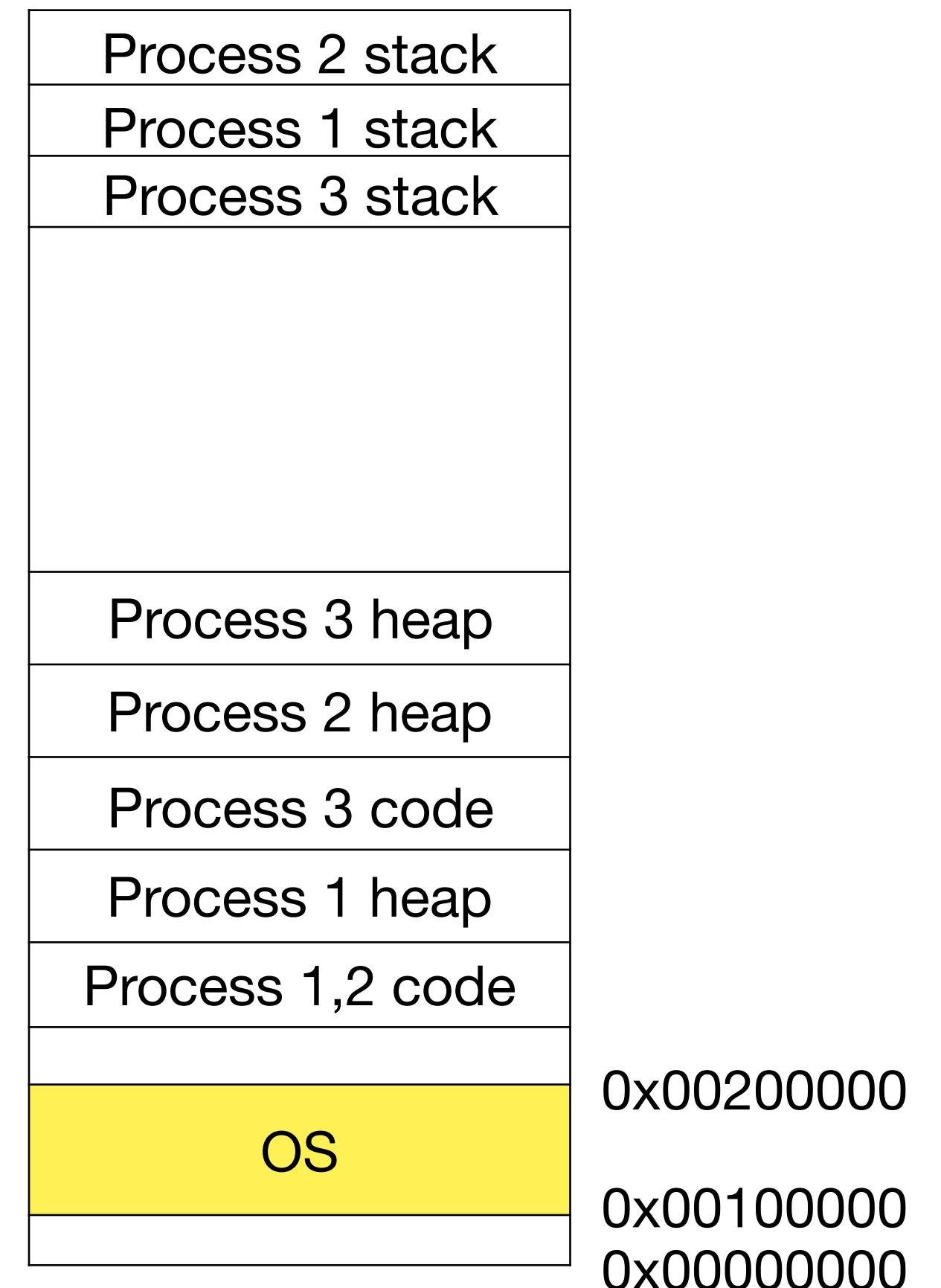
Allocating memory to a new process

- Find free spaces in physical memory.
- Create new entries in GDT for the new process.



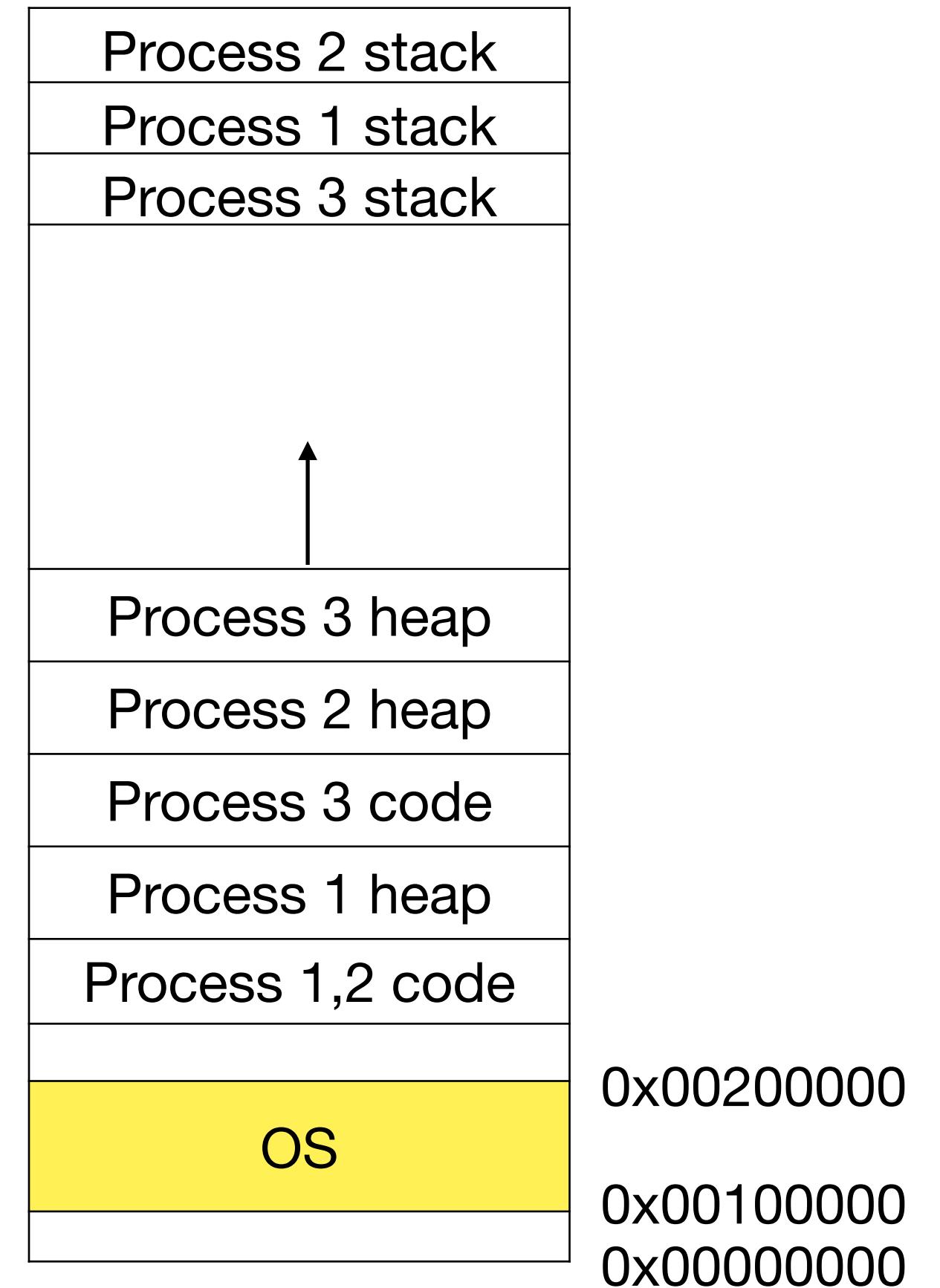
Allocating memory to a new process

- Find free spaces in physical memory.
- Create new entries in GDT for the new process.



Growing heap

- sbrk can grow heap segment



External fragmentation

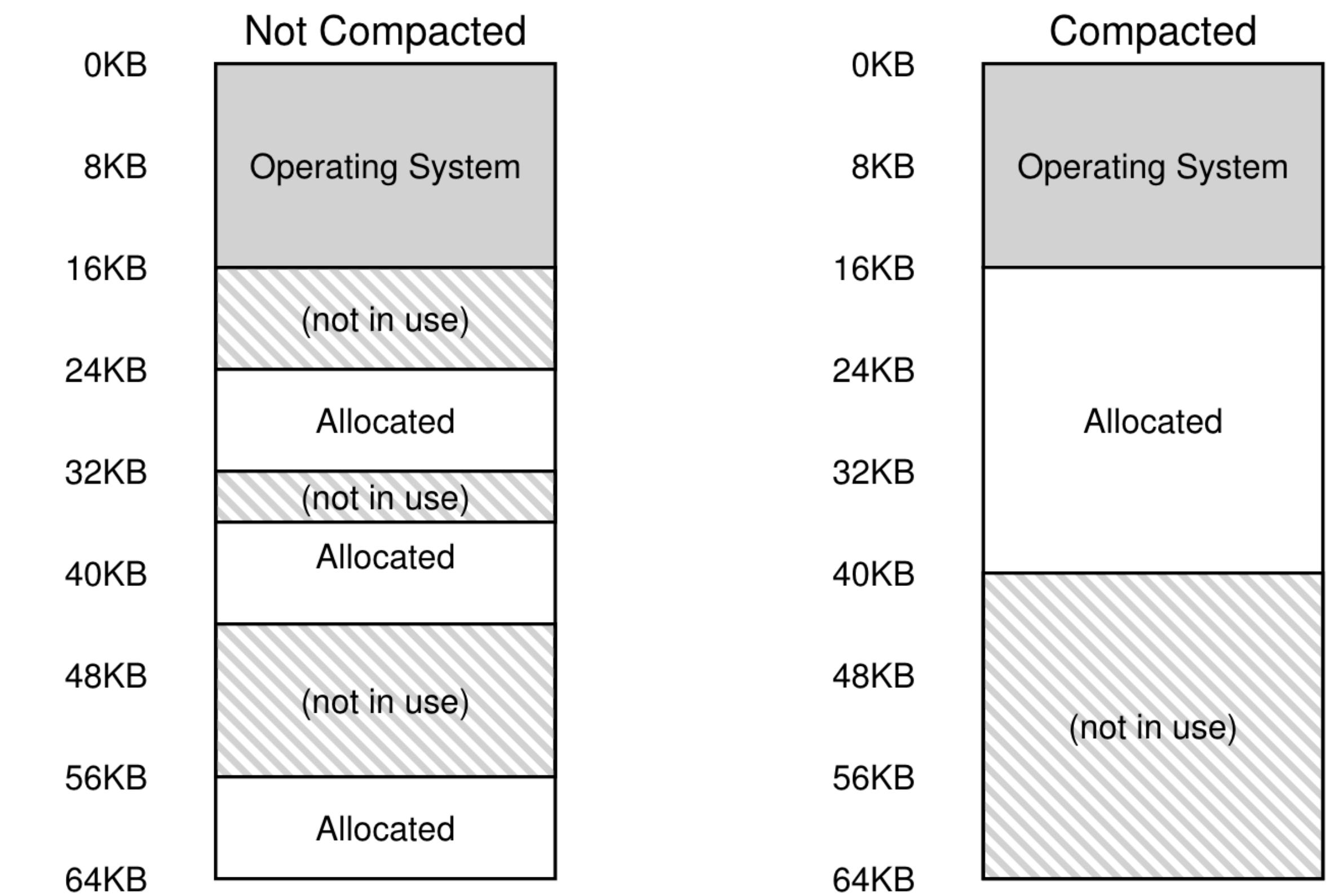


Figure 16.6: Non-compacted and Compacted Memory

External fragmentation

- After many processes start and exit, memory might become “fragmented” (similar to disk)
 - Example: cannot allocate 20 KB segment

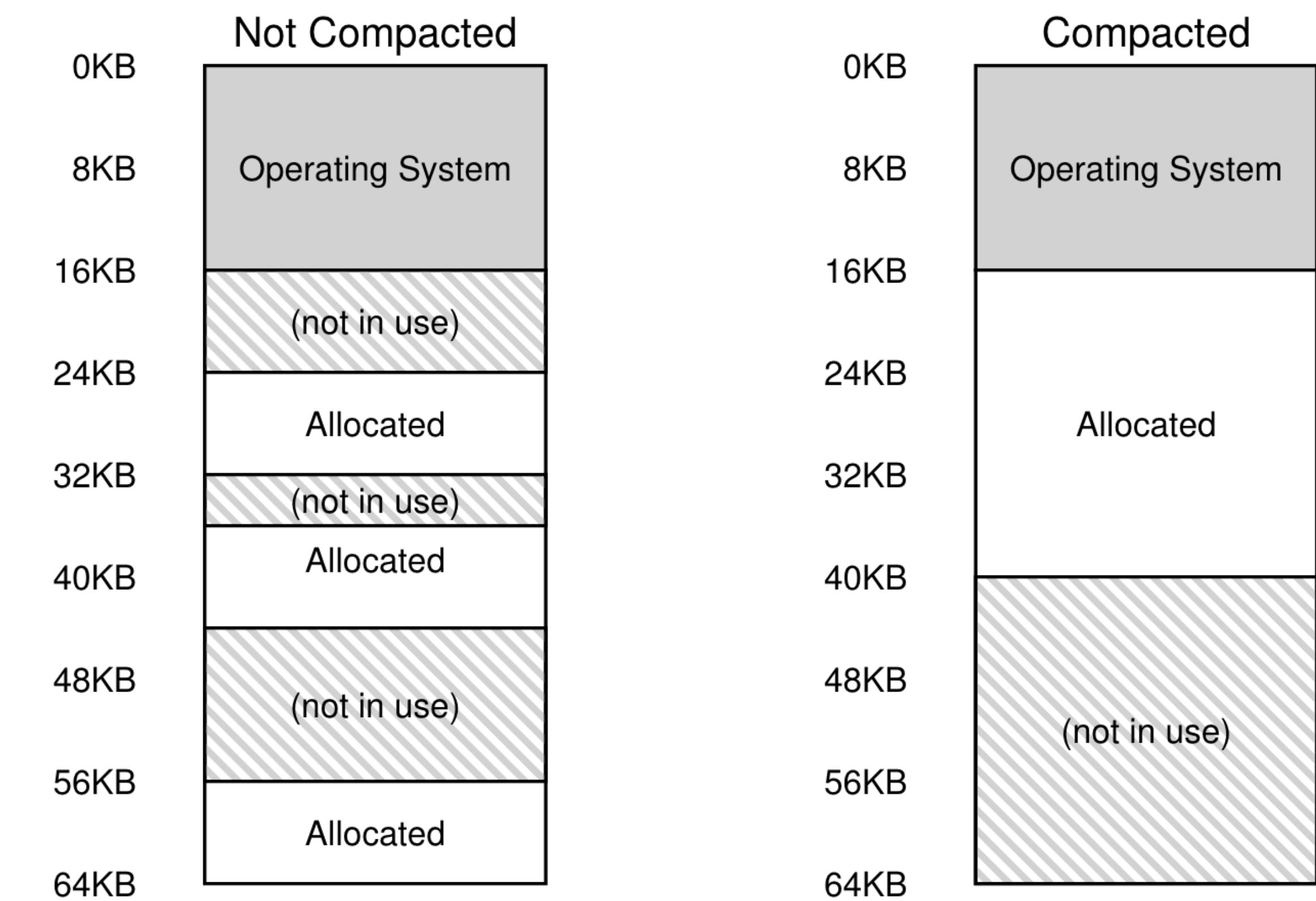


Figure 16.6: Non-compacted and Compacted Memory

External fragmentation

- After many processes start and exit, memory might become “fragmented” (similar to disk)
 - Example: cannot allocate 20 KB segment
 - Compaction: copy all allocated regions contiguously, update segment base and bound registers
 - Copying is expensive
 - Growing heap becomes not possible

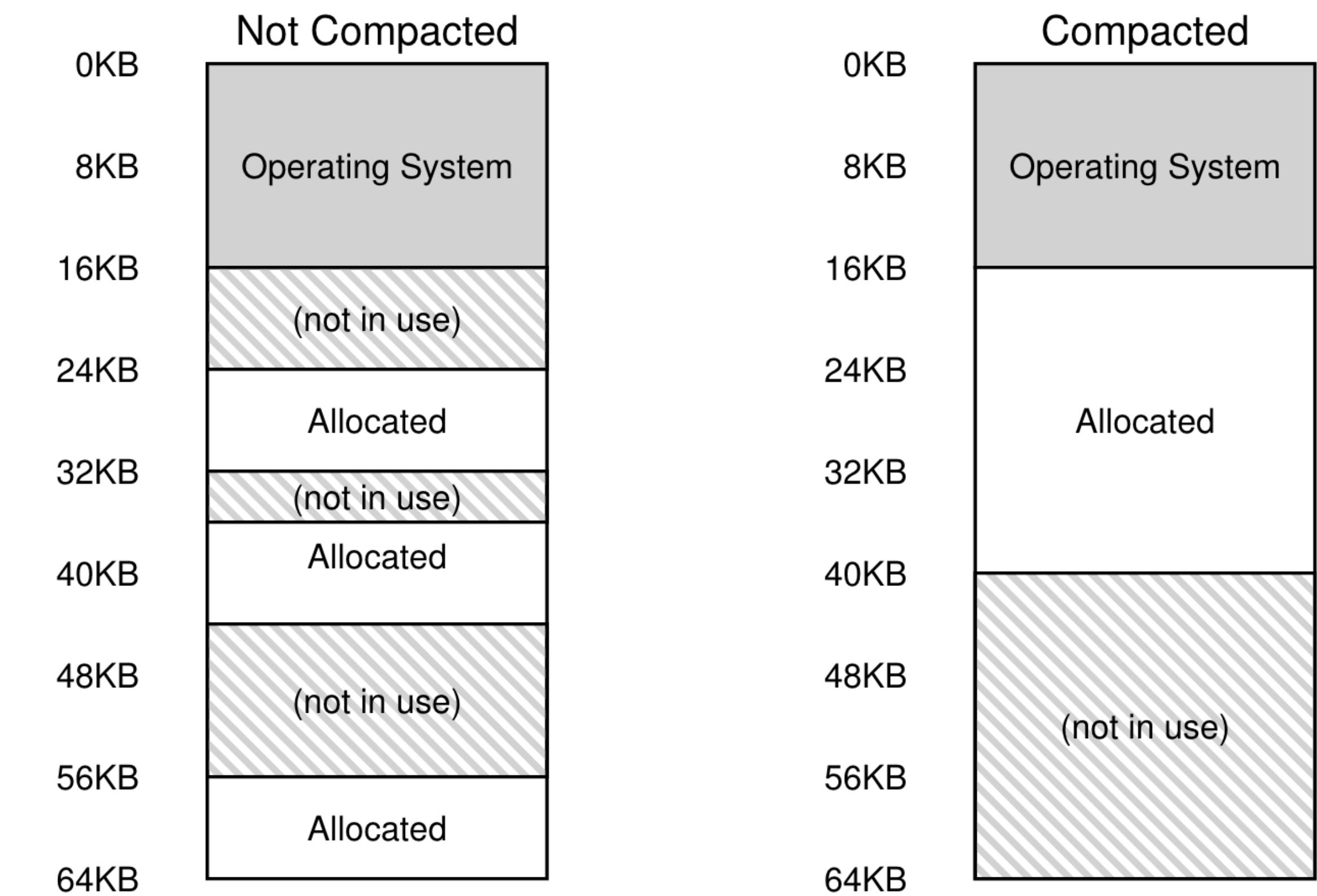
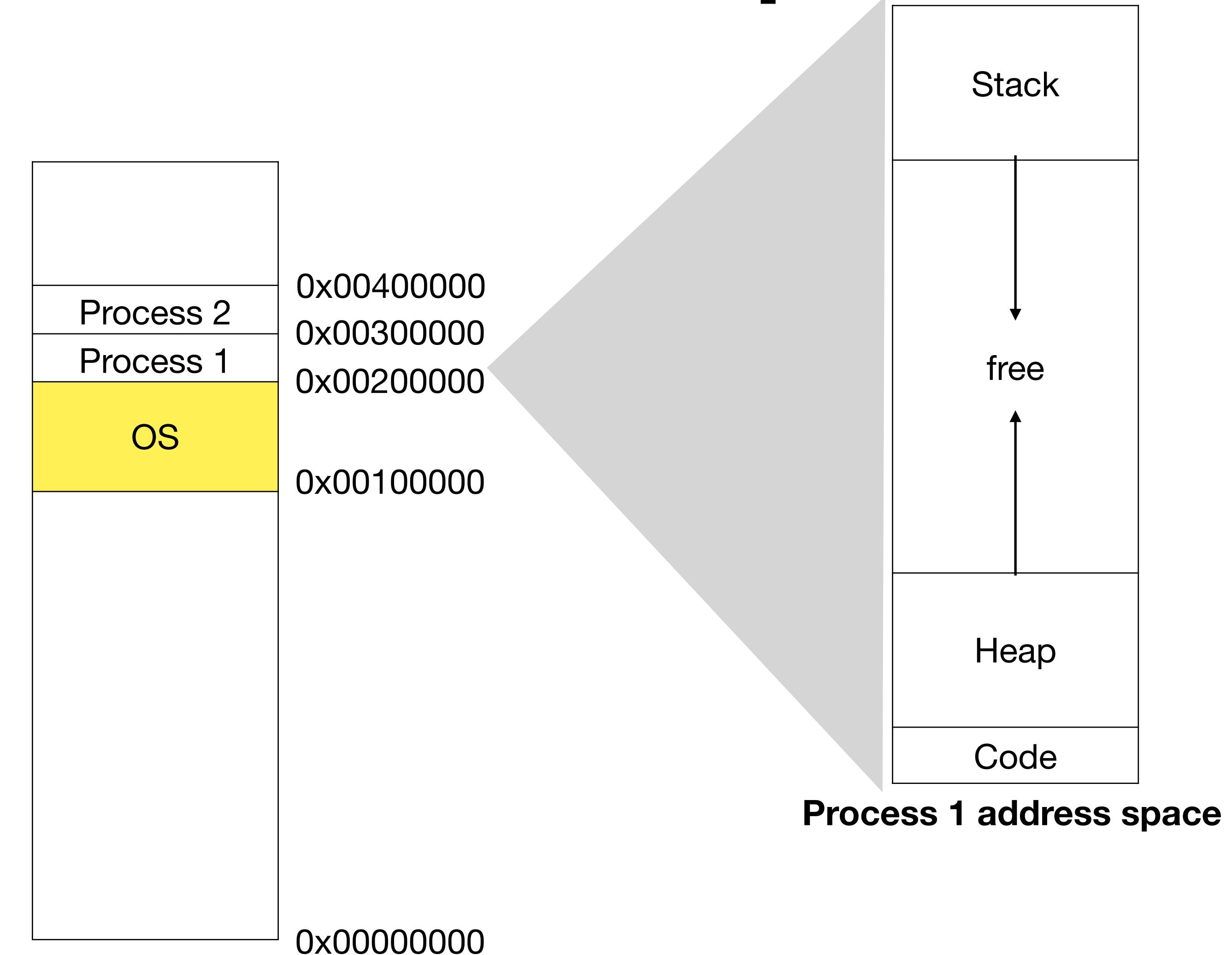


Figure 16.6: Non-compacted and Compacted Memory

Processes in action

xv6 Ch. 3: system calls, x86 protection, trap handlers

Memory isolation and address space



Protection

Protection

- Process cannot modify OS code since it is not in process' *address space*

Protection

- Process cannot modify OS code since it is not in process' *address space*
- Process cannot modify GDT and IDT entries since GDT, IDT are not in its address space

Protection

- Process cannot modify OS code since it is not in process' *address space*
- Process cannot modify GDT and IDT entries since GDT, IDT are not in its address space
- What stops process from calling lgdt and lidt instructions?

Protection

- Process cannot modify OS code since it is not in process' *address space*
- Process cannot modify GDT and IDT entries since GDT, IDT are not in its address space
- What stops process from calling lgdt and lidt instructions?

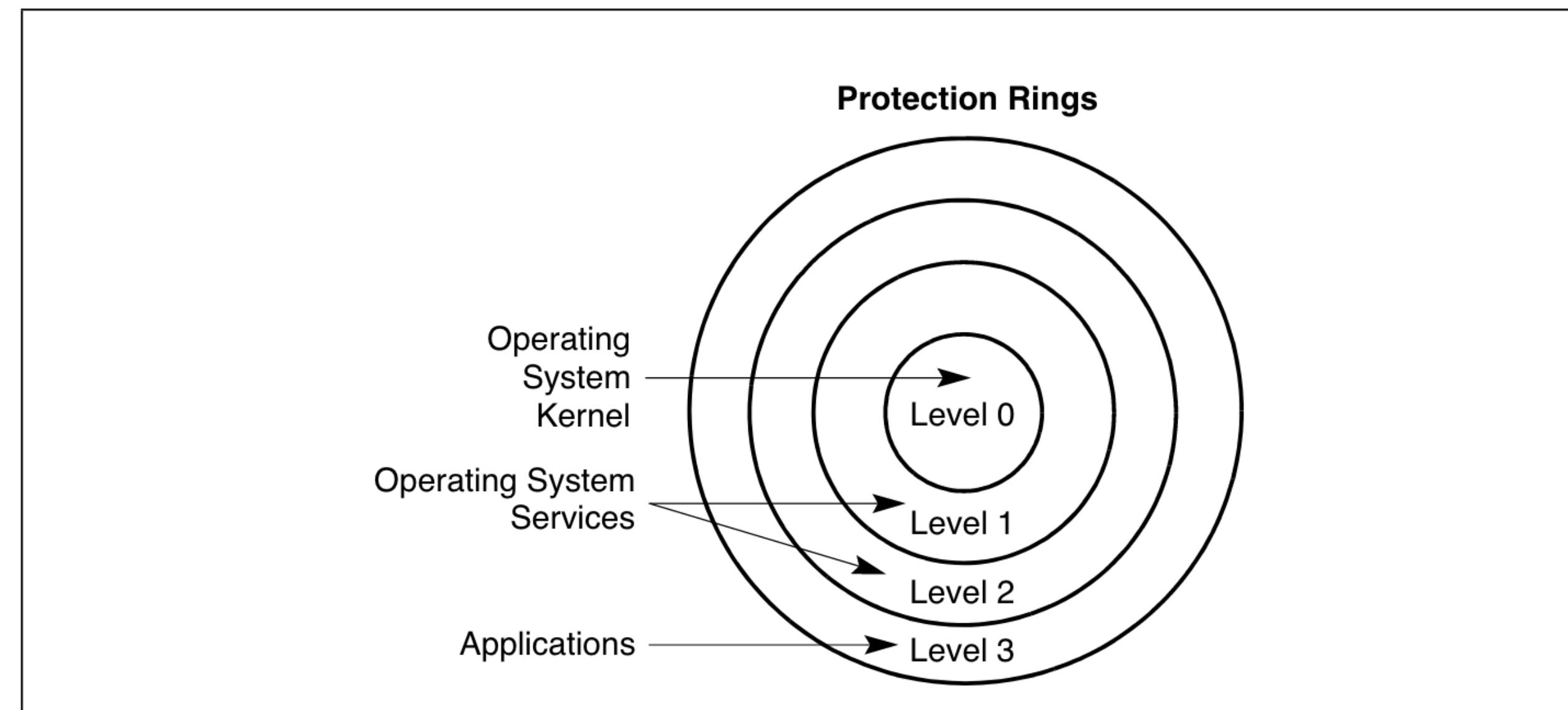


Figure 5-3. Protection Rings

Protection

- Process cannot modify OS code since it is not in process' *address space*
- Process cannot modify GDT and IDT entries since GDT, IDT are not in its address space
- What stops process from calling lgdt and lidt instructions?
 - Ring 0: kernel mode. Ring 3: user mode

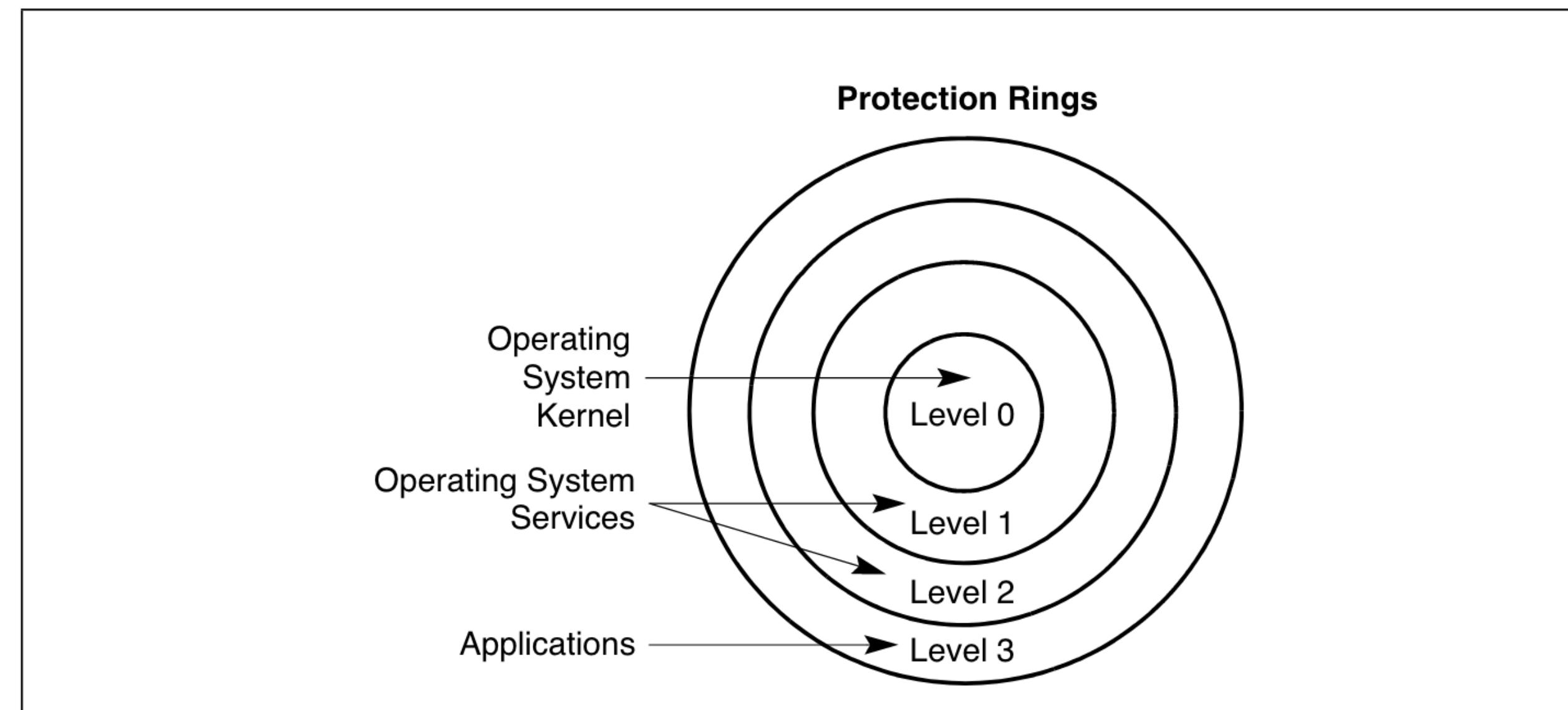


Figure 5-3. Protection Rings

Protection

- Process cannot modify OS code since it is not in process' *address space*
- Process cannot modify GDT and IDT entries since GDT, IDT are not in its address space
- What stops process from calling lgdt and lidt instructions?
 - Ring 0: kernel mode. Ring 3: user mode
 - lgdt and lidt are *privileged instructions* only callable from “ring 0”

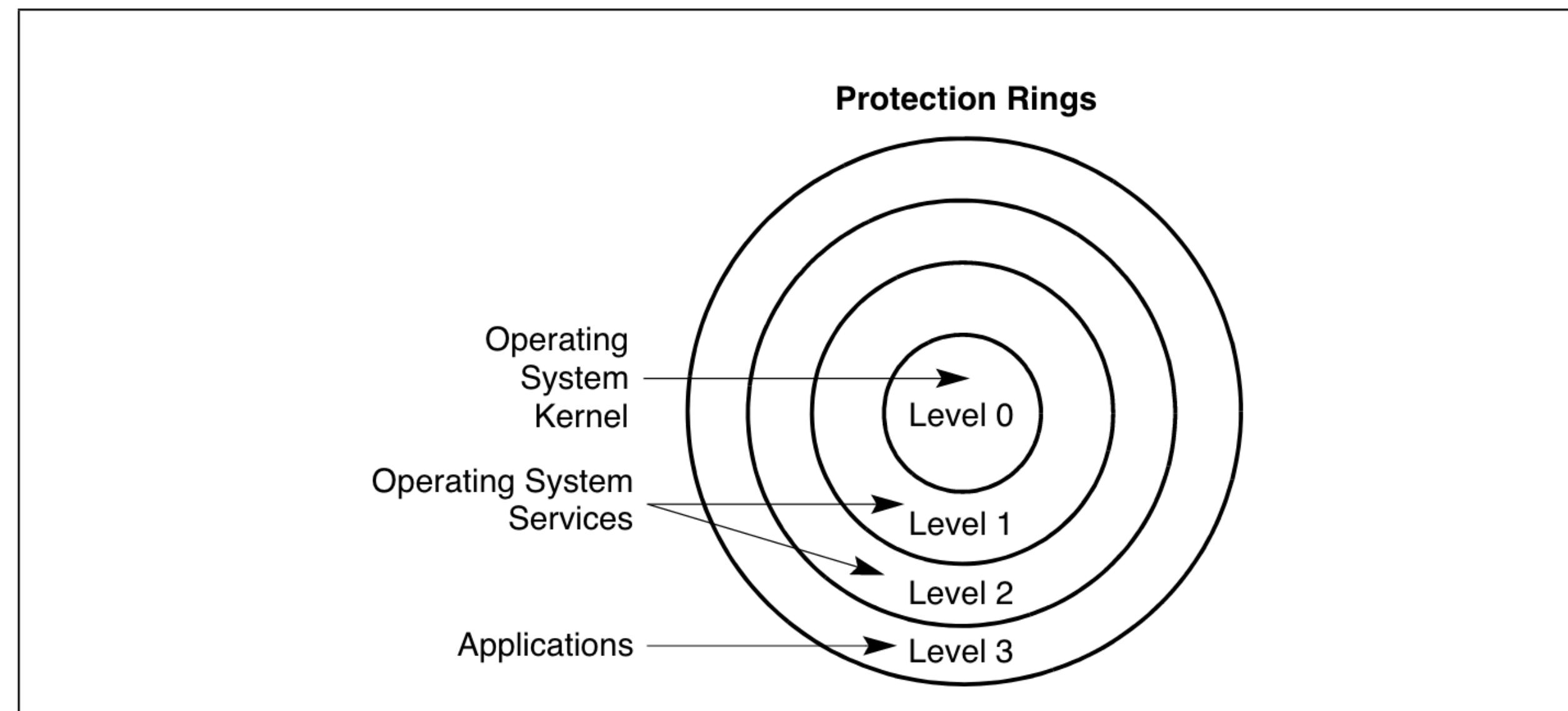


Figure 5-3. Protection Rings

How does hardware know the current privilege level?

How does hardware know the current privilege level?

- Two LSBs of %cs determine “current privilege level” (CPL)

How does hardware know the current privilege level?

- Two LSBs of %cs determine “current privilege level” (CPL)
- Two LSBs of other segment selectors specify “required privilege level” (RPL)

How does hardware know the current privilege level?

- Two LSBs of %cs determine “current privilege level” (CPL)
- Two LSBs of other segment selectors specify “required privilege level” (RPL)

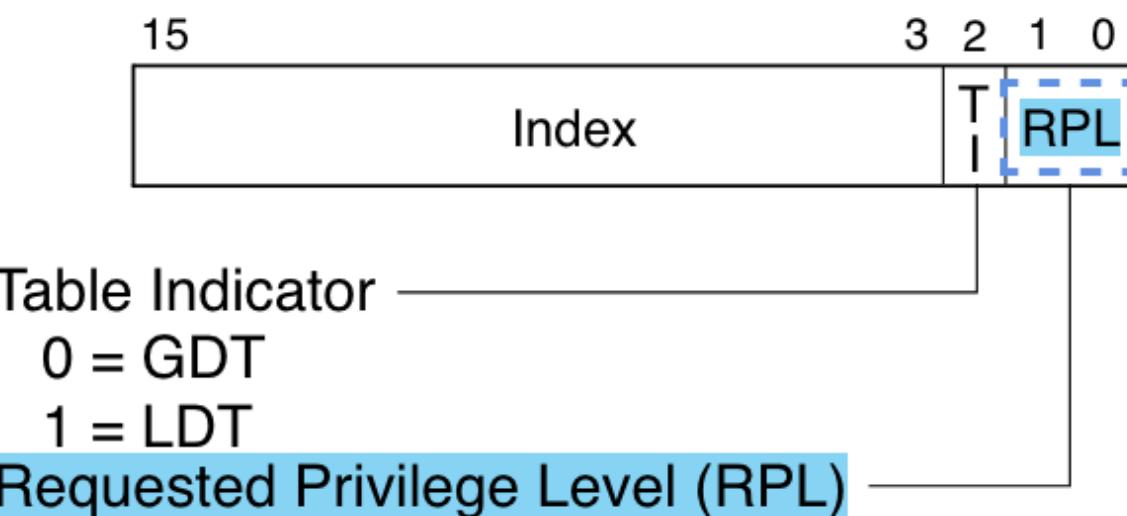


Figure 3-6. Segment Selector

How does hardware know the current privilege level?

- Two LSBs of %cs determine “current privilege level” (CPL)
- Two LSBs of other segment selectors specify “required privilege level” (RPL)
- Segment descriptor specifies “descriptor privilege level” (DPL)

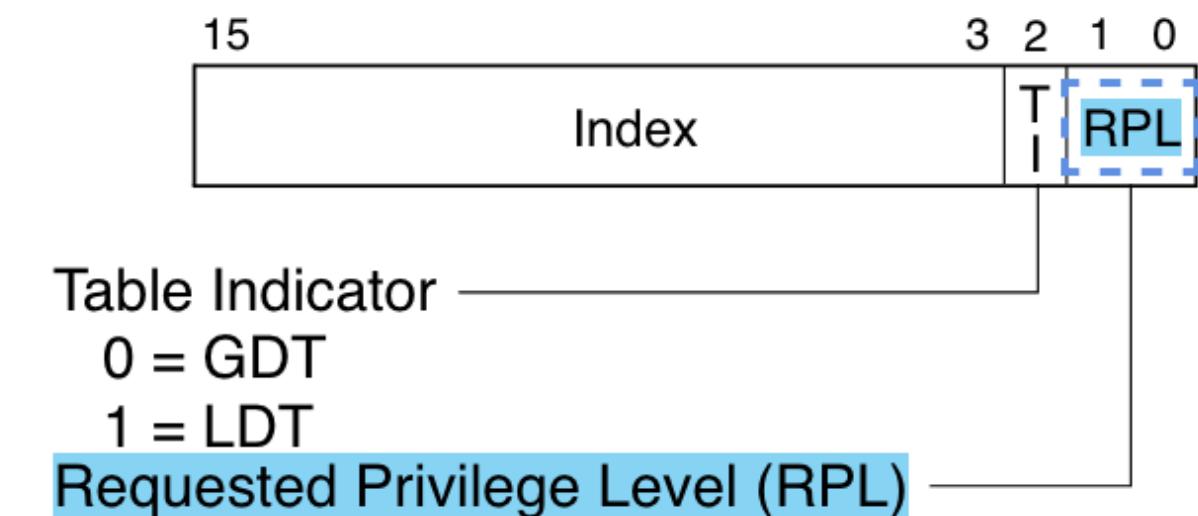
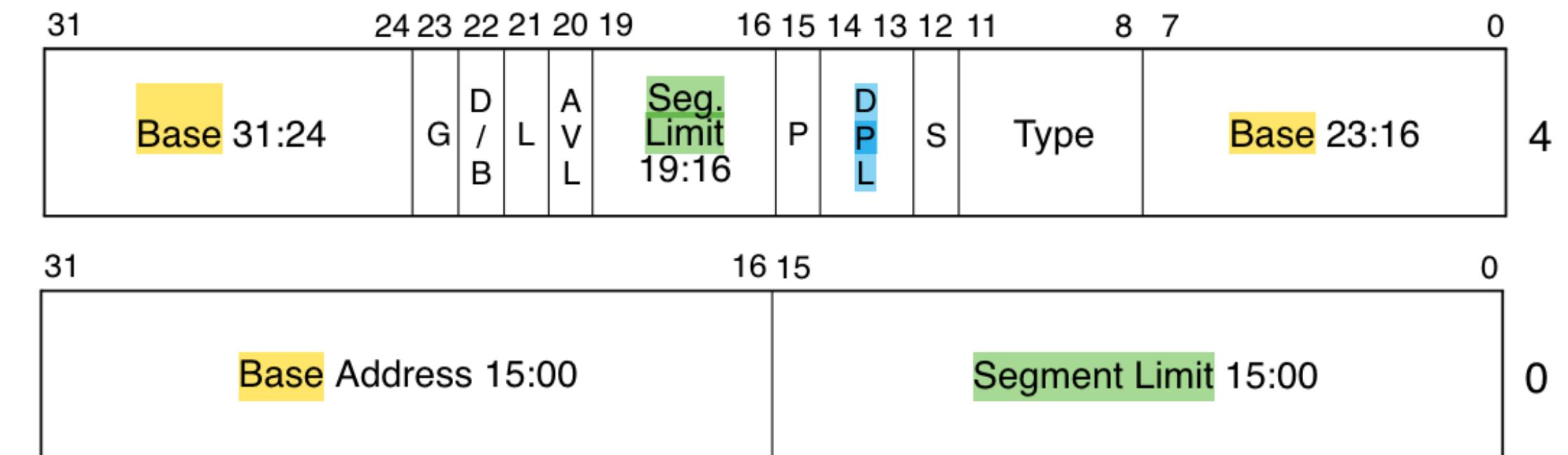


Figure 3-6. Segment Selector

How does hardware know the current privilege level?

- Two LSBs of %cs determine “current privilege level” (CPL)
- Two LSBs of other segment selectors specify “required privilege level” (RPL)
- Segment descriptor specifies “descriptor privilege level” (DPL)



Legend:
L — 64-bit code segment (IA-32e mode only)
AVL — Available for use by system software
BASE — Segment base address
D/B — Default operation size (0 = 16-bit segment; 1 = 32-bit segment)
DPL — Descriptor privilege level
G — Granularity
LIMIT — Segment Limit
P — Segment present
S — Descriptor type (0 = system; 1 = code or data)
TYPE — Segment type

Figure 3-8. Segment Descriptor

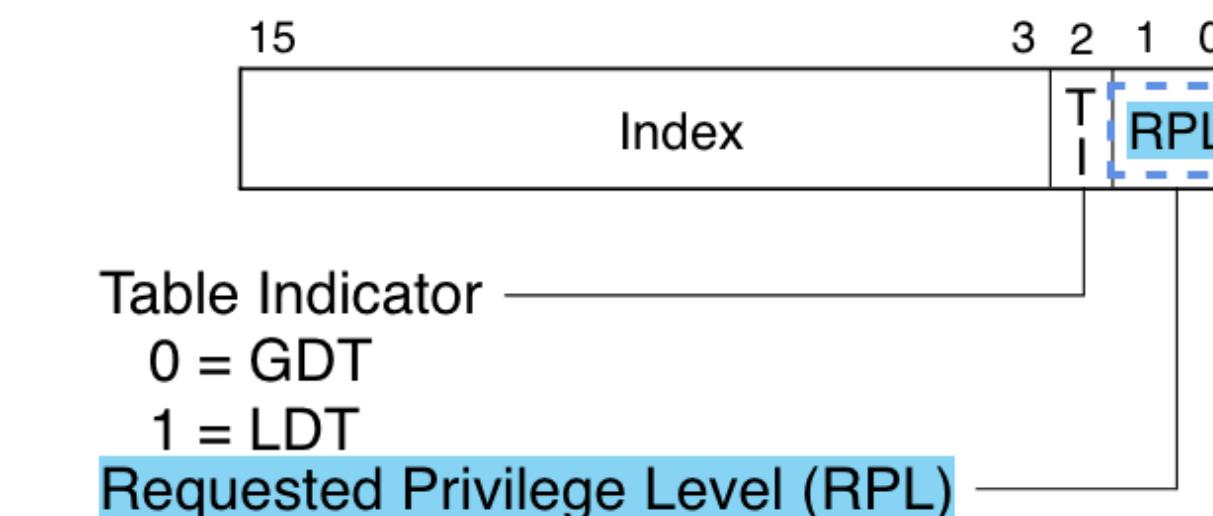


Figure 3-6. Segment Selector

How does hardware enforce the current privilege level?

How does hardware enforce the current privilege level?

- Privileged instructions can only be called when CPL=0

How does hardware enforce the current privilege level?

- Privileged instructions can only be called when CPL=0
- Programs can change their segment selectors

How does hardware enforce the current privilege level?

- Privileged instructions can only be called when CPL=0
- Programs can change their segment selectors
 - Programs cannot lower their CPL directly

How does hardware enforce the current privilege level?

- Privileged instructions can only be called when CPL=0
- Programs can change their segment selectors
 - Programs cannot lower their CPL directly
 - INT instruction causes a software interrupt to set ring = 0

How does hardware enforce the current privilege level?

- Privileged instructions can only be called when CPL=0
- Programs can change their segment selectors
 - Programs cannot lower their CPL directly
 - INT instruction causes a software interrupt to set ring = 0
 - CPL <= DPL and RPL <= DPL

How does hardware enforce the current privilege level?

- Privileged instructions can only be called when CPL=0
- Programs can change their segment selectors
 - Programs cannot lower their CPL directly
 - INT instruction causes a software interrupt to set ring = 0
 - CPL <= DPL and RPL <= DPL

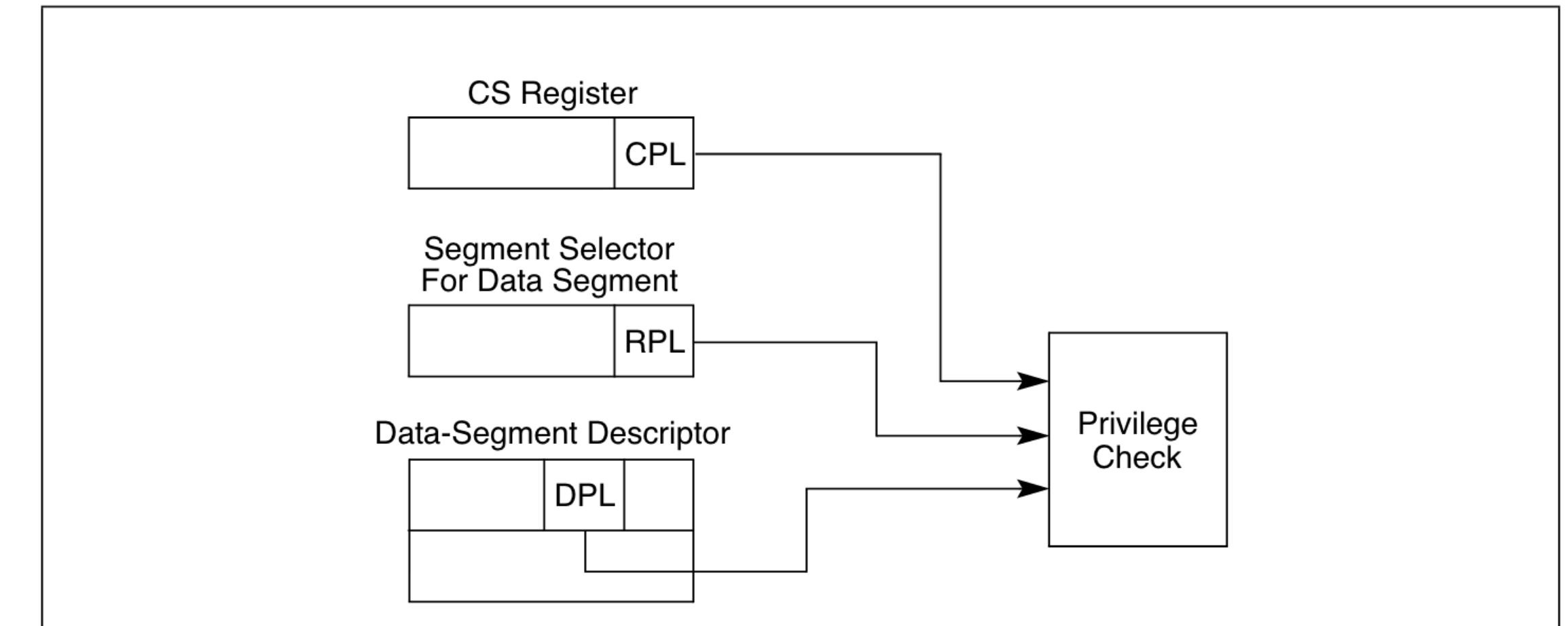


Figure 5-4. Privilege Check for Data Access

How does hardware enforce the current privilege level?

- Privileged instructions can only be called when CPL=0
- Programs can change their segment selectors
 - Programs cannot lower their CPL directly
 - INT instruction causes a software interrupt to set ring = 0
 - CPL <= DPL and RPL <= DPL

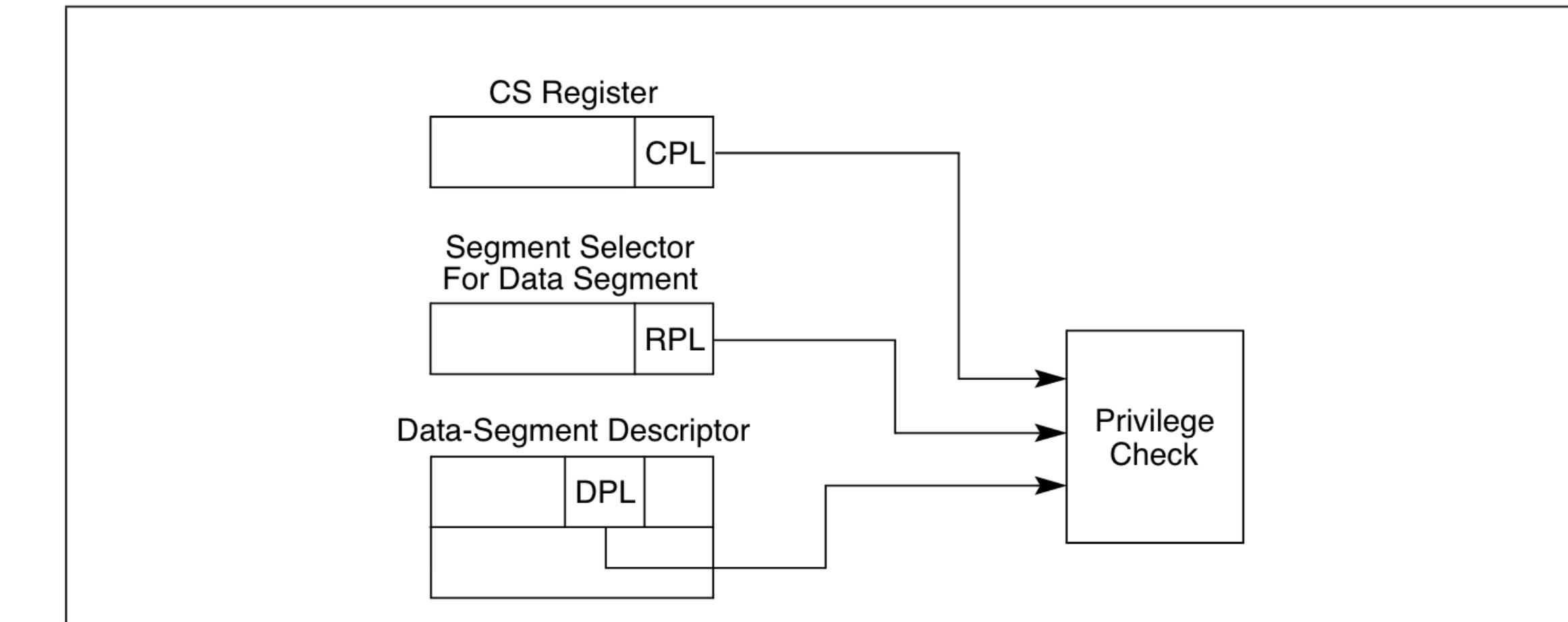


Figure 5-4. Privilege Check for Data Access

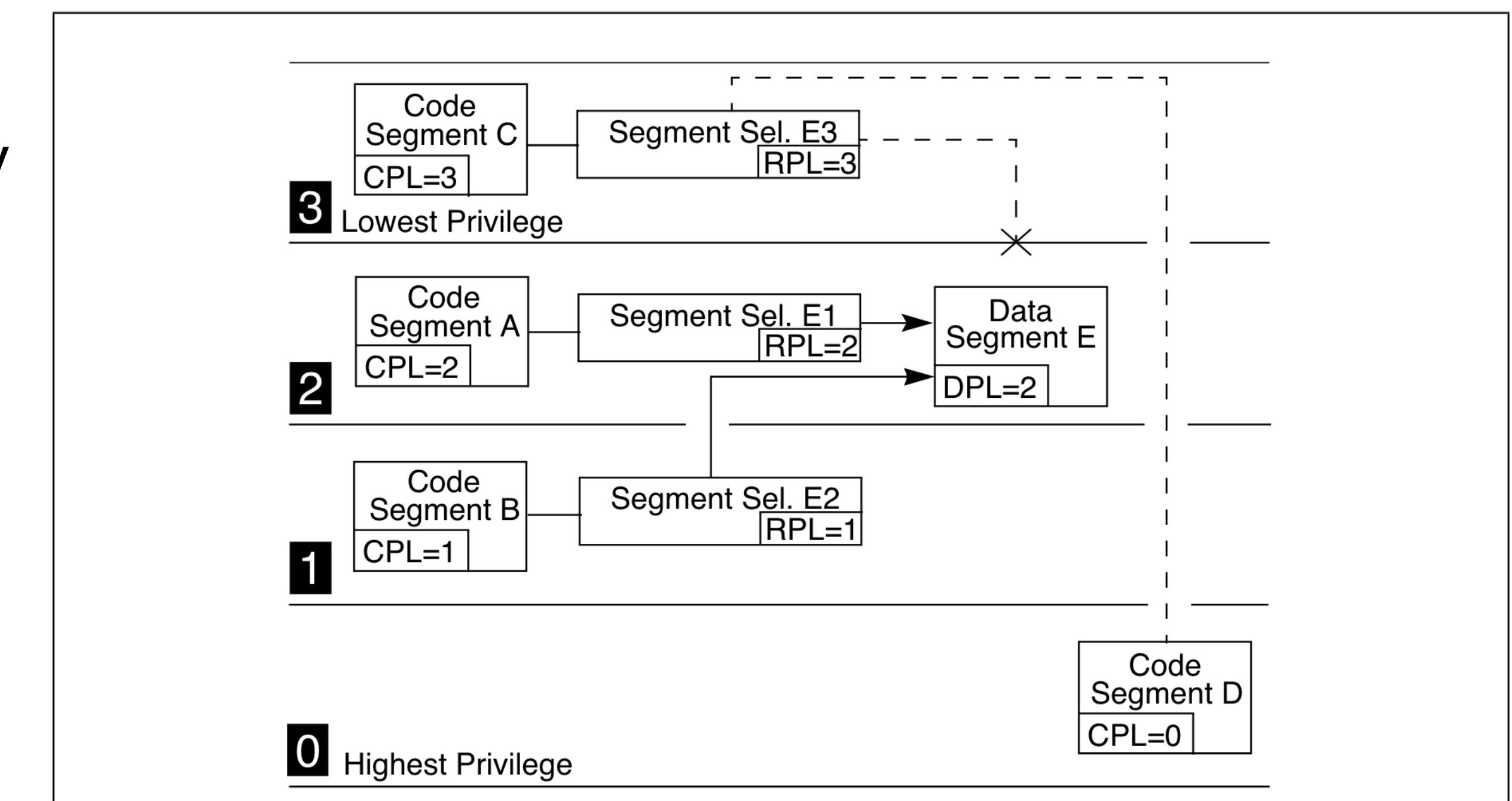
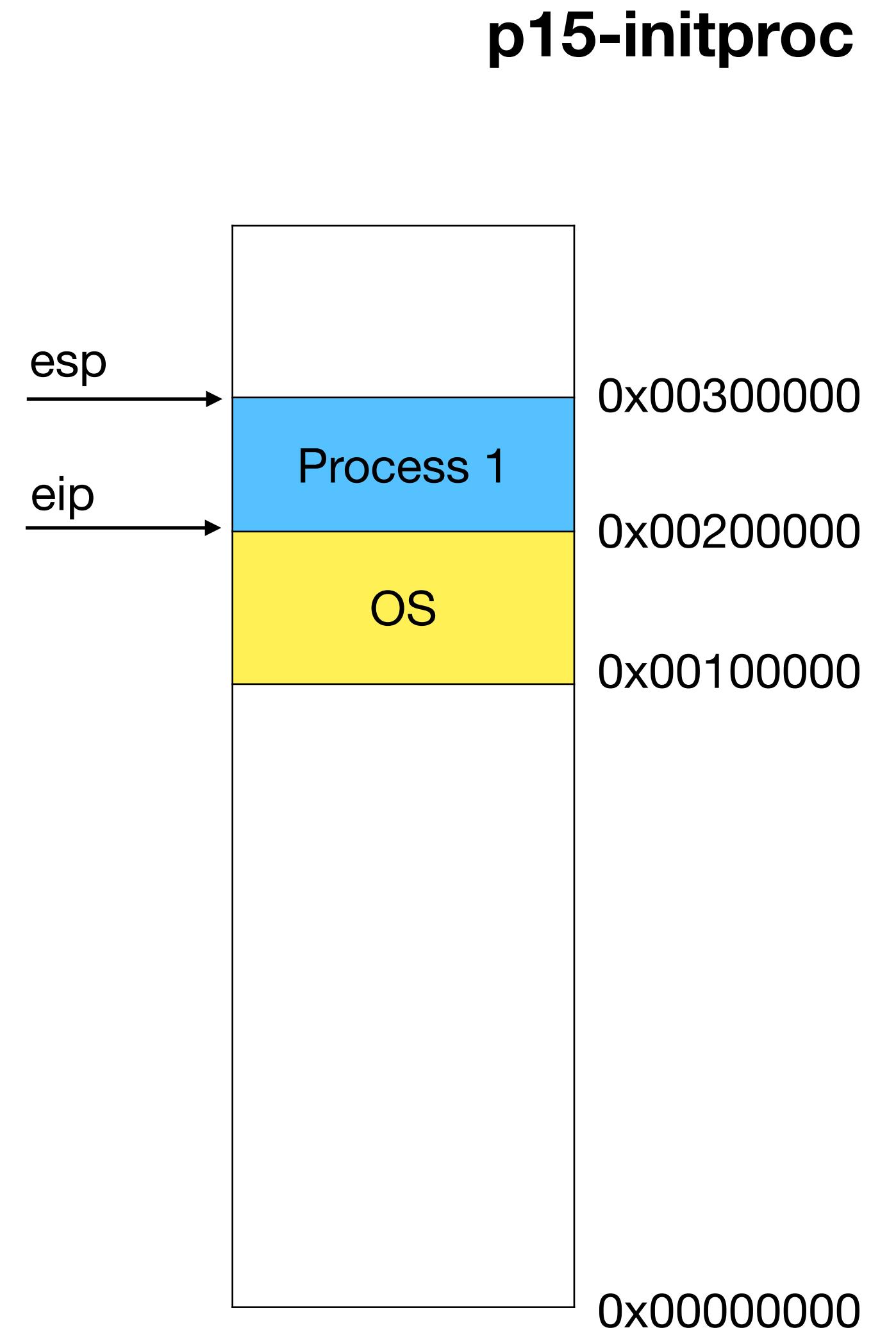


Figure 5-5. Examples of Accessing Data Segments From Various Privilege Levels

Setting up our first process in xv6!

- main.c calls seginit, then pinit, then scheduler.
- seginit in vm.c creates code (UCODE) and data segments (UDATA) from STARTPROC (2MB) to 3MB. Flat memory model!
- pinit in proc.c copies program binary to STARTPROC and sets cs,ds,ss,es to the program's segments. Last two bits are set to DPL_USER. Sets eip=0, esp=1MB. Enables interrupt flag in eflags. proc.c maintains list of processes in ptable. We are just starting one process for now.
- scheduler in proc.c selects RUNNABLE process and switches to it.
- Process cannot change GDT entries, IDT entries since they are not in address space of the process! Process cannot call lgdt, lidt since it will run in ring 3.

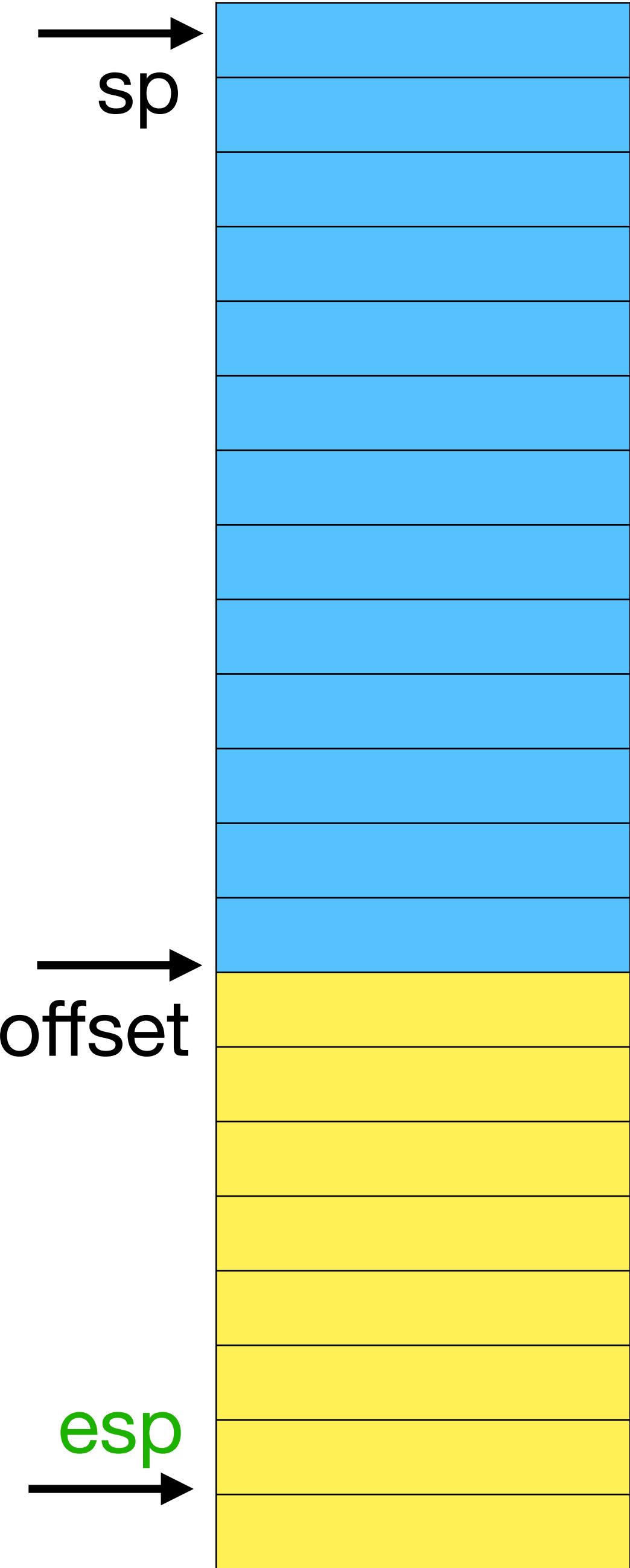


Understanding swtch

```
pinit(){  
    p = allocproc();  
    memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}  
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

eip

```
scheduler() {  
    ...  
    swtch(p->context);  
}  
swtch:  
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret  
.globl trapret  
trapret:  
    popal  
    popl %gs  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```



Understanding swtch

```
pinit(){  
    p = allocproc();  
    memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}
```

```
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
eip → sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

```
scheduler() {  
    ...  
    swtch(p->context);  
}  
  
.globl trapret  
trapret:  
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret  
  
    .globl _start  
_start:  
    popal  
    popl %gs  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```



Understanding swtch

```
pinit(){  
    p = allocproc();  
    memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}  
allocproc() {
```

eip
→

```
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;
```

}

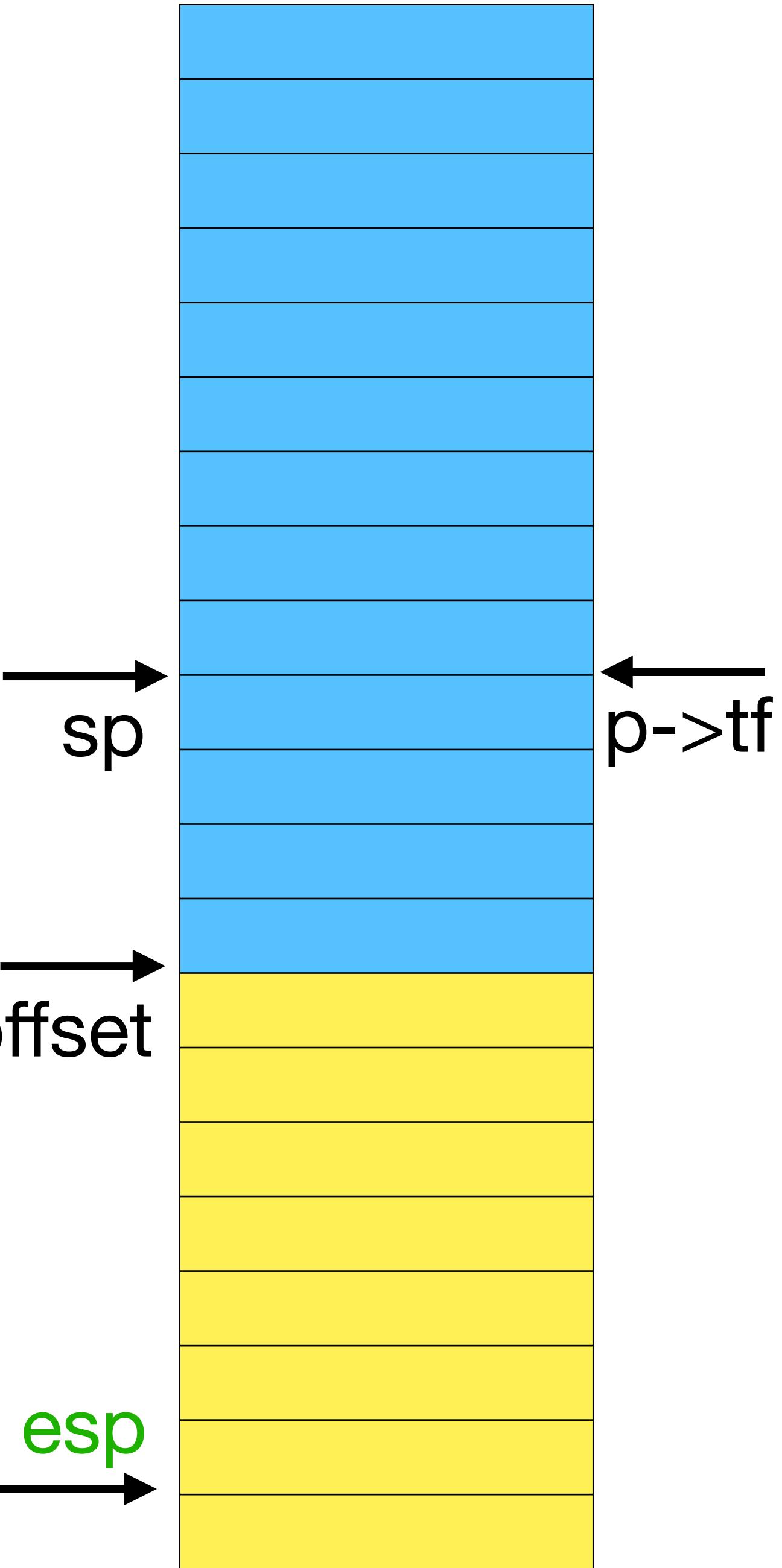
```
scheduler() {  
    ...  
    swtch(p->context);  
}  
swtch:  
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret  
.globl trapret  
trapret:  
    popal  
    popl %gs  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```



Understanding swtch

```
pinit(){  
    p = allocproc();  
  
    memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}  
  
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
  
    eip → p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->tf;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

```
scheduler() {  
    ...  
    swtch(p->context);  
}  
  
swtch:  
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret  
  
.globl trapret  
trapret:  
    popal  
    popl %gs  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```

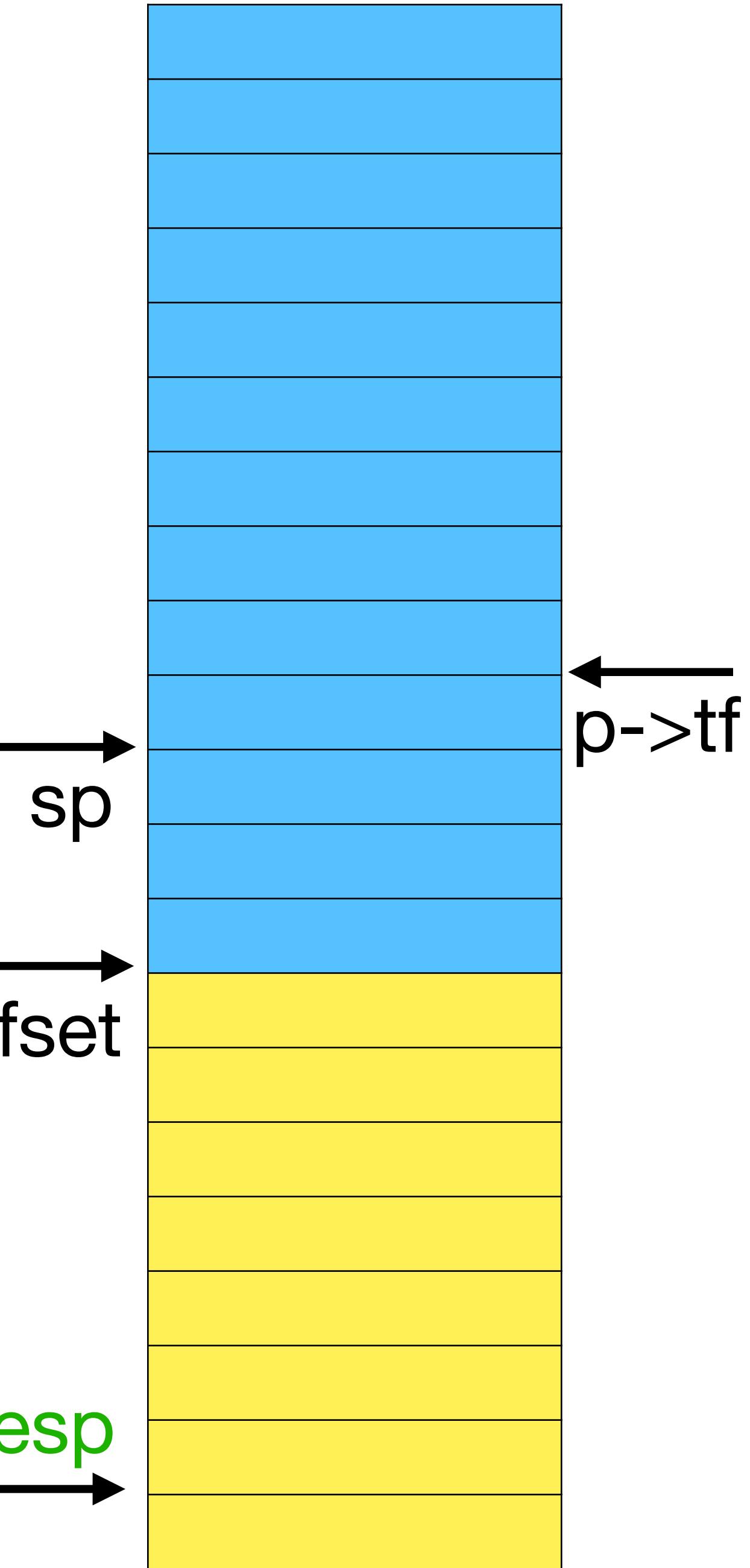


Understanding swtch

```
pinit(){  
    p = allocproc();  
  
    memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}
```

```
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

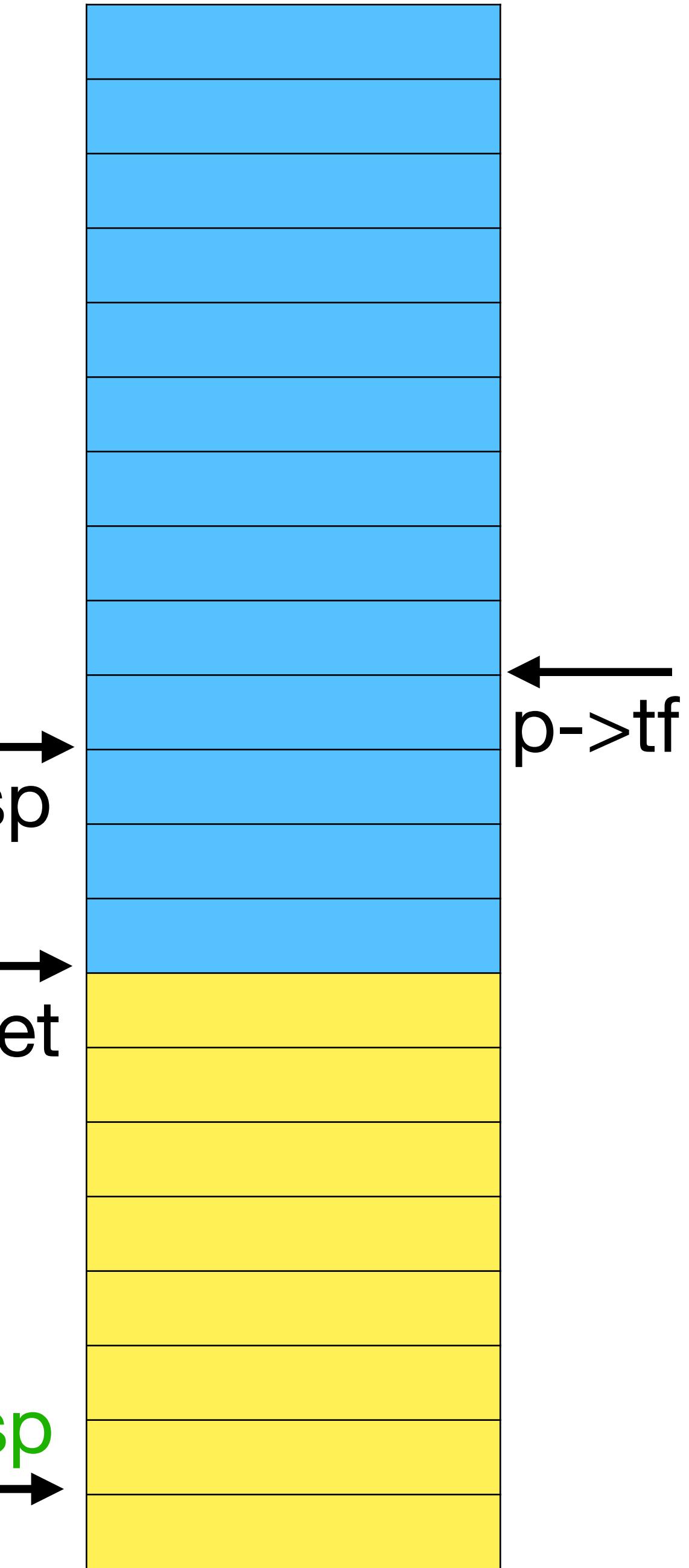
```
scheduler() {  
    ...  
    swtch(p->context);  
}  
  
swtch:  
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret  
  
.globl trapret  
trapret:  
    popal  
    popl %gs  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```



Understanding swtch

```
pinit(){  
    p = allocproc();  
  
    memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}  
  
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    eip → p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

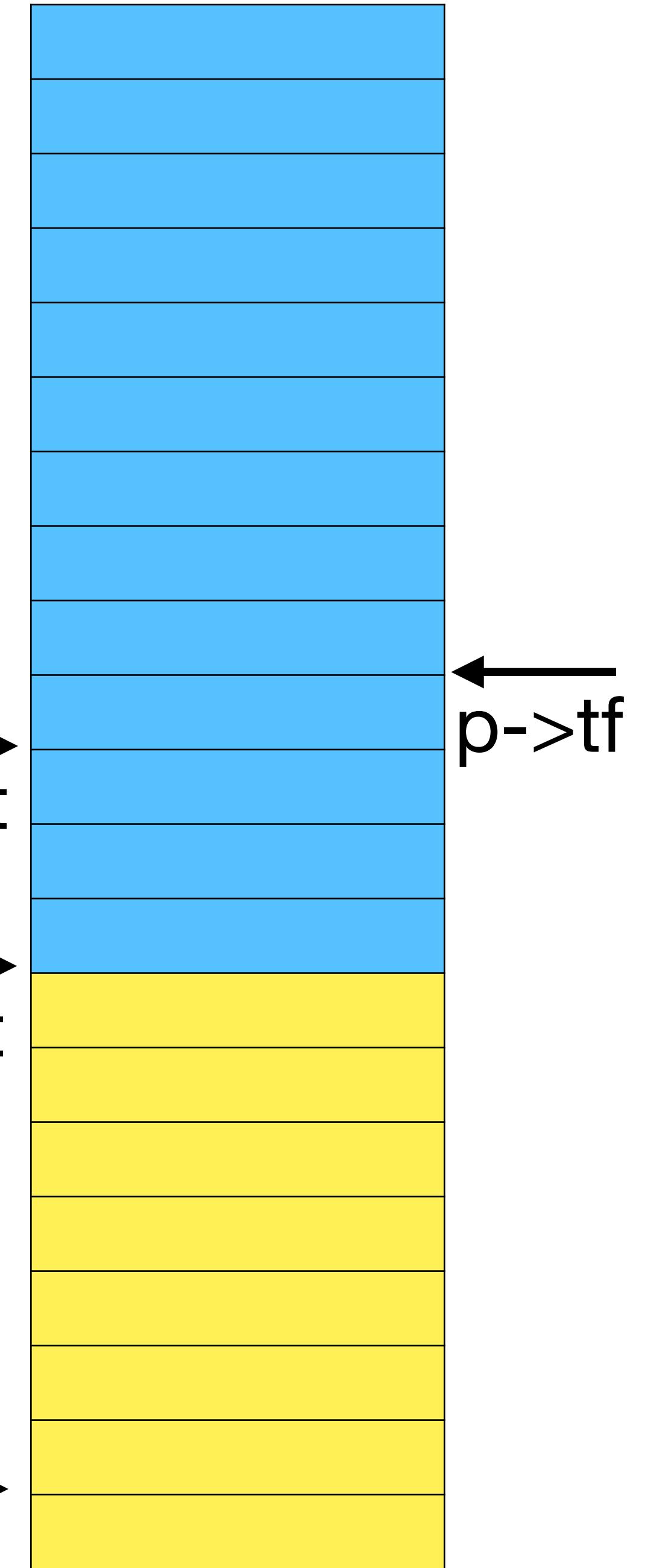
```
scheduler() {  
    ...  
    swtch(p->context);  
}  
  
swtch:  
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret  
  
.globl trapret  
trapret:  
    popal  
    popl %gs  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```



Understanding swtch

```
pinit(){  
    p = allocproc();  
  
    memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}  
  
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    eip → p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

```
scheduler() {  
    ...  
    swtch(p->context);  
}  
  
swtch:  
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret  
  
.globl trapret p->context  
trapret:  
    popal  
    popl %gs p->offset  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```



Understanding swtch

```
pinit(){  
    p = allocproc();  
    memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}
```

```
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

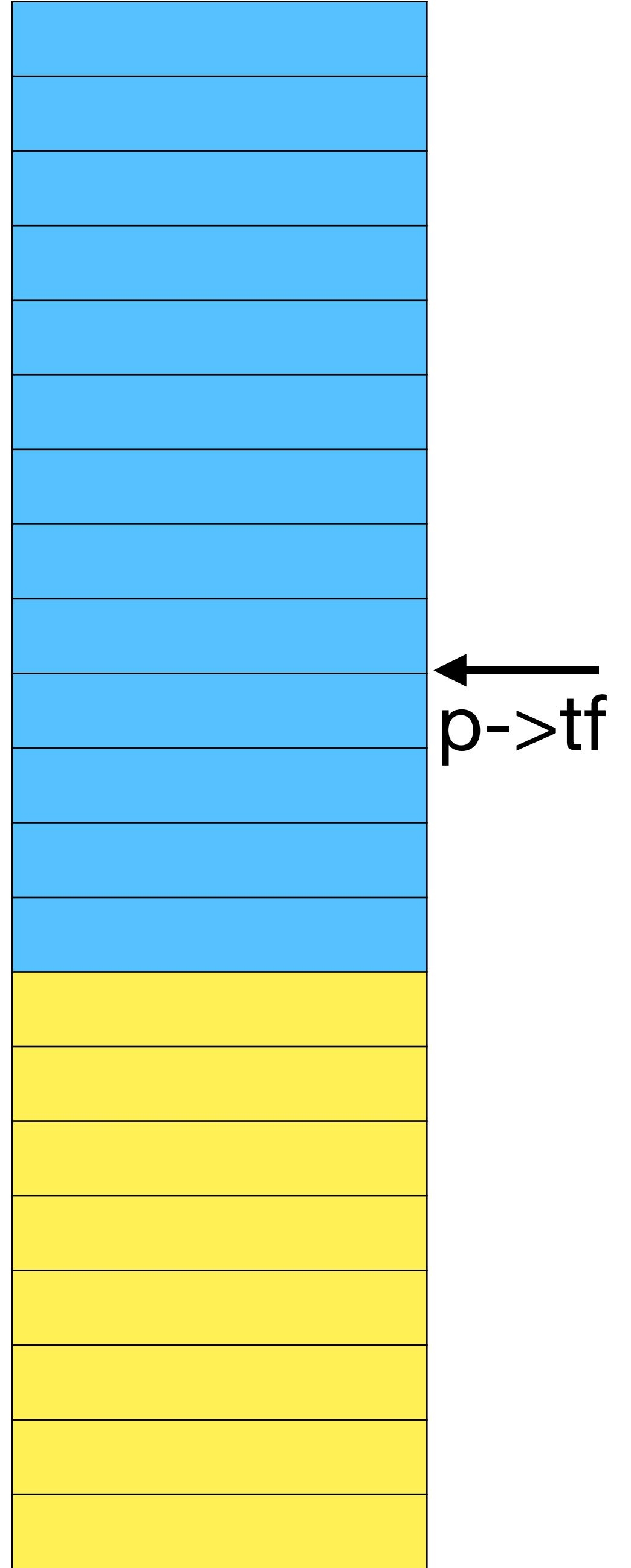
```
scheduler() {  
    ...  
    swtch(p->context);  
}
```

```
swtch:  
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret
```

```
.globl trapret    p->context
```

```
trapret:
```

```
    popal  
    popl %gs  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```



Understanding swtch

```
pinit(){  
    p = allocproc();  
  
    memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}  
}
```

```
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

```
scheduler() {
```

```
...
```

```
    swtch(p->context);  
}
```

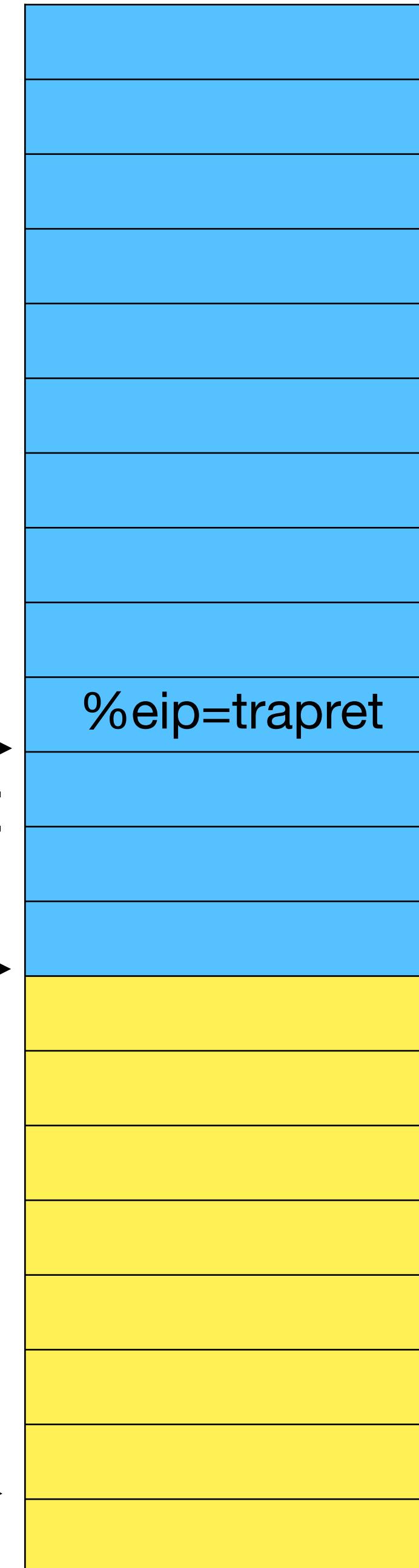
```
swtch:
```

```
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret
```

```
.globl trapret    p->context
```

```
trapret:
```

```
    popal  
    popl %gs  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```



eip

→ p->context->eip = (uint)trapret;

Understanding swtch

```
pinit(){  
    p = allocproc();  
    memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}  
allocproc() {
```

```
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    → return p;  
}
```

```
scheduler() {
```

```
...
```

```
    swtch(p->context);  
}
```

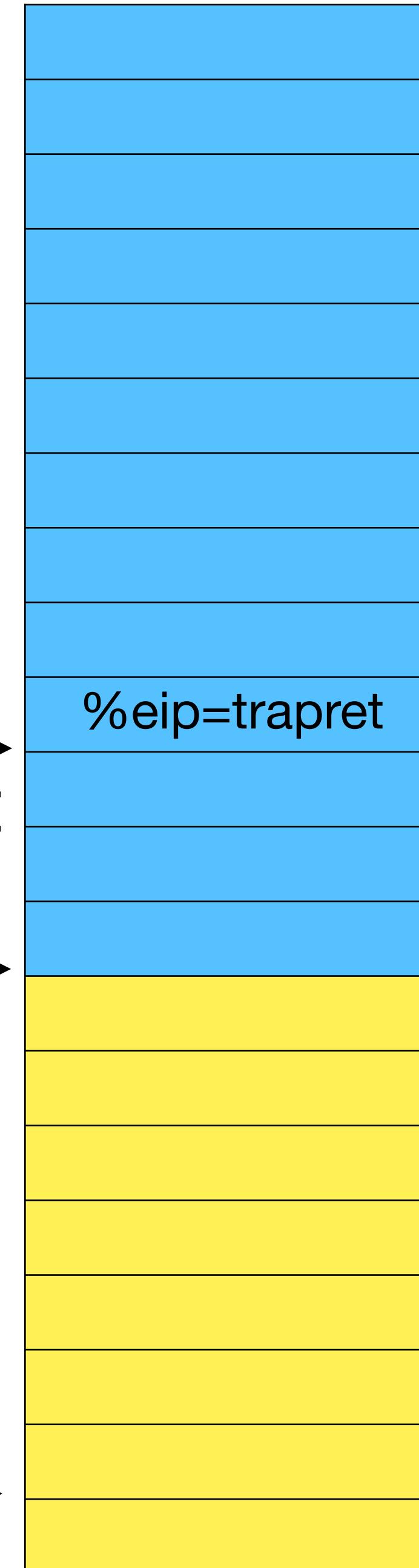
```
swtch:
```

```
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret
```

```
.globl trapret → p->context
```

```
trapret:
```

```
    popal → p->offset  
    popl %gs  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```



Understanding swtch

```
pinit(){  
    p = allocproc();  
    eip → memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}
```

```
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

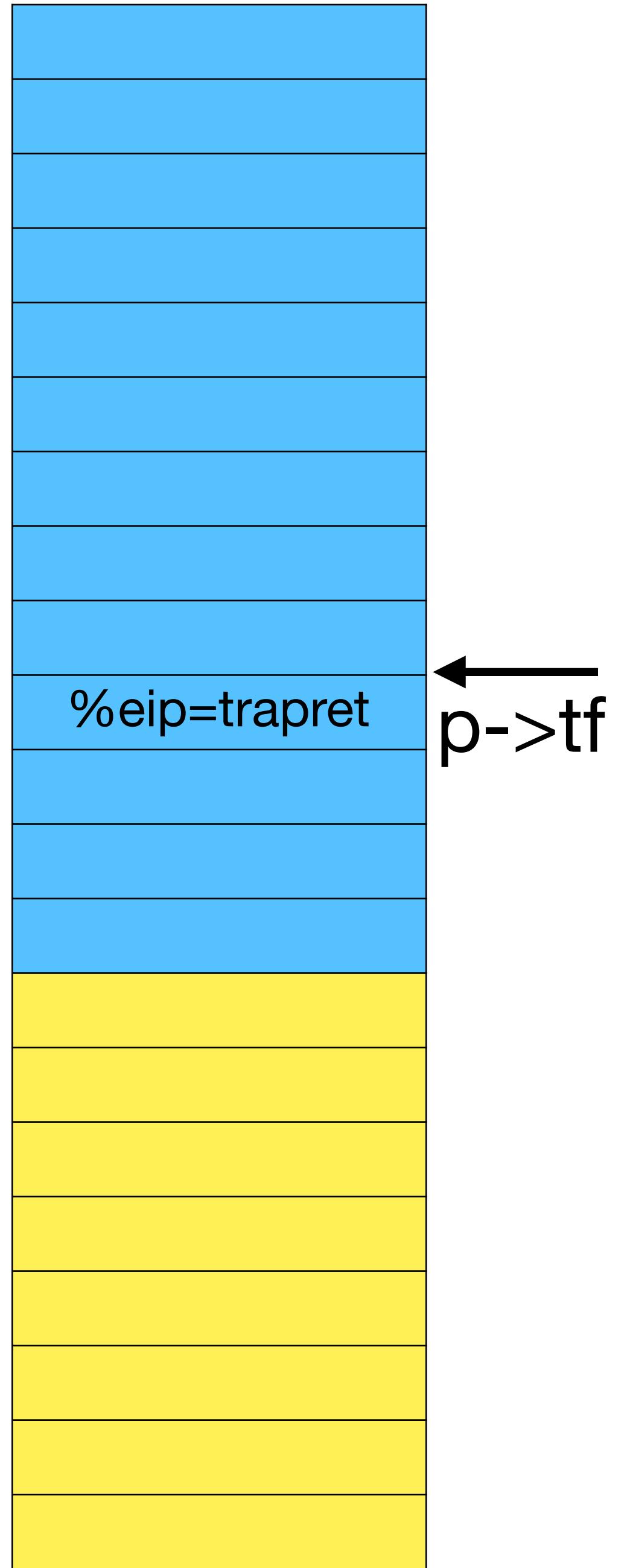
```
scheduler() {  
    ...  
    swtch(p->context);  
}
```

```
swtch:  
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret
```

```
.globl trapret p->context
```

```
trapret:
```

```
    popal  
    popl %gs  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```

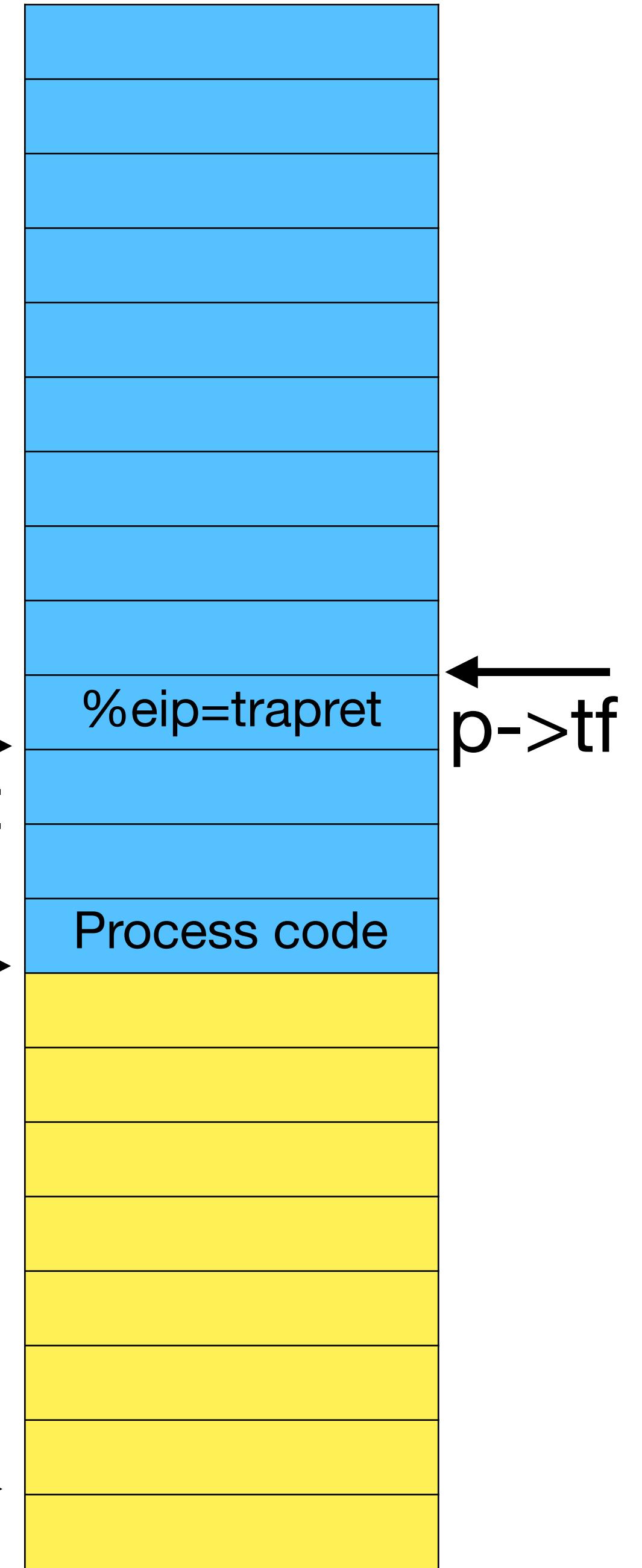


Understanding swtch

```
pinit(){  
    p = allocproc();  
    eip → memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}
```

```
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

```
scheduler() {  
    ...  
    swtch(p->context);  
}  
  
swtch:  
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret  
  
.globl trapret p->context  
trapret:  
    popal  
    popl %gs p->offset  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```



Understanding swtch

```
pinit(){  
    p = allocproc();  
  
    memmove(p->offset, _binary_initcode_start,);  
  
    eip → p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}  
}
```

```
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

```
scheduler()
```

```
...
```

```
    swtch(p->context);  
}
```

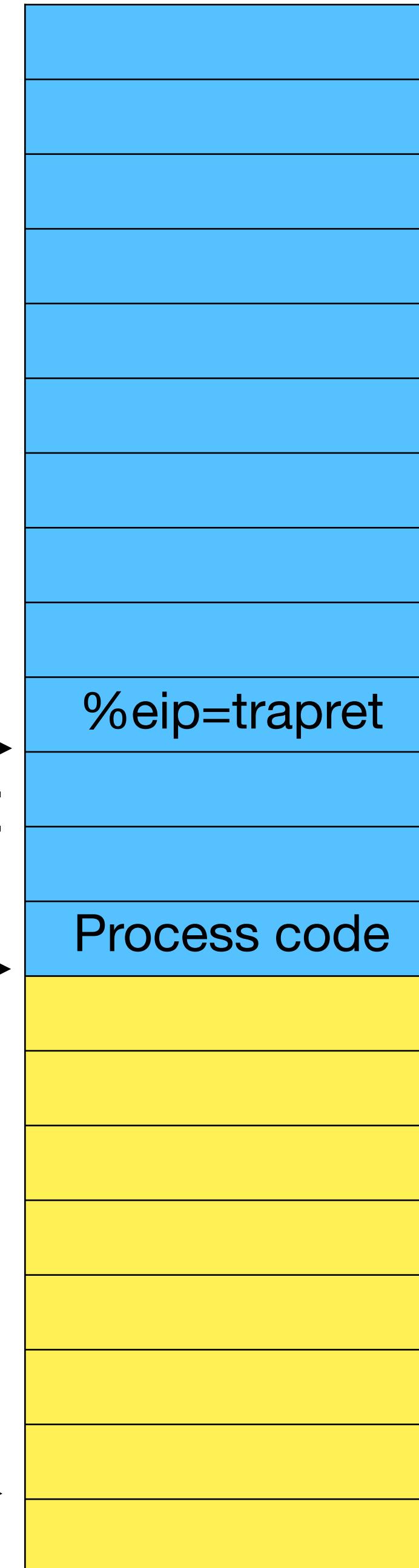
```
swtch:
```

```
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret
```

```
.globl trapret    p->context
```

```
trapret:
```

```
    popal  
    popl %gs  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```



Understanding swtch

```
pinit(){  
    p = allocproc();  
  
    memmove(p->offset, _binary_initcode_start,);  
  
    eip → p->tf->ds, es, ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}  
}
```

```
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

```
scheduler()
```

```
...
```

```
    swtch(p->context);  
}
```

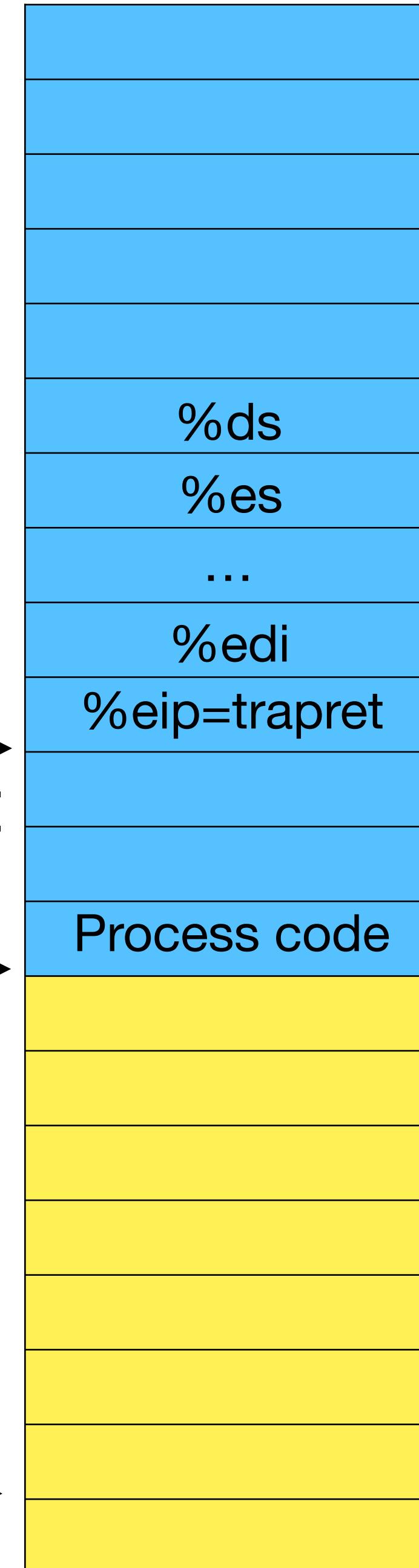
```
swtch:
```

```
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret
```

```
.globl trapret    p->context
```

```
trapret:
```

```
    popal  
    popl %gs  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```

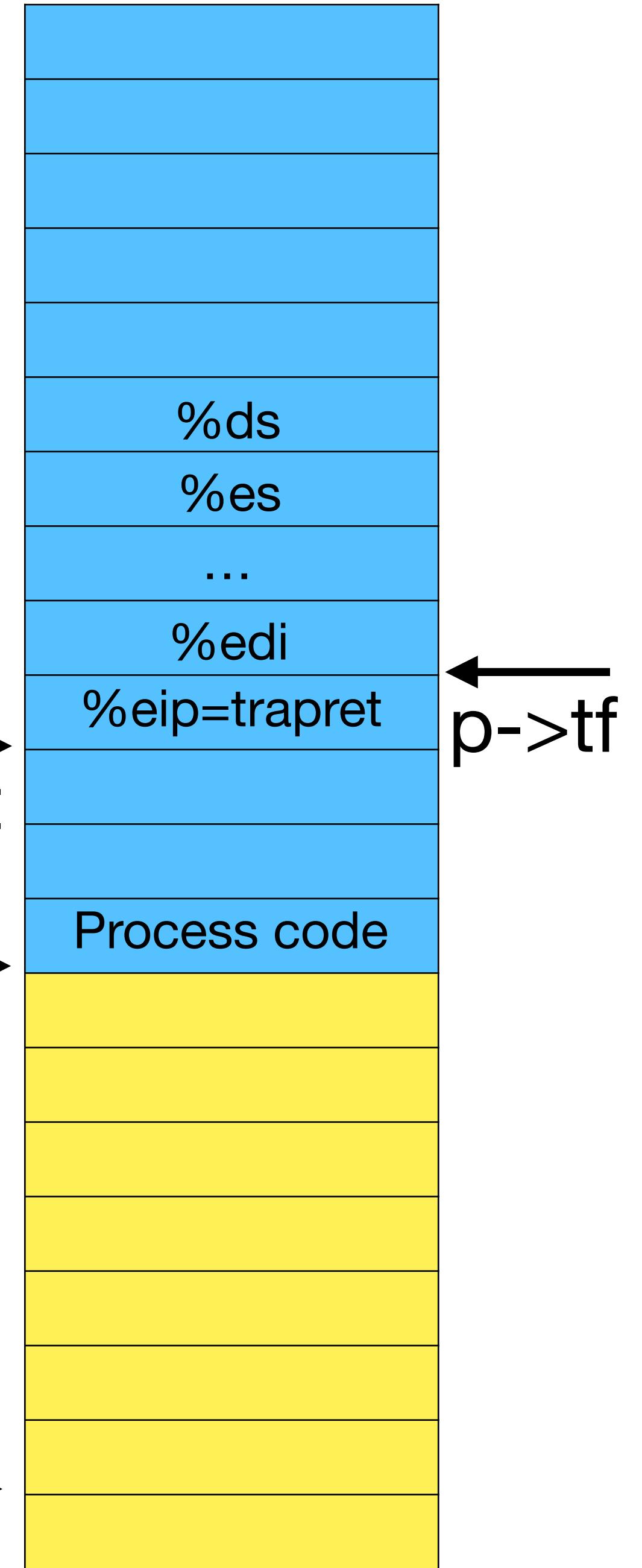


Understanding swtch

```
pinit(){  
    p = allocproc();  
    memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    eip → p->tf->eip = 0;  
}
```

```
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

```
scheduler() {  
    ...  
    swtch(p->context);  
}  
  
swtch:  
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret  
  
.globl trapret → p->context  
trapret:  
    popal  
    popl %gs → p->offset  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```



Understanding swtch

```
pinit(){  
    p = allocproc();  
    memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    eip → p->tf->eip = 0;  
}  
→
```

```
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

```
scheduler() {
```

```
...
```

```
    swtch(p->context);  
}
```

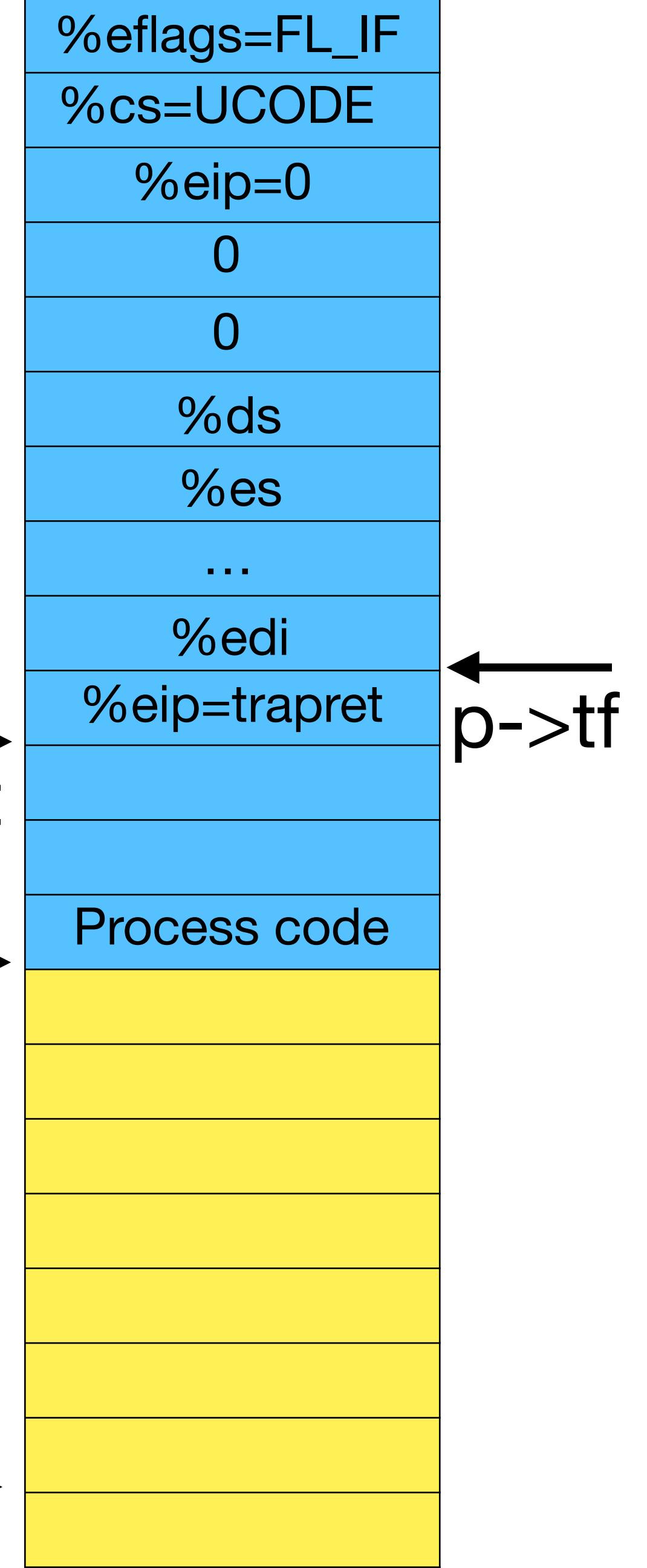
```
swtch:
```

```
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret
```

```
.globl trapret → p->context
```

```
trapret:
```

```
    popal  
    popl %gs → p->offset  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```



Understanding swtch

```
pinit() {
```

```
    p = allocproc();
```

```
    memmove(p->offset, _binary_initcode_start,); swtch:
```

```
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;
```

```
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;
```

```
    p->tf->eflags = FL_IF;
```

```
    p->tf->eip = 0;
```

```
}
```

```
allocproc() {
```

```
    sp = (char*)(STARTPROC + (PROCSIZE>>12));
```

```
    sp -= sizeof *p->tf;
```

```
    p->tf = (struct trapframe*)sp;
```

```
    sp -= sizeof *p->context;
```

```
    p->context = (struct context*)sp;
```

```
    p->context->eip = (uint)trapret;
```

```
    return p;
```

```
scheduler() {
```

eip
...
swtch(p->context);
}

```
    movl 4(%esp), %eax
```

```
    movl %eax, %esp
```

```
    movl $0, %eax
```

```
    ret
```

.globl trapret p->context
trapret:

```
    popal
```

```
    popl %gs
```

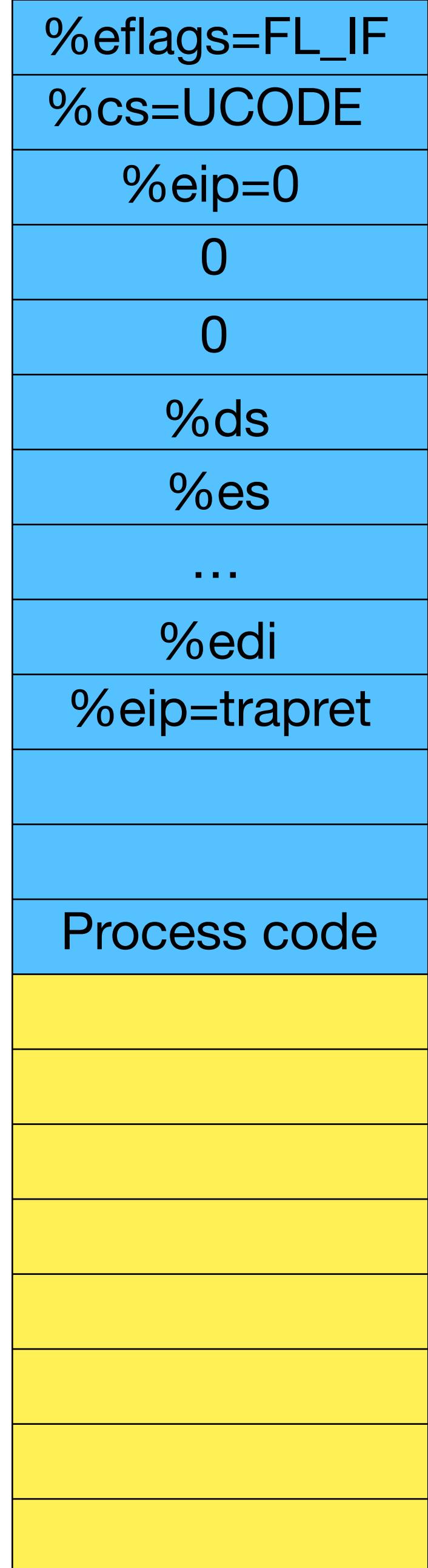
```
    popl %fs
```

```
    popl %es
```

```
    popl %ds
```

```
    addl $0x8, %esp
```

```
    iret
```



Understanding swtch

```
pinit() {
```

```
    p = allocproc();
```

```
    memmove(p->offset, _binary_initcode_start,); swtch:
```

```
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;
```

```
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;
```

```
    p->tf->eflags = FL_IF;
```

```
    p->tf->eip = 0;
```

```
}
```

```
allocproc() {
```

```
    sp = (char*)(STARTPROC + (PROCSIZE>>12));
```

```
    sp -= sizeof *p->tf;
```

```
    p->tf = (struct trapframe*)sp;
```

```
    sp -= sizeof *p->context;
```

```
    p->context = (struct context*)sp;
```

```
    p->context->eip = (uint)trapret;
```

```
    return p;
```

```
}
```

```
scheduler() {
```

eip
→

```
    ...  
    swtch(p->context);
```

}

```
swtch:
```

```
    movl 4(%esp), %eax
```

```
    movl %eax, %esp
```

```
    movl $0, %eax
```

```
    ret
```

.globl trapret → p->context

```
trapret:
```

```
    popal
```

```
    popl %gs → p->offset
```

```
    popl %fs
```

```
    popl %es
```

```
    popl %ds
```

```
    addl $0x8, %esp
```

```
    iret
```

esp
→

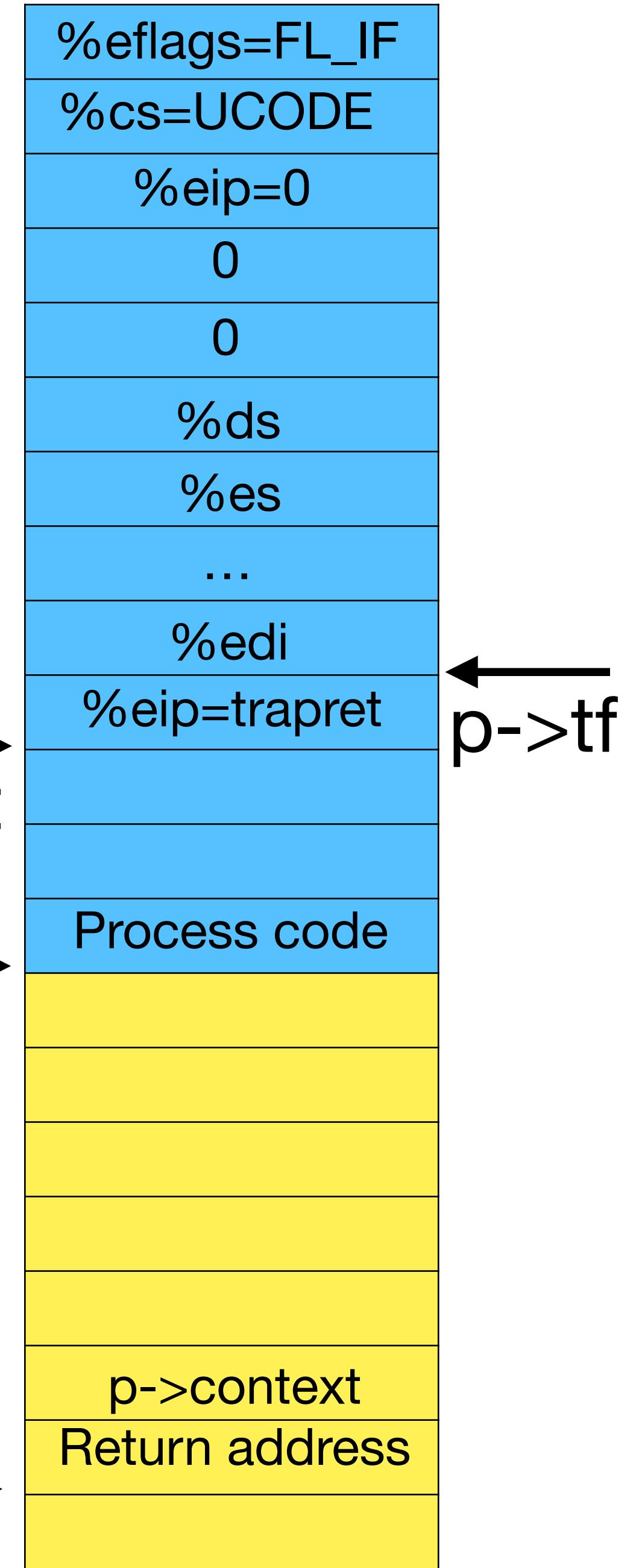
%eflags=FL_IF
%cs=UCODE
%eip=0
0
0
%ds
%es
...
%edi
%eip=trapret
Process code
p->context
Return address

← p->tf

Understanding swtch

```
pinit(){  
    p = allocproc();  
  
    memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}  
  
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

```
scheduler() {  
    ...  
    swtch(p->context);  
}  
  
swtch:  
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret  
  
.globl trapret  p->context  
trapret:  
    popal  
    popl %gs  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```



Understanding swtch

```
pinit(){  
    p = allocproc();  
  
    memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}  
  
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

```
scheduler() {
```

```
...
```

```
    swtch(p->context);
```

```
}
```

```
swtch:
```

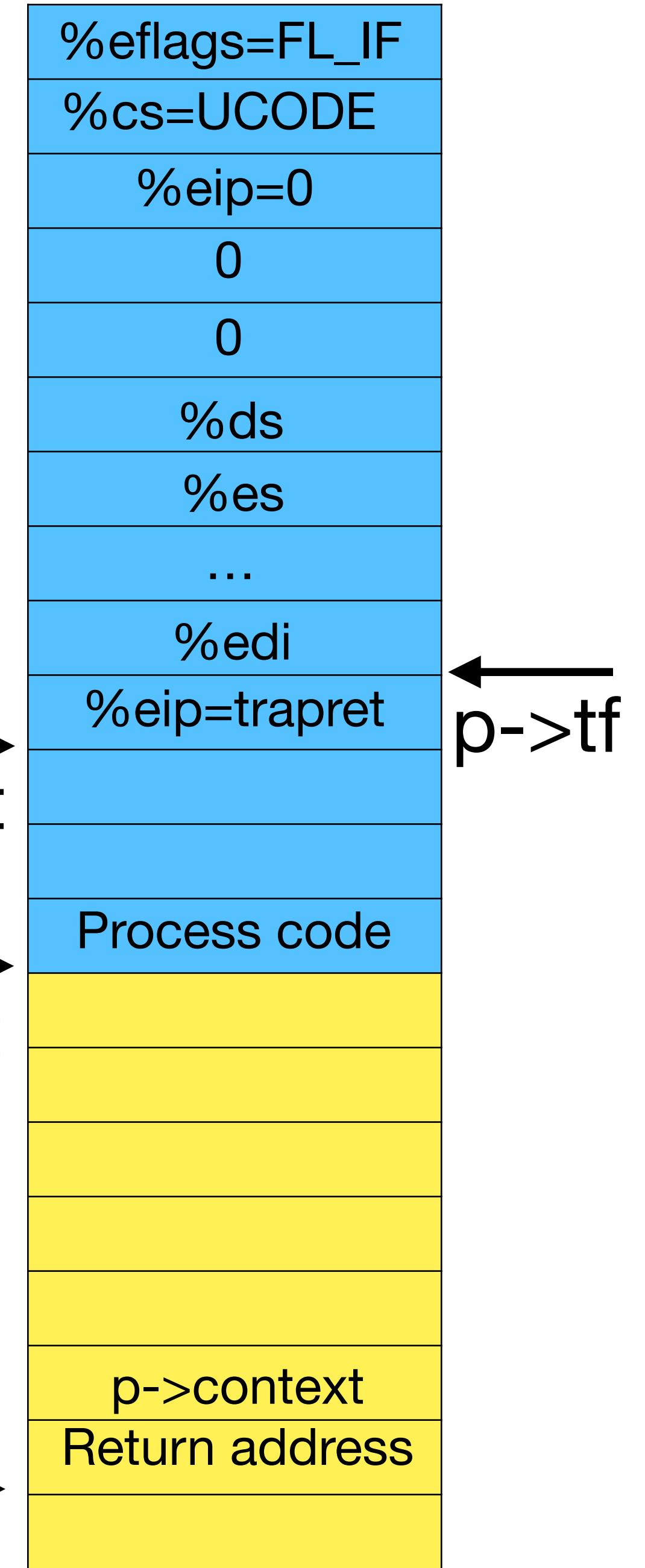
```
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret
```

```
.globl trapret    p->context
```

```
trapret:
```

```
    popal  
    popl %gs  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```

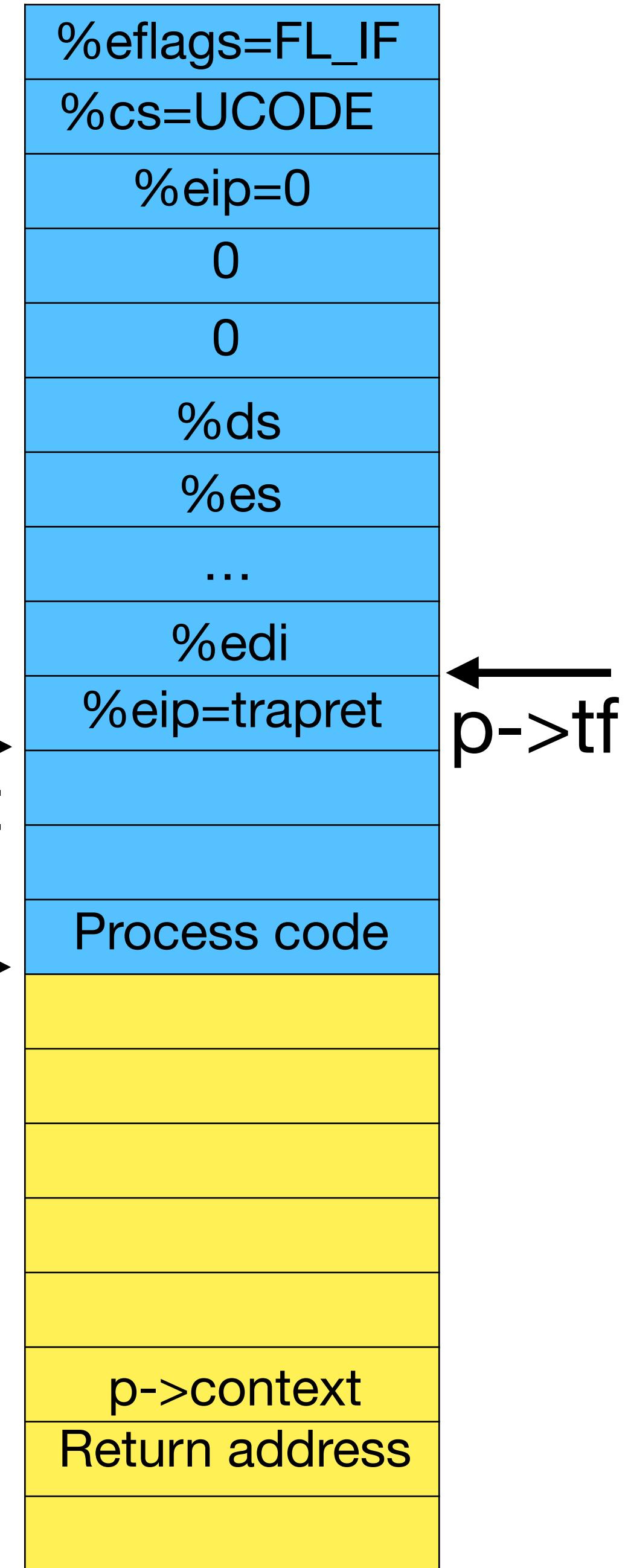
```
esp
```



Understanding swtch

```
pinit(){  
    p = allocproc();  
  
    memmove(p->offset, _binary_initcode_start,);  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}  
  
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}
```

```
scheduler() {  
    ...  
    swtch(p->context);  
}  
  
swtch:  
    movl 4(%esp), %eax  
    movl %eax, %esp  
    movl $0, %eax  
    ret  
  
.globl trapret    p->context  
trapret:  
    popal  
    popl %gs  
    popl %fs  
    popl %es  
    popl %ds  
    addl $0x8, %esp  
    iret
```



Understanding swtch

```
pinit(){
```

```
    p = allocproc();
```

```
    memmove(p->offset, _binary_initcode_start,);
```

```
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;
```

```
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;
```

```
    p->tf->eflags = FL_IF;
```

```
    p->tf->eip = 0;
```

```
}
```

```
allocproc() {
```

```
    sp = (char*)(STARTPROC + (PROCSIZE>>12));
```

```
    sp -= sizeof *p->tf;
```

```
    p->tf = (struct trapframe*)sp;
```

```
    sp -= sizeof *p->context;
```

```
    p->context = (struct context*)sp;
```

```
    p->context->eip = (uint)trapret;
```

```
    return p;
```

```
}
```

```
scheduler() {
```

```
...
```

```
    swtch(p->context);
```

```
}
```

```
swtch:
```

```
    movl 4(%esp), %eax
```

eip
→

```
    movl %eax, %esp
```

```
    movl $0, %eax
```

```
    ret
```

```
.globl trapret
```

esp
→
p->context

```
trapret:
```

```
    popal
```

p->offset
→

```
    popl %gs
```

```
    popl %fs
```

```
    popl %es
```

```
    popl %ds
```

```
    addl $0x8, %esp
```

```
    iret
```

%eflags=FL_IF

%cs=UCODE

%eip=0

0

0

%ds

%es

...

%edi

%eip=trapret

p->tf

Process code

p->context

p->context

Return address

Return address

Understanding swtch

```
pinit(){
```

```
    p = allocproc();
```

```
    memmove(p->offset, _binary_initcode_start,);
```

```
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;
```

```
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;
```

```
    p->tf->eflags = FL_IF;
```

```
    p->tf->eip = 0;
```

```
}
```

```
allocproc() {
```

```
    sp = (char*)(STARTPROC + (PROCSIZE>>12));
```

```
    sp -= sizeof *p->tf;
```

```
    p->tf = (struct trapframe*)sp;
```

```
    sp -= sizeof *p->context;
```

```
    p->context = (struct context*)sp;
```

```
    p->context->eip = (uint)trapret;
```

```
    return p;
```

```
}
```

```
scheduler() {
```

```
...
```

```
    swtch(p->context);
```

```
}
```

```
swtch:
```

```
    movl 4(%esp), %eax
```

```
    movl %eax, %esp
```

```
    movl $0, %eax
```

eip → ret

```
.globl trapret    p->context
```

```
trapret:
```

```
    popal
```

```
    popl %gs
```

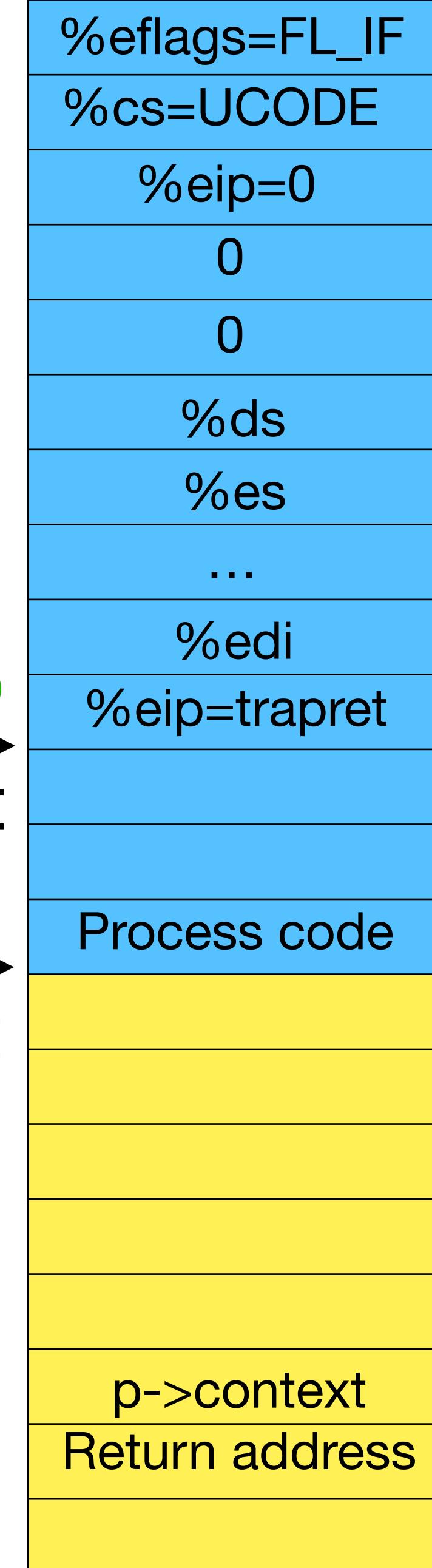
```
    popl %fs
```

```
    popl %es
```

```
    popl %ds
```

```
    addl $0x8, %esp
```

```
    iret
```



Understanding swtch

```
pinit(){
```

```
    p = allocproc();
```

```
    memmove(p->offset, _binary_initcode_start,);
```

```
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;
```

```
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;
```

```
    p->tf->eflags = FL_IF;
```

```
    p->tf->eip = 0;
```

```
}
```

```
allocproc() {
```

```
    sp = (char*)(STARTPROC + (PROCSIZE>>12));
```

```
    sp -= sizeof *p->tf;
```

```
    p->tf = (struct trapframe*)sp;
```

```
    sp -= sizeof *p->context;
```

```
    p->context = (struct context*)sp;
```

```
    p->context->eip = (uint)trapret;
```

```
    return p;
```

```
}
```

```
scheduler() {
```

```
...
```

```
    swtch(p->context);
```

```
}
```

```
swtch:
```

```
    movl 4(%esp), %eax
```

```
    movl %eax, %esp
```

```
    movl $0, %eax
```

```
eip → ret
```

```
.globl trapret p->context
```

```
trapret:
```

```
    popal
```

```
    popl %gs
```

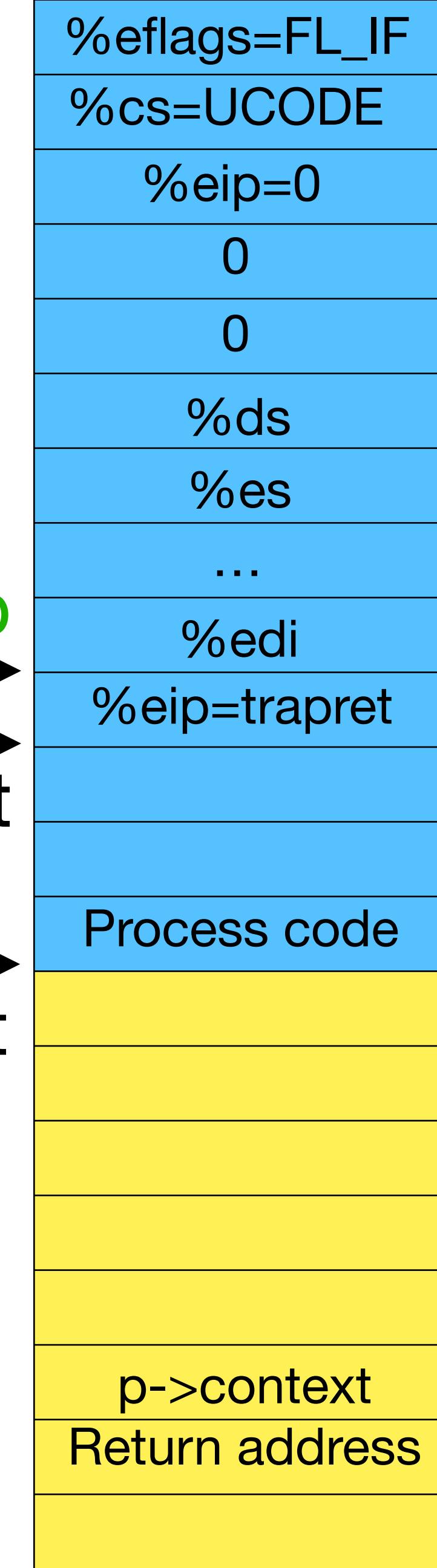
```
    popl %fs
```

```
    popl %es
```

```
    popl %ds
```

```
    addl $0x8, %esp
```

```
    iret
```



Understanding swtch

```
pinit(){
```

```
    p = allocproc();
```

```
    memmove(p->offset, _binary_initcode_start,);
```

```
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;
```

```
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;
```

```
    p->tf->eflags = FL_IF;
```

```
    p->tf->eip = 0;
```

```
}
```

```
allocproc() {
```

```
    sp = (char*)(STARTPROC + (PROCSIZE>>12));
```

```
    sp -= sizeof *p->tf;
```

```
    p->tf = (struct trapframe*)sp;
```

```
    sp -= sizeof *p->context;
```

```
    p->context = (struct context*)sp;
```

```
    p->context->eip = (uint)trapret;
```

```
    return p;
```

```
}
```

```
scheduler() {
```

```
...
```

```
    swtch(p->context);
```

```
}
```

```
swtch:
```

```
    movl 4(%esp), %eax
```

```
    movl %eax, %esp
```

```
    movl $0, %eax
```

```
    ret
```

```
.globl trapret
```

```
trapret:
```

```
eip → popal
```

```
popl %gs
```

```
popl %fs
```

```
popl %es
```

```
popl %ds
```

```
addl $0x8, %esp
```

```
iret
```

%eflags=FL_IF

%cs=UCODE

%eip=0

0

0

%ds

%es

...

%edi

%eip=trapret

esp

→

p->context

eip

→ popal

p->offset

popl %gs

popl %fs

popl %es

popl %ds

Process code

Return address

← p->tf

Understanding swtch

```
pinit(){
```

```
    p = allocproc();
```

```
    memmove(p->offset, _binary_initcode_start,);
```

```
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;
```

```
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;
```

```
    p->tf->eflags = FL_IF;
```

```
    p->tf->eip = 0;
```

```
}
```

```
allocproc() {
```

```
    sp = (char*)(STARTPROC + (PROCSIZE>>12));
```

```
    sp -= sizeof *p->tf;
```

```
    p->tf = (struct trapframe*)sp;
```

```
    sp -= sizeof *p->context;
```

```
    p->context = (struct context*)sp;
```

```
    p->context->eip = (uint)trapret;
```

```
    return p;
```

```
}
```

```
scheduler() {
```

```
...
```

```
    swtch(p->context);
```

```
}
```

```
swtch:
```

```
    movl 4(%esp), %eax
```

```
    movl %eax, %esp
```

```
    movl $0, %eax
```

```
    ret
```

```
.globl trapret
```

```
trapret:
```

```
eip → popal
```

```
popl %gs
```

```
popl %fs
```

```
popl %es
```

```
popl %ds
```

```
addl $0x8, %esp
```

```
iret
```

esp

p->context

eip

p->offset

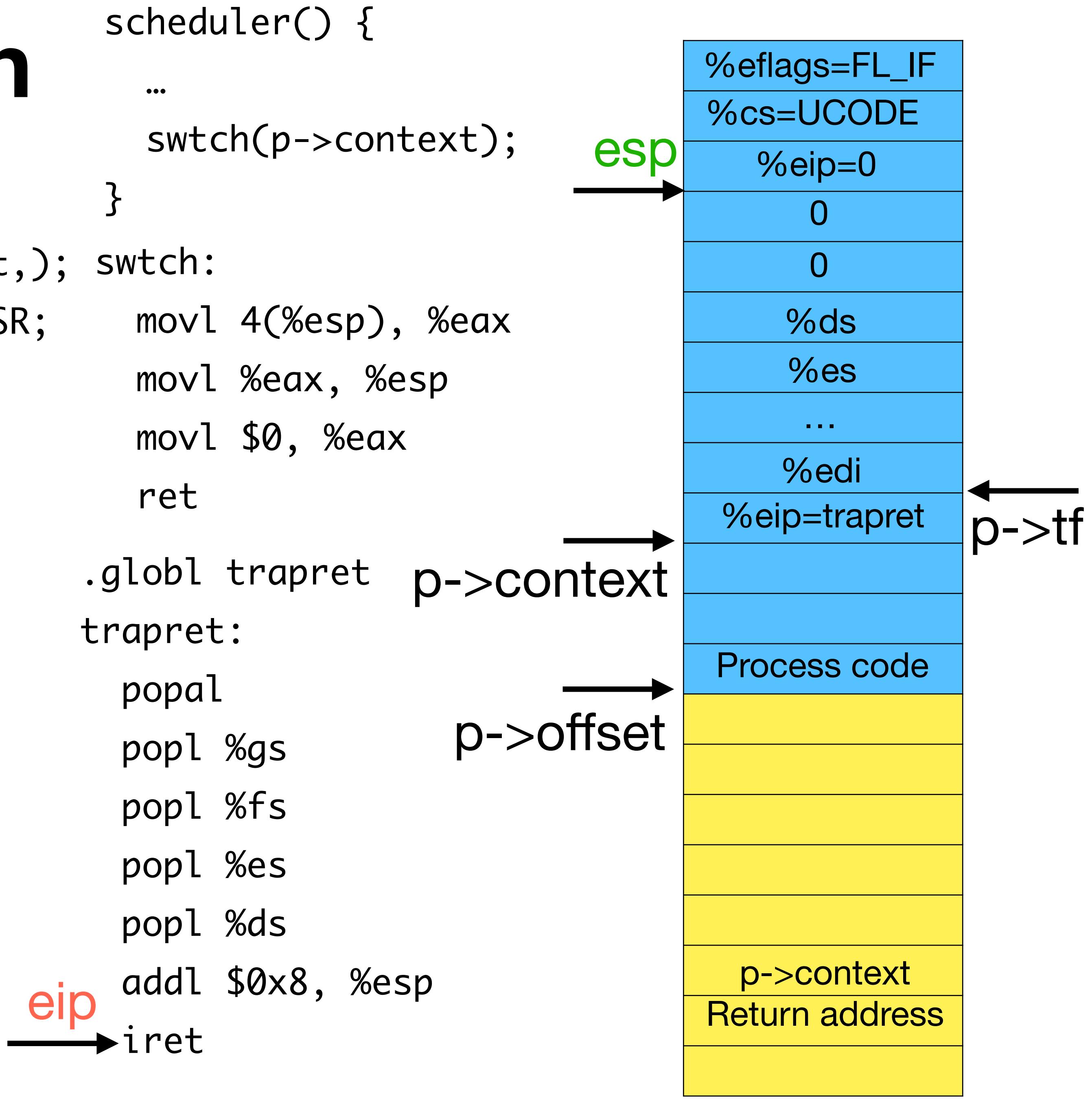
%eflags=FL_IF	
%cs=UCODE	
%eip=0	
0	
0	
%ds	
%es	
...	
%edi	
%eip=trapret	
Process code	
p->context	
Return address	

p->tf

Understanding swtch

```
pinit(){  
    p = allocproc();  
  
    memmove(p->offset, _binary_initcode_start,);  
  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}  
}
```

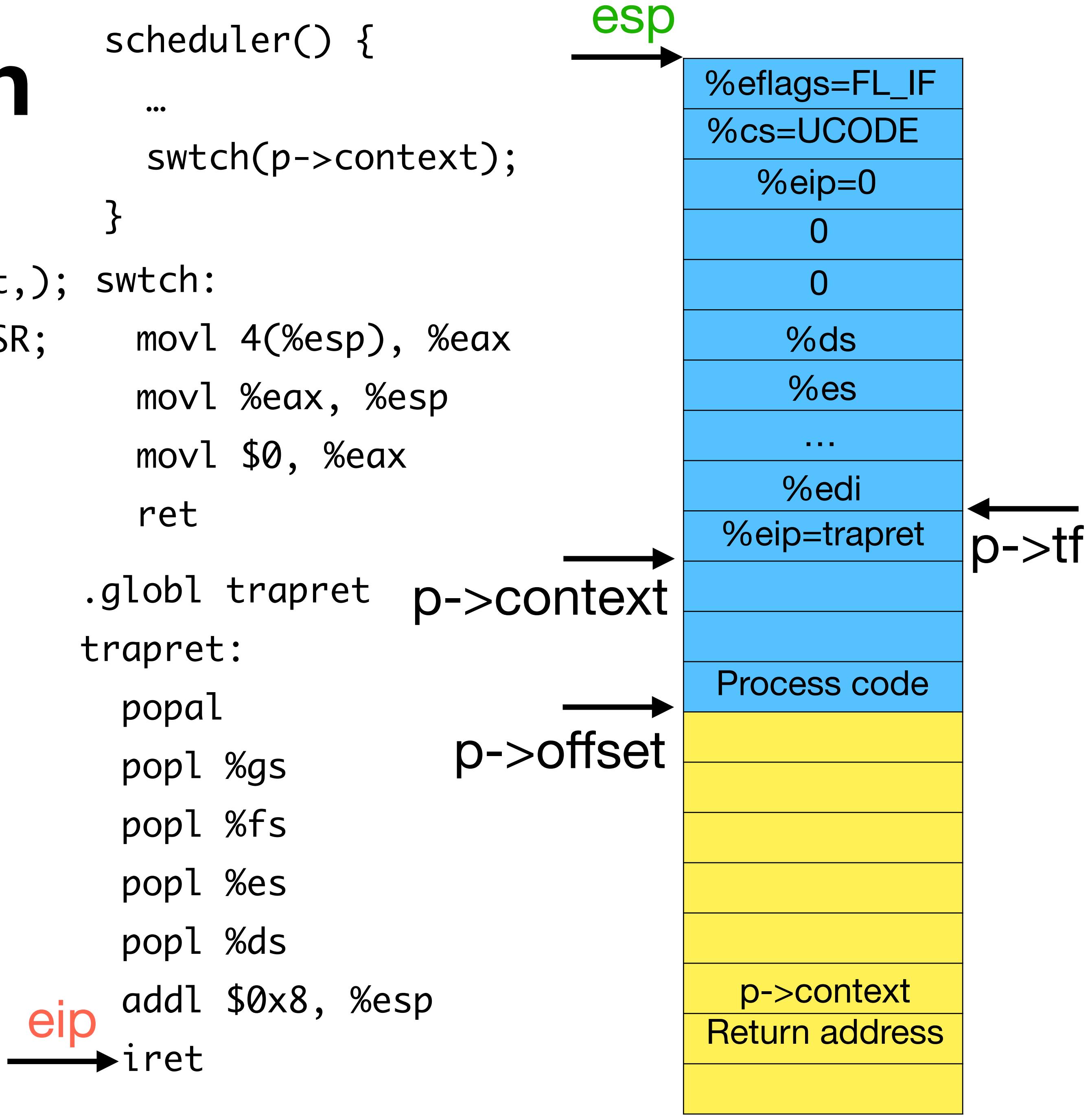
```
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    sp -= sizeof *p->tf;  
    p->tf = (struct trapframe*)sp;  
    sp -= sizeof *p->context;  
    p->context = (struct context*)sp;  
    p->context->eip = (uint)trapret;  
    return p;  
}  
}
```



Understanding swtch

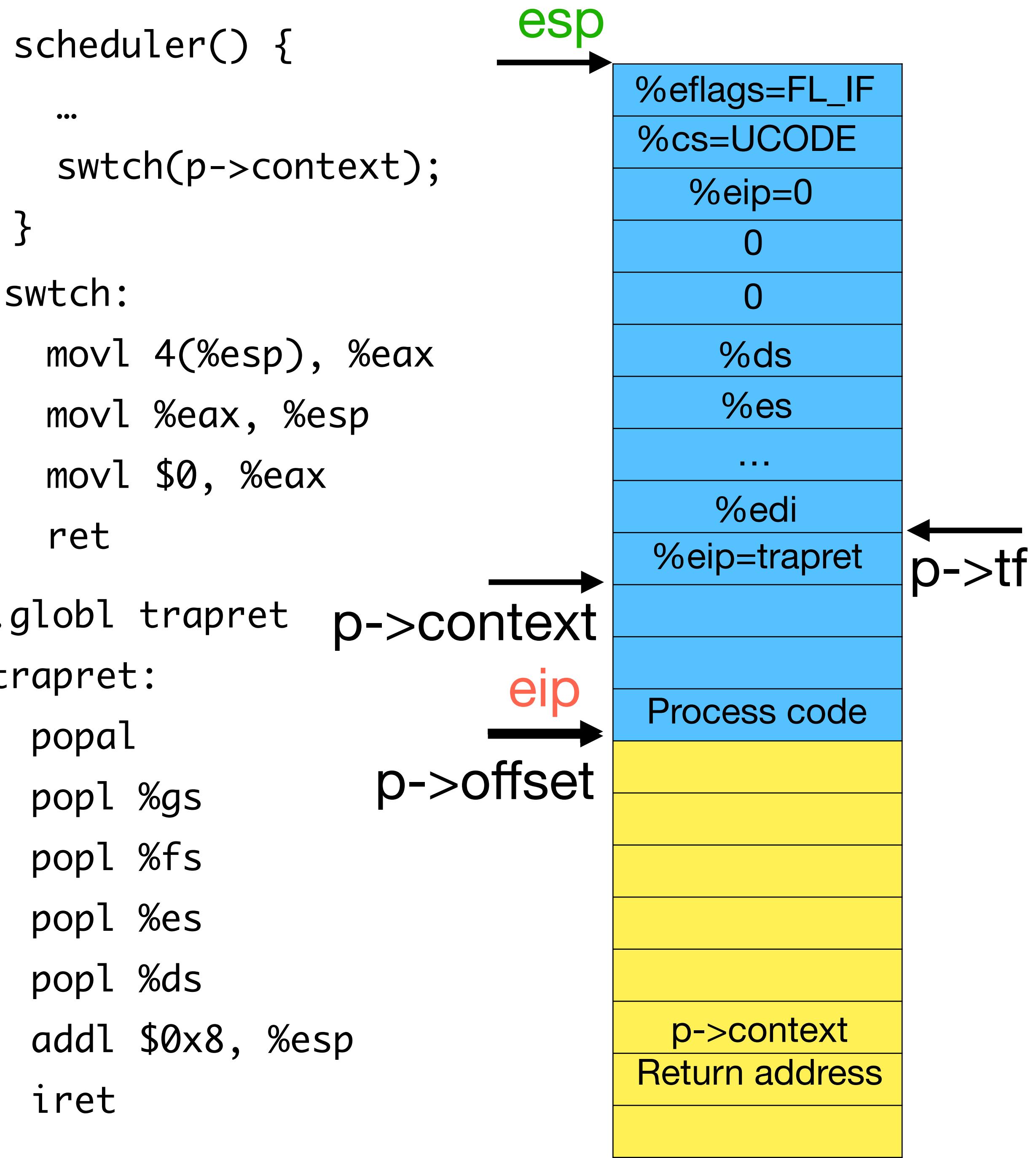
```
pinit(){  
    p = allocproc();  
  
    memmove(p->offset, _binary_initcode_start,);  
  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}  
}
```

```
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
  
    sp -= sizeof *p->tf;  
  
    p->tf = (struct trapframe*)sp;  
  
    sp -= sizeof *p->context;  
  
    p->context = (struct context*)sp;  
  
    p->context->eip = (uint)trapret;  
  
    return p;  
}  
}
```



Understanding swtch

```
pinit(){  
    p = allocproc();  
  
    memmove(p->offset, _binary_initcode_start,);  
  
    p->tf->ds,es,ss = (SEG_UDATA<<3) | DPL_USR;  
    p->tf->cs = (SEG_UCODE<<3) | DPL_USR;  
    p->tf->eflags = FL_IF;  
    p->tf->eip = 0;  
}  
  
allocproc() {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
  
    sp -= sizeof *p->tf;  
  
    p->tf = (struct trapframe*)sp;  
  
    sp -= sizeof *p->context;  
  
    p->context = (struct context*)sp;  
  
    p->context->eip = (uint)trapret;  
  
    return p;  
}
```



Interrupt handling revisited

```
eip → for(;;)
;
trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
return
```

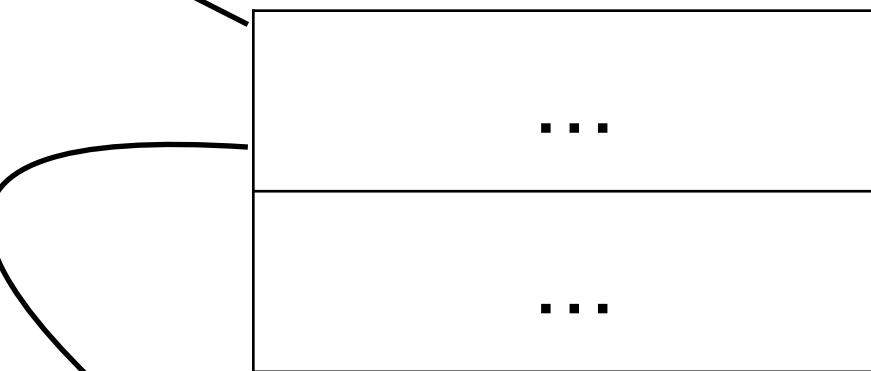
vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

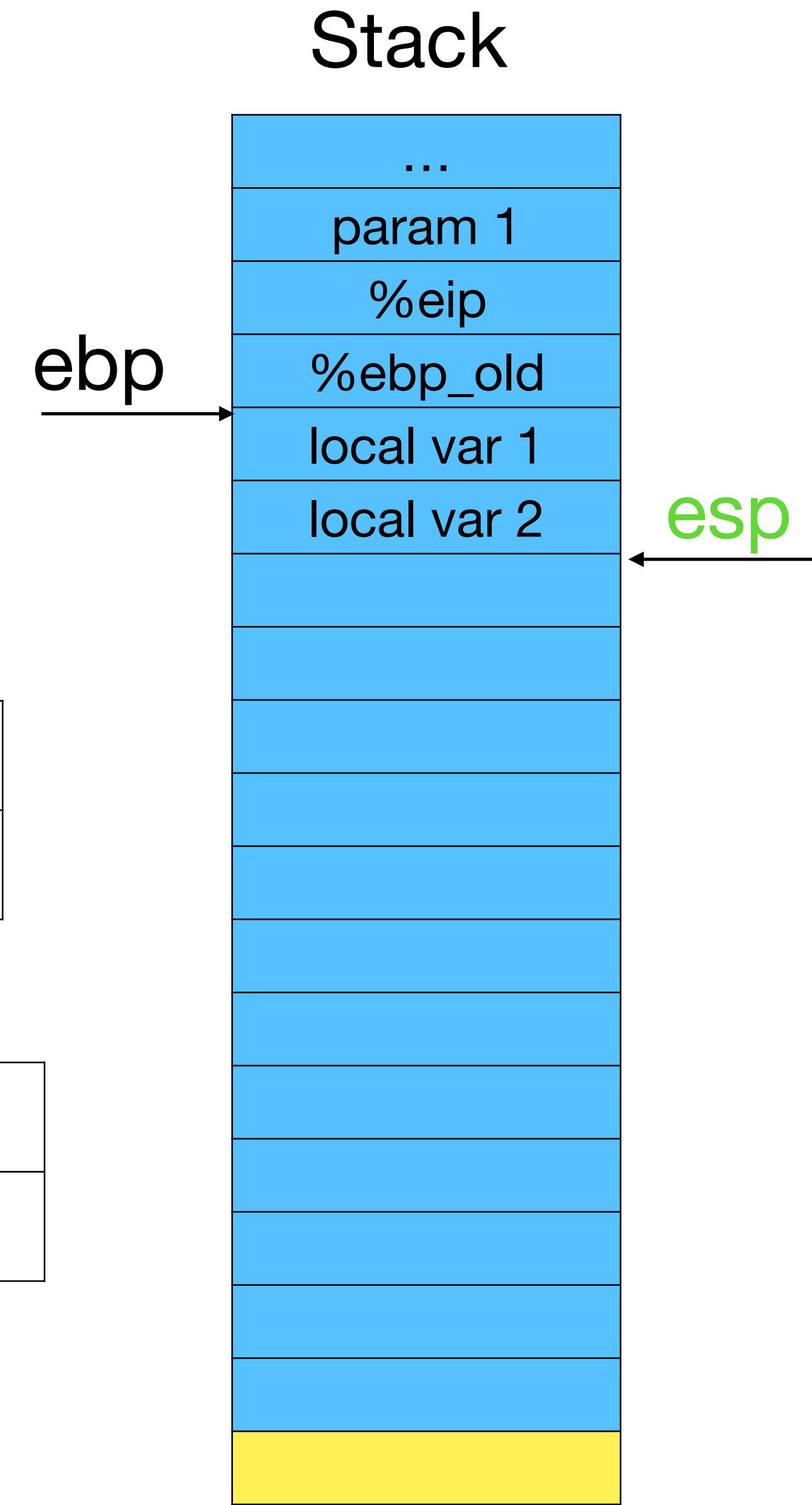
trapasm.S

```
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

IDT



GDT



Interrupt handling revisited

```
eip → for(;;)  
;  
  
trap.c  
void  
trap(struct trapframe *tf)  
{  
    switch(tf->trapno){  
        case T_IRQ0 + IRQ_TIMER:  
            ticks++;  
            cprintf("Tick! %d\n", ticks);  
            lapiceoi();  
            ...  
    }  
    return
```

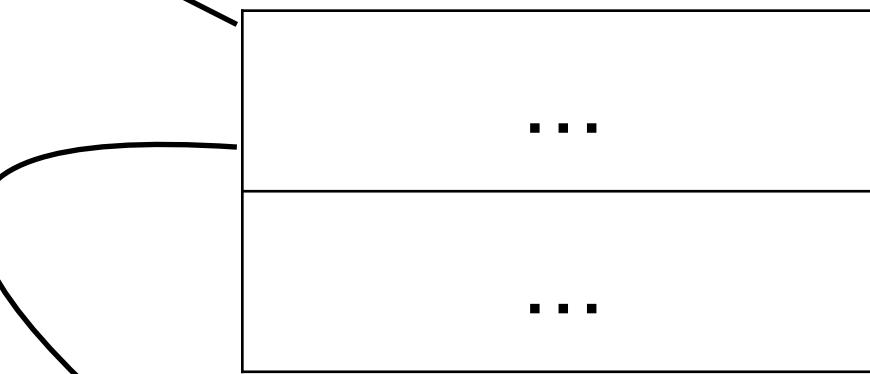
vectors.S

```
.globl vector0  
vector0:  
    pushl $0  
    pushl $0  
    jmp alltraps
```

trapasm.S

```
alltraps:  
    pushal  
    pushl %esp  
    call trap  
    addl $4, %esp  
    popal  
    addl $0x8, %esp  
    iret
```

IDT



GDT



Interrupt handling revisited

eip → for(;;)
;

trap.c
void
trap(struct trapframe *tf)
{
 switch(tf->trapno){
 case T_IRQ0 + IRQ_TIMER:
 ticks++;
 cprintf("Tick! %d\n", ticks);
 lapiceoi();
 ..
 }
 return;

```
vectors.S  
.globl vector0  
vector0:  
    pushl $0  
    pushl $0  
    jmp alltraps
```

trapasm.S

alltraps:

pushal

pushl %esp

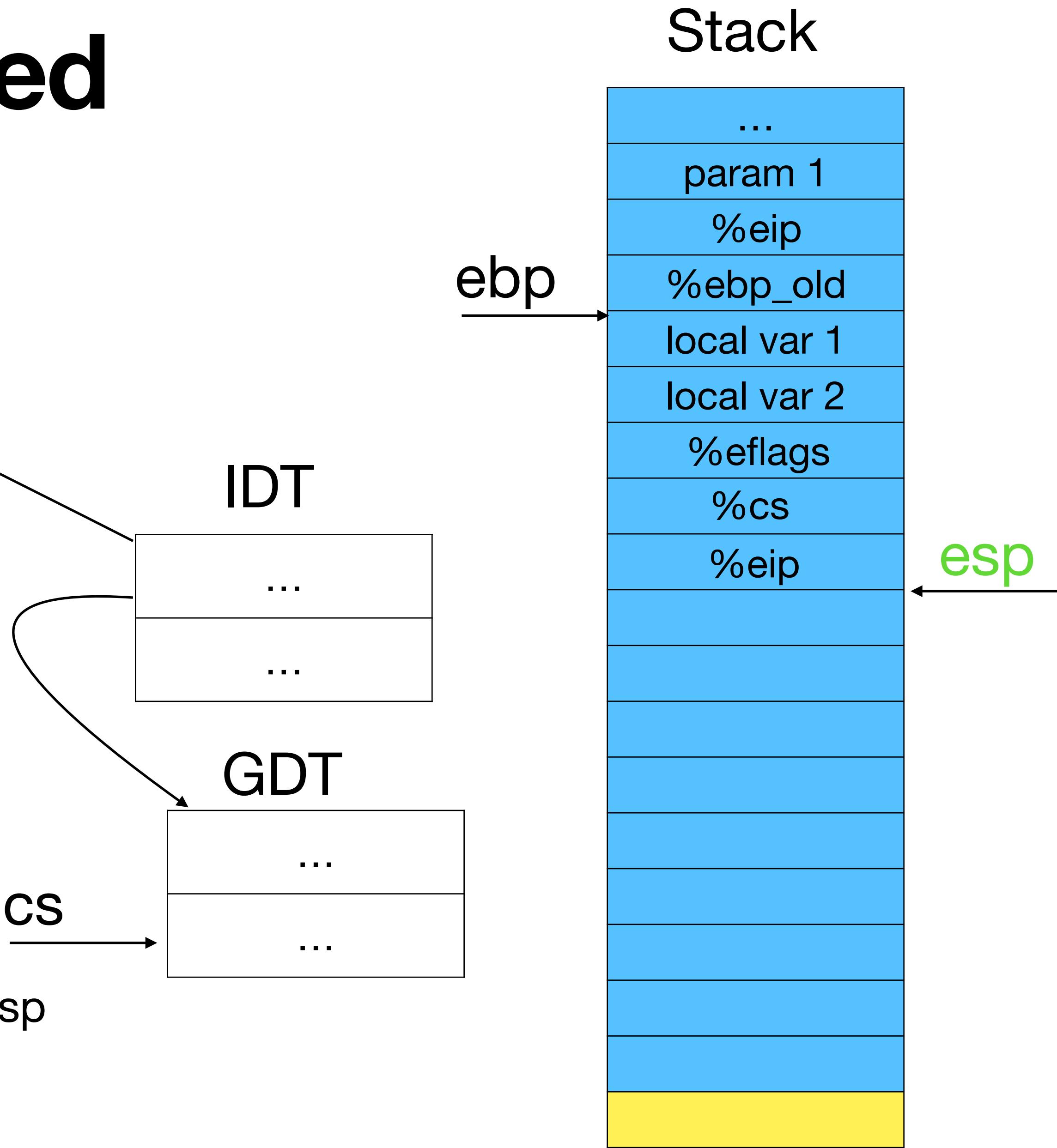
call trap

addl \$4, %esp cs

popal

addl \$0x8, %esp

iret



Interrupt handling revisited

```
eip → for(;;)  
;  
  
trap.c  
void  
trap(struct trapframe *tf)  
{  
    switch(tf->trapno){  
        case T_IRQ0 + IRQ_TIMER:  
            ticks++;  
            cprintf("Tick! %d\n0", ticks);  
            lapiceoi();  
            ...  
    }  
    return
```

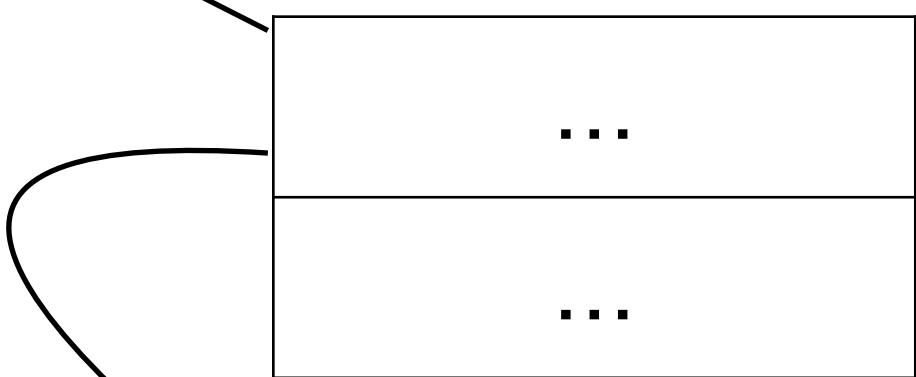
vectors.S

```
.globl vector0  
vector0:  
    pushl $0  
    pushl $0  
    jmp alltraps
```

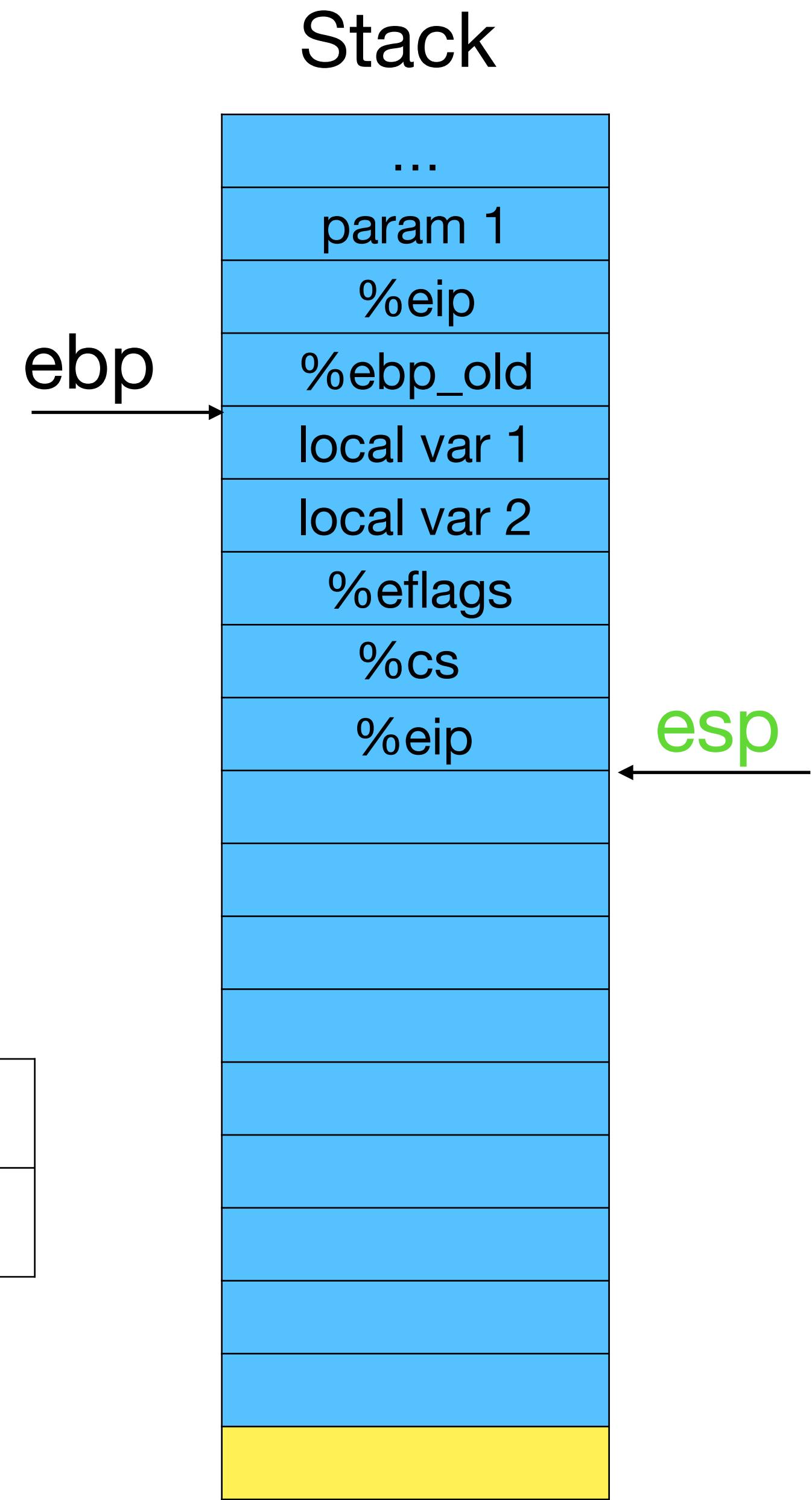
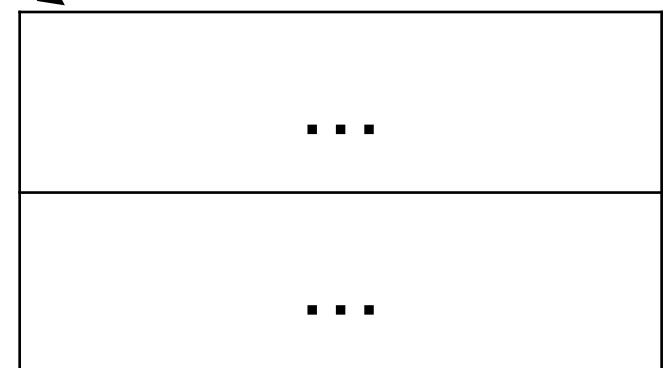
trapasm.S

```
alltraps:  
    pushal  
    pushl %esp  
    call trap  
    addl $4, %esp  
    popal  
    addl $0x8, %esp  
    iret
```

IDT



GDT



Interrupt handling revisited

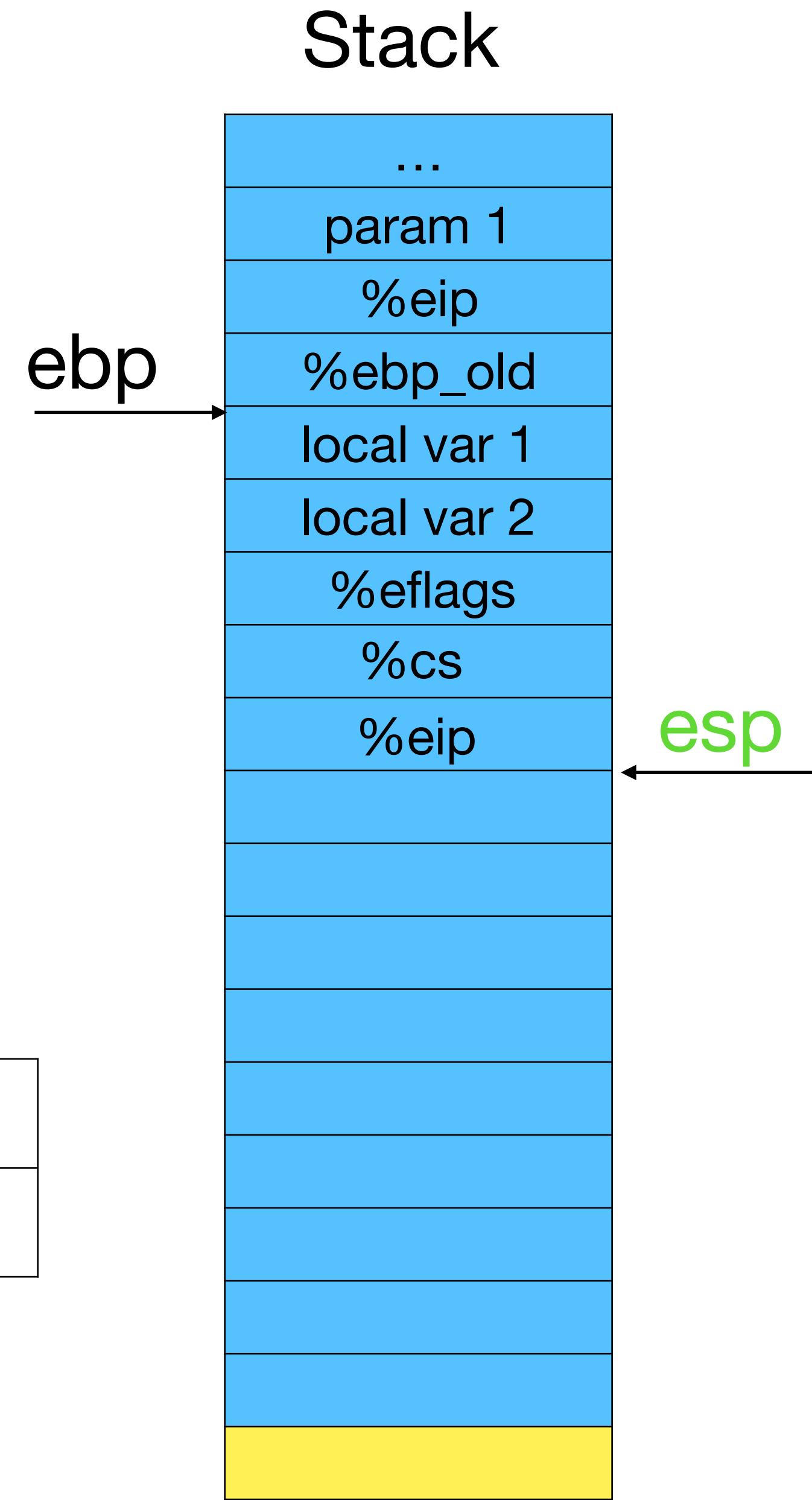
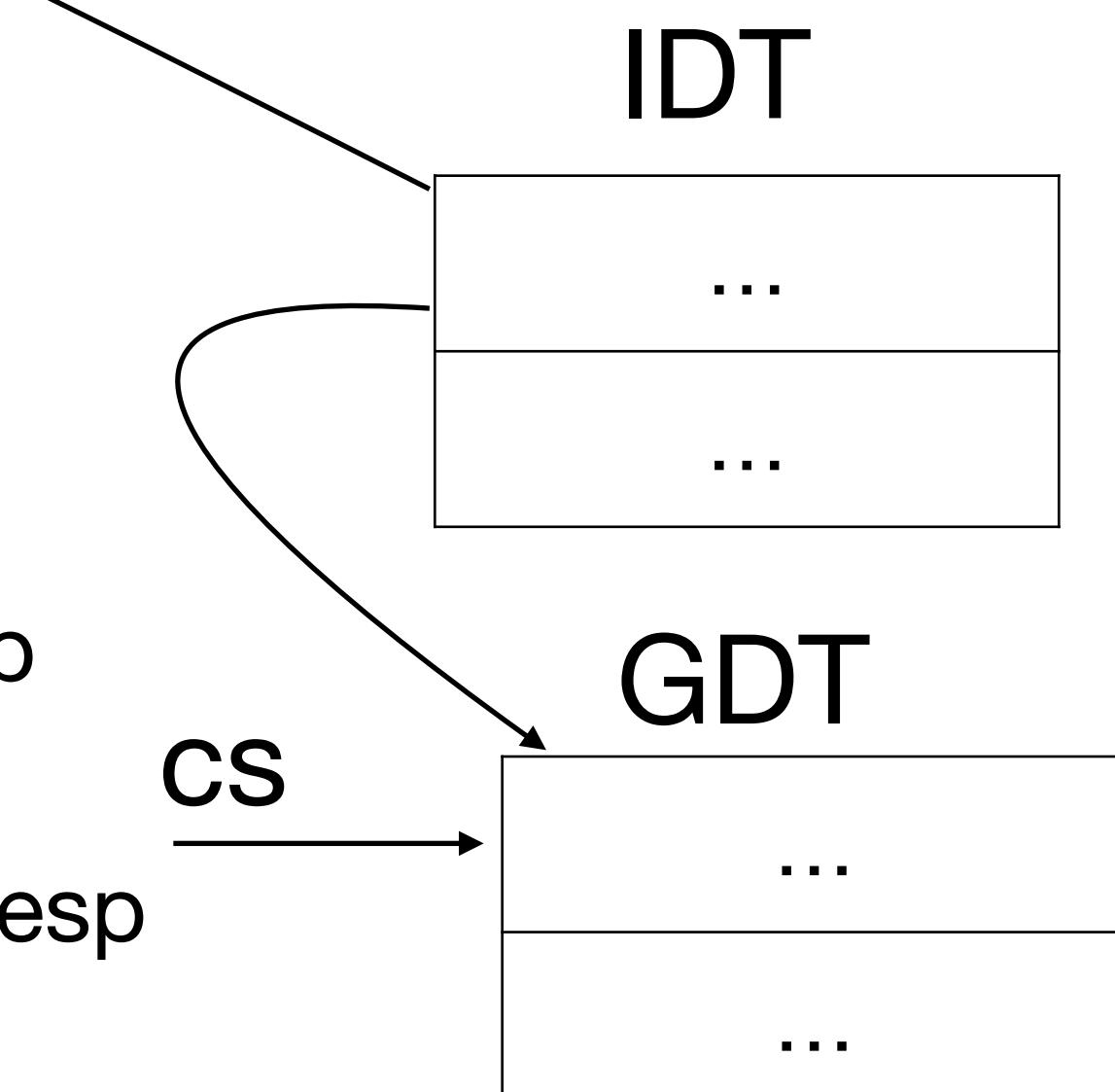
```
eip → for(;;)
;
trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
return
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

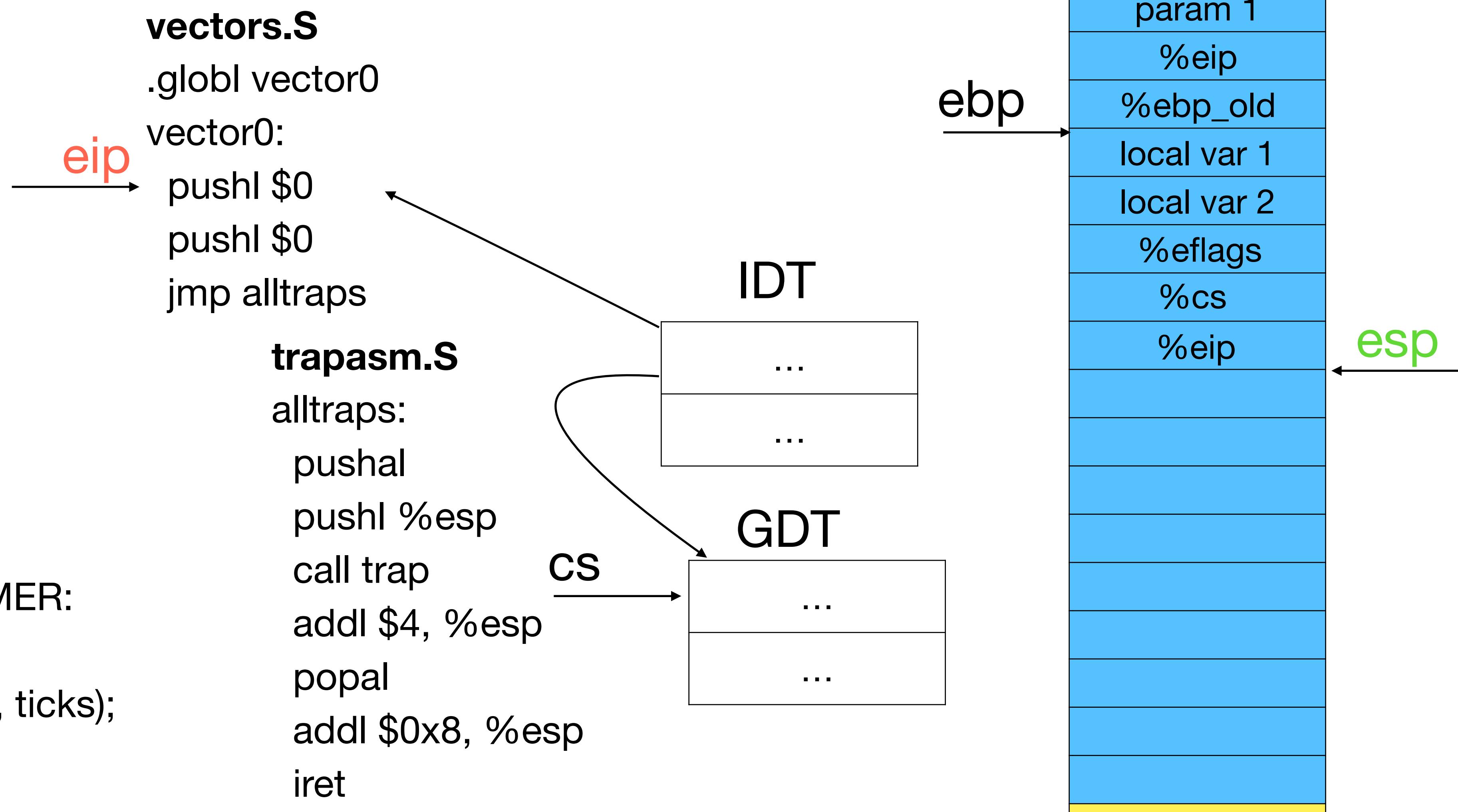
trapasm.S

```
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```



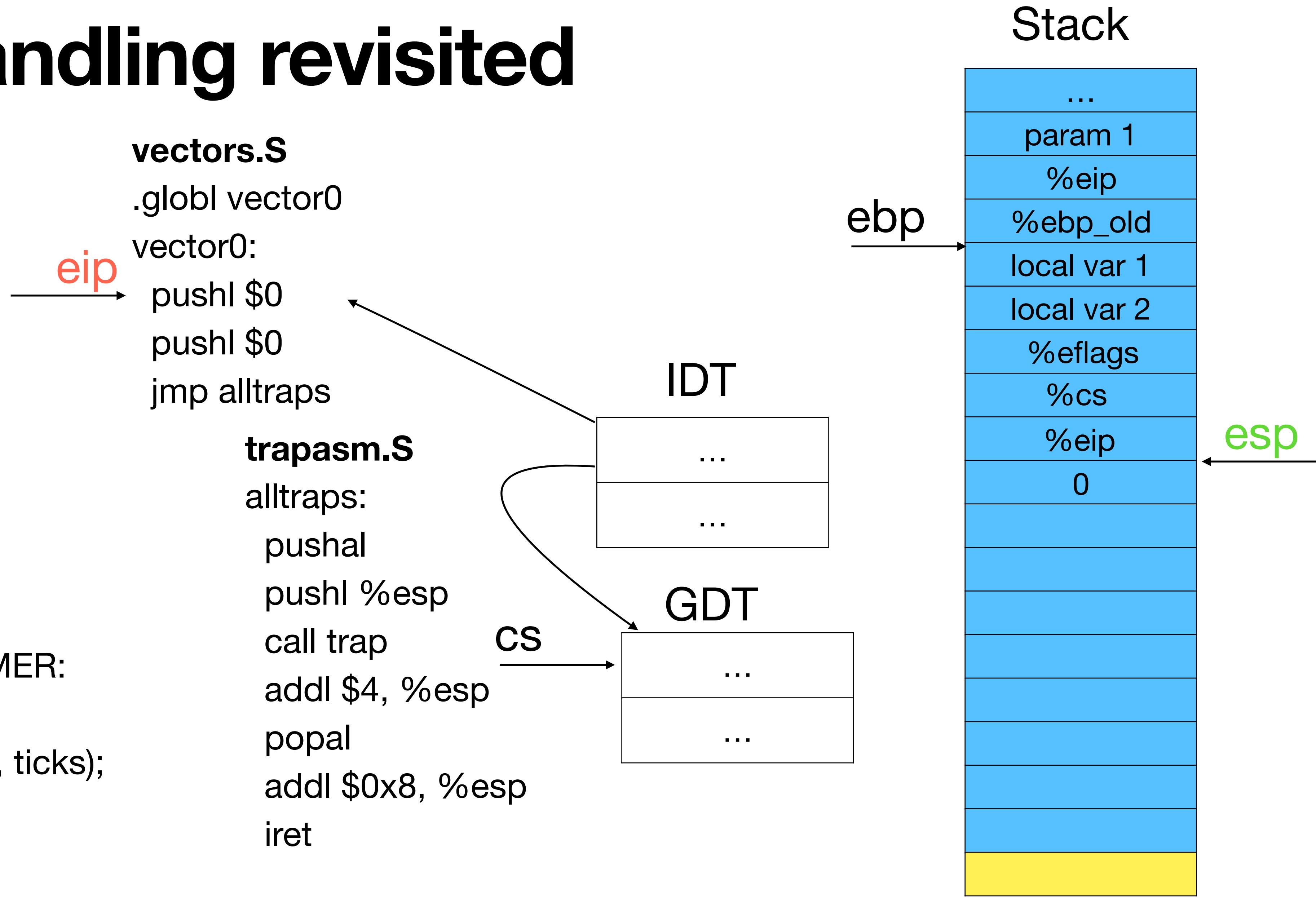
Interrupt handling revisited

```
for(;;)
;
trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
return
```



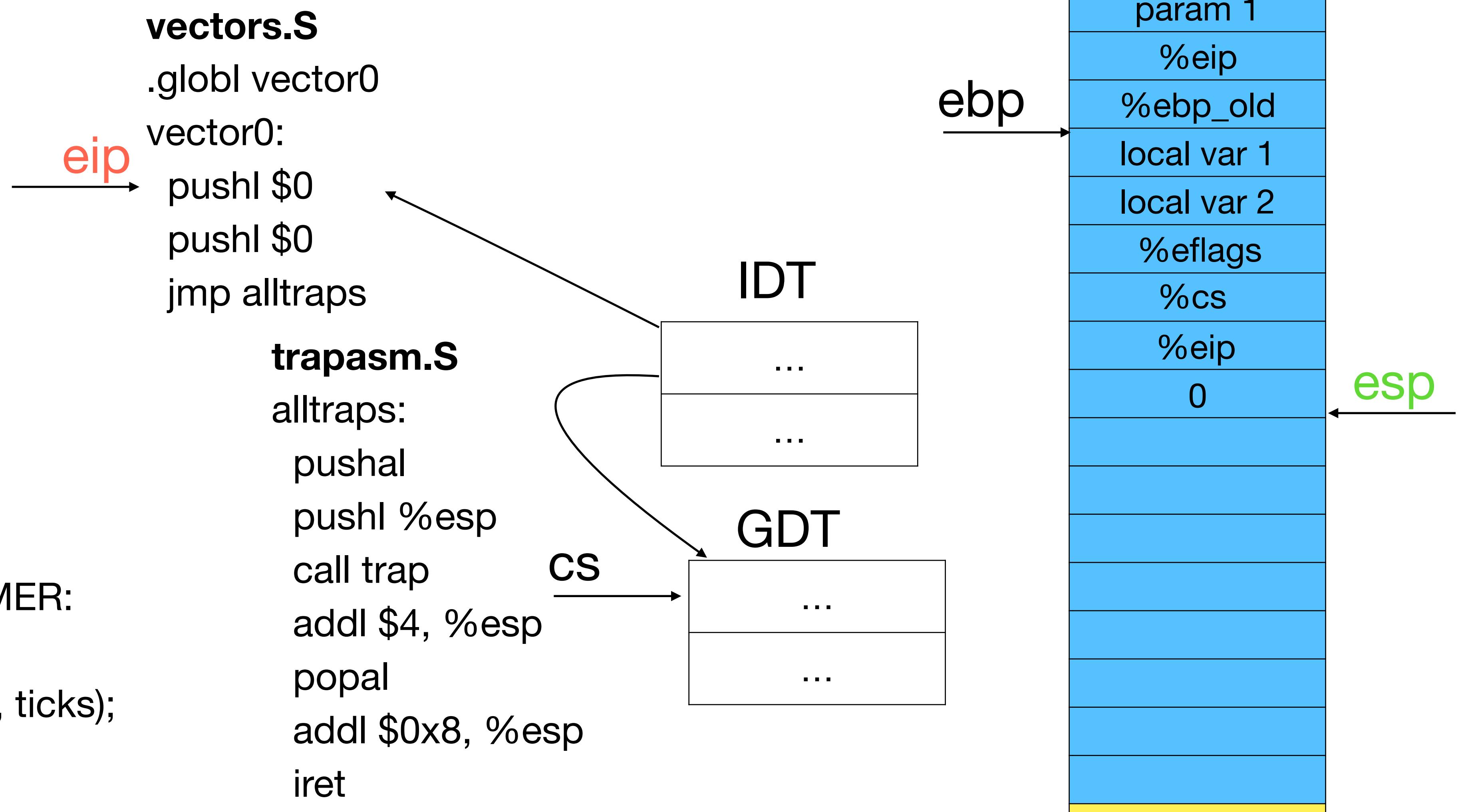
Interrupt handling revisited

```
for(;;)
;
trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
return
```



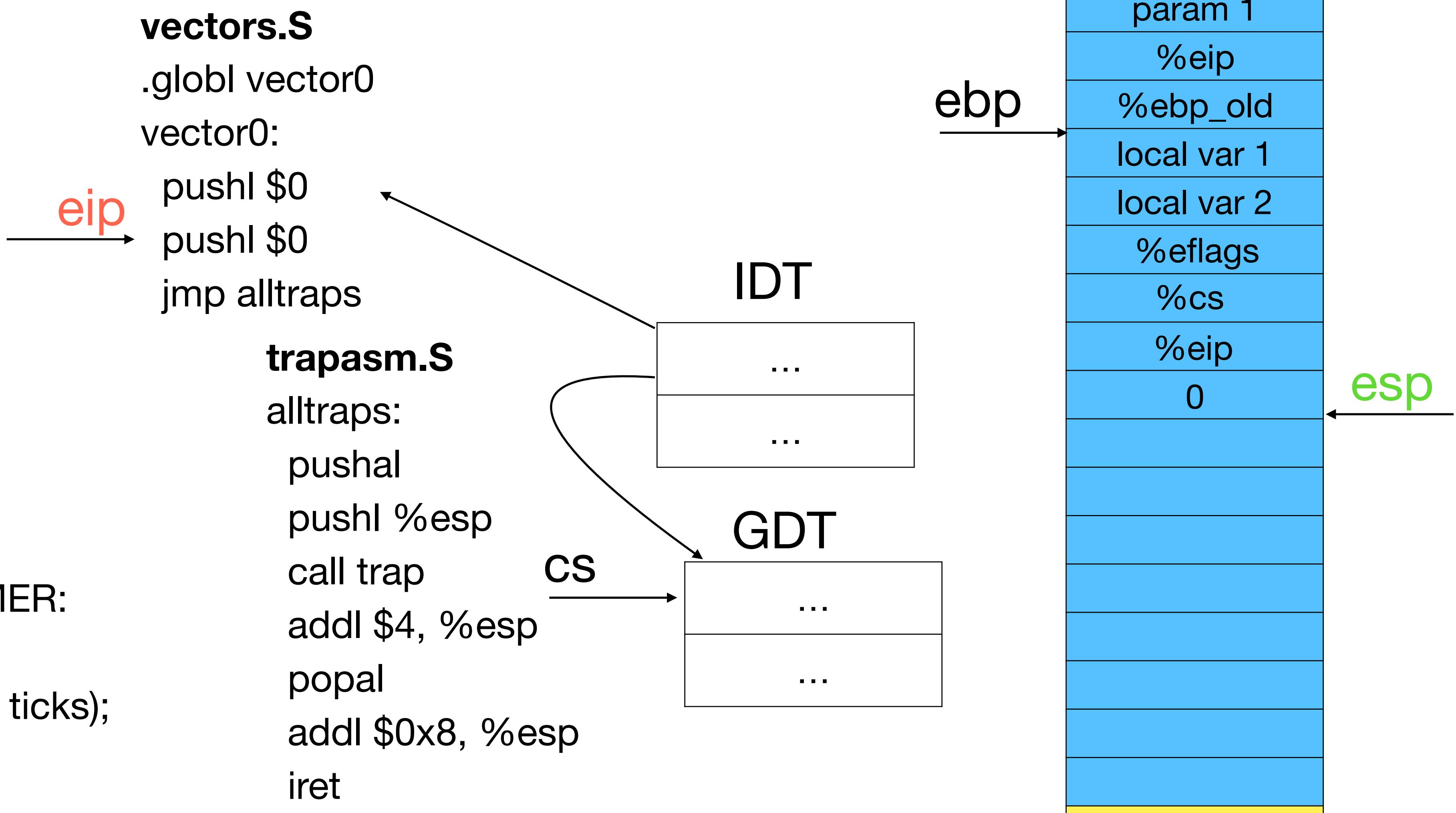
Interrupt handling revisited

```
for(;;)
;
trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
return
```



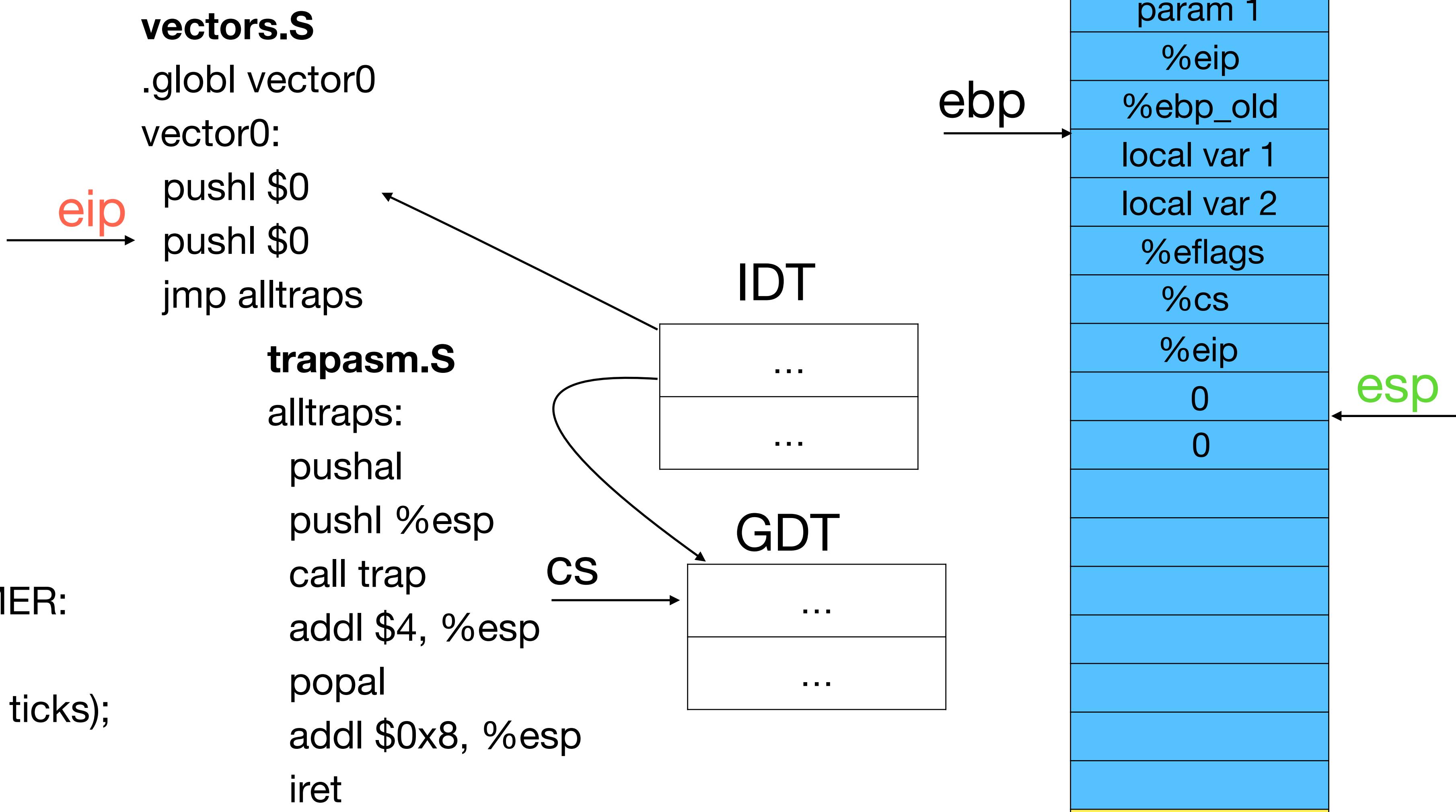
Interrupt handling revisited

```
for(;;)
;
trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
return
```



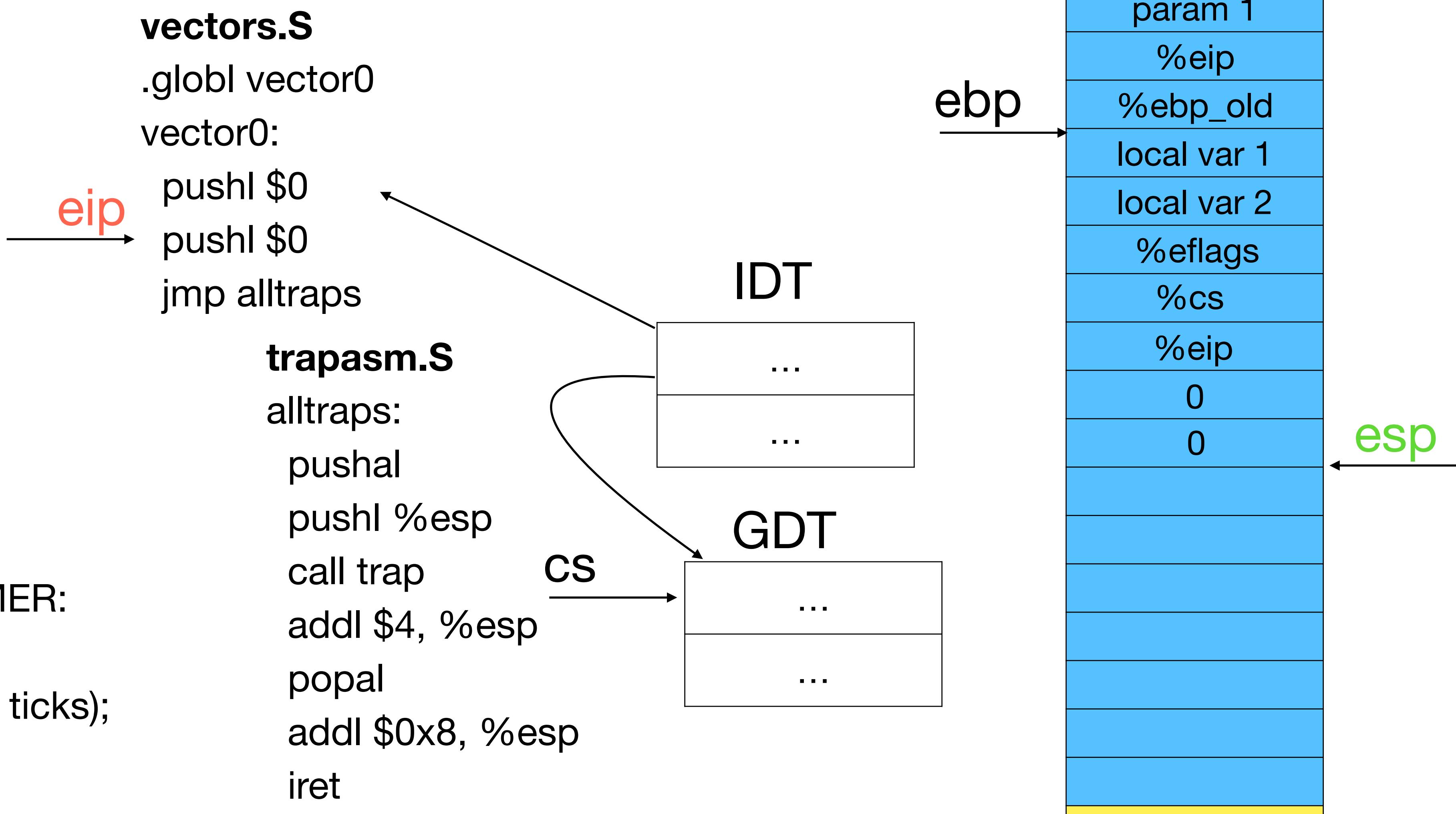
Interrupt handling revisited

```
for(;;)
;
trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
return
```



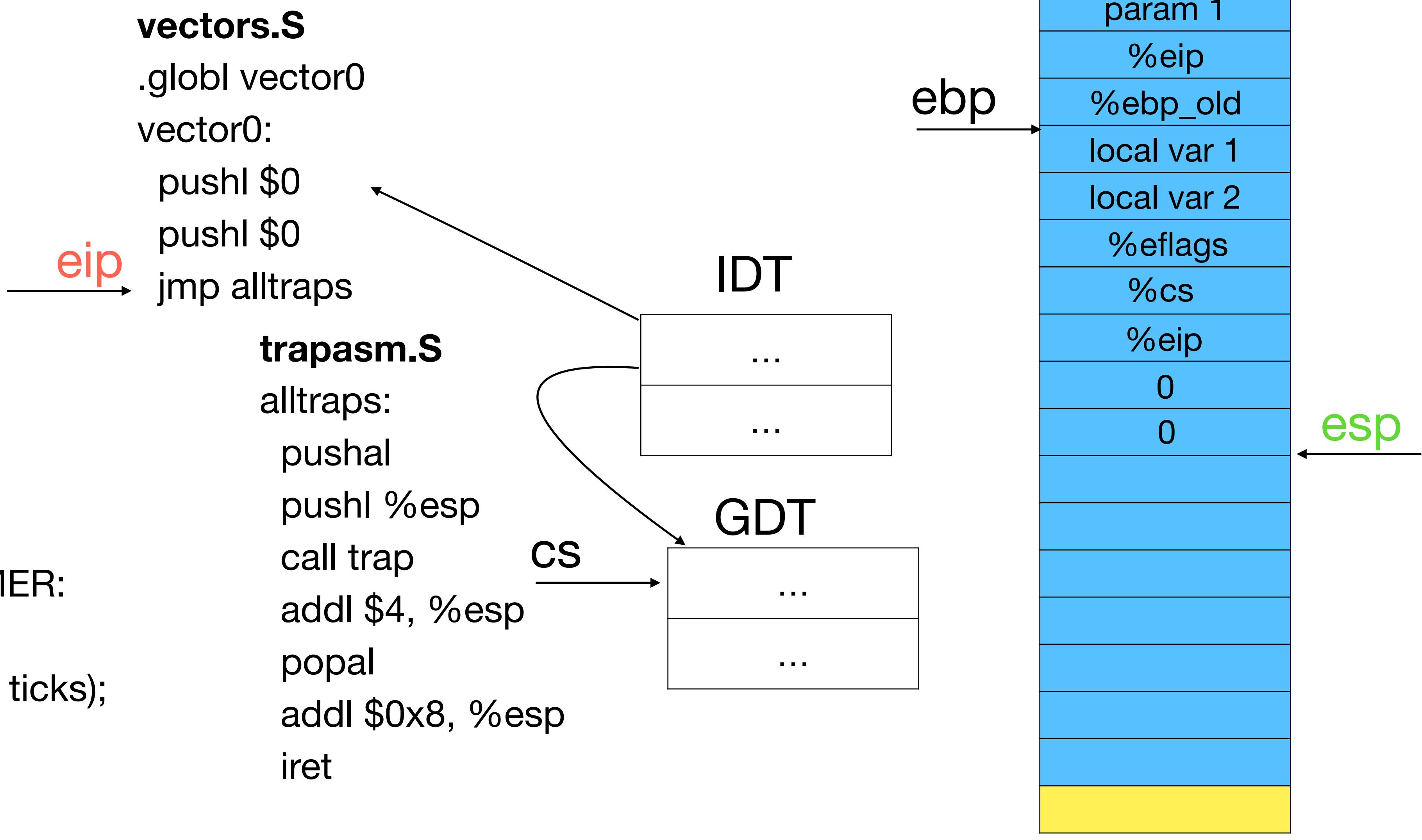
Interrupt handling revisited

```
for(;;)
;
trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
return
```



Interrupt handling revisited

```
for(;;)
;
trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
return
```



Interrupt handling revisited

```
for(;;)
;

trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
}

return
```

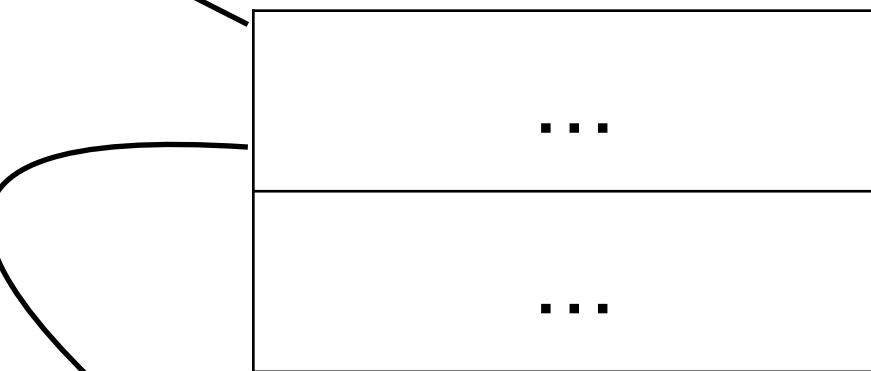
vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

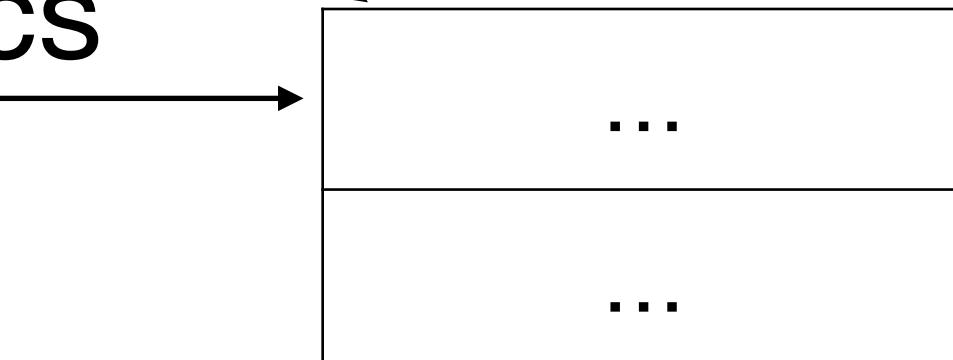
trapasm.S

```
eip → alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

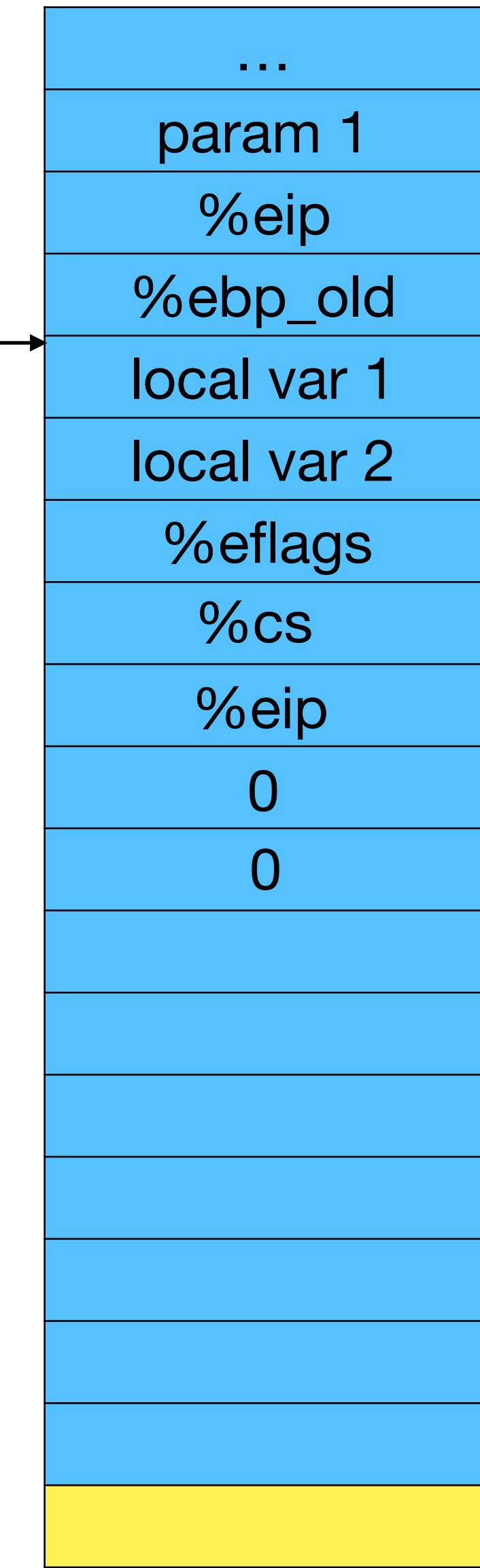
IDT



GDT



ebp



esp

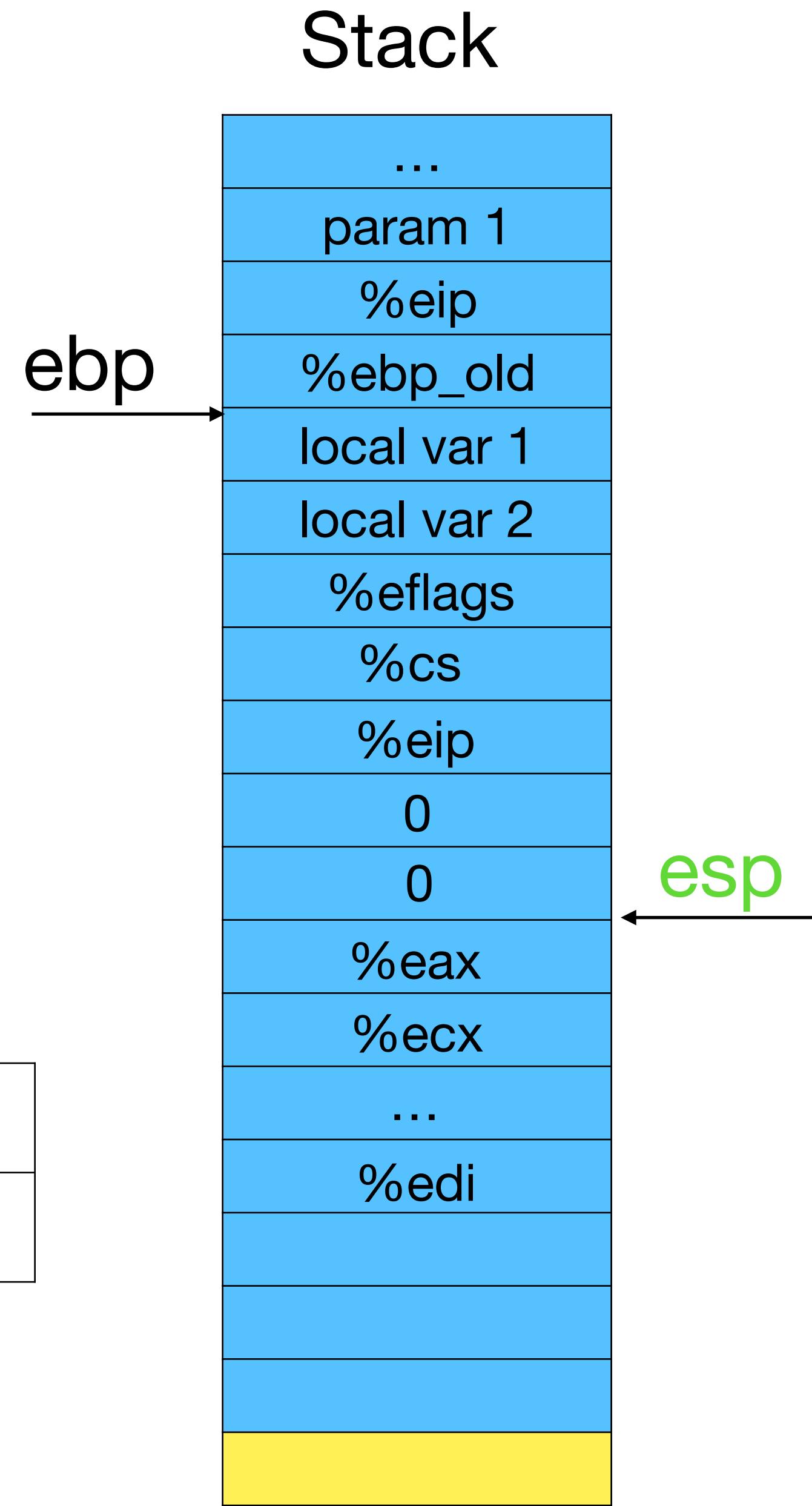
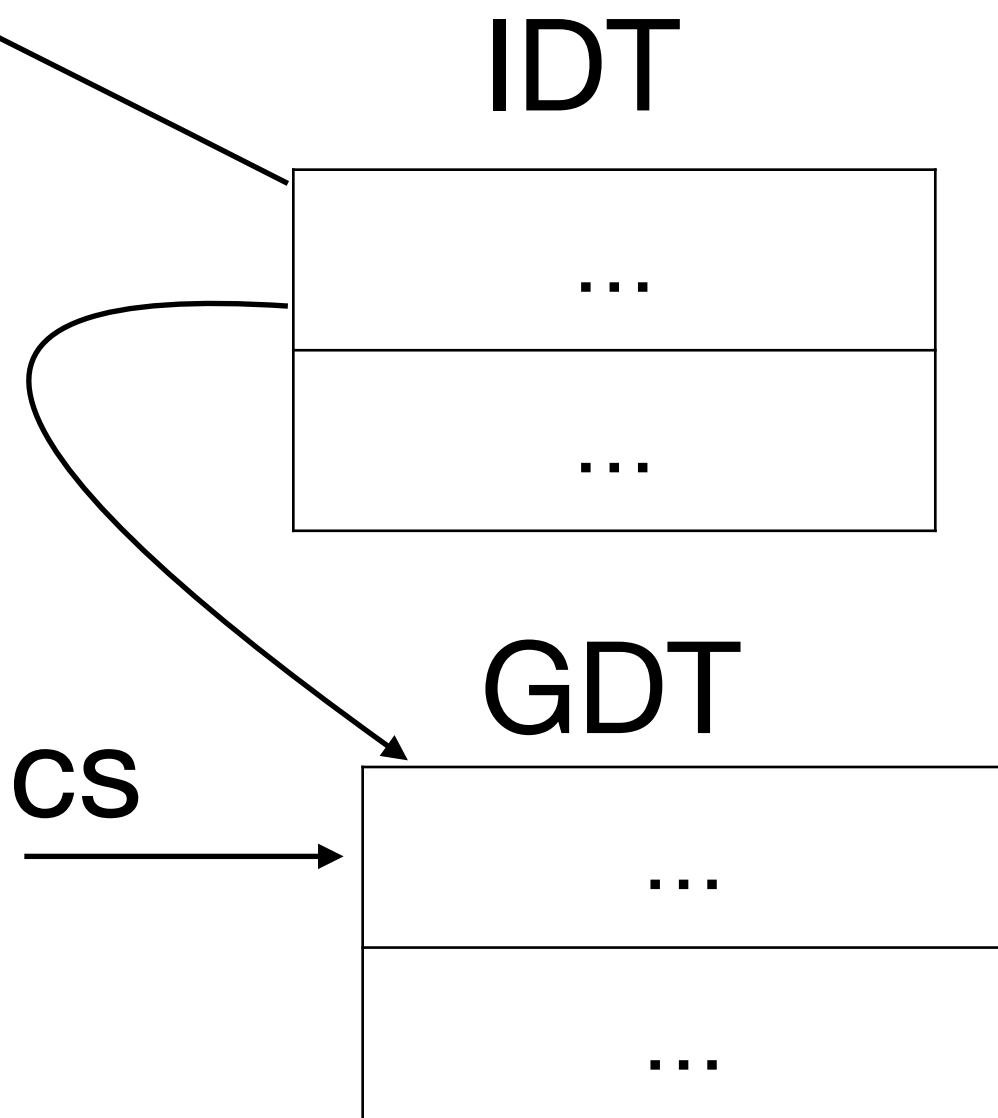
Interrupt handling revisited

```
for(;;)
;

trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
        ...
    return
}
```

```
vectors.S
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

```
trapasm.S
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```



Interrupt handling revisited

```
for(;;)
;

trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
    return
}
```

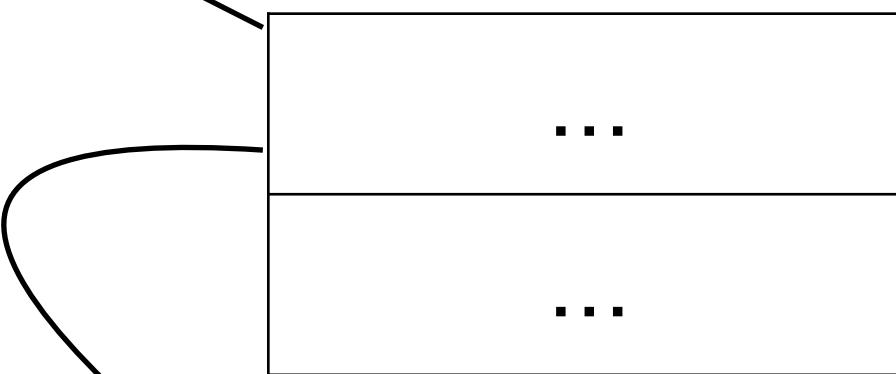
vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

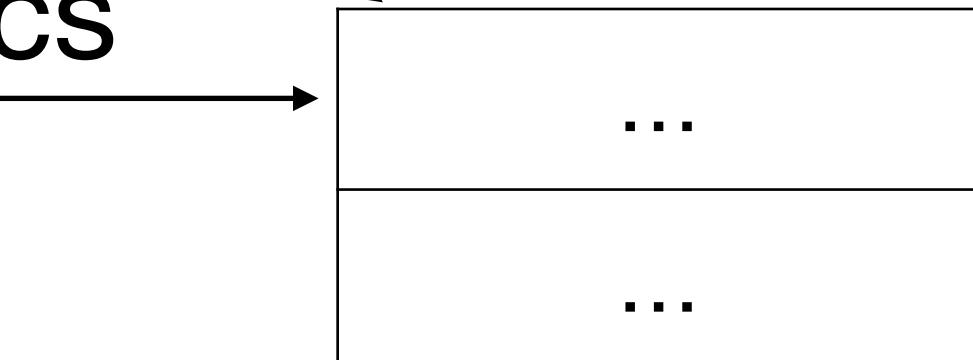
trapasm.S

```
eip → alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

IDT



GDT



ebp

Stack
...
param 1
%eip
%ebp_old
local var 1
local var 2
%eflags
%cs
%eip
0
0
%eax
%ecx
...
%edi
esp

Interrupt handling revisited

```
for(;;)
;

trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
}

return
```

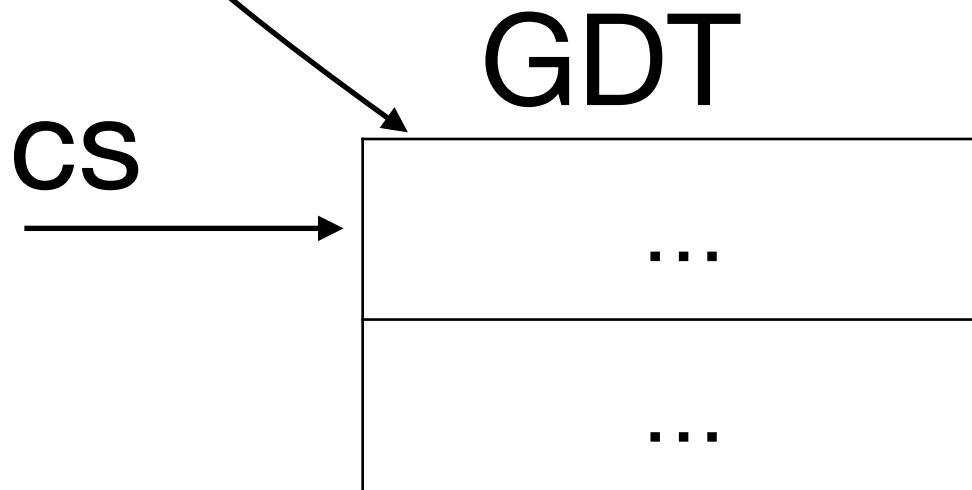
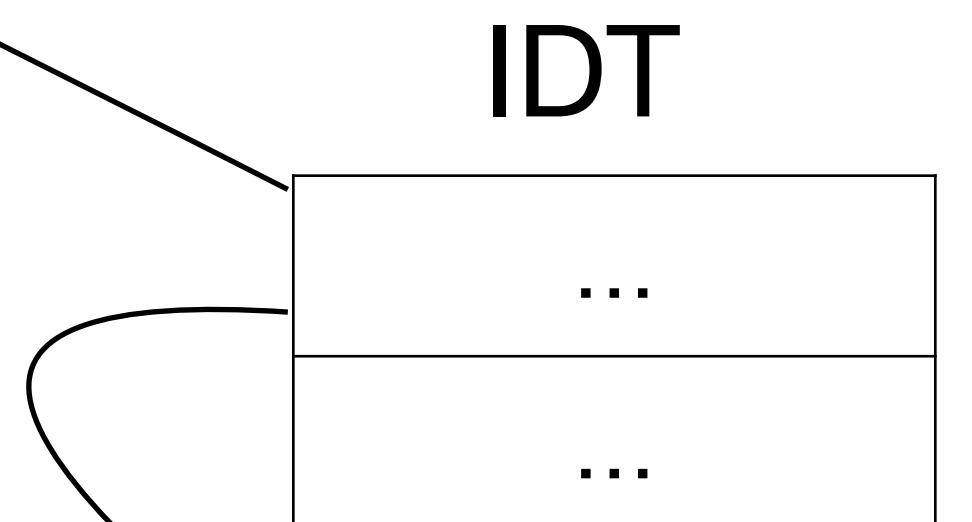
vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

trapasm.S

```
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

eip



ebp

Stack

...
param 1
%eip
%ebp_old
local var 1
local var 2
%eflags
%cs
%eip
0
0
%eax
%ecx
...
%edi
esp

esp

Interrupt handling revisited

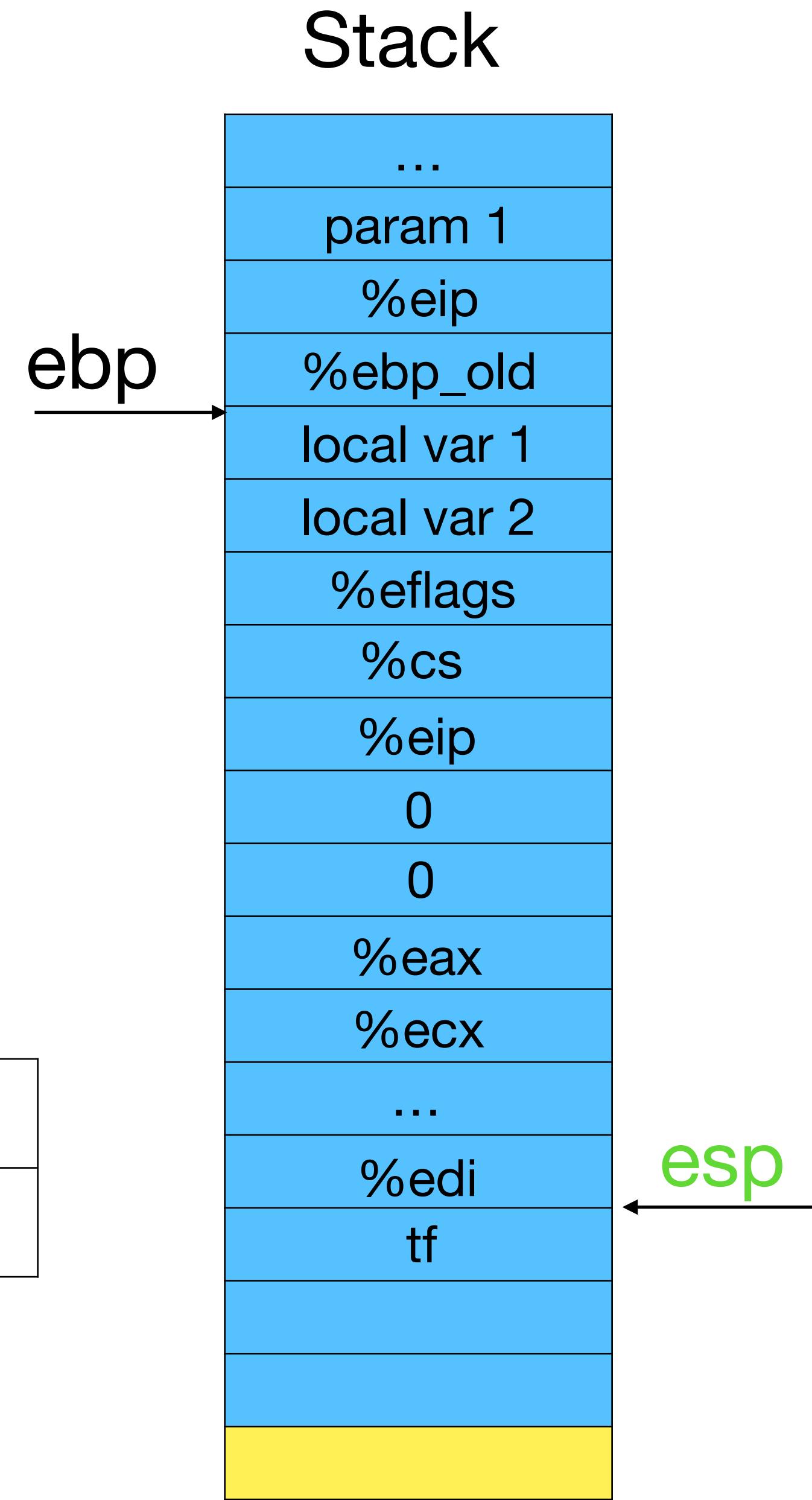
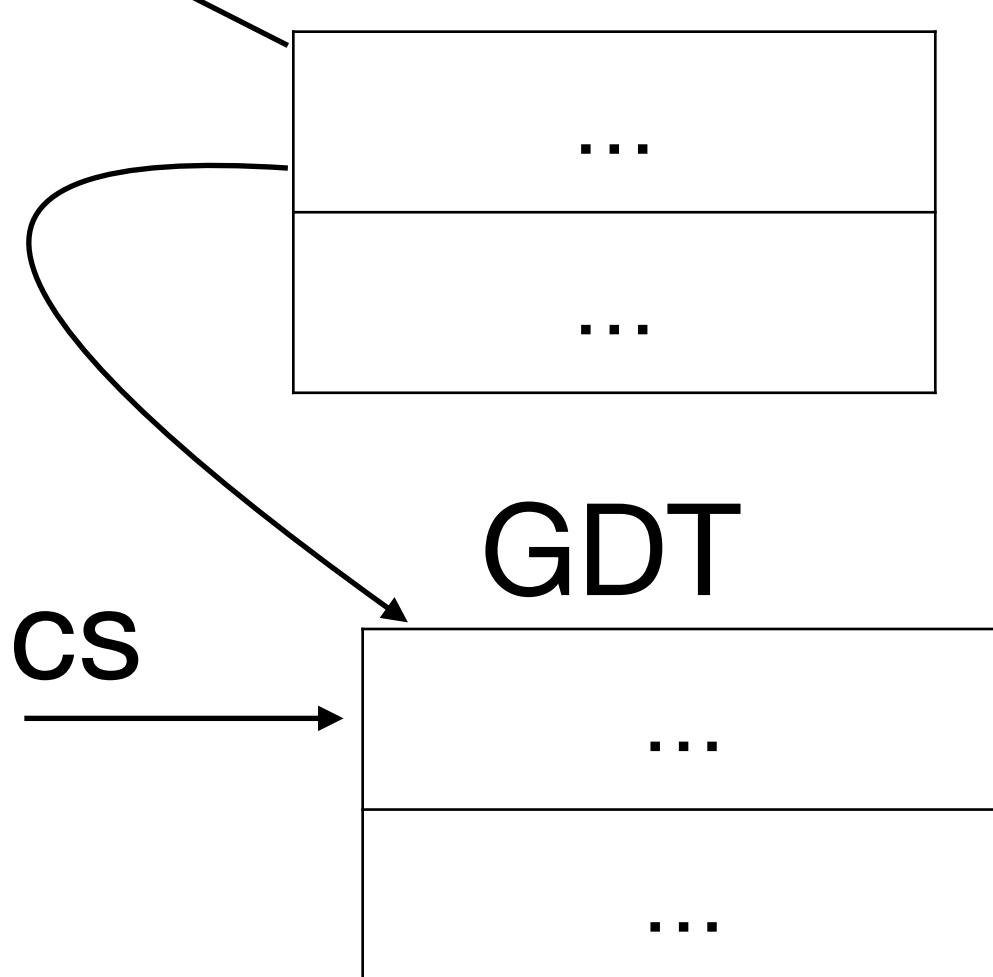
```
for(;;)
;

trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
    return
}
```

```
vectors.S
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

```
trapasm.S
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

eip



Interrupt handling revisited

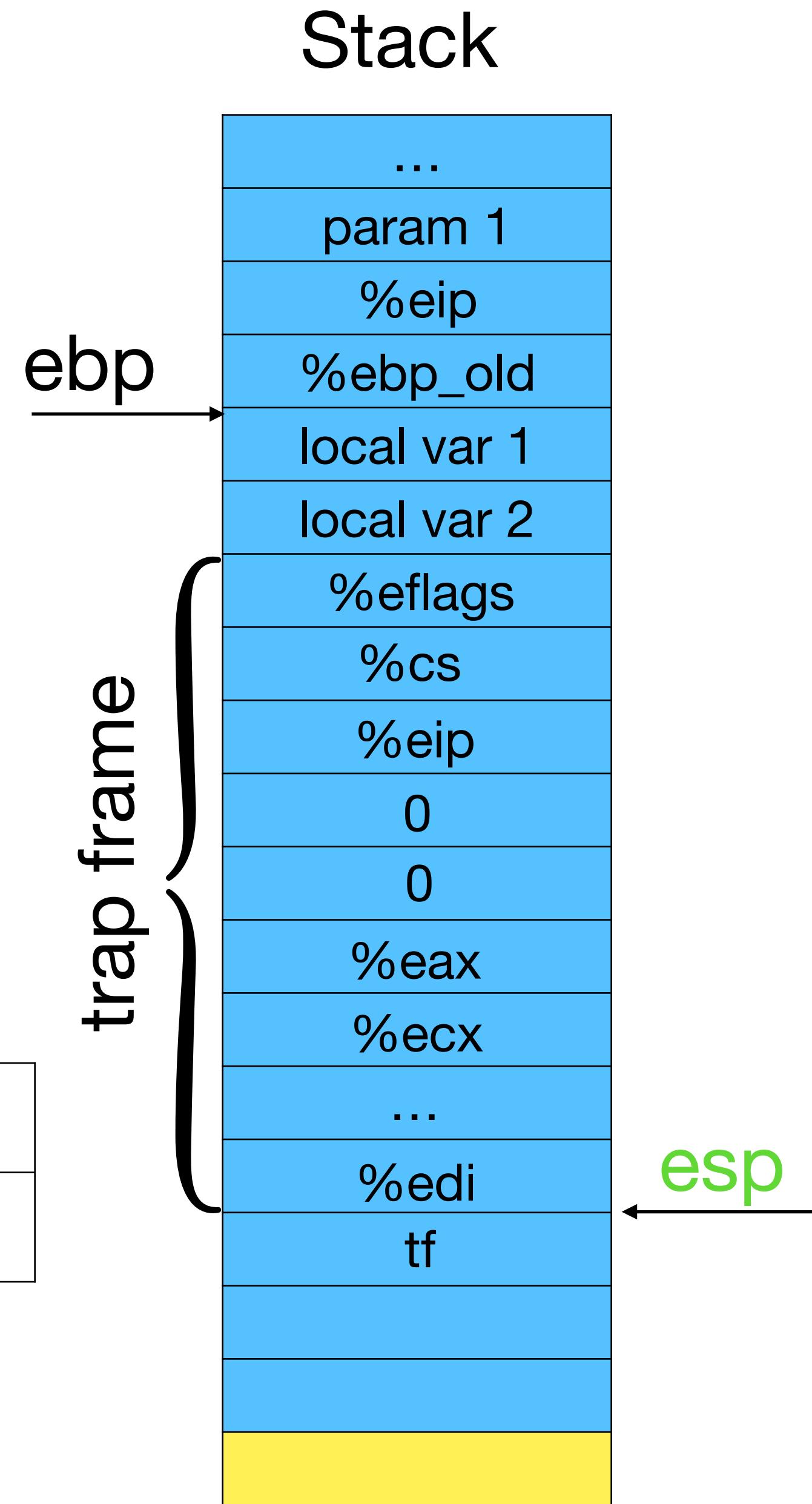
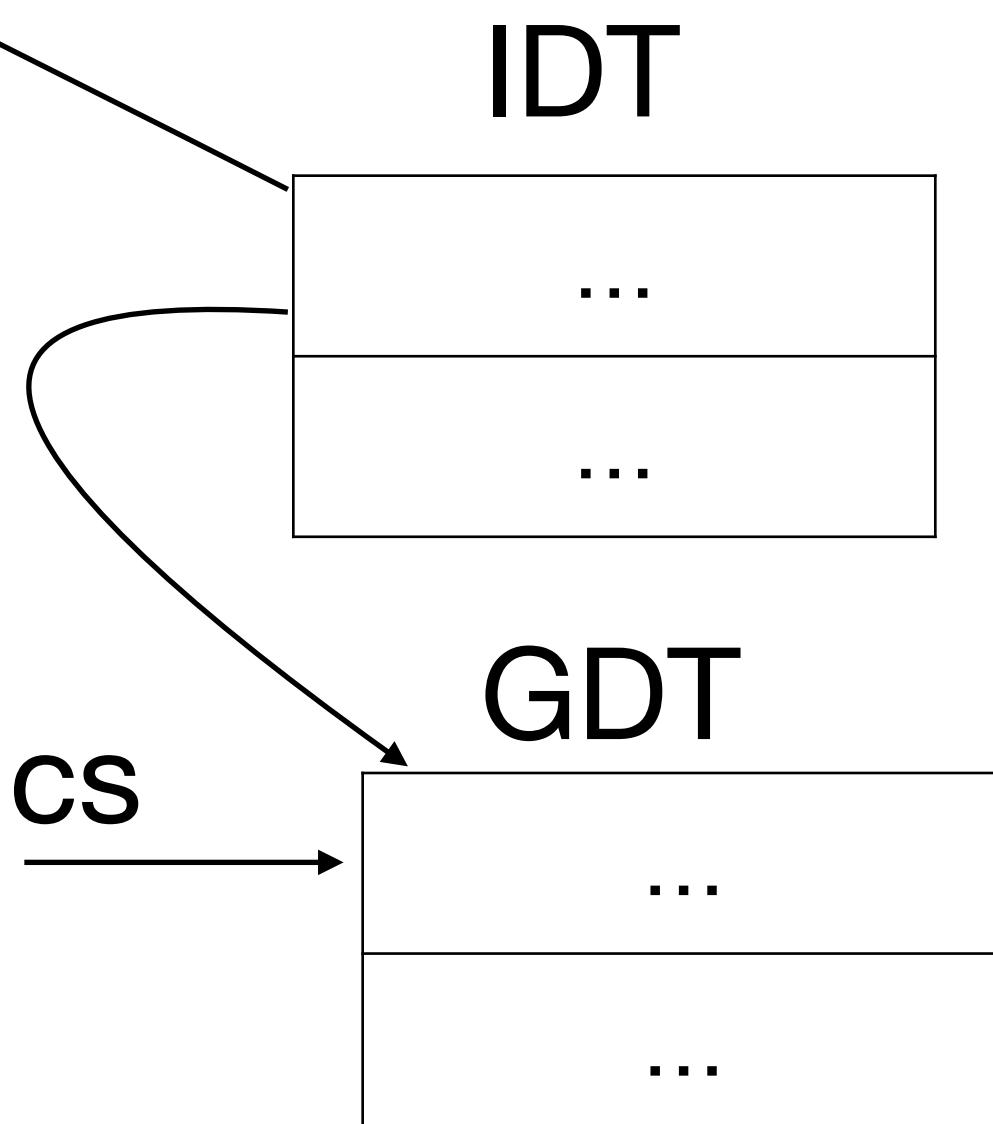
```
for(;;)
;

trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
    return
}
```

```
vectors.S
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

```
trapasm.S
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

eip



Interrupt handling revisited

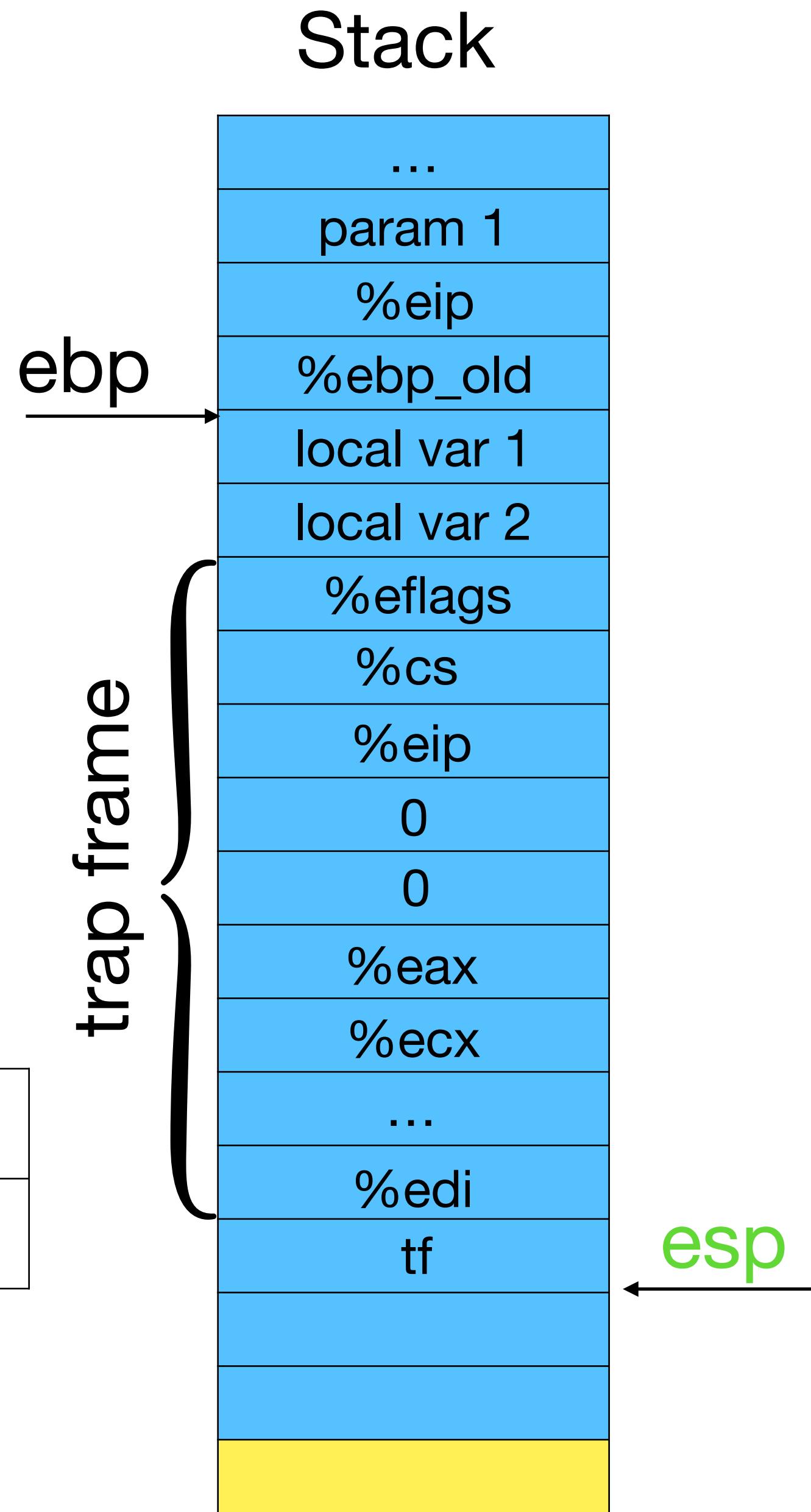
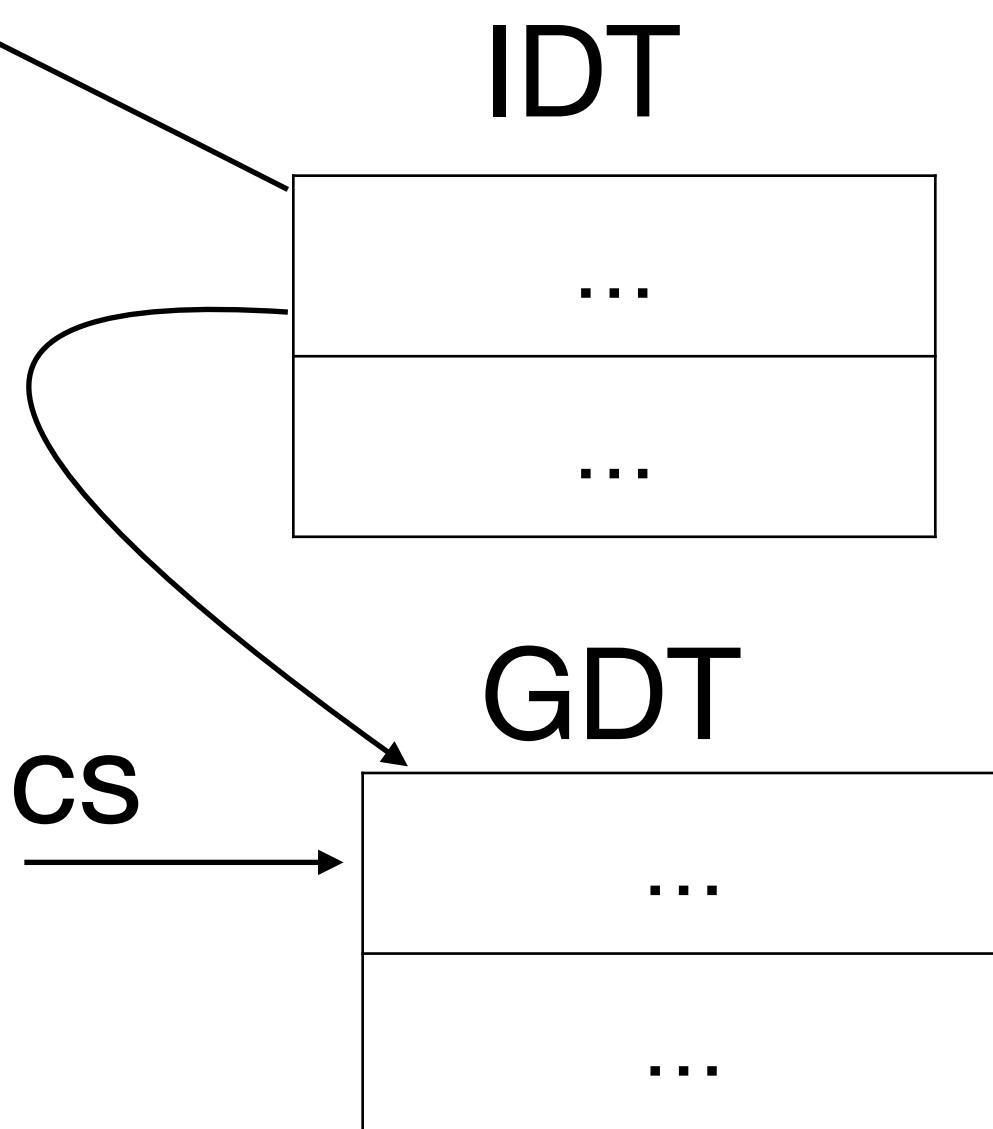
```
for(;;)
;

trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
    return
}
```

```
vectors.S
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

```
trapasm.S
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

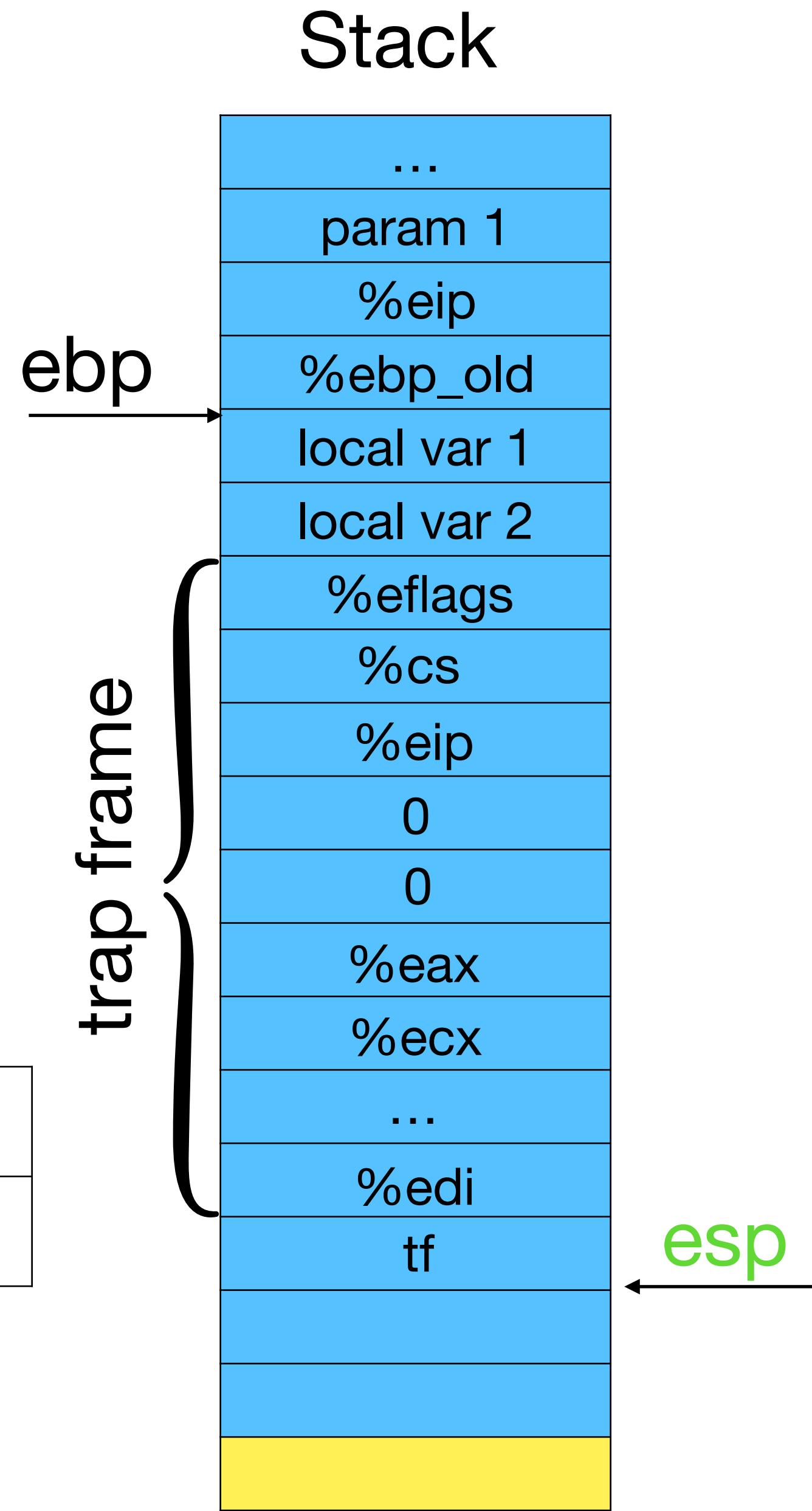
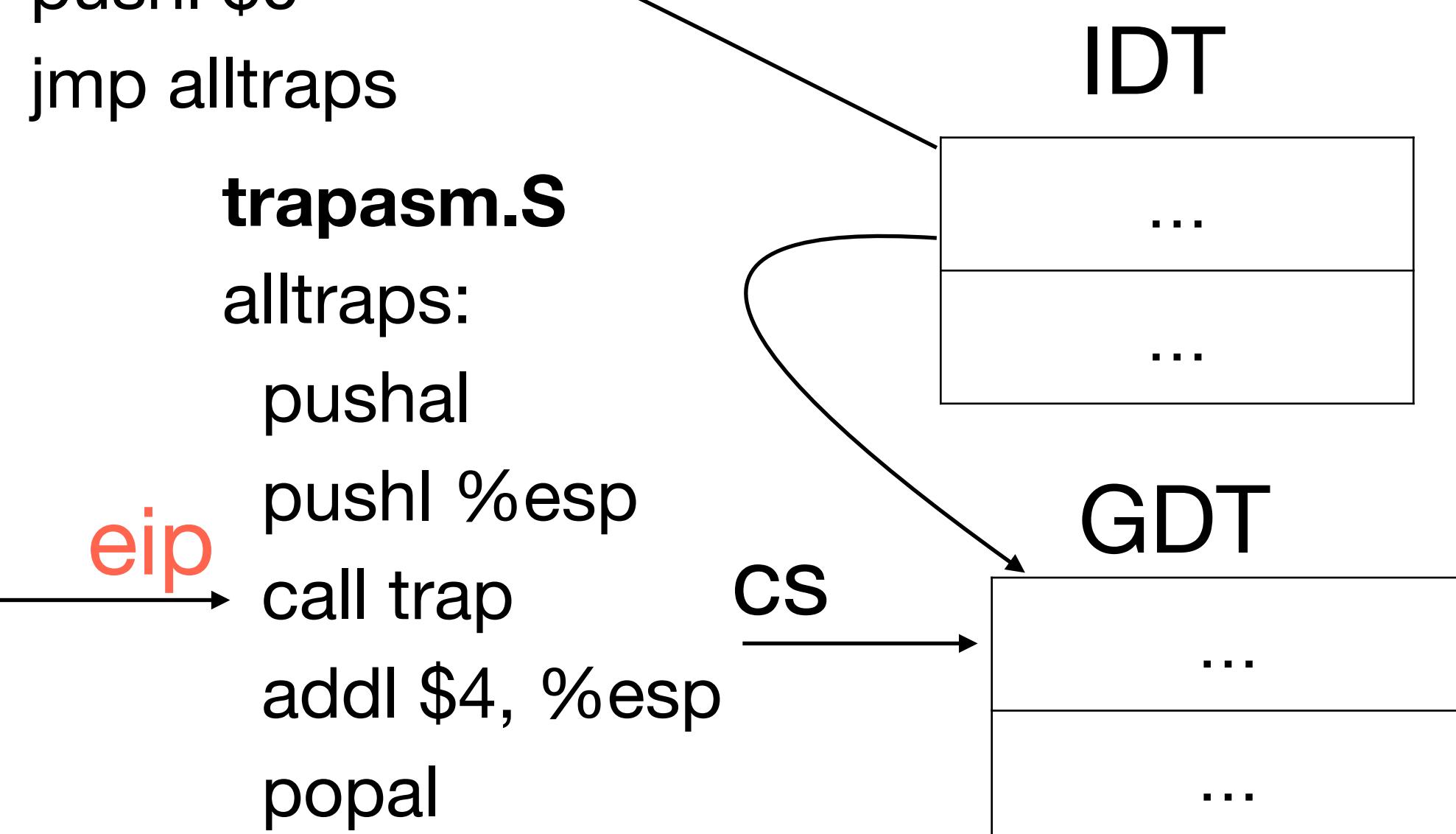
eip



Interrupt handling revisited

```
for(;;)  
;  
  
trap.c  
void  
trap(struct trapframe *tf)  
{  
    switch(tf->trapno){  
        case T_IRQ0 + IRQ_TIMER:  
            ticks++;  
            cprintf("Tick! %d\n", ticks);  
            lapiceoi();  
            ...  
    }  
    return  
}
```

```
vectors.S  
.globl vector0  
vector0:  
    pushl $0  
    pushl $0  
    jmp alltraps  
  
trapasm.S  
alltraps:  
    pushal  
    pushl %esp  
    call trap  
    addl $4, %esp  
    popal  
    addl $0x8, %esp  
    iret
```



Interrupt handling revisited

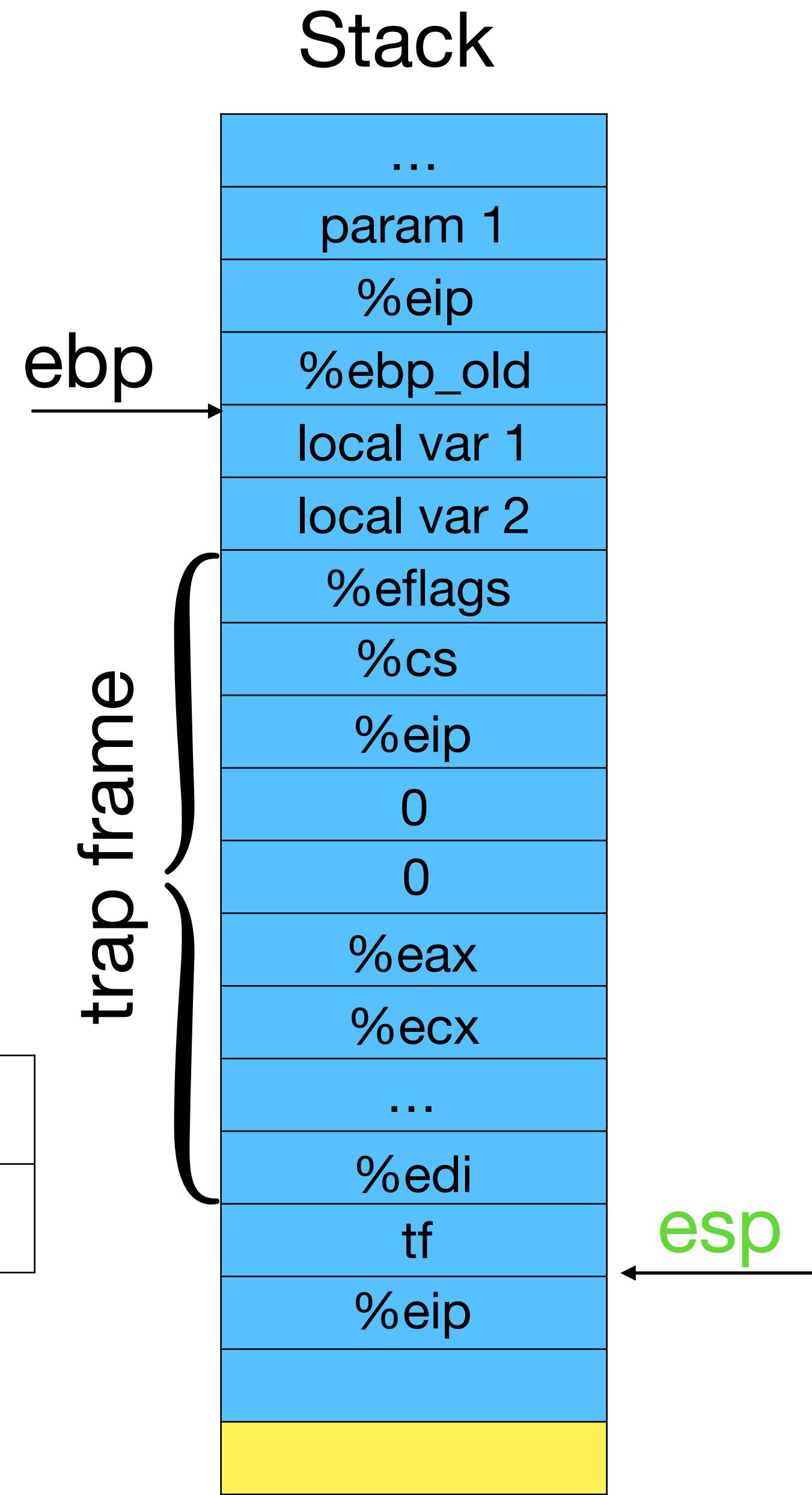
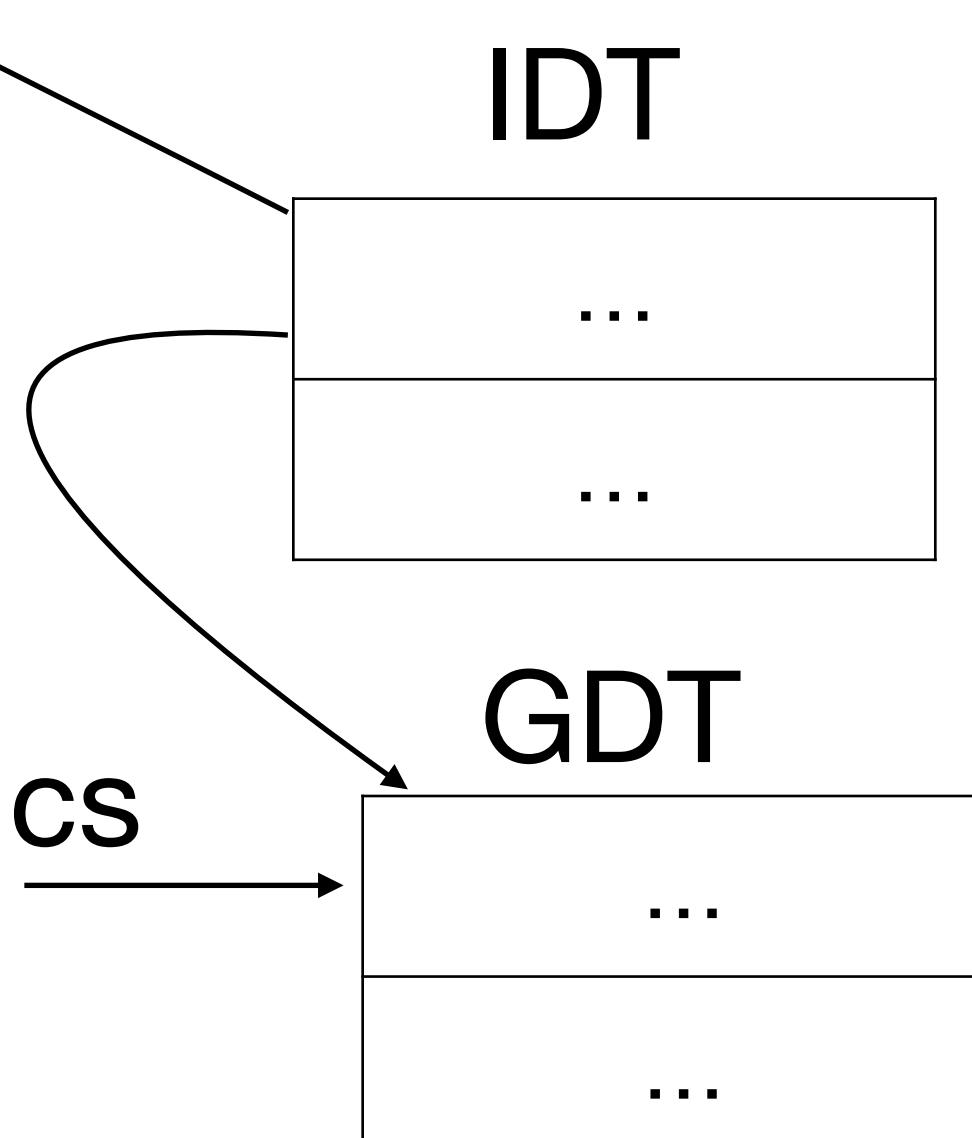
```
for(;;)
;

trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
    return
}
```

```
vectors.S
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

```
trapasm.S
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

eip



Interrupt handling revisited

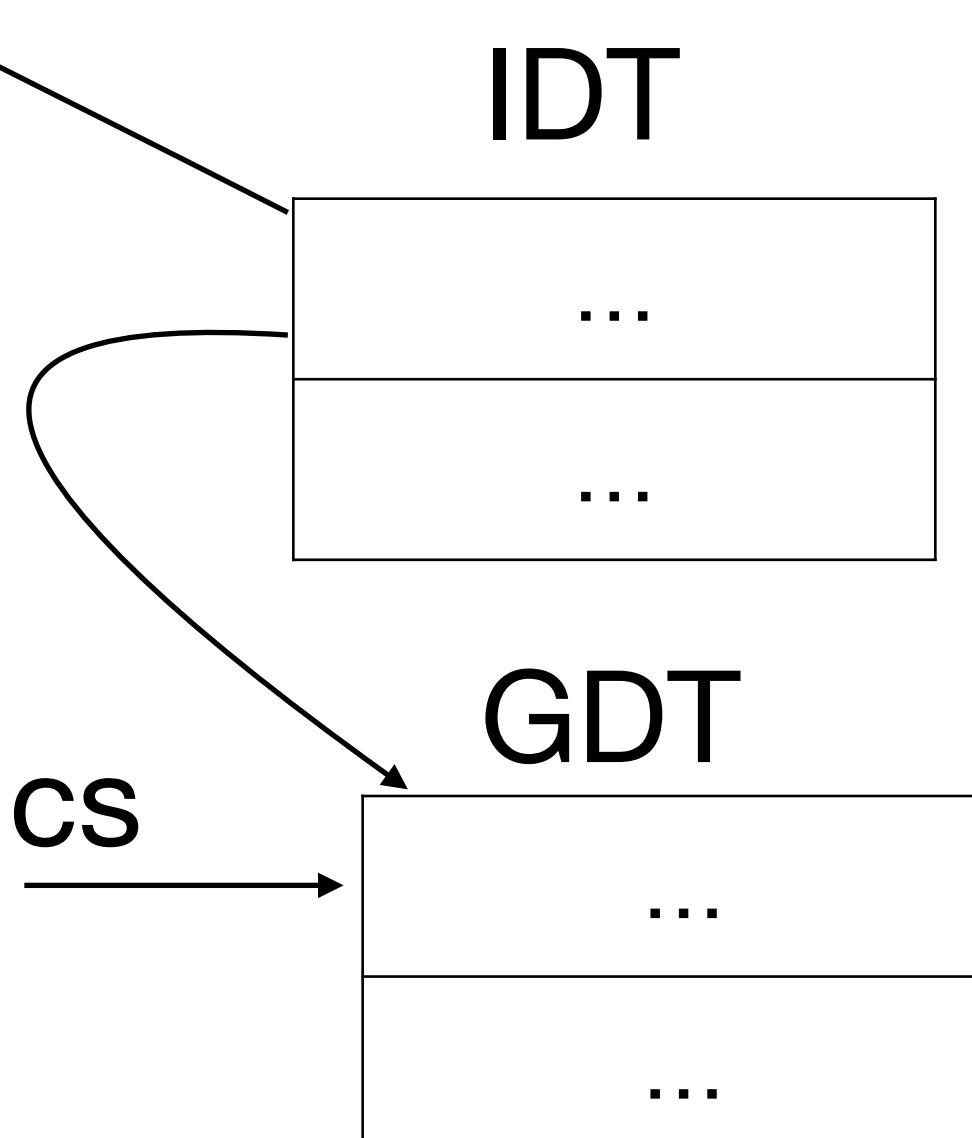
```
for(;;)
;

trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
    return
}
```

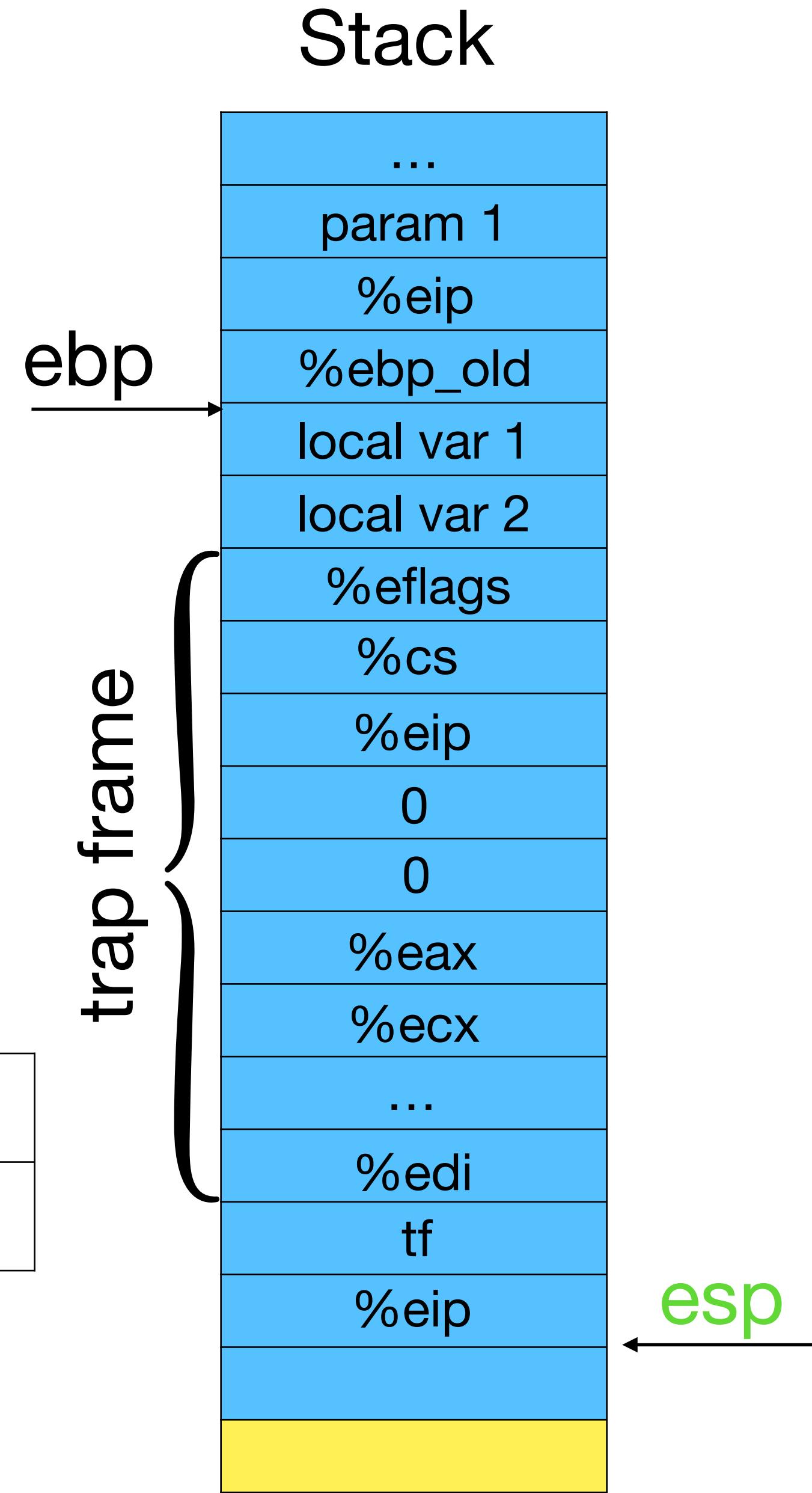
```
vectors.S
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

```
trapasm.S
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

eip



CS



esp

Interrupt handling revisited

```
for(;;)
;

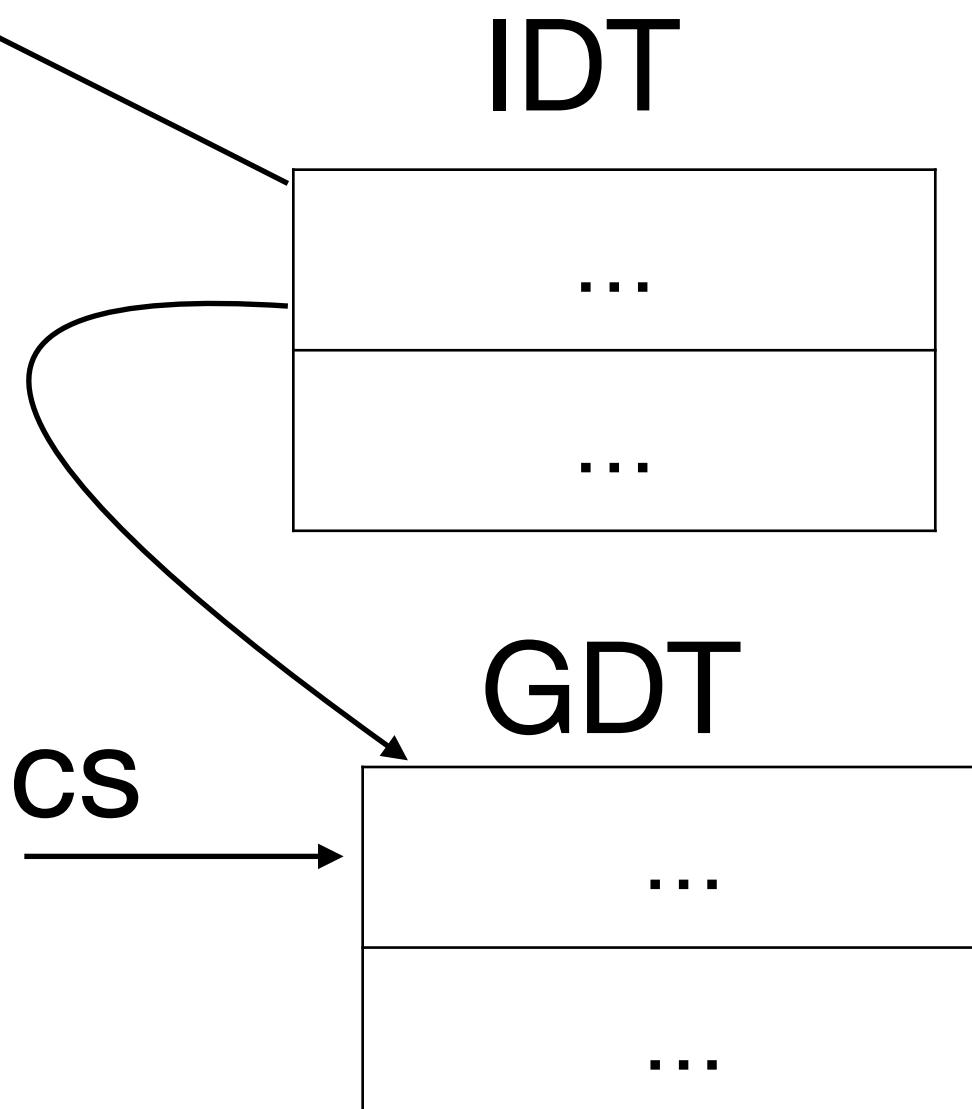
trap.c
void
trap(struct trapframe *tf)
{
    eip → switch(tf->trapno){
        case T_IRQ0 + IRQ_TIMER:
            ticks++;
            cprintf("Tick! %d\n", ticks);
            lapiceoi();
    ...
    return
}
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

trapasm.S

```
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```



Stack

...
param 1
%eip
%ebp_old
local var 1
local var 2
%eflags
%cs
%eip
0
0
%eax
%ecx
...
%edi
tf
%eip

Interrupt handling revisited

```
for(;;)
;

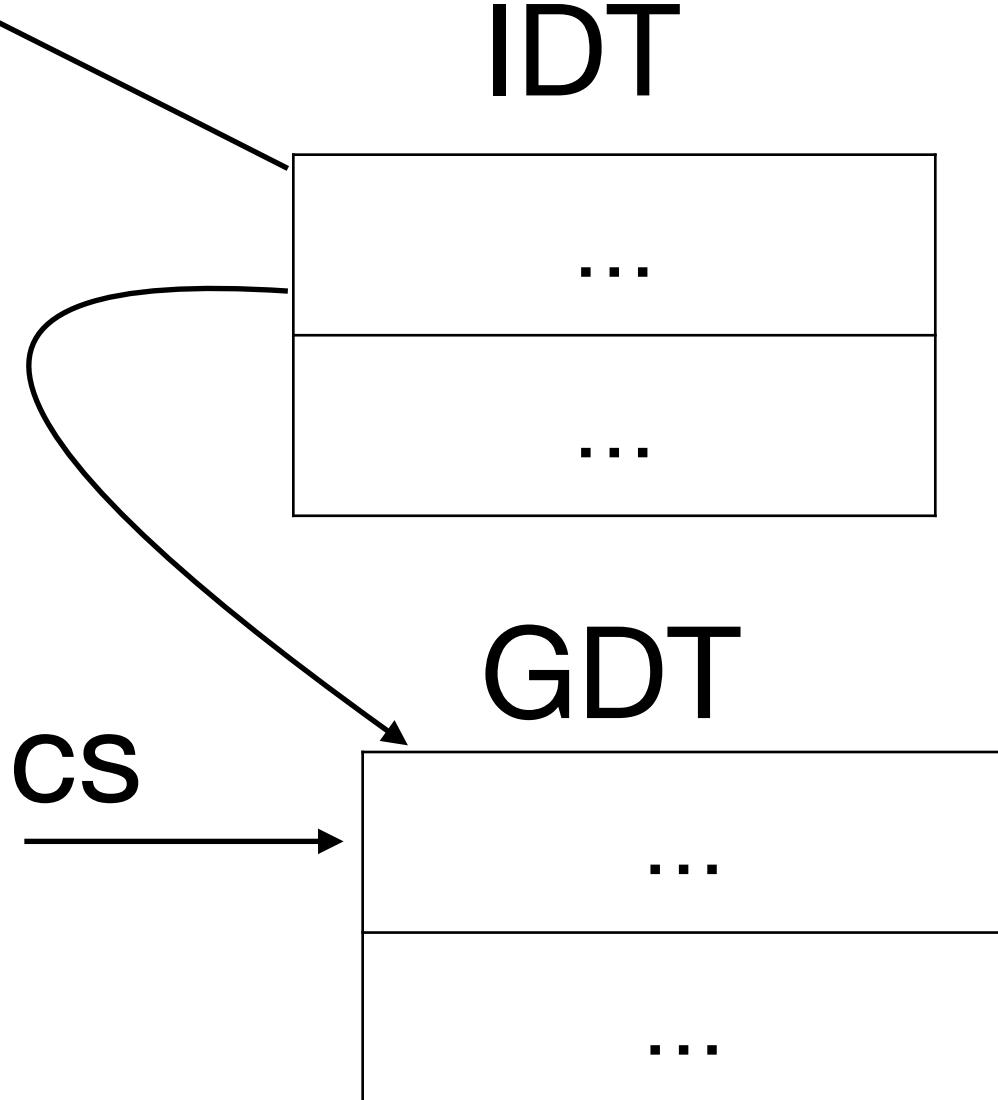
trap.c
void
trap(struct trapframe *tf)
{
    eip → switch(tf->trapno){
        case T_IRQ0 + IRQ_TIMER:
            ticks++;
            cprintf("Tick! %d\n", ticks);
            lapiceoi();
    ...
    return
}
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

trapasm.S

```
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```



Stack

...
param 1
%eip
%ebp_old
local var 1
local var 2
%eflags
%cs
%eip
0
0
%eax
%ecx
...
%edi
tf
%eip

esp

Interrupt handling revisited

```
for(;;)
;

trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
}

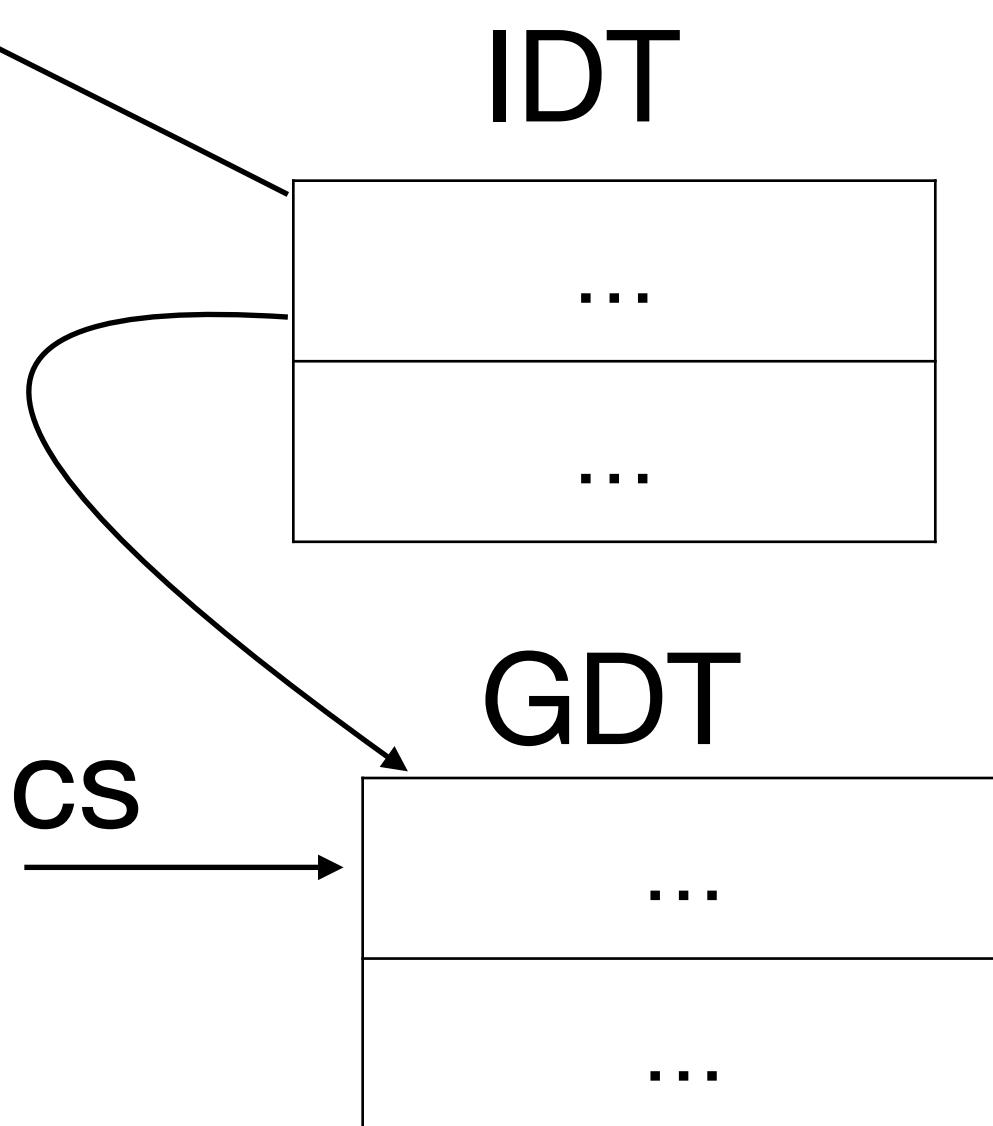
eip → return
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

trapasm.S

```
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```



Stack

...
param 1
%eip
%ebp_old
local var 1
local var 2
%eflags
%cs
%eip
0
0
%eax
%ecx
...
%edi
tf
%eip
...

ebp

trap frame

esp

Interrupt handling revisited

```
for(;;)
;

trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
}

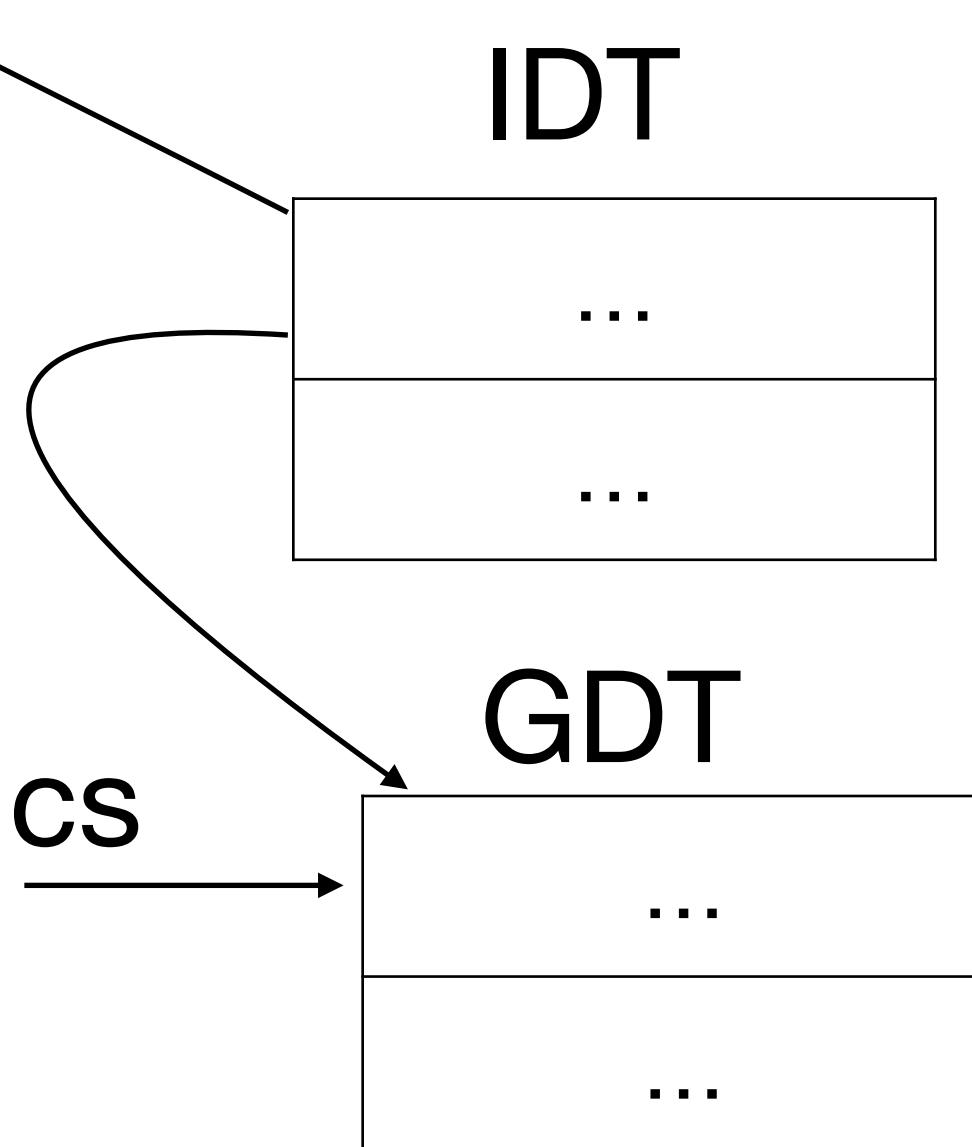
eip → return
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

trapasm.S

```
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```



Stack

...
param 1
%eip
%ebp_old
local var 1
local var 2
%eflags
%cs
%eip
0
0
%eax
%ecx
...
%edi
tf
%eip
esp

Interrupt handling revisited

```
for(;;)
;

trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
        case T_IRQ0 + IRQ_TIMER:
            ticks++;
            cprintf("Tick! %d\n", ticks);
            lapiceoi();
            ...
    }
    return
}
```

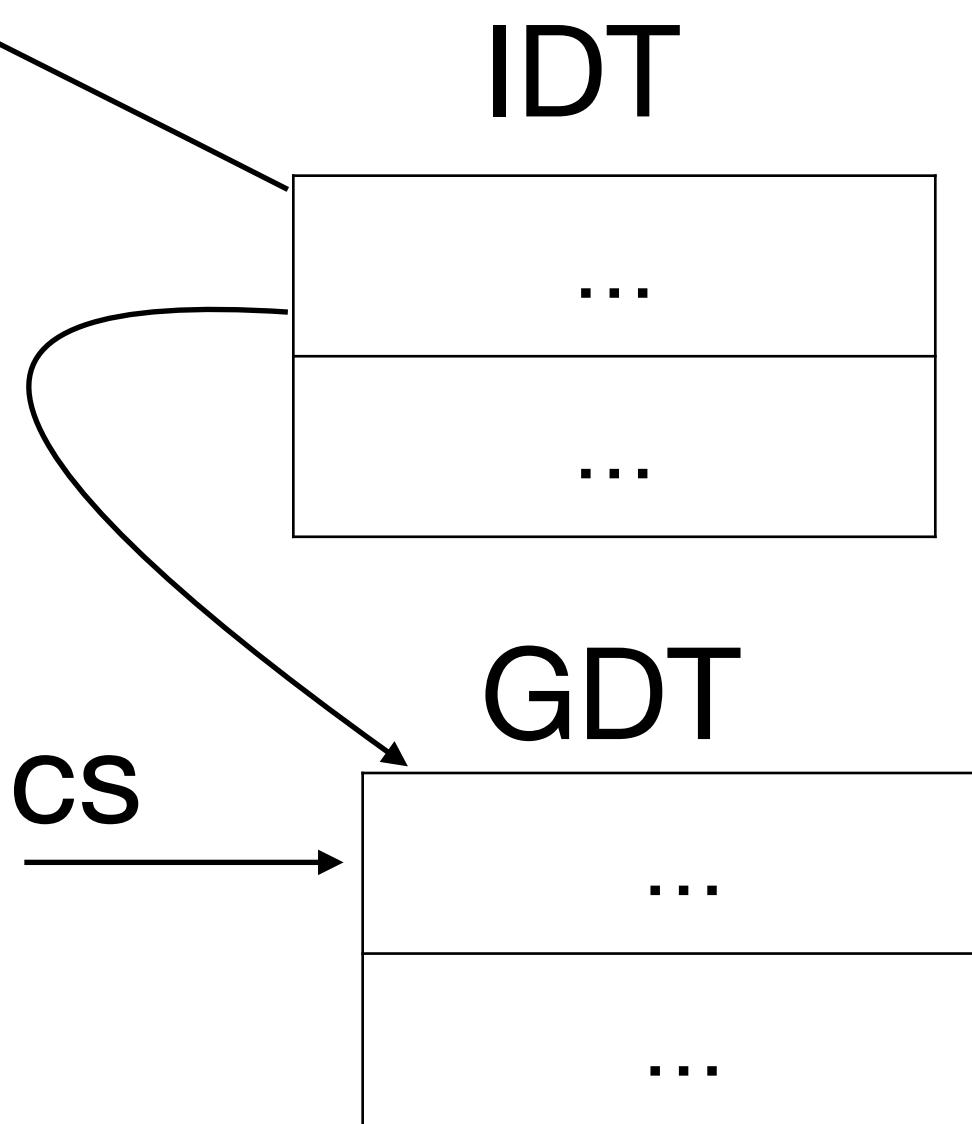
vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

trapasm.S

```
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

eip



Stack

...
param 1
%eip
%ebp_old
local var 1
local var 2
%eflags
%cs
%eip
0
0
%eax
%ecx
...
%edi
tf
%eip
esp

Interrupt handling revisited

```
for(;;)
;

trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
        case T_IRQ0 + IRQ_TIMER:
            ticks++;
            cprintf("Tick! %d\n", ticks);
            lapiceoi();
            ...
    }
    return
}
```

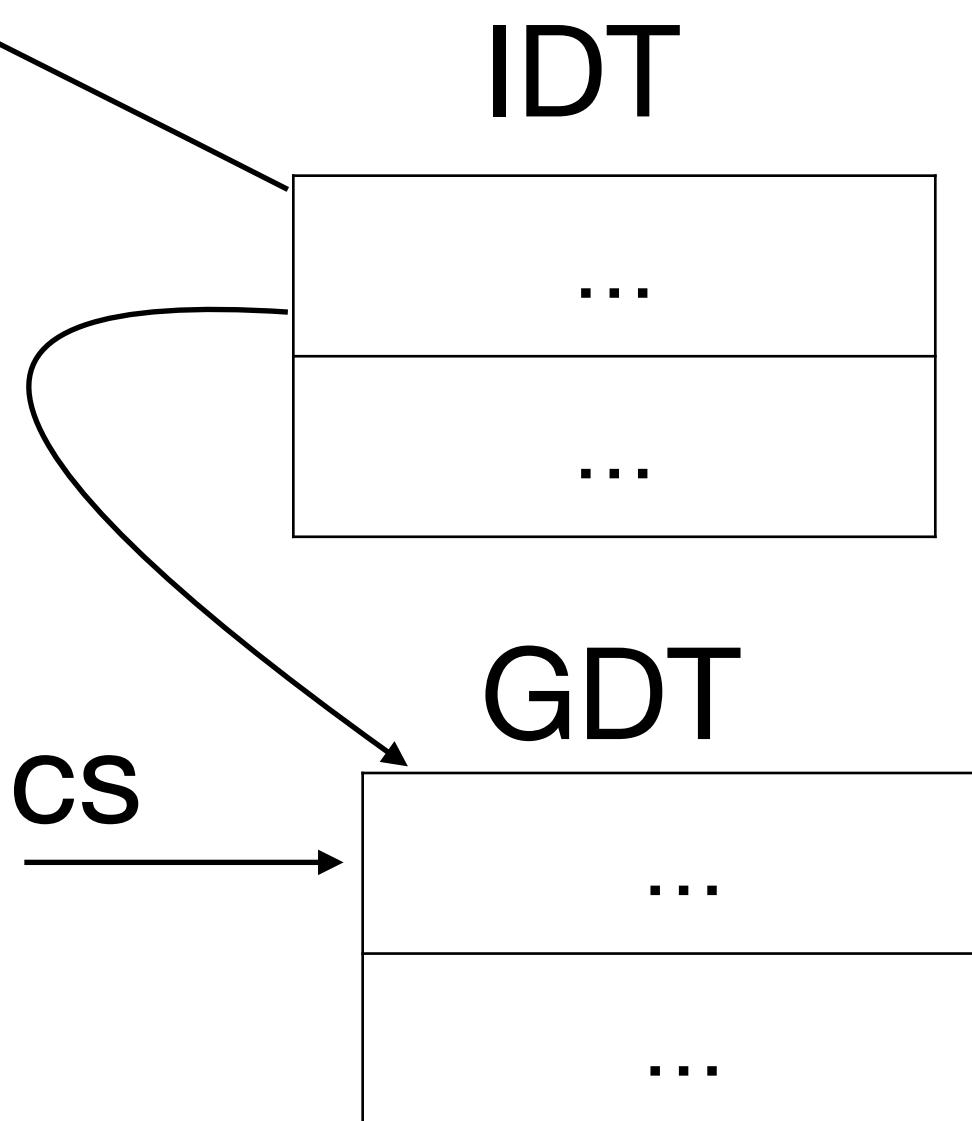
vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

trapasm.S

```
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

eip



Stack

...
param 1
%eip
%ebp_old
local var 1
local var 2
%eflags
%cs
%eip
0
0
%eax
%ecx
...
%edi
tf
%eip

ebp

trap frame

esp

Interrupt handling revisited

```
for(;;)
;

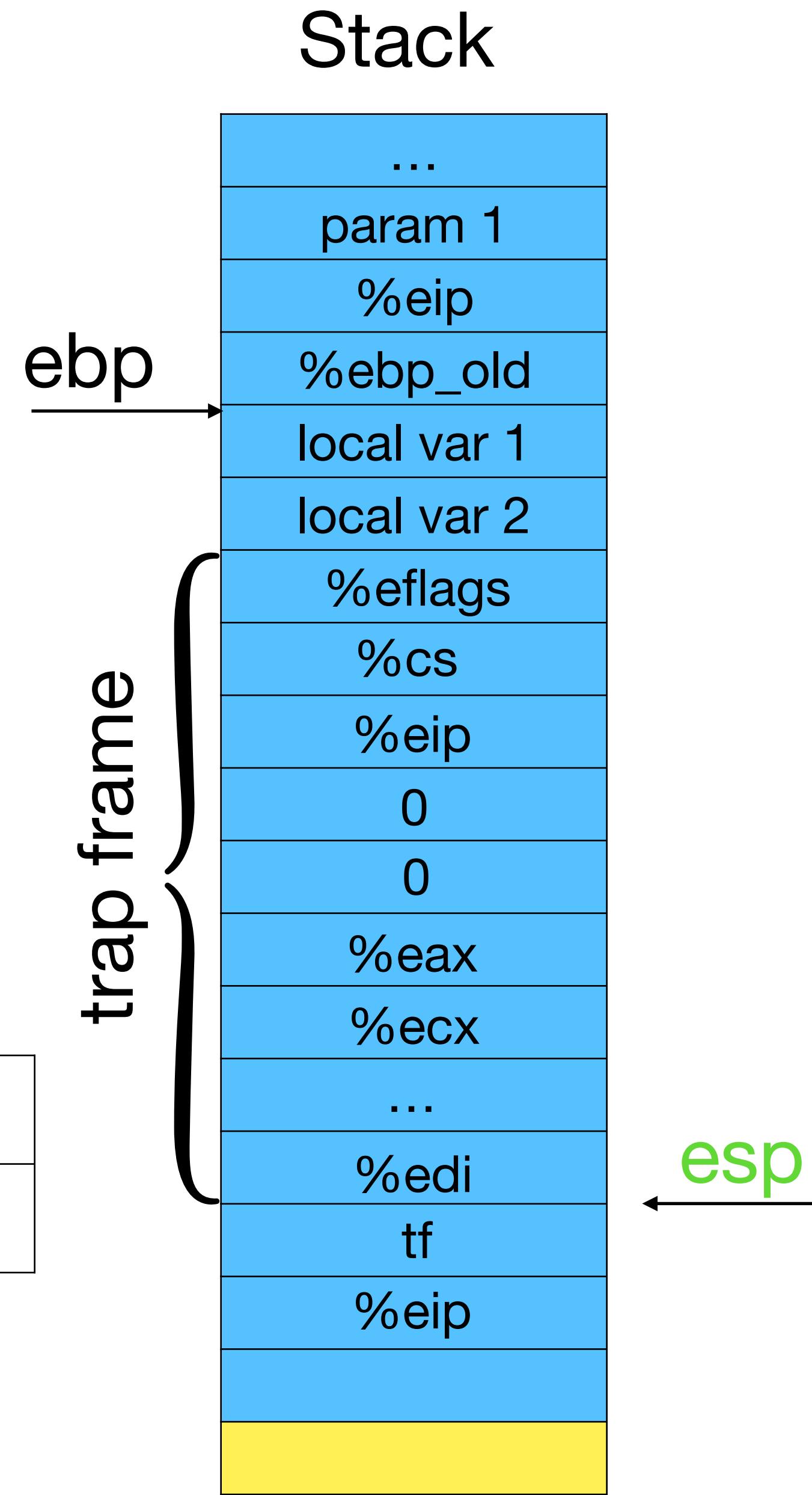
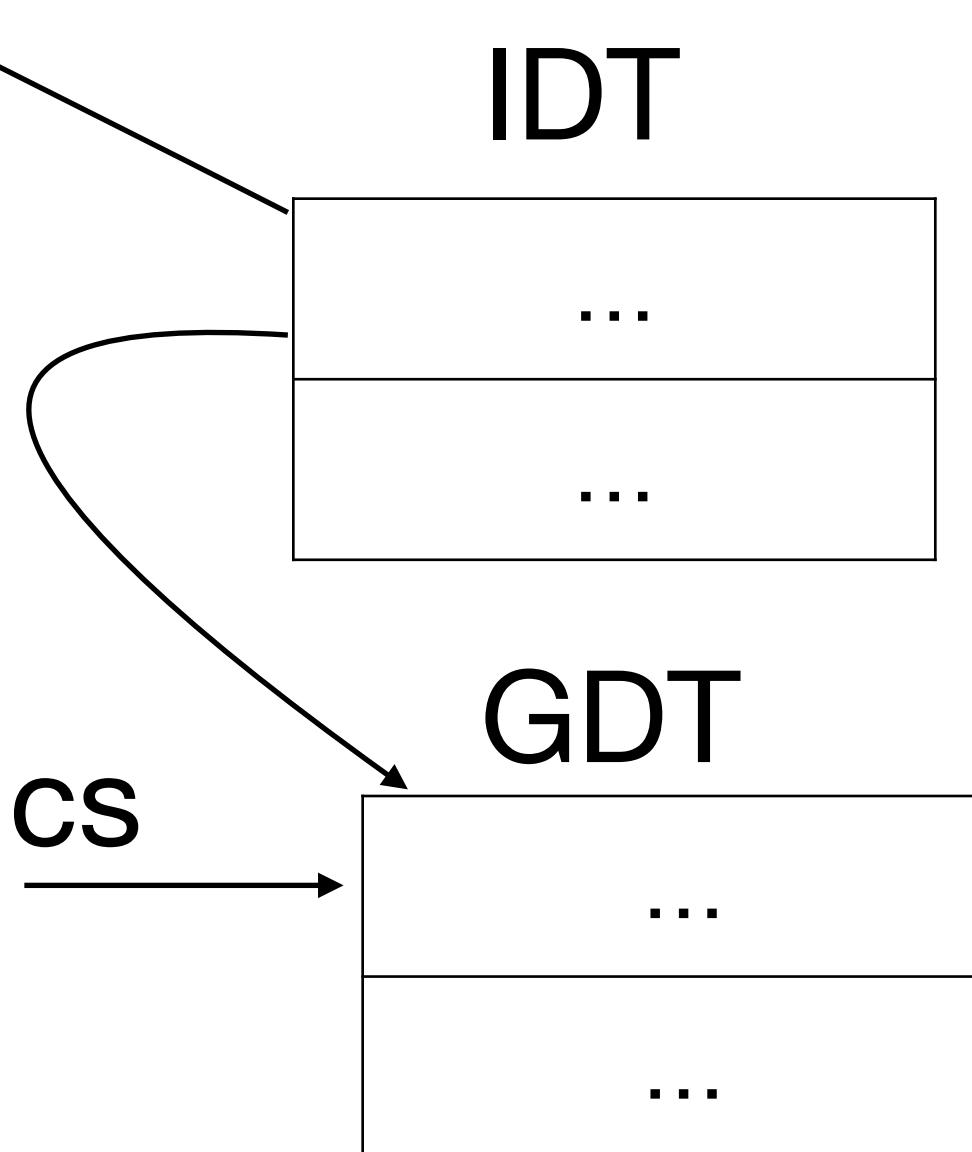
trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
    return
}
```

```
vectors.S
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

```
trapasm.S
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

eip

CS



Interrupt handling revisited

```
for(;;)
;

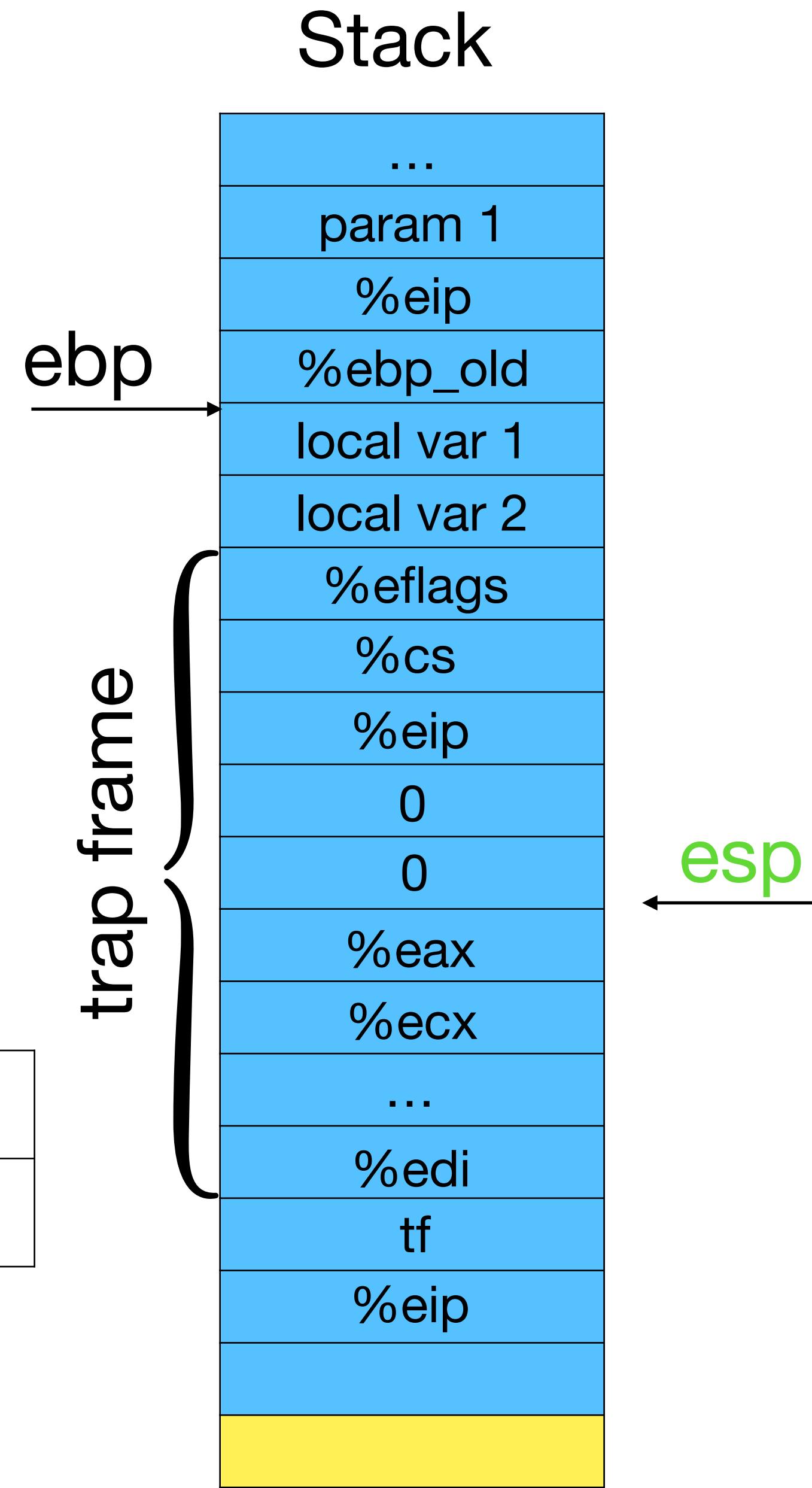
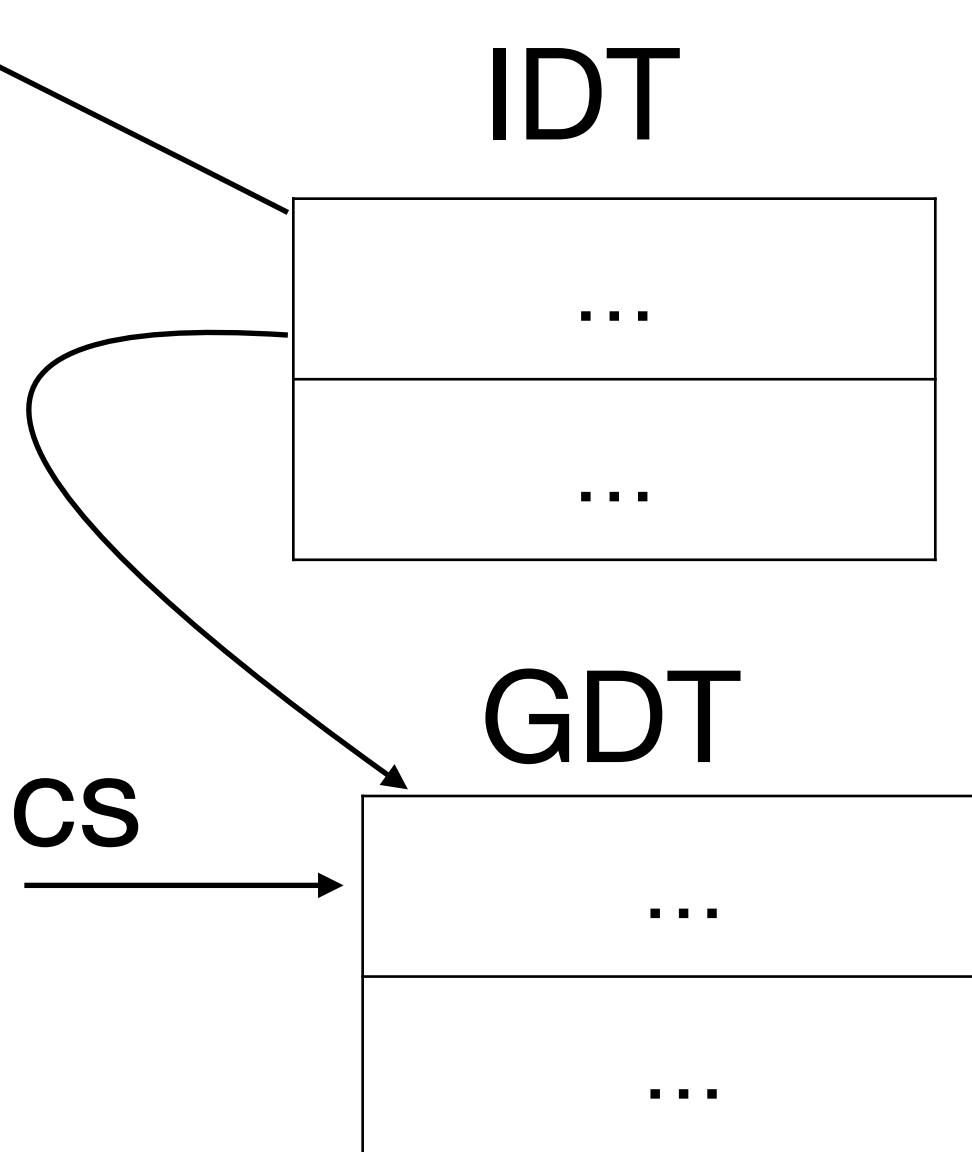
trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
    return
}
```

```
vectors.S
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

```
trapasm.S
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

eip

CS



Interrupt handling revisited

```
for(;;)
;

trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
    return
}
```

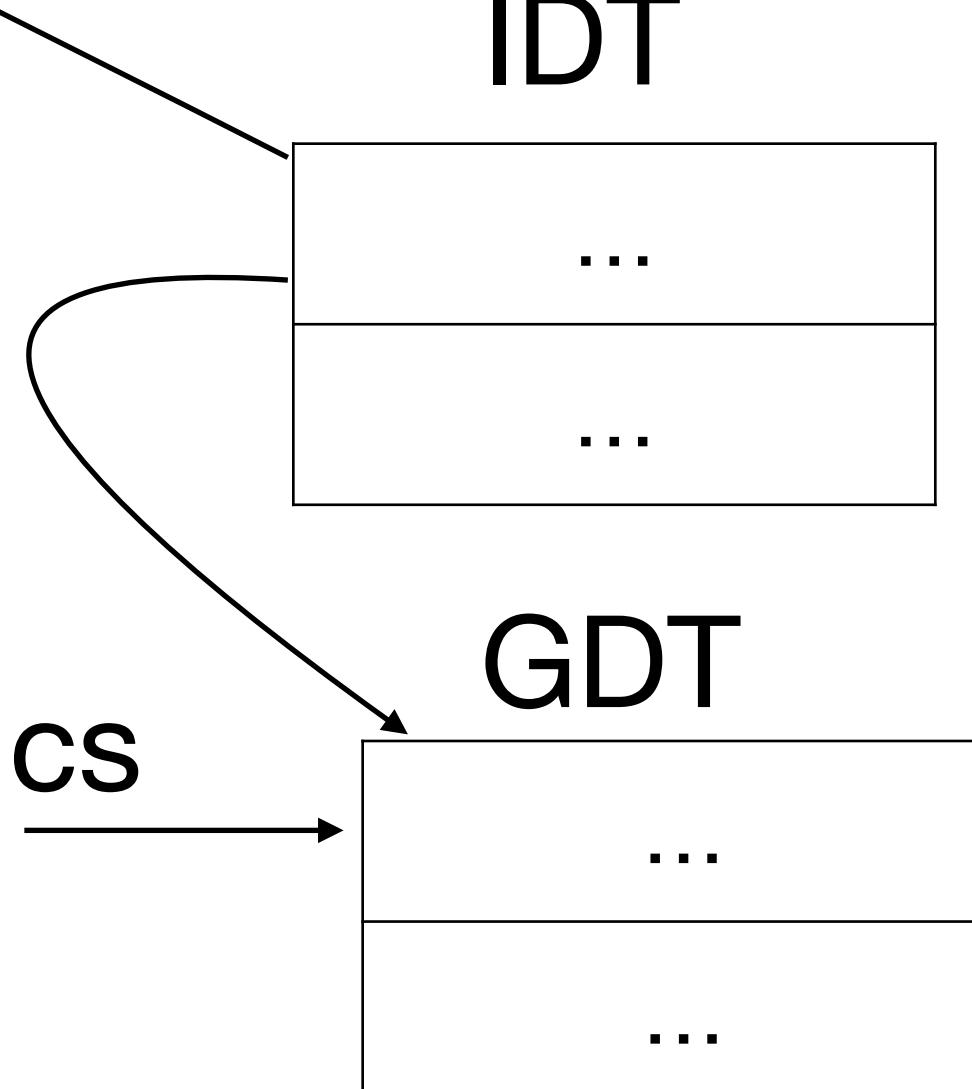
vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

trapasm.S

```
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

eip



Stack

...
param 1
%eip
%ebp_old
local var 1
local var 2
%eflags
%cs
%eip
0
0
%eax
%ecx
...
%edi
tf
%eip

ebp

trap frame

esp

Interrupt handling revisited

```
for(;;)
;

trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
    return
}
```

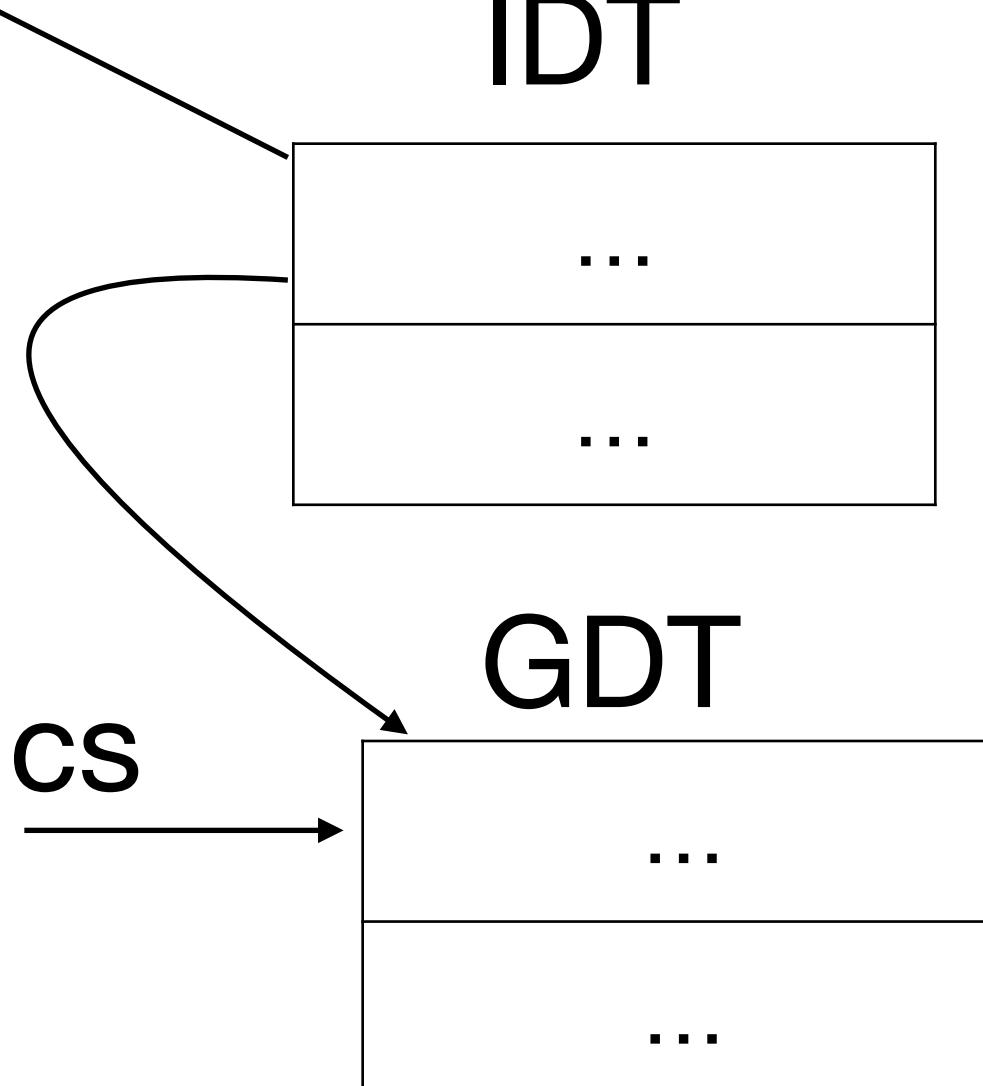
vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

trapasm.S

```
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

eip



Stack

...
param 1
%eip
%ebp_old
local var 1
local var 2
%eflags
%cs
%eip
0
0
%eax
%ecx
...
%edi
tf
%eip

trap frame

ebp

esp

Interrupt handling revisited

```
for(;;)
;

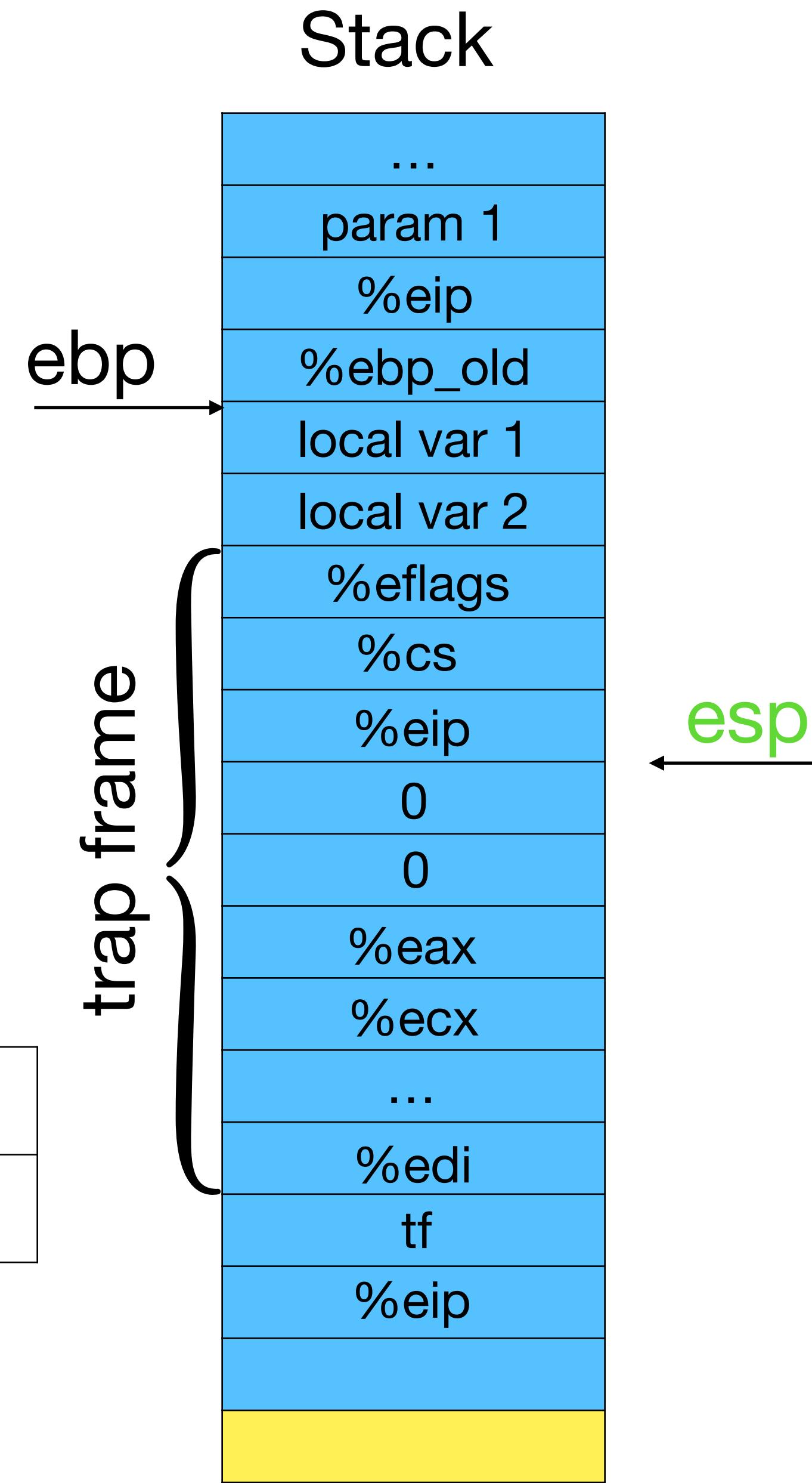
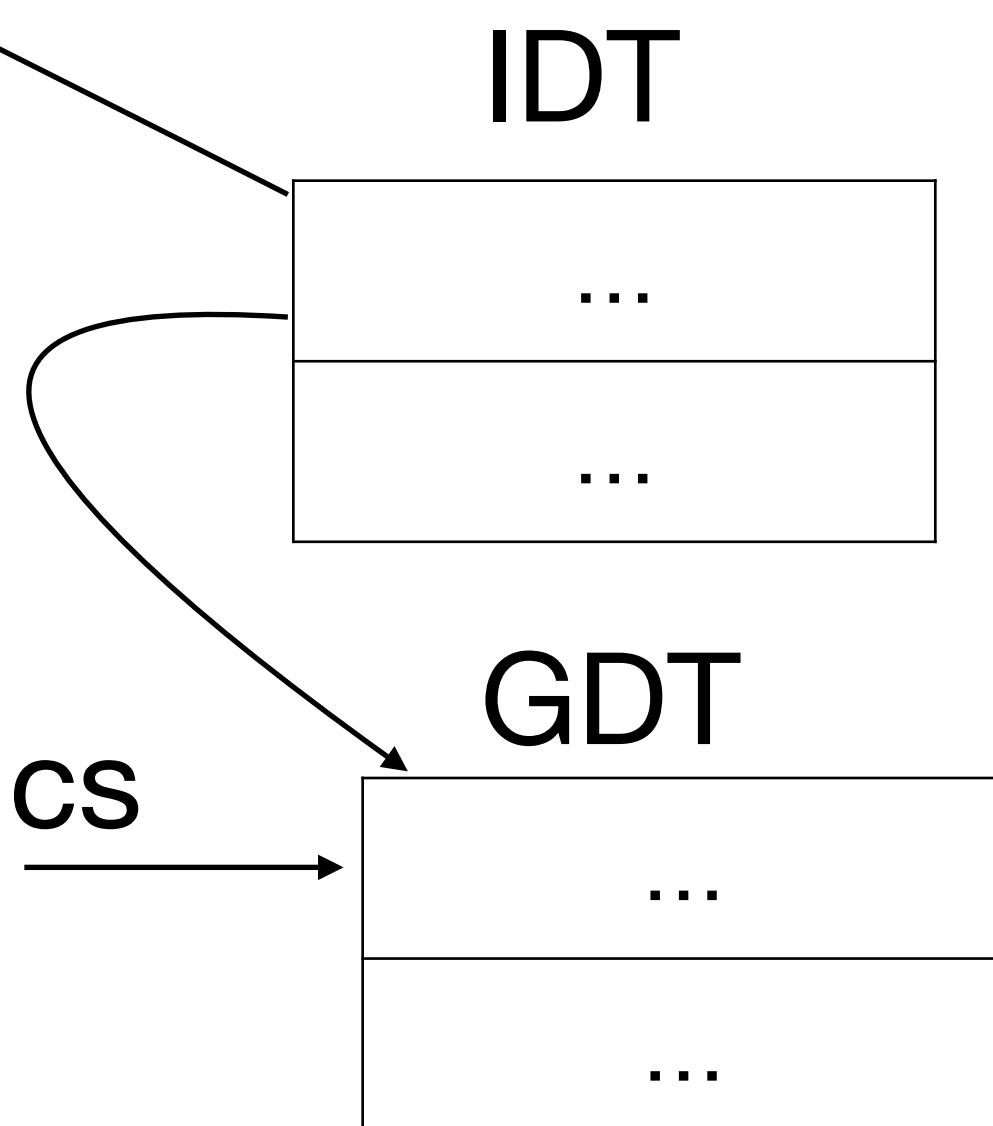
trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
    return
}
```

```
vectors.S
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

trapasm.S

```
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

eip



Interrupt handling revisited

```
for(;;)
;

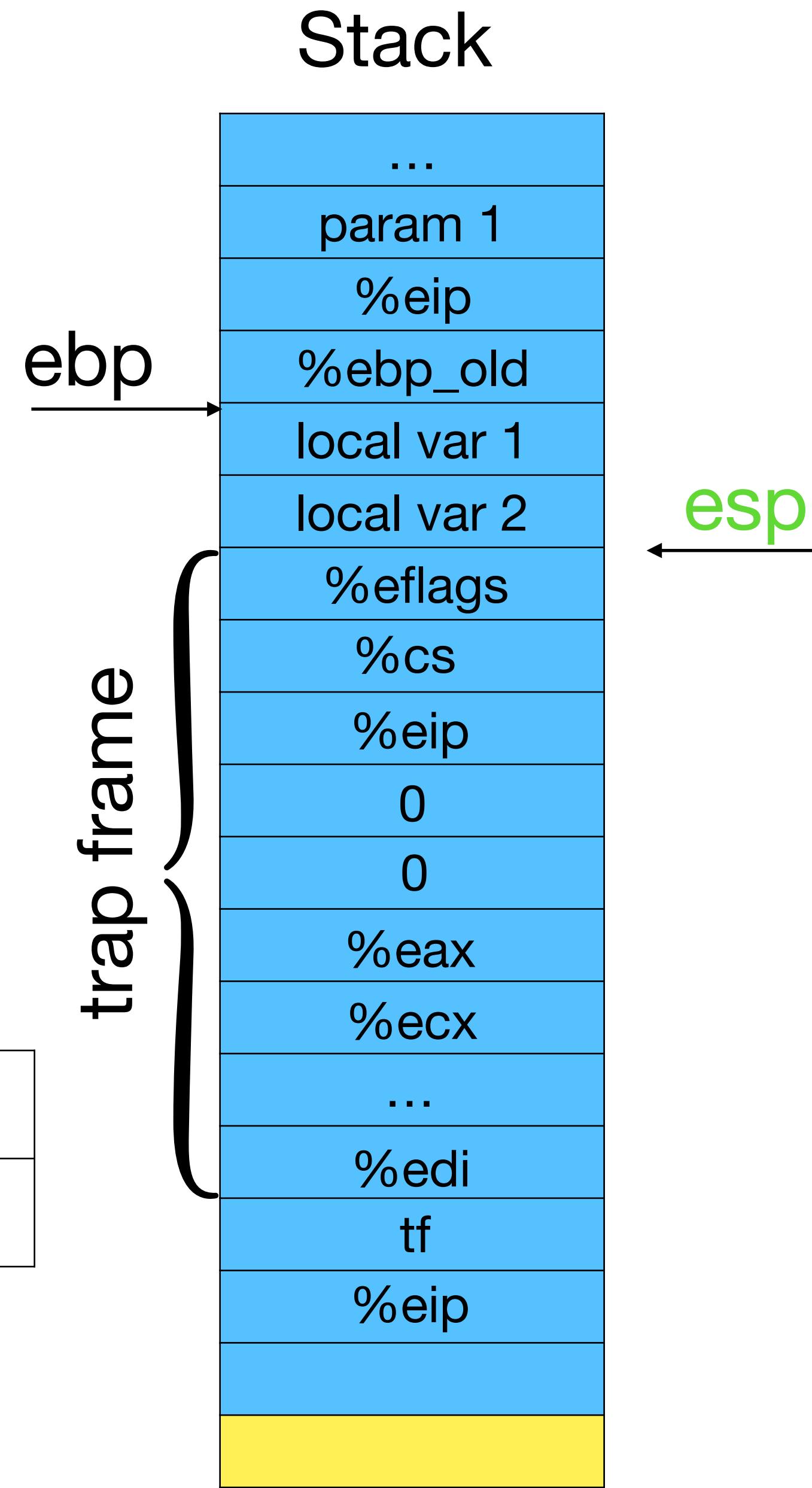
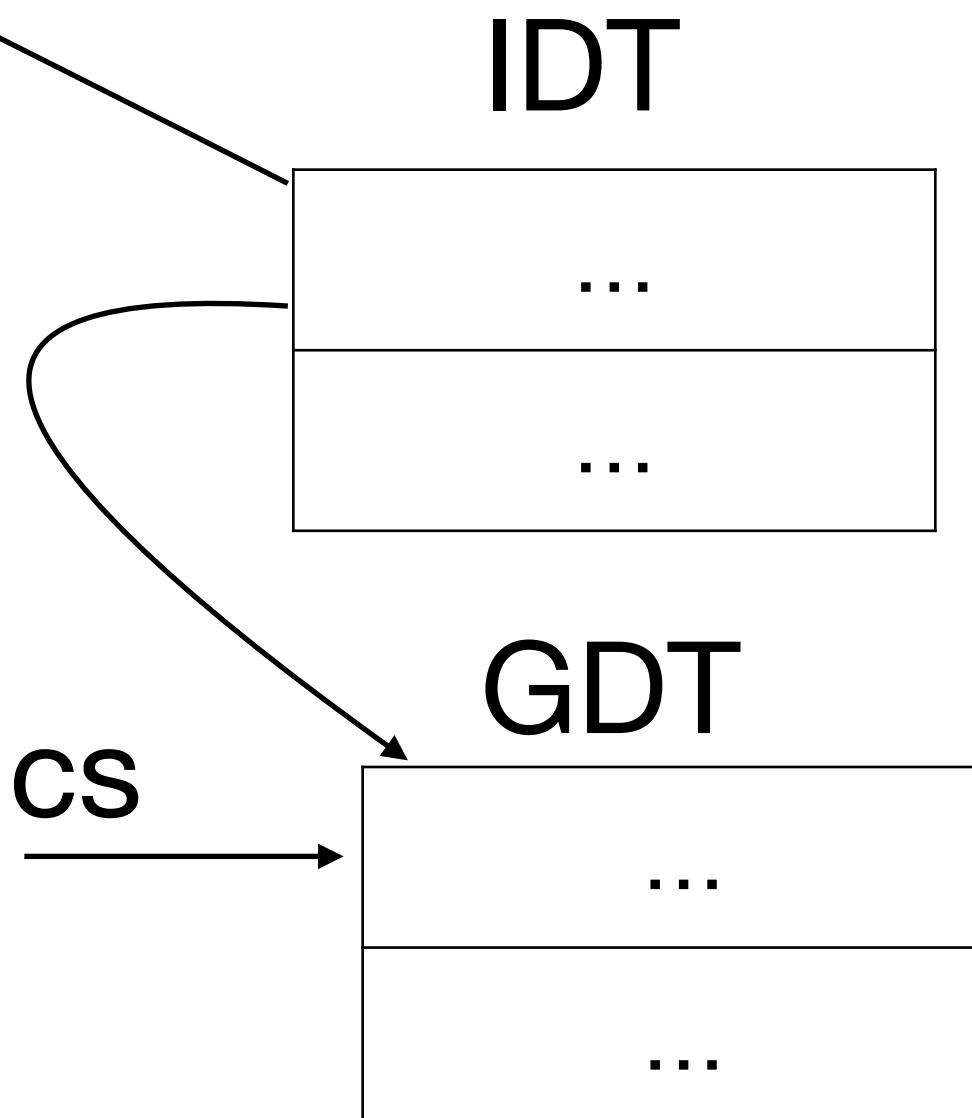
trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
    return
}
```

```
vectors.S
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

trapasm.S

```
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

eip



Interrupt handling revisited

```
for(;;)
;
trap.c
void
trap(struct trapframe *tf)
{
    switch(tf->trapno){
    case T_IRQ0 + IRQ_TIMER:
        ticks++;
        cprintf("Tick! %d\n", ticks);
        lapiceoi();
    ...
    return
```

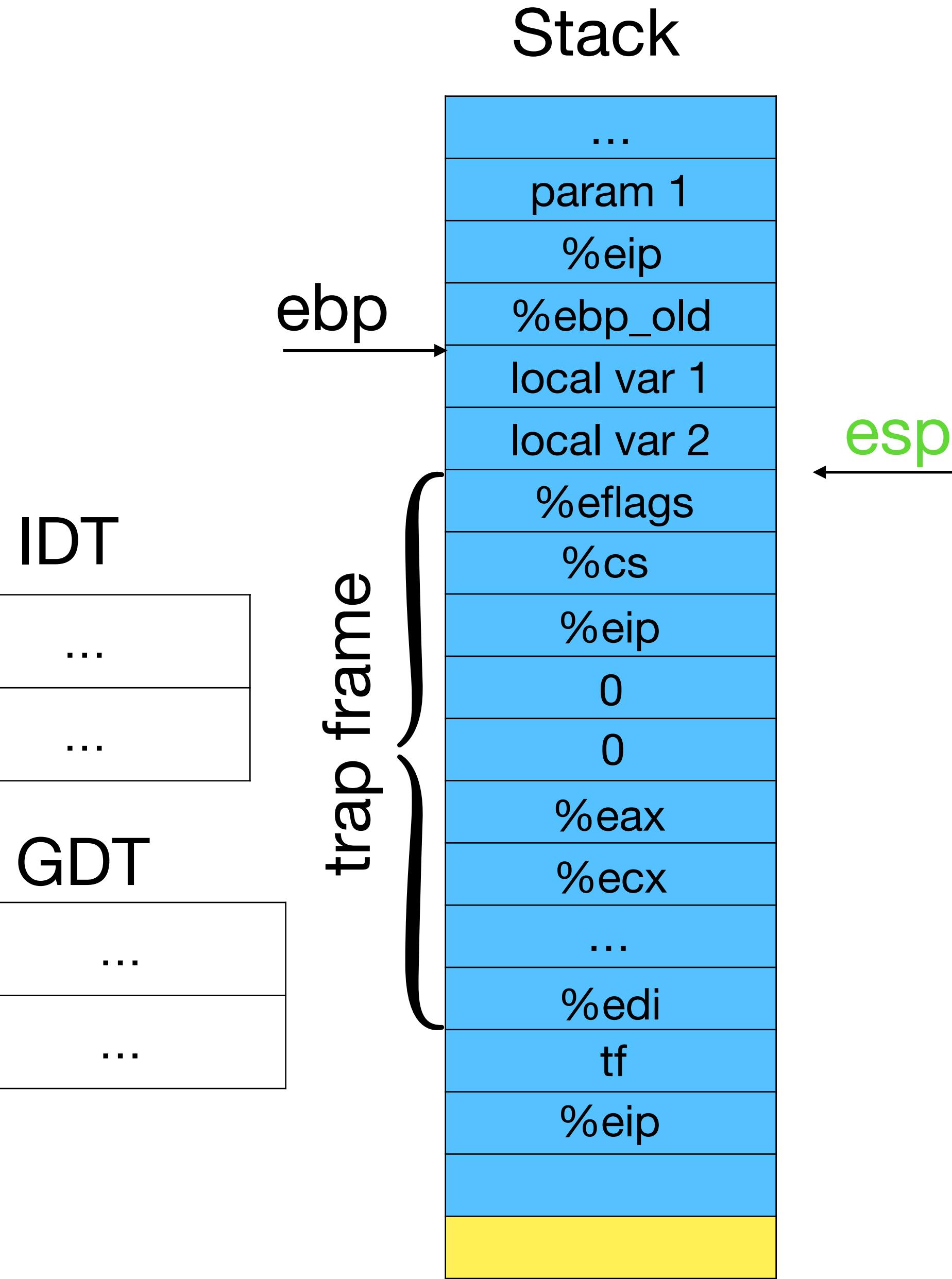
vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

trapasm.S

```
alltraps:
    pushal
    pushl %esp
    call trap
    addl $4, %esp
    popal
    addl $0x8, %esp
    iret
```

eip



Interrupt handling revisited

```
eip → for(;;)  
;  
  
trap.c  
void  
trap(struct trapframe *tf)  
{  
    switch(tf->trapno){  
        case T_IRQ0 + IRQ_TIMER:  
            ticks++;  
            cprintf("Tick! %d\n", ticks);  
            lapiceoi();  
            ...  
    }  
    return
```

vectors.S

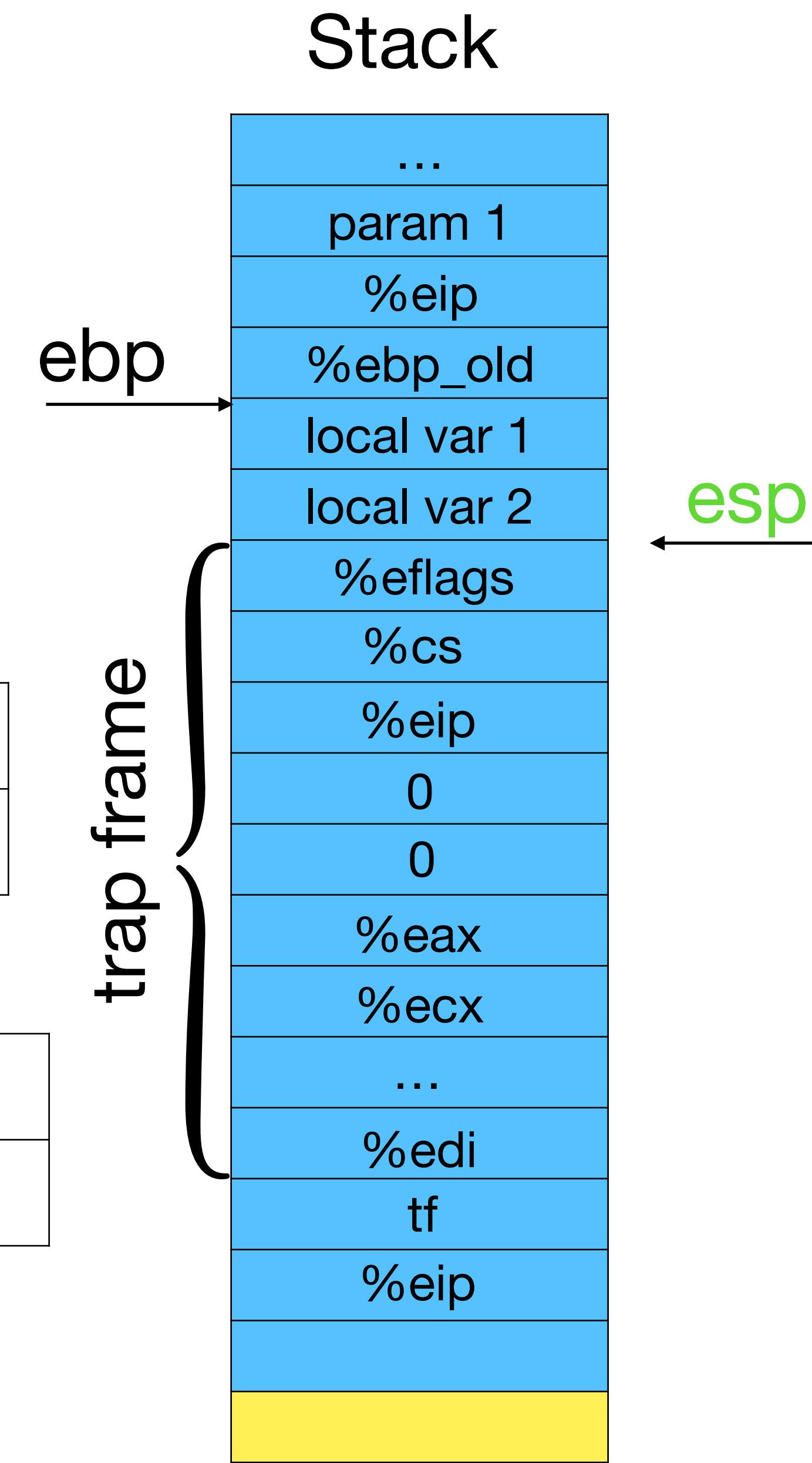
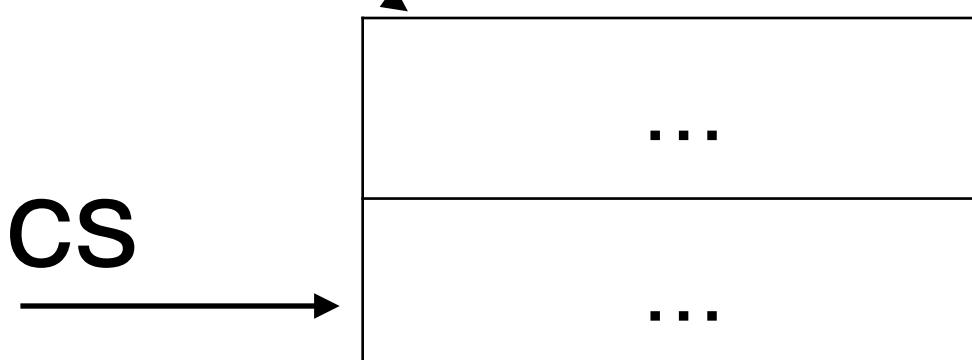
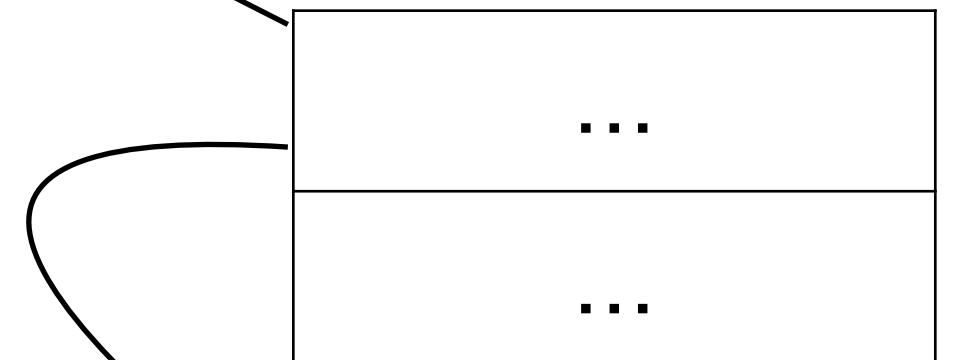
```
.globl vector0  
vector0:  
    pushl $0  
    pushl $0  
    jmp alltraps
```

trapasm.S

```
alltraps:  
    pushal  
    pushl %esp  
    call trap  
    addl $4, %esp  
    popal  
    addl $0x8, %esp  
    iret
```

IDT

GDT



Interrupt handling revisited

```
eip → for(;;)  
;  
  
trap.c  
void  
trap(struct trapframe *tf)  
{  
    switch(tf->trapno){  
        case T_IRQ0 + IRQ_TIMER:  
            ticks++;  
            cprintf("Tick! %d\n", ticks);  
            lapiceoi();  
            ...  
    }  
    return
```

vectors.S

```
.globl vector0  
vector0:  
    pushl $0  
    pushl $0  
    jmp alltraps
```

trapasm.S

```
alltraps:  
    pushal  
    pushl %esp  
    call trap  
    addl $4, %esp  
    popal  
    addl $0x8, %esp  
    iret
```

IDT

GDT

ebp

trap frame

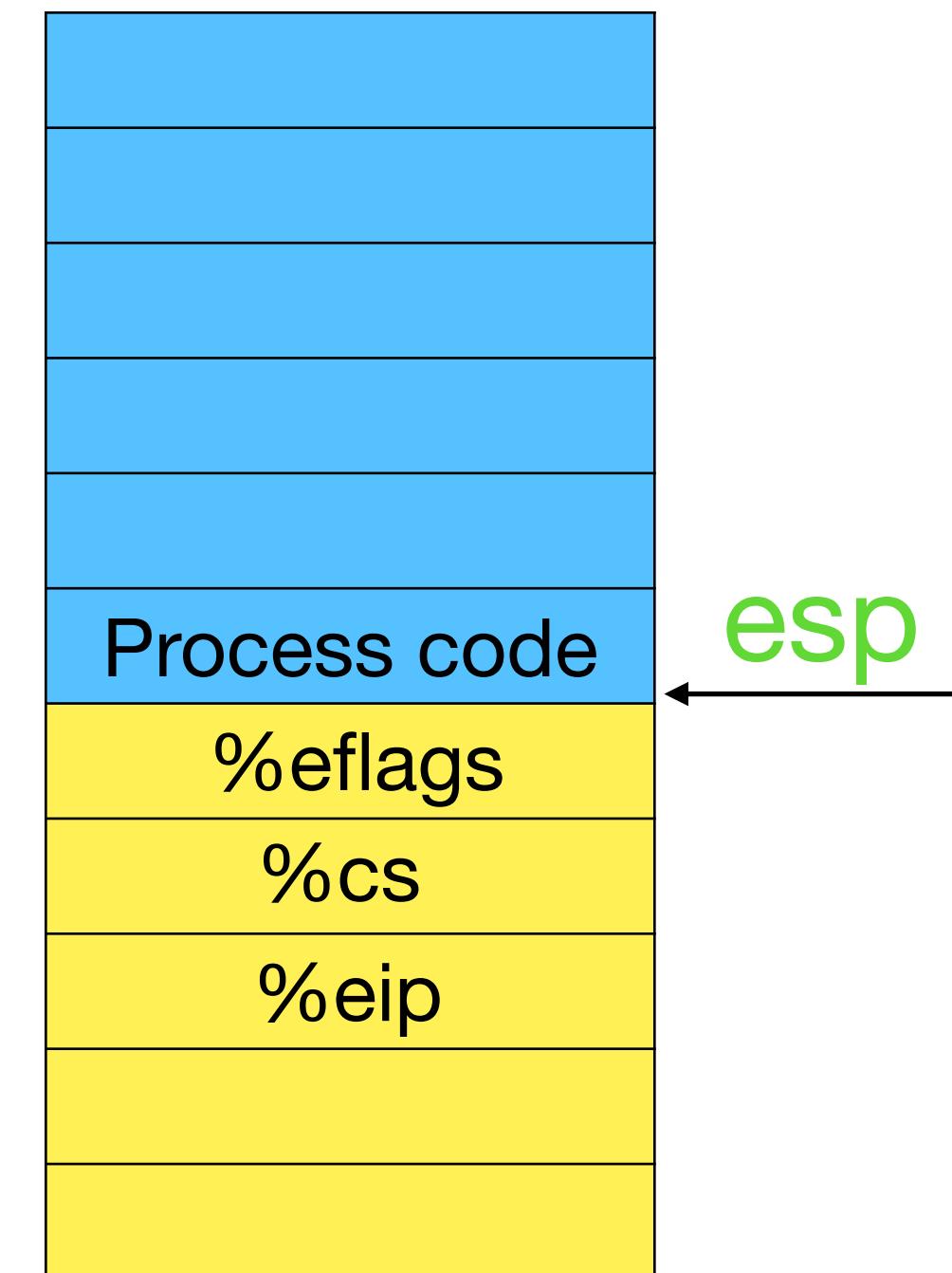
esp

Stack

...
param 1
%eip
%ebp_old
local var 1
local var 2
%eflags
%cs
%eip
0
0
%eax
%ecx
...
%edi
tf
%eip

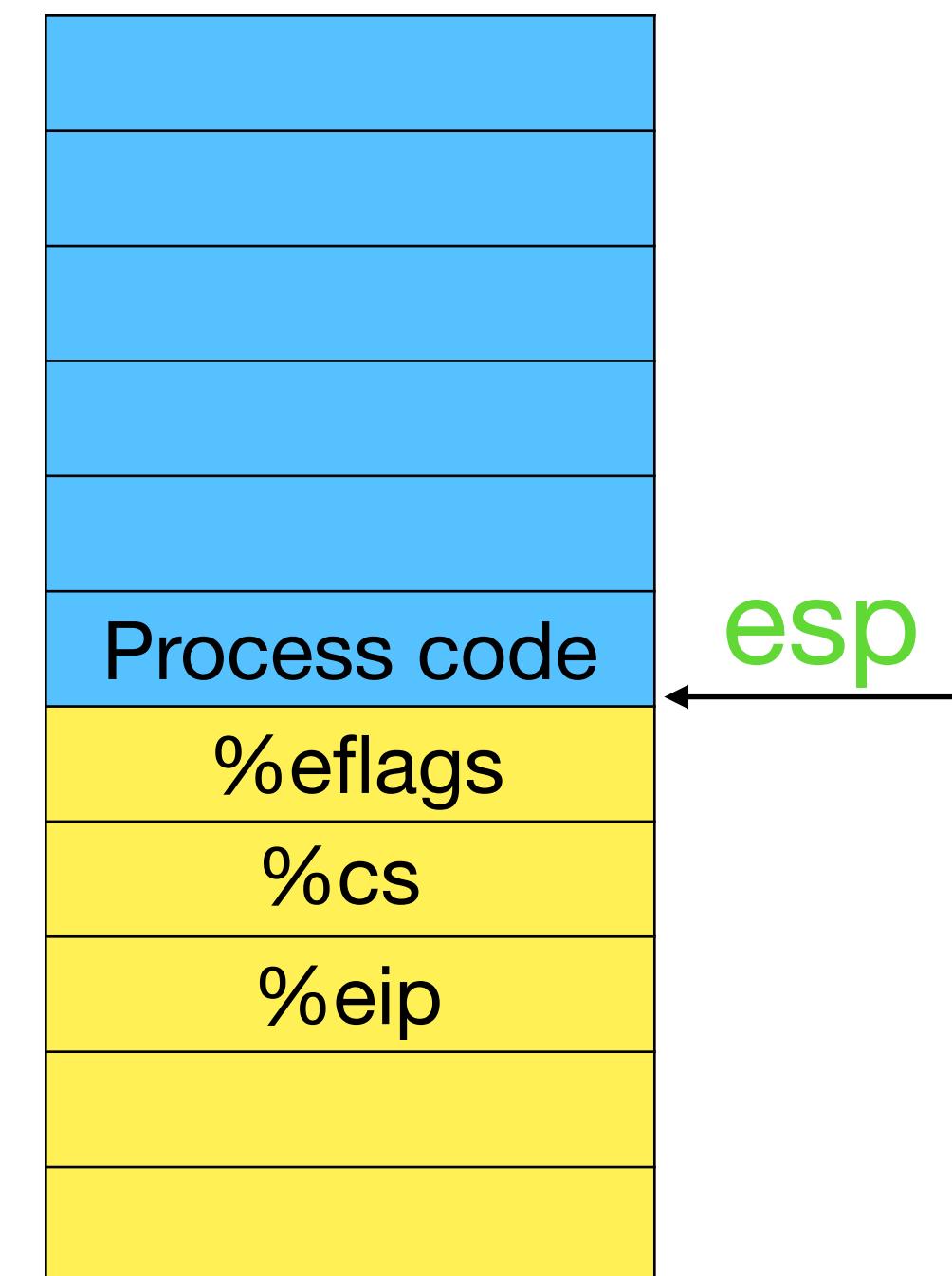
Interrupt handling problem

- We cannot trust `%esp` of the process. Hardware might write (`%eflags`, `%cs`, `%eip`) into another process' address space or into OS memory.



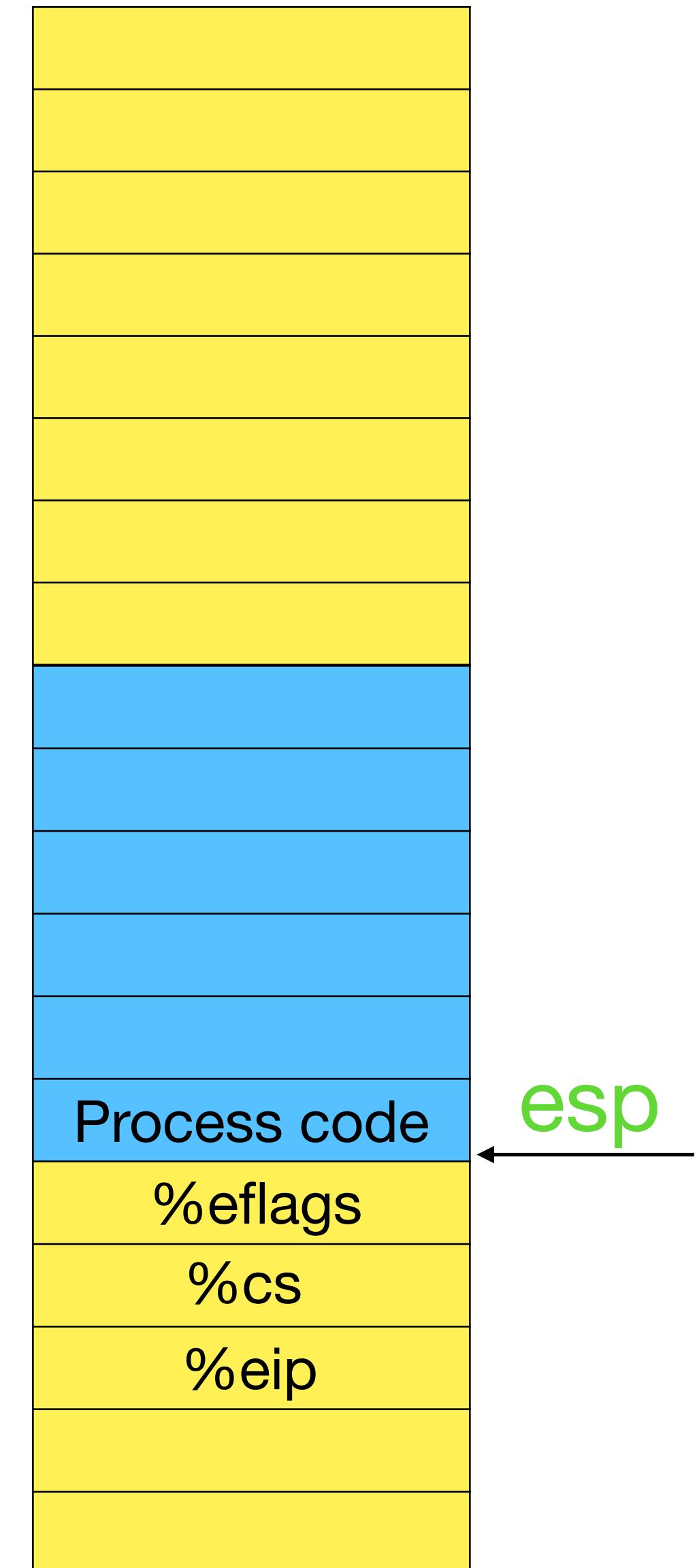
Interrupt handling problem

- We cannot trust %esp of the process. Hardware might write (%eflags, %cs, %eip) into another process' address space or into OS memory.
 - Use a separate stack set up by the OS!



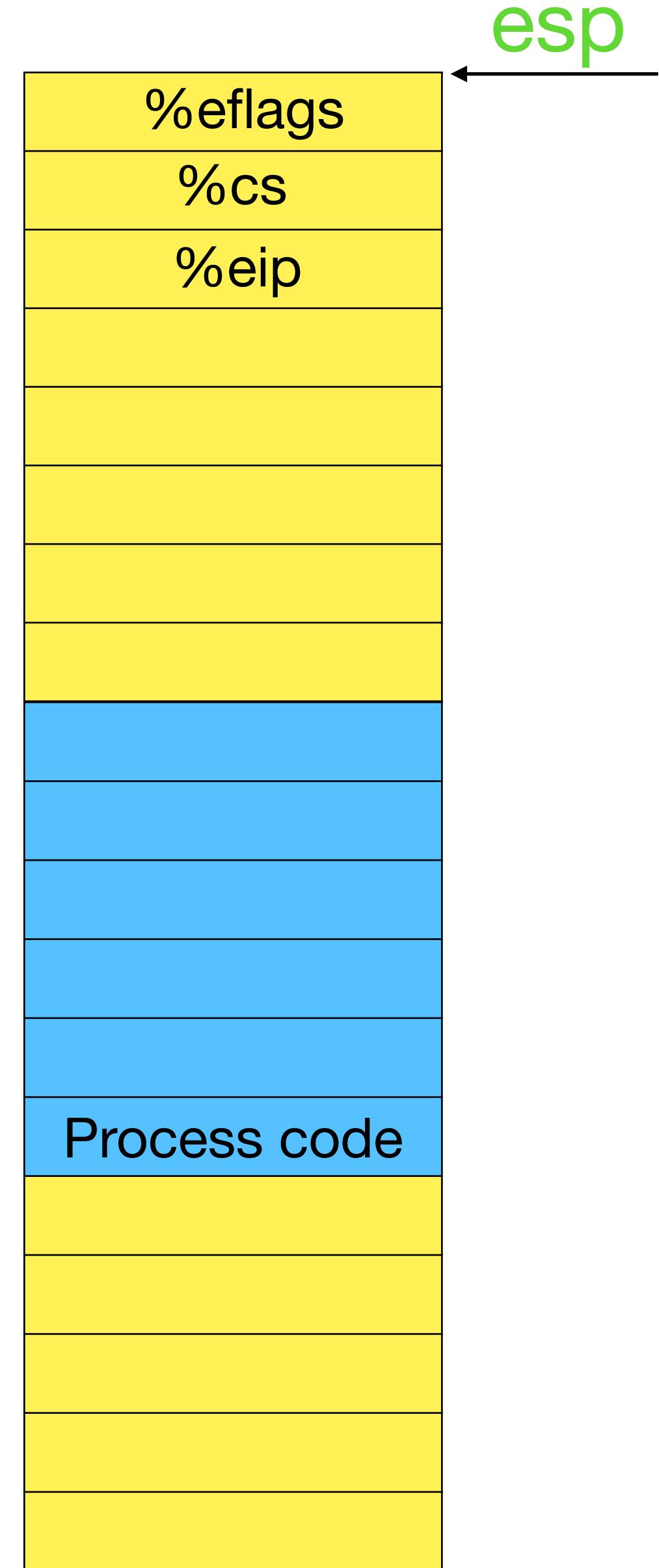
Interrupt handling problem

- We cannot trust `%esp` of the process. Hardware might write (`%eflags`, `%cs`, `%eip`) into another process' address space or into OS memory.
- Use a separate stack set up by the OS!



Interrupt handling problem

- We cannot trust %esp of the process. Hardware might write (%eflags, %cs, %eip) into another process' address space or into OS memory.
- Use a separate stack set up by the OS!



OS tells hardware where the kernel stack is

- In task-state segment (TSS), OS sets up
 - Stack segment for transition to ring 0 in SS0 and
 - Stack pointer for transition to ring 0 in ESP0

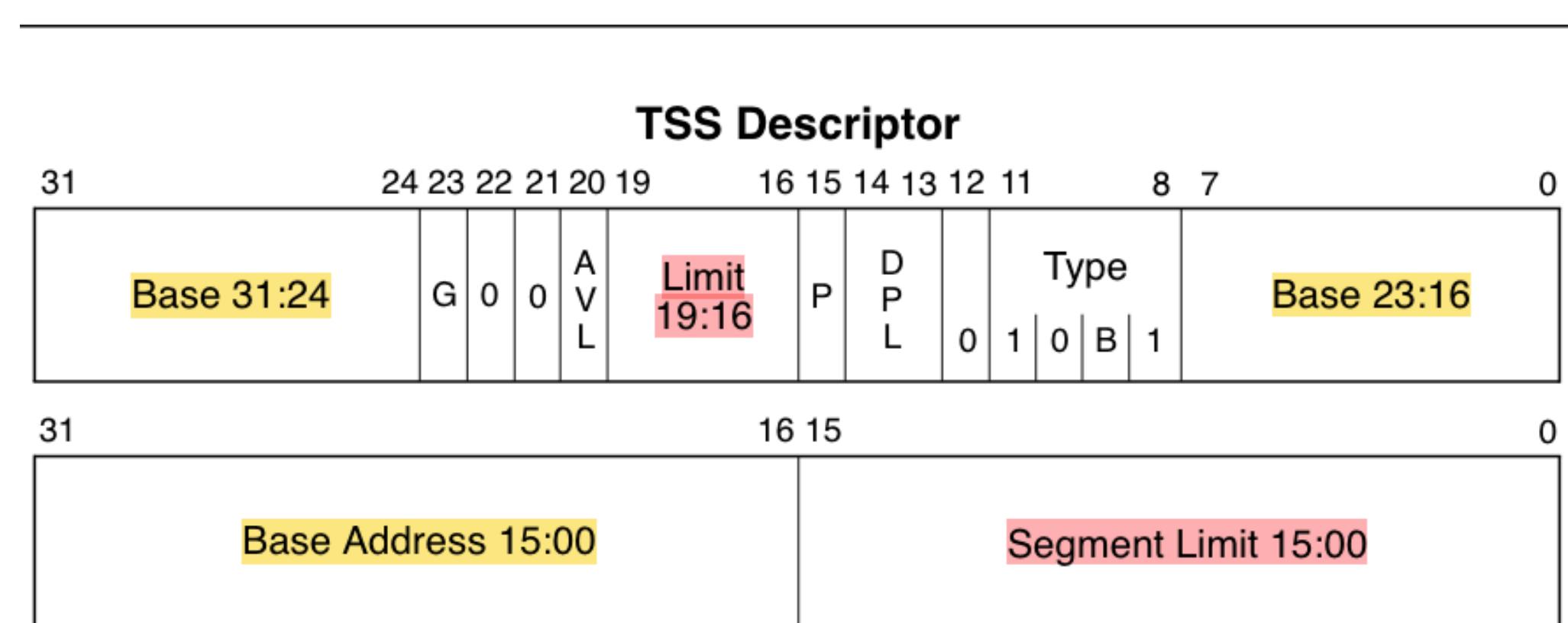
31	15	0
I/O Map Base Address	Reserved	T 100
Reserved	LDT Segment Selector	96
Reserved	GS	92
Reserved	FS	88
Reserved	DS	84
Reserved	SS	80
Reserved	CS	76
Reserved	ES	72
	EDI	68
	ESI	64
	EBP	60
	ESP	56
	EBX	52
	EDX	48
	ECX	44
	EAX	40
	EFLAGS	36
	EIP	32
	CR3 (PDBR)	28
Reserved	SS2	24
	ESP2	20
Reserved	SS1	16
	ESP1	12
Reserved	SS0	8
	ESP0	4
Reserved	Previous Task Link	0

 Reserved bits. Set to 0.

Figure 7-2. 32-Bit Task-State Segment (TSS)

OS tells hardware where the kernel stack is (2)

- OS sets up an entry in GDT that contains location of TSS
 - LTR instruction changes task register



AVL	Available for use by system software
B	Busy flag
BASE	Segment Base Address
DPL	Descriptor Privilege Level
G	Granularity
LIMIT	Segment Limit
P	Segment Present
TYPE	Segment Type

Figure 7-3. TSS Descriptor

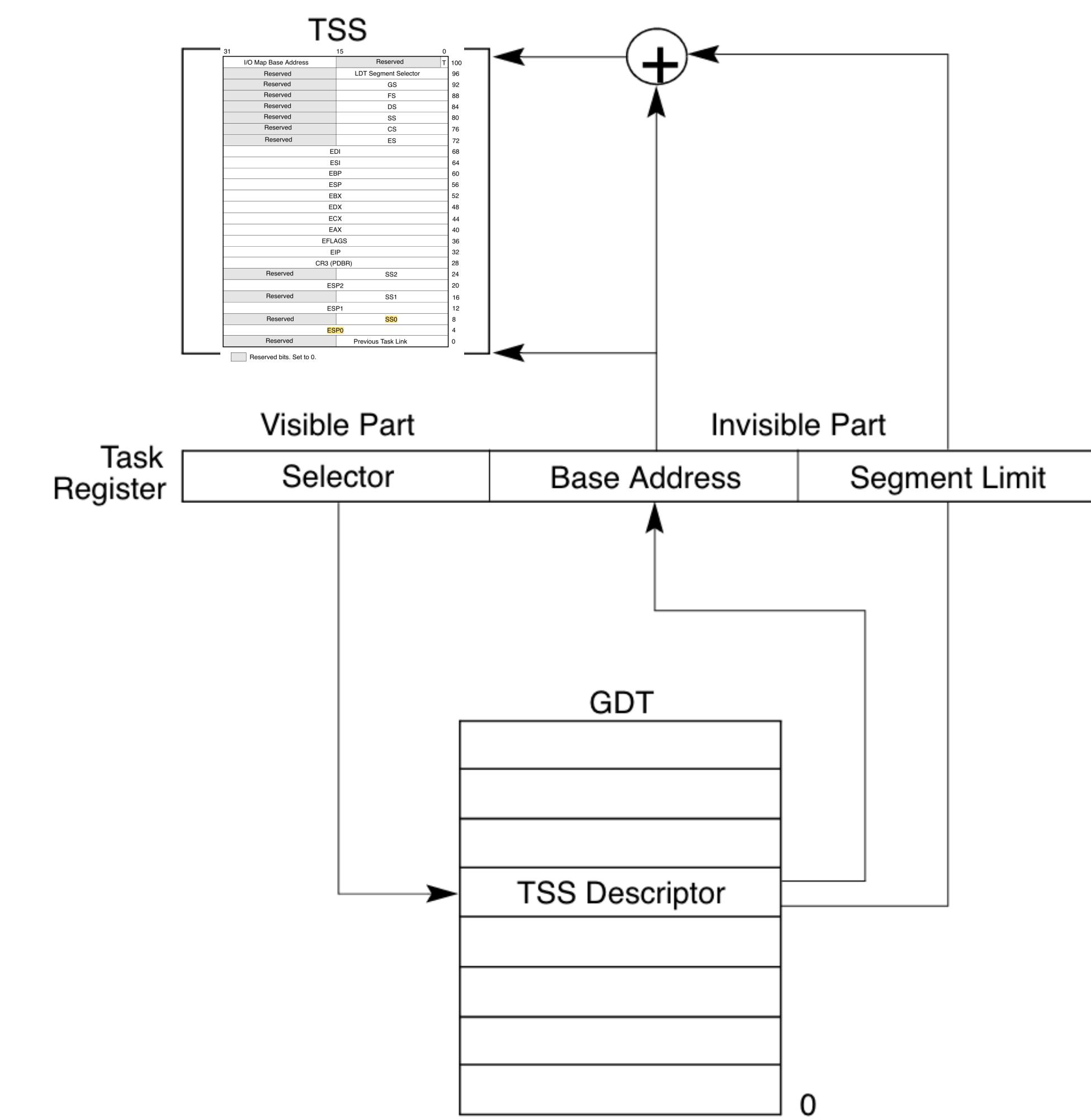
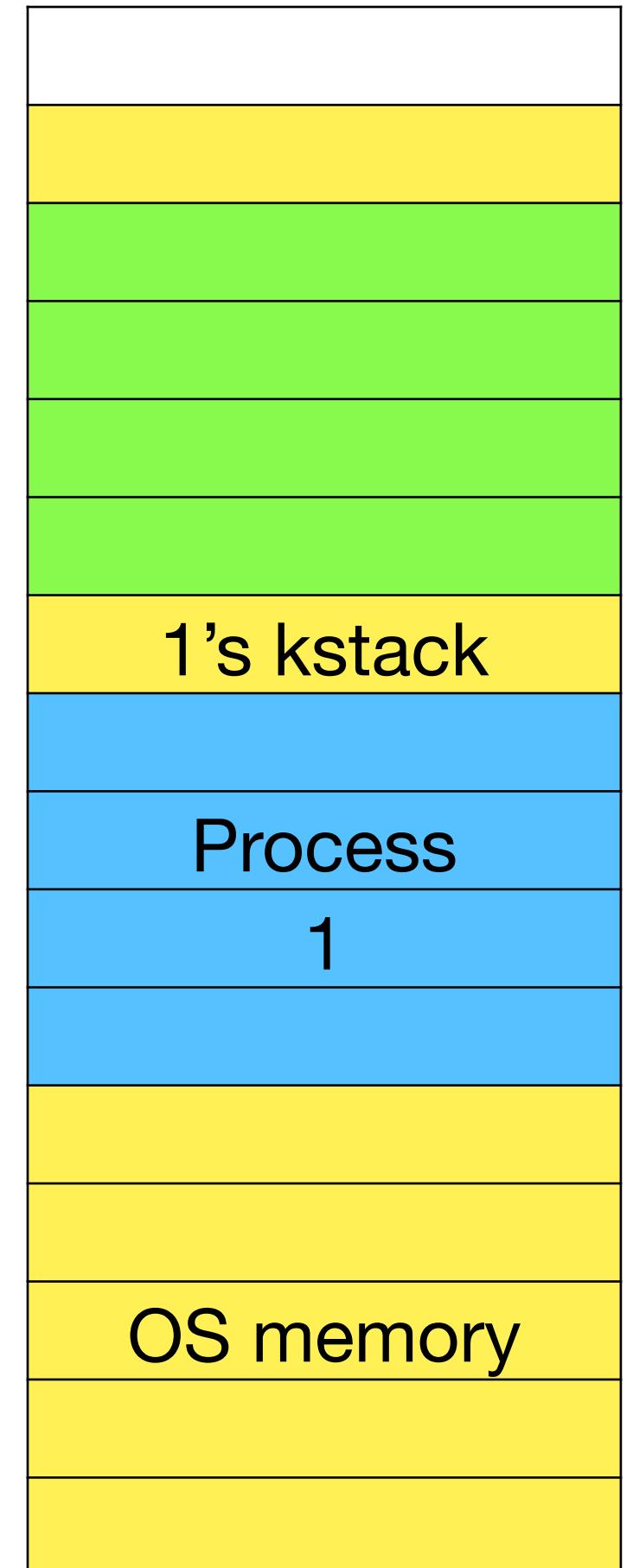


Figure 7-5. Task Register

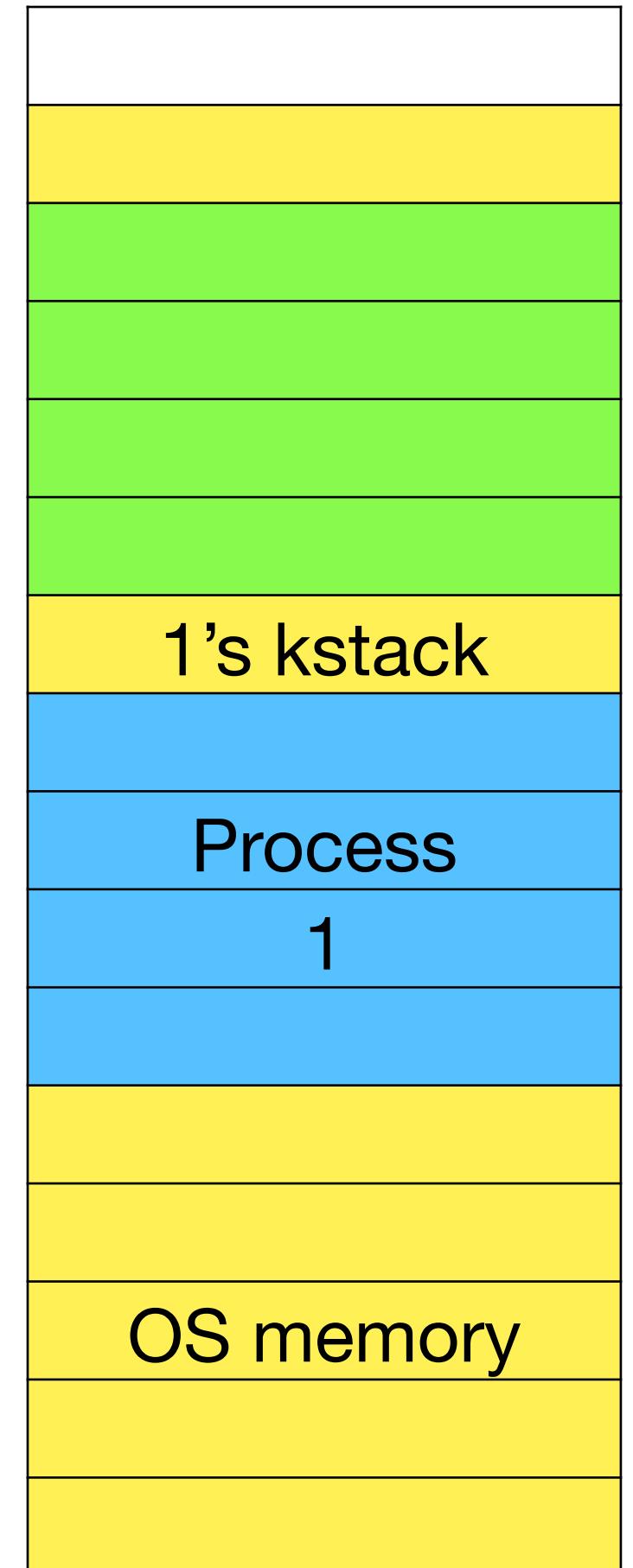
Code walkthrough: setting up kernel stack p16-tss

```
void seginit(void) {  
    c->gdt[SEG_UCODE] = SEG(.., STARTPROC, (PROCSIZE << 12) - 1, DPL_USER);  
    c->gdt[SEG_UDATA] = SEG(.., STARTPROC, (PROCSIZE << 12) - 1, DPL_USER);  
}
```



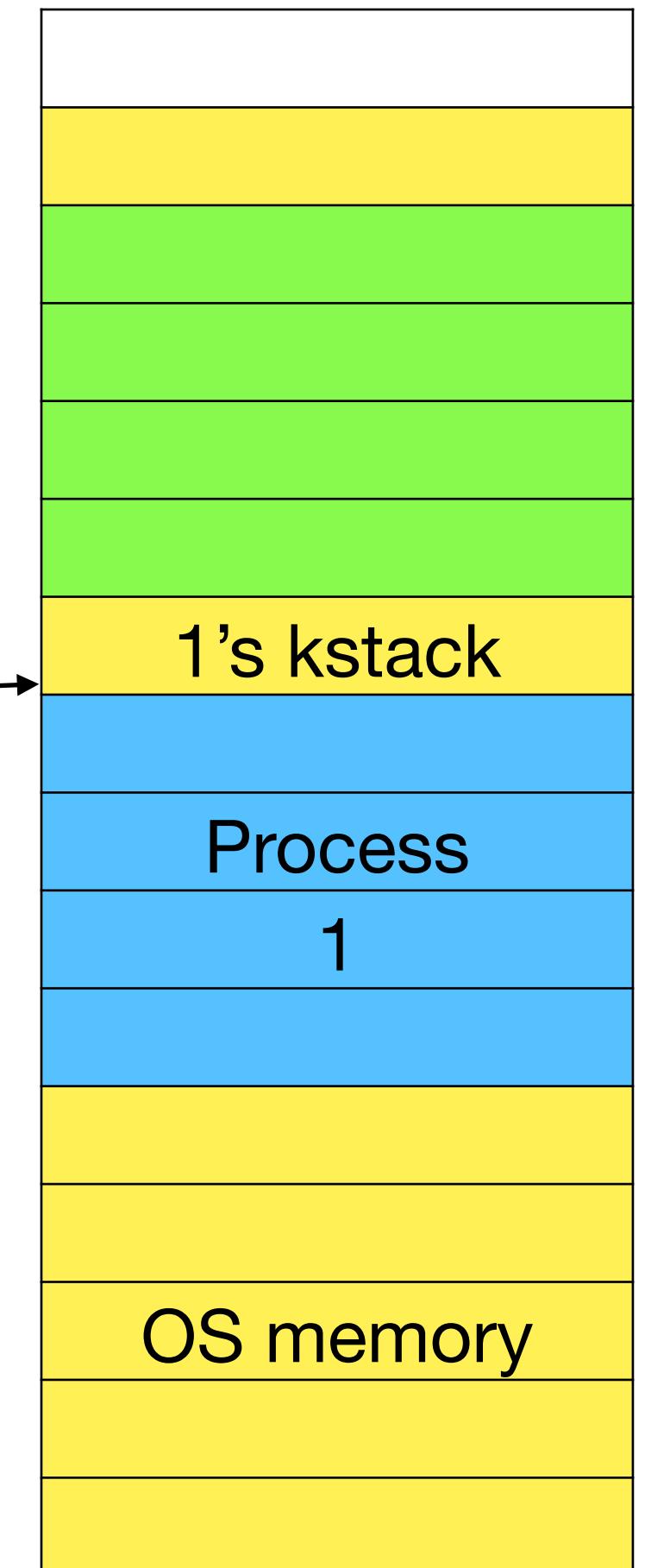
Code walkthrough: setting up kernel stack p16-tss

```
void seginit(void) {  
    c->gdt[SEG_UCODE] = SEG(.., STARTPROC, (PROCSIZE << 12) - 1, DPL_USER);  
    c->gdt[SEG_UDATA] = SEG(.., STARTPROC, (PROCSIZE << 12) - 1, DPL_USER);  
}  
  
static struct proc* allocproc(void) {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    p->kstack = sp - KSTACKSIZE;  
}
```



Code walkthrough: setting up kernel stack p16-tss

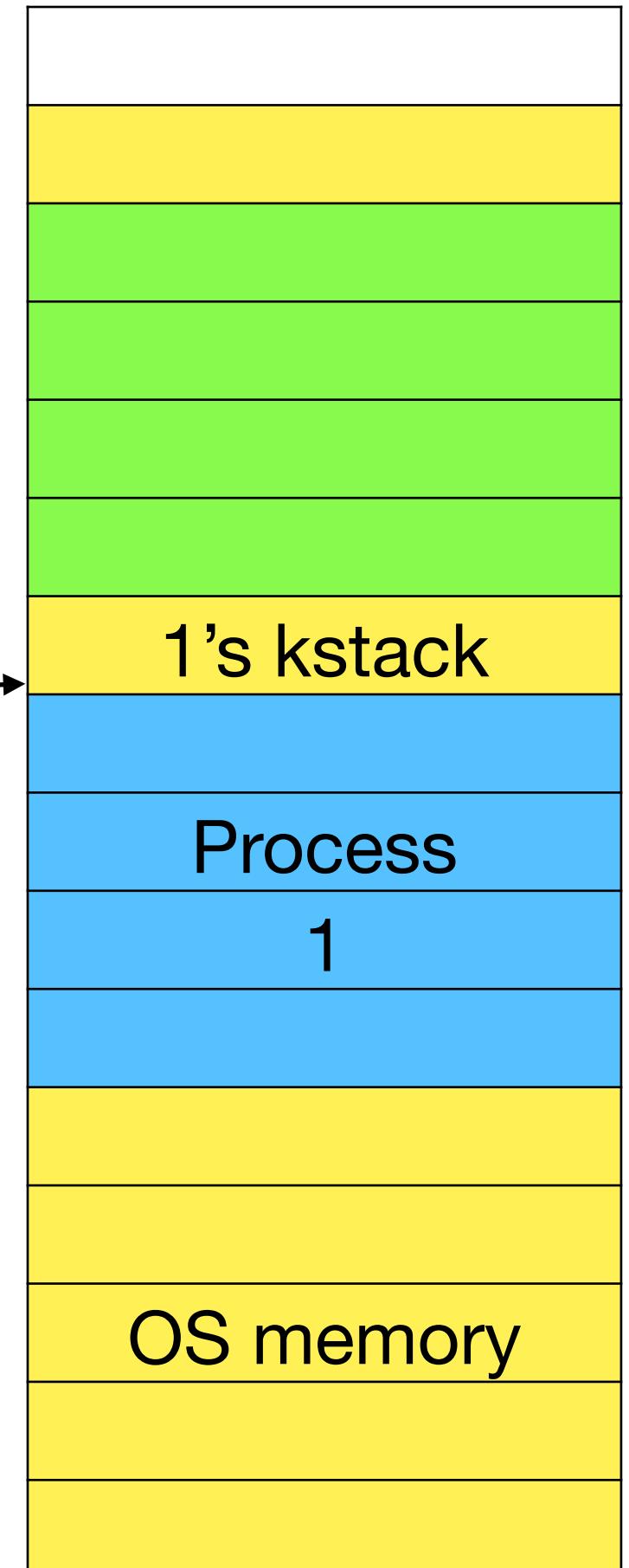
```
void seginit(void) {  
    c->gdt[SEG_UCODE] = SEG(.., STARTPROC, (PROCSIZE << 12) - 1, DPL_USER);  
    c->gdt[SEG_UDATA] = SEG(.., STARTPROC, (PROCSIZE << 12) - 1, DPL_USER);  
}  
  
static struct proc* allocproc(void) {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    p->kstack = sp - KSTACKSIZE;  
}
```



Code walkthrough: setting up kernel stack p16-tss

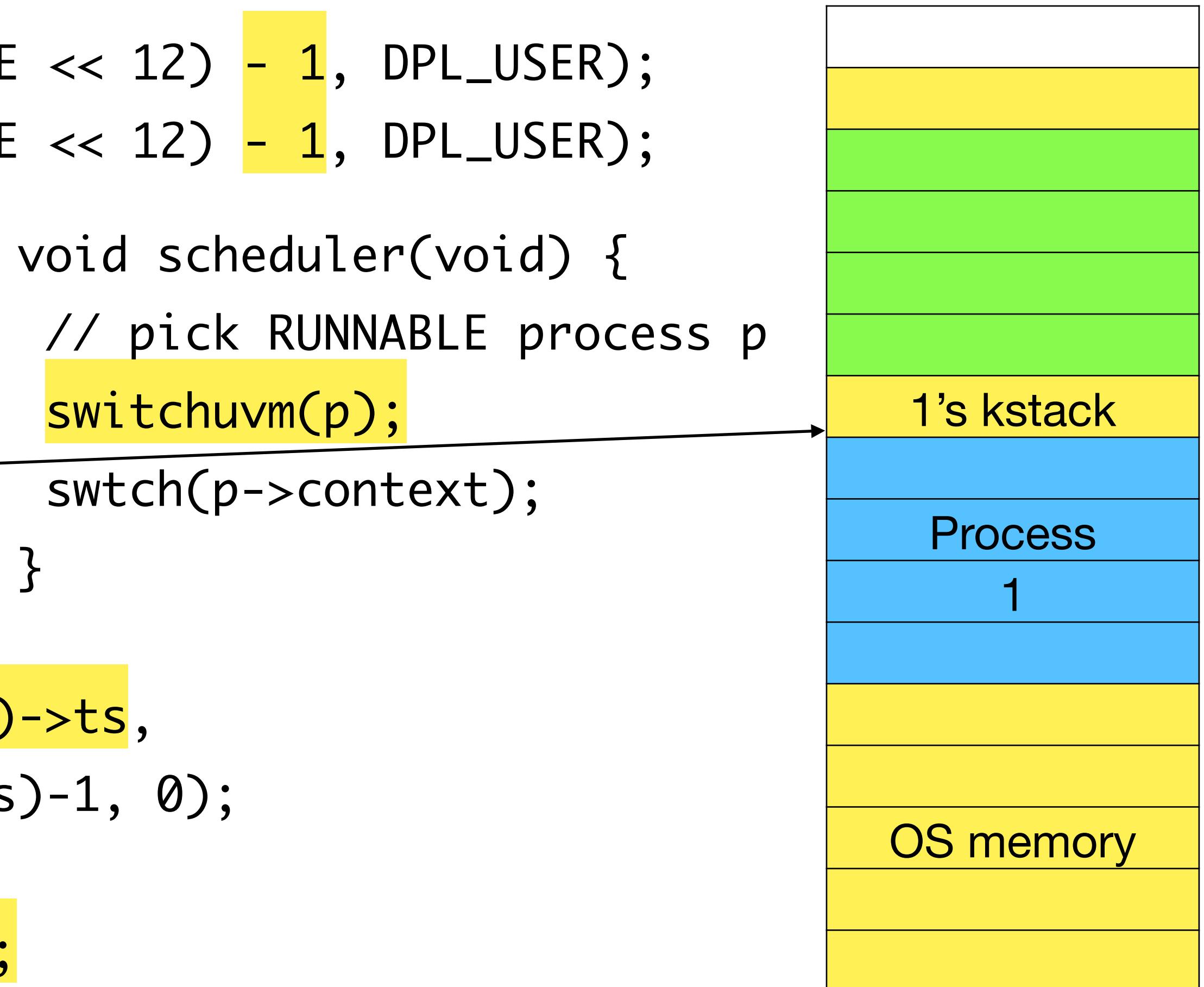
```
void seginit(void) {  
    c->gdt[SEG_UCODE] = SEG(.., STARTPROC, (PROCSIZE << 12) - 1, DPL_USER);  
    c->gdt[SEG_UDATA] = SEG(.., STARTPROC, (PROCSIZE << 12) - 1, DPL_USER);  
}  
  
static struct proc* allocproc(void) {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    p->kstack = sp - KSTACKSIZE;  
}
```

```
void scheduler(void) {  
    // pick RUNNABLE process p  
    switchuvm(p);  
    swtch(p->context);  
}
```



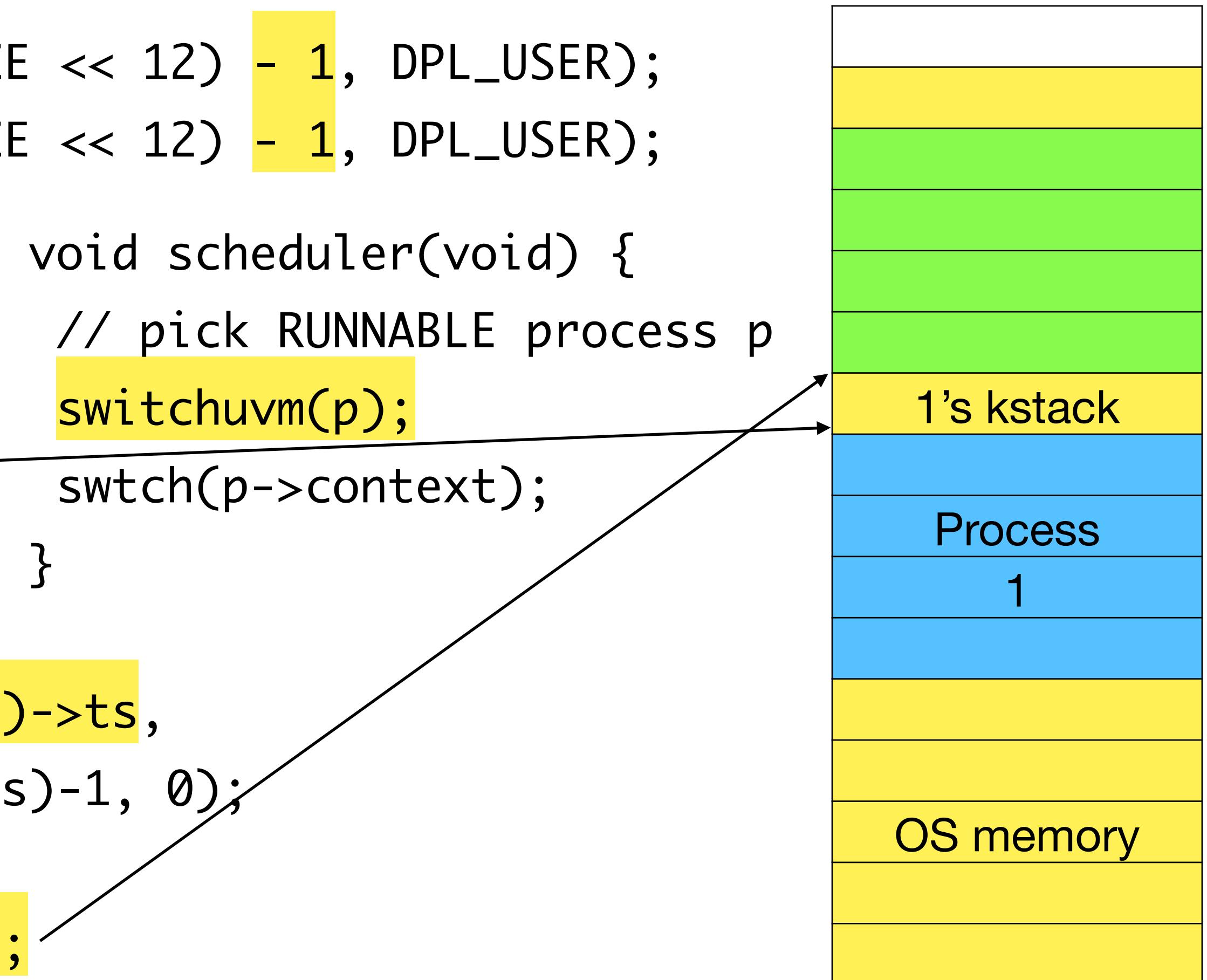
Code walkthrough: setting up kernel stack p16-tss

```
void seginit(void) {  
    c->gdt[SEG_UCODE] = SEG(.., STARTPROC, (PROCSIZE << 12) - 1, DPL_USER);  
    c->gdt[SEG_UDATA] = SEG(.., STARTPROC, (PROCSIZE << 12) - 1, DPL_USER);  
}  
  
static struct proc* allocproc(void) {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    p->kstack = sp - KSTACKSIZE;  
}  
  
void switchuvm(struct proc *p) {  
    mycpu()->gdt[SEG_TSS] = SEG16(STS_T32A, &mycpu()->ts,  
                                    sizeof(mycpu()->ts)-1, 0);  
    mycpu()->ts.ss0 = SEG_KDATA << 3;  
    mycpu()->ts.esp0 = (uint)p->kstack + KSTACKSIZE;  
    ltr(SEG_TSS << 3);  
}
```

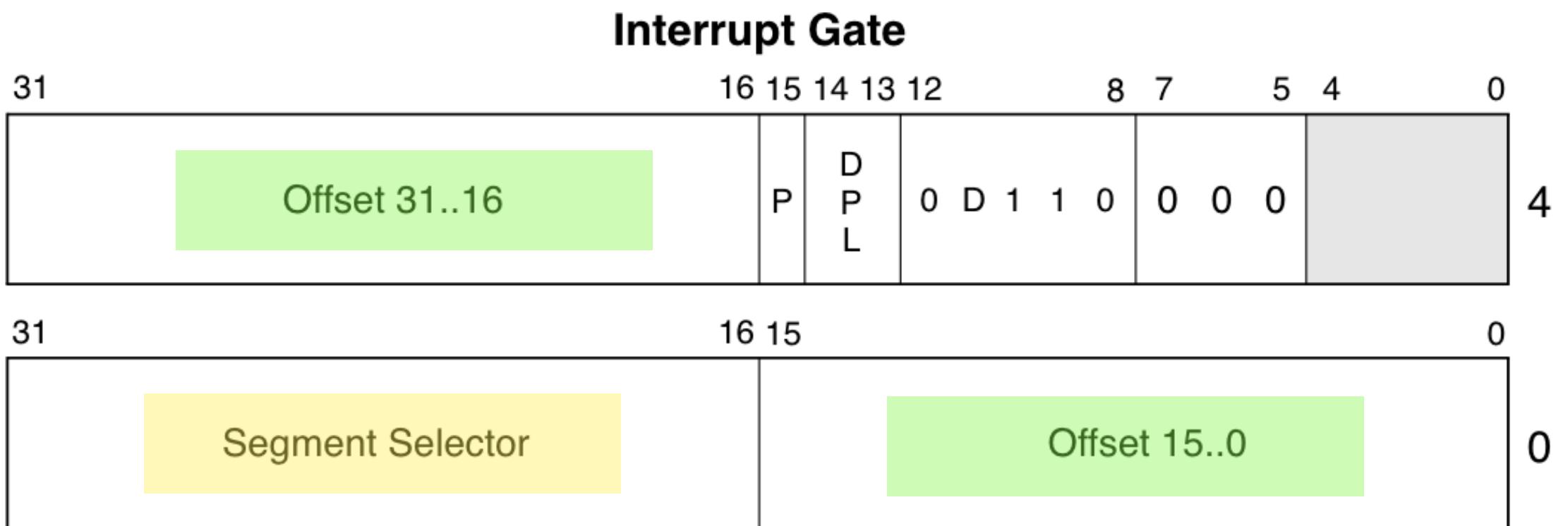


Code walkthrough: setting up kernel stack p16-tss

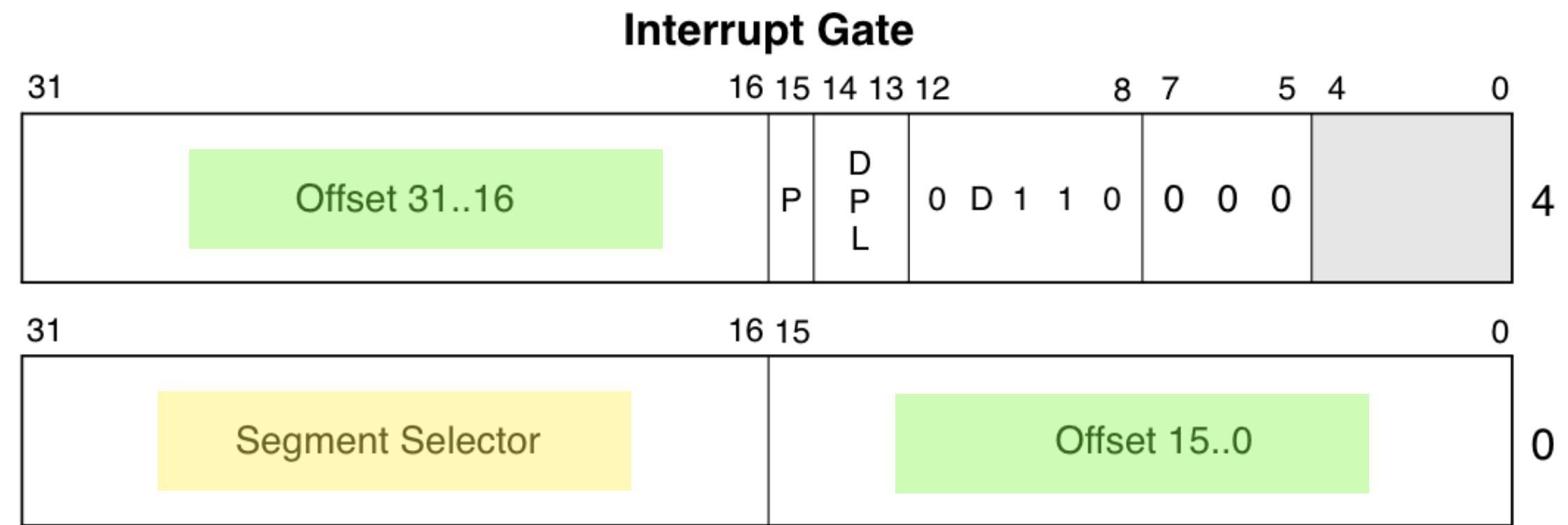
```
void seginit(void) {  
    c->gdt[SEG_UCODE] = SEG(.., STARTPROC, (PROCSIZE << 12) - 1, DPL_USER);  
    c->gdt[SEG_UDATA] = SEG(.., STARTPROC, (PROCSIZE << 12) - 1, DPL_USER);  
}  
  
static struct proc* allocproc(void) {  
    sp = (char*)(STARTPROC + (PROCSIZE>>12));  
    p->kstack = sp - KSTACKSIZE;  
}  
  
void switchuvm(struct proc *p) {  
    mycpu()->gdt[SEG_TSS] = SEG16(STS_T32A, &mycpu()->ts,  
                                    sizeof(mycpu()->ts)-1, 0);  
    mycpu()->ts.ss0 = SEG_KDATA << 3;  
    mycpu()->ts.esp0 = (uint)p->kstack + KSTACKSIZE;  
    ltr(SEG_TSS << 3);  
}
```



Interrupt handling

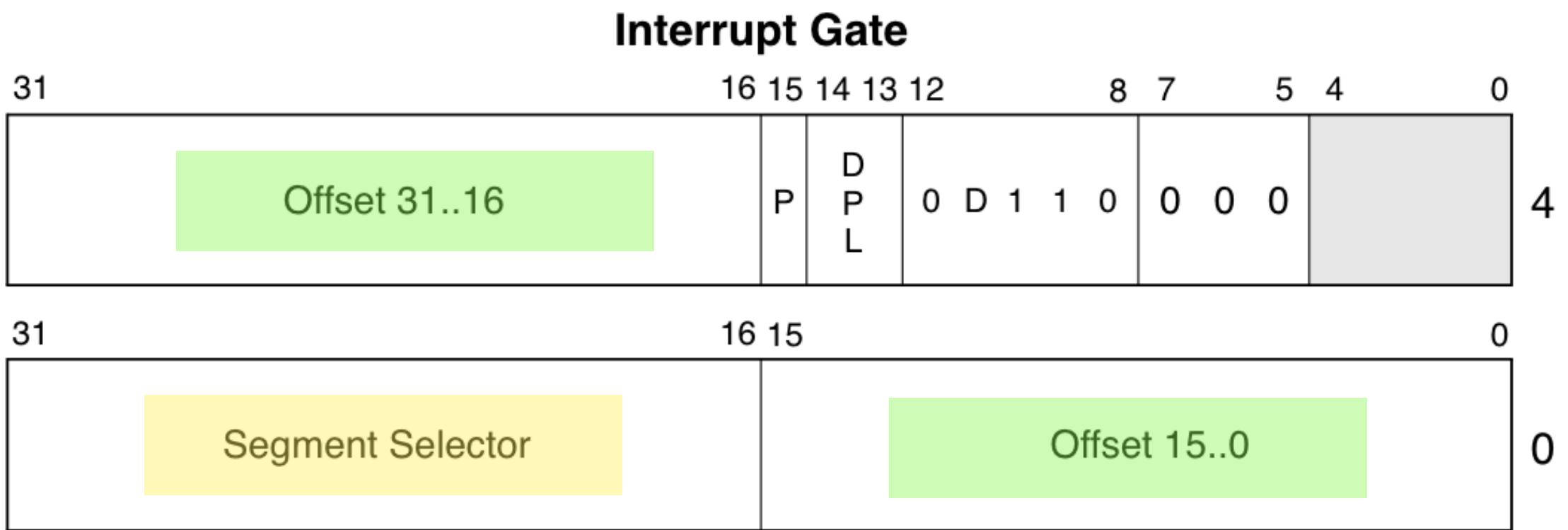


Interrupt handling



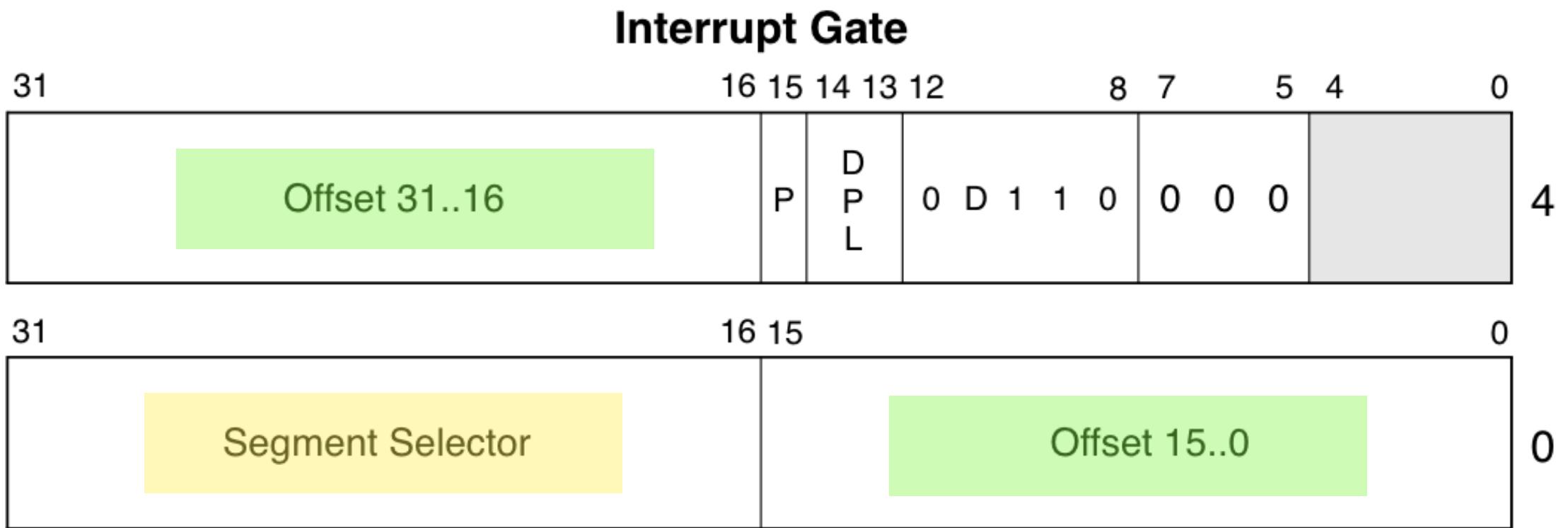
- Each IDT entry is 64-bits. Contains code segment and eip

Interrupt handling



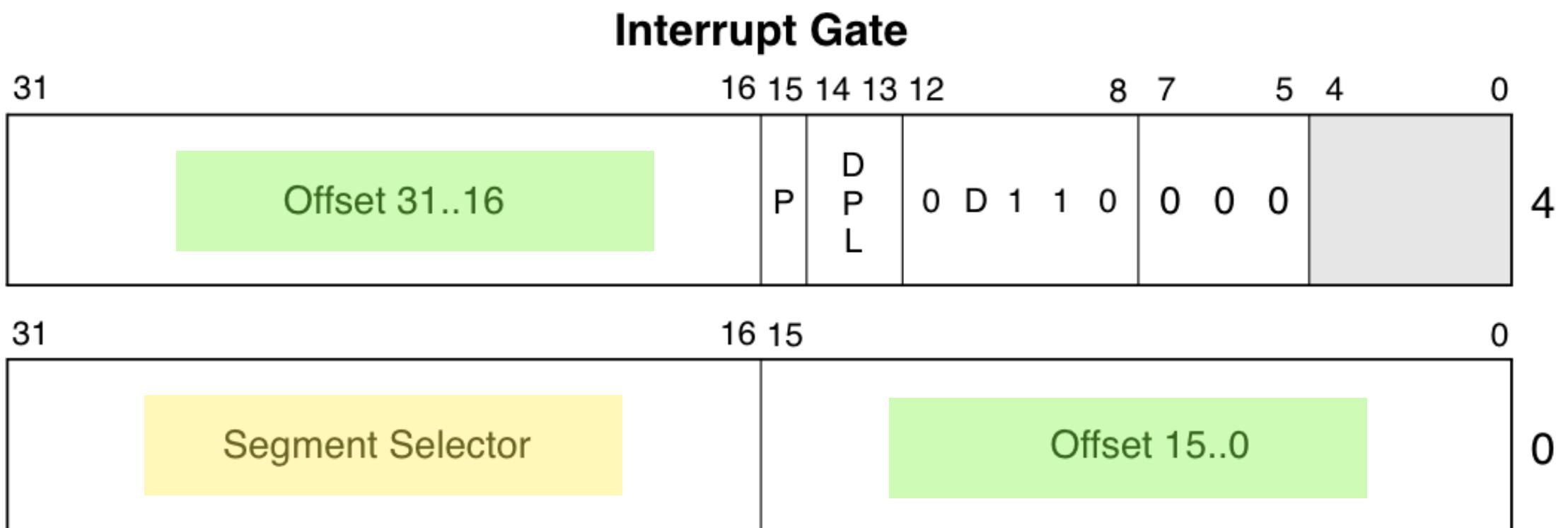
- Each IDT entry is 64-bits. Contains code segment and eip
- When interrupt appears,

Interrupt handling



- Each IDT entry is 64-bits. Contains code segment and eip
- When interrupt appears,
 - Hardware changes SS and ESP according to the TSS

Interrupt handling



- Each IDT entry is 64-bits. Contains code segment and eip
- When interrupt appears,
 - Hardware changes SS and ESP according to the TSS
 - Hardware changes CS and EIP to the one pointed by IDT entry.

Interrupt handling

- Each IDT entry is 64-bits. Contains code segment and eip
- When interrupt appears,
 - Hardware changes SS and ESP according to the TSS
 - Hardware changes CS and EIP to the one pointed by IDT entry.

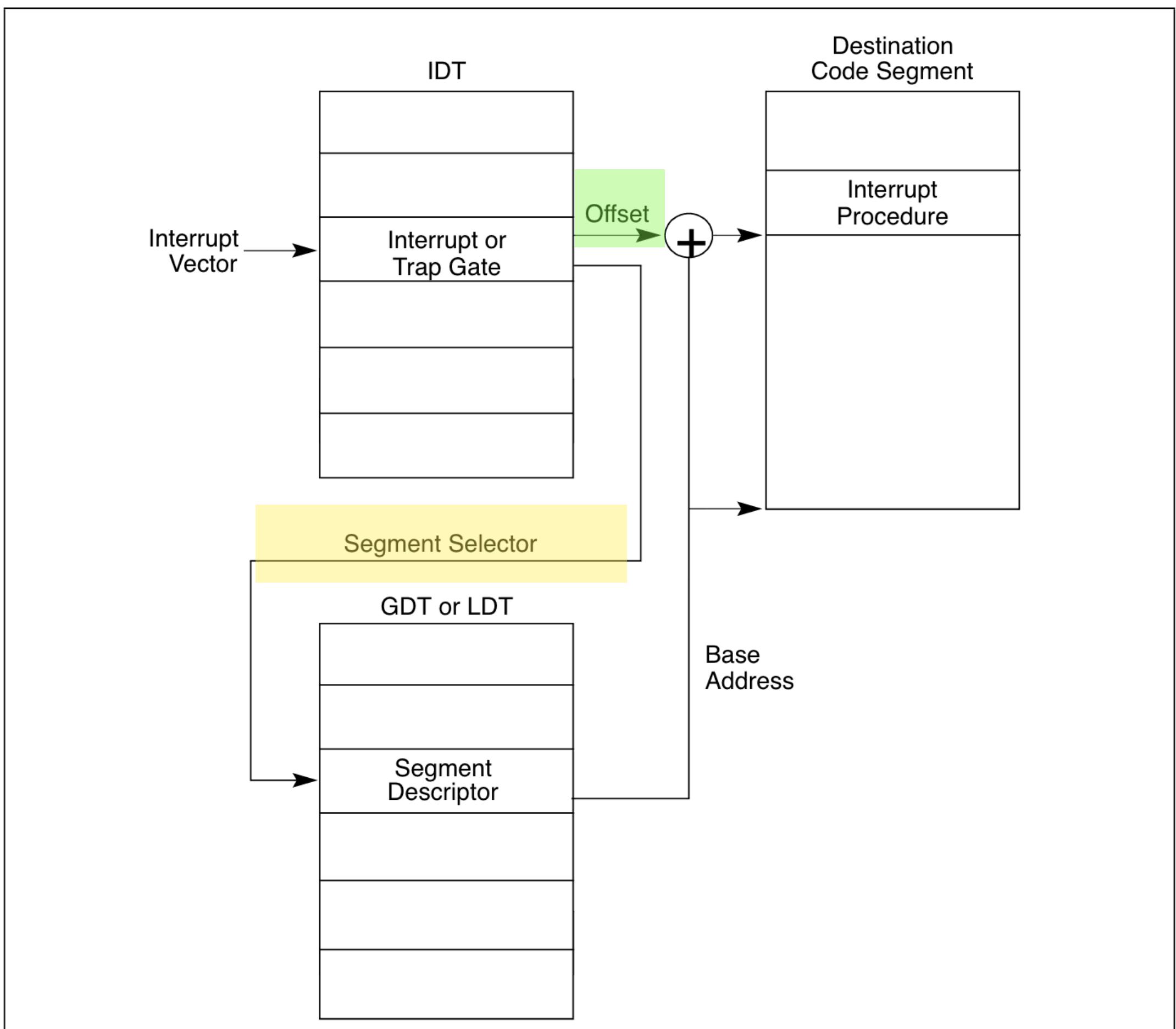
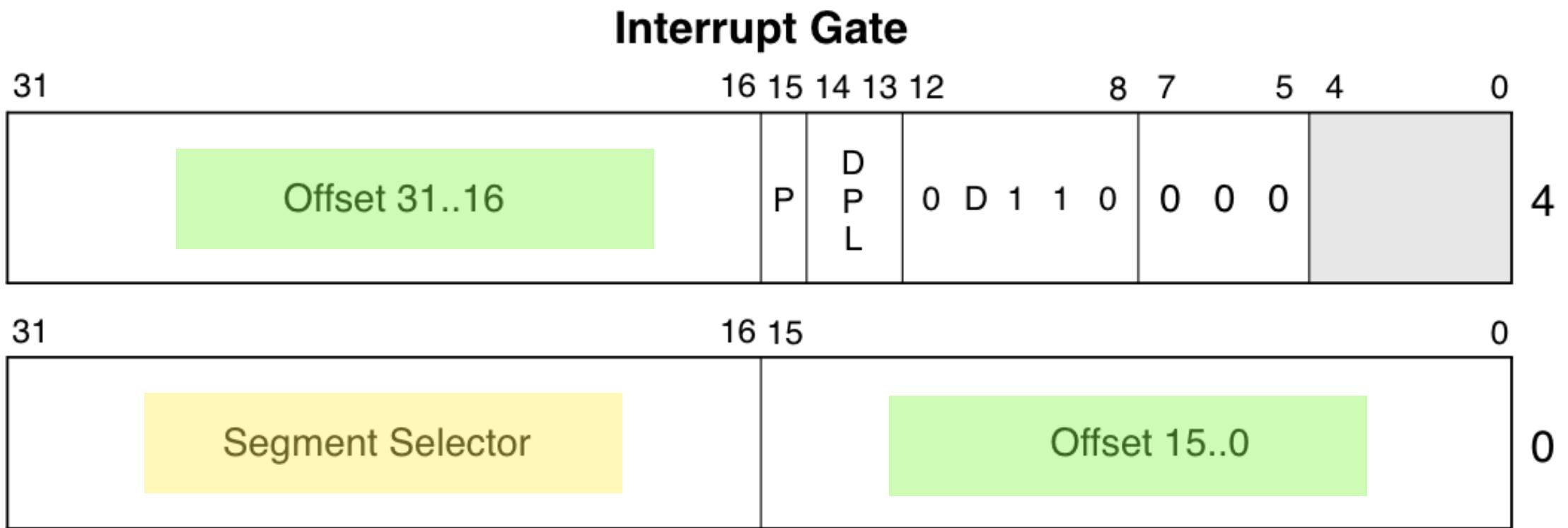


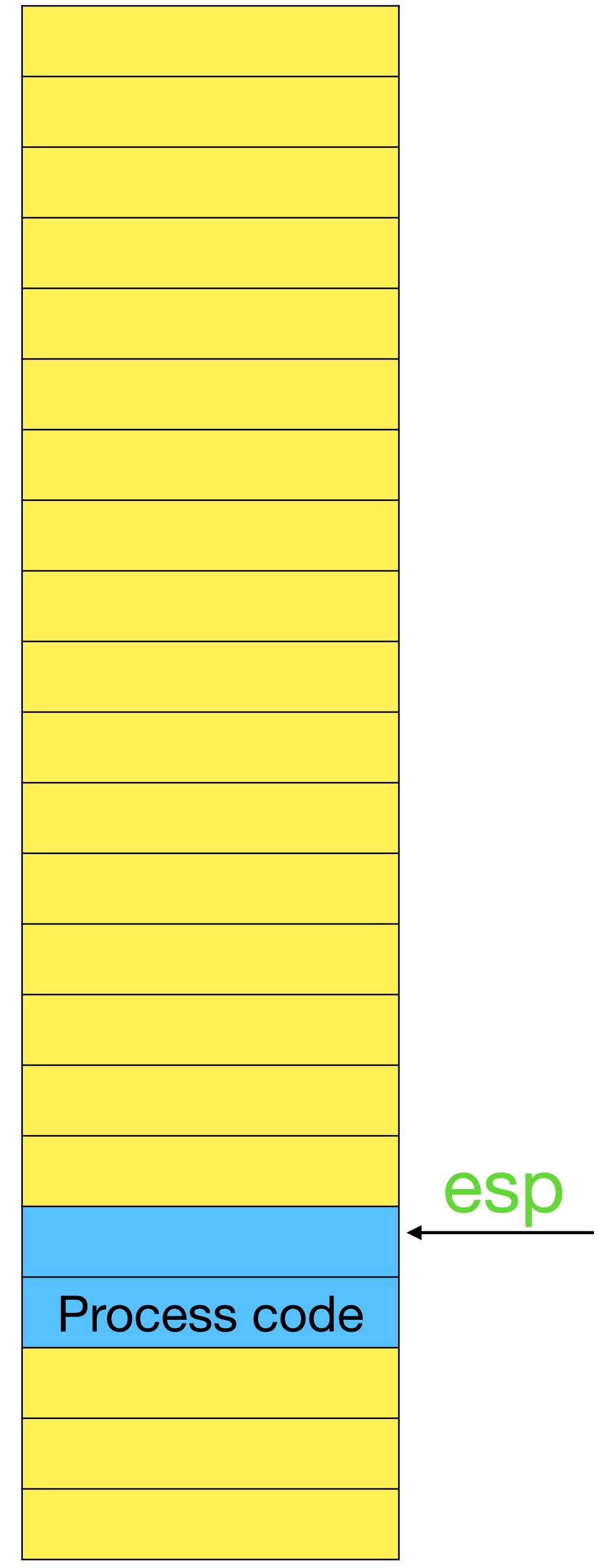
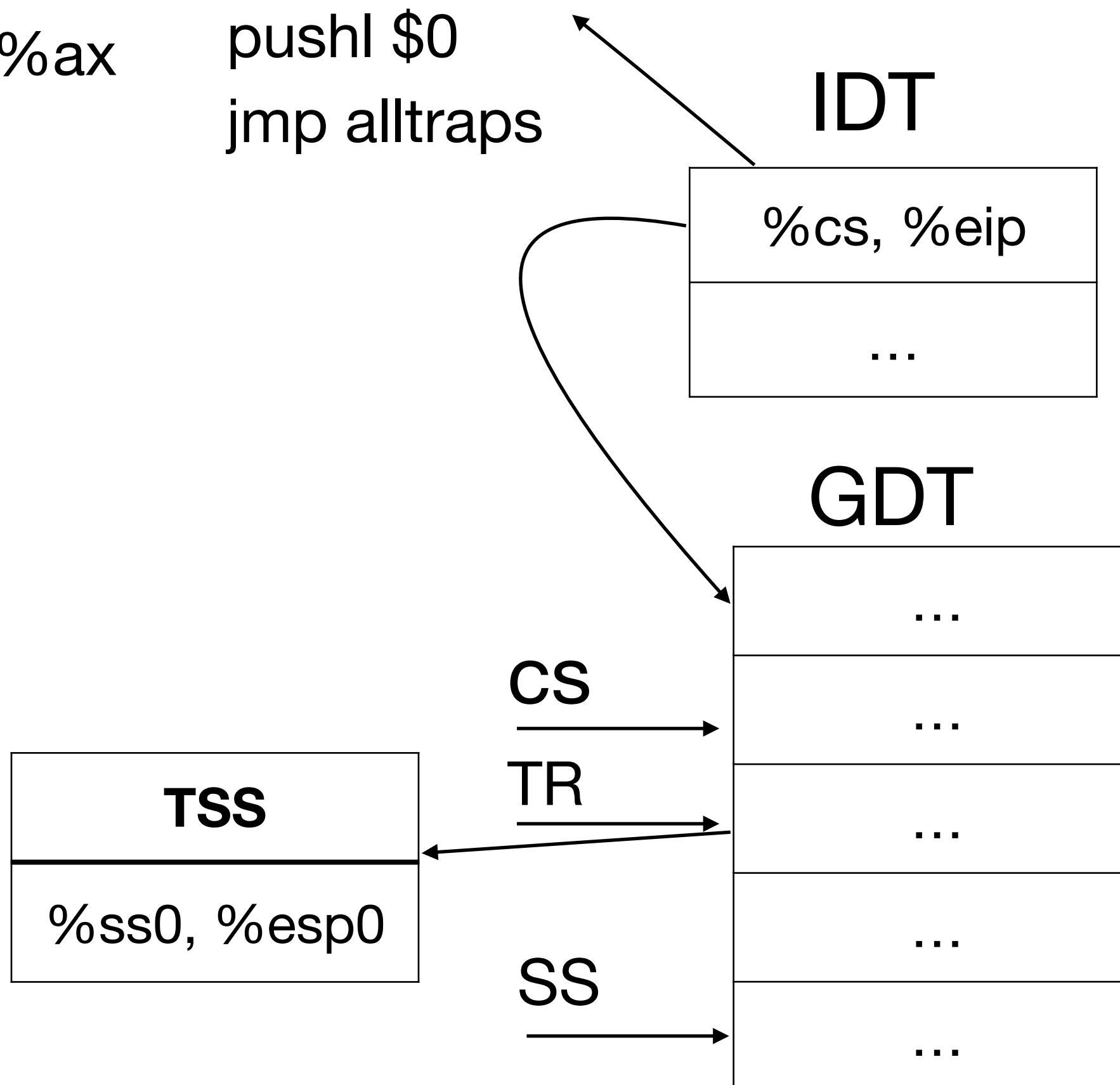
Figure 6-3. Interrupt Procedure Call

Interrupt handling with user process running

```
eip → for(;;)
      ;           trapasm.S
      alltraps:
      pushl %ds..
      pushal
      movw $(SEG_KDATA<<3), %ax
      movw %ax, %ds..
      pushl %esp
      call trap
      addl $4, %esp
      popal
      popl %ds..
      addl $0x8, %esp
      iret
```

vectors.S

```
.globl vector0
vector0:
pushl $0
pushl $0
jmp alltraps
```

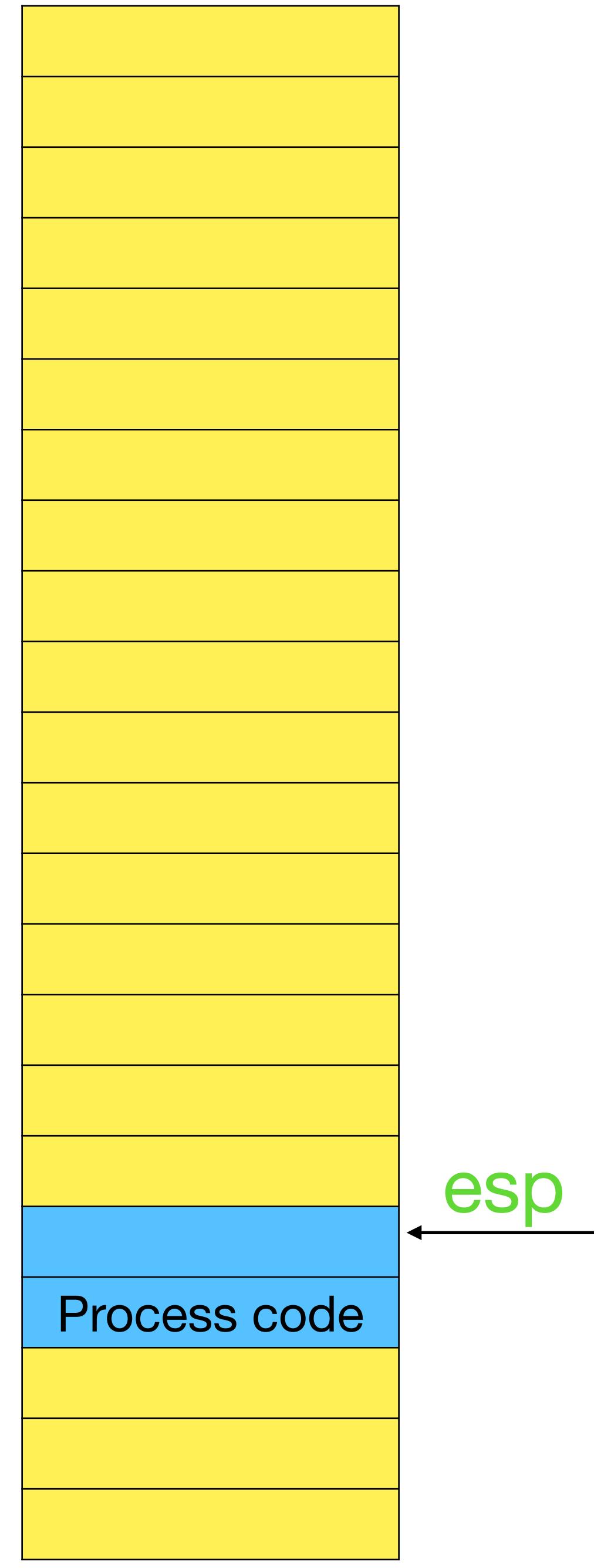
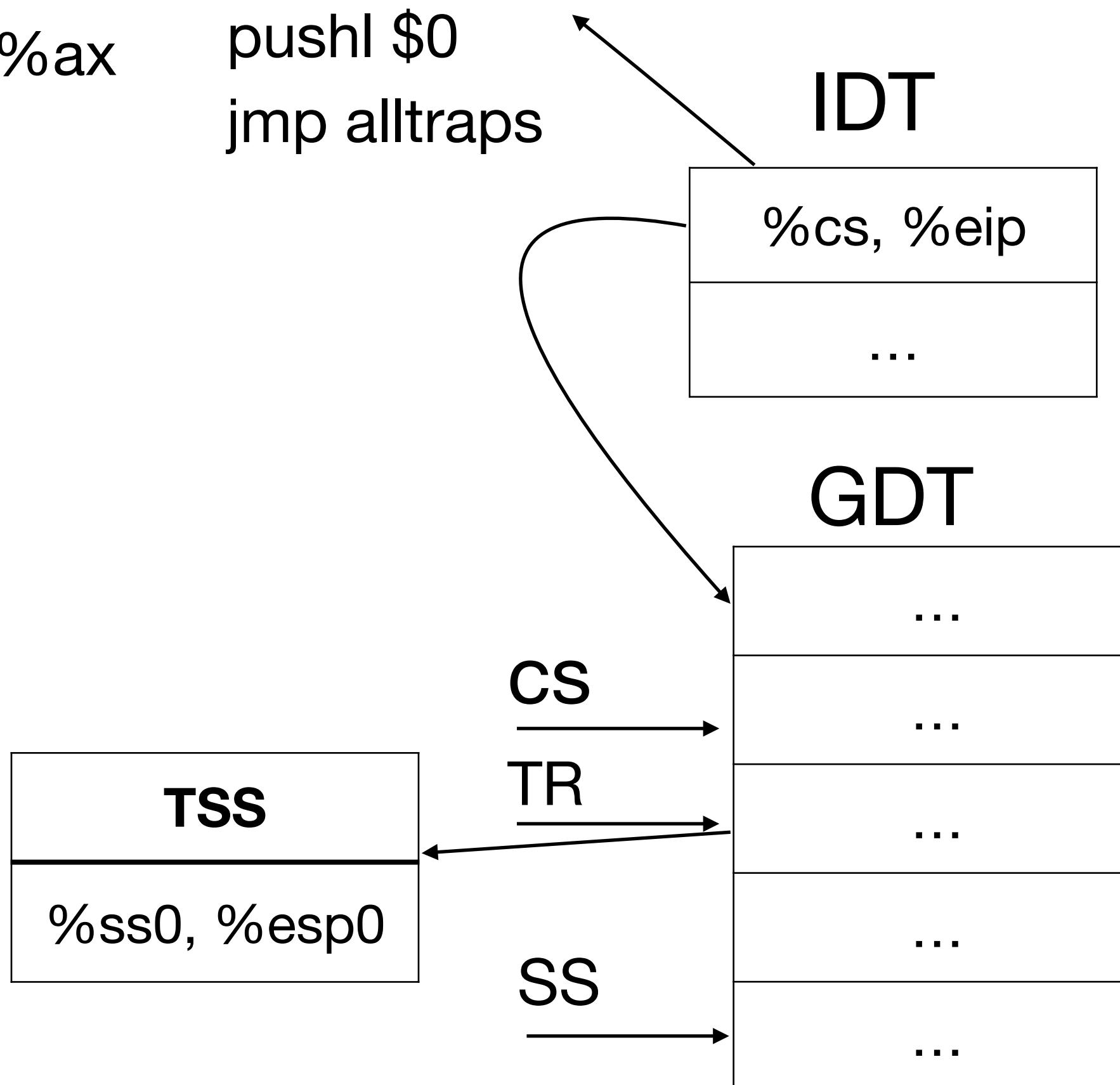


Interrupt handling with user process running

```
eip → for(;;)
for(;;)
;
trapasm.S
alltraps:
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

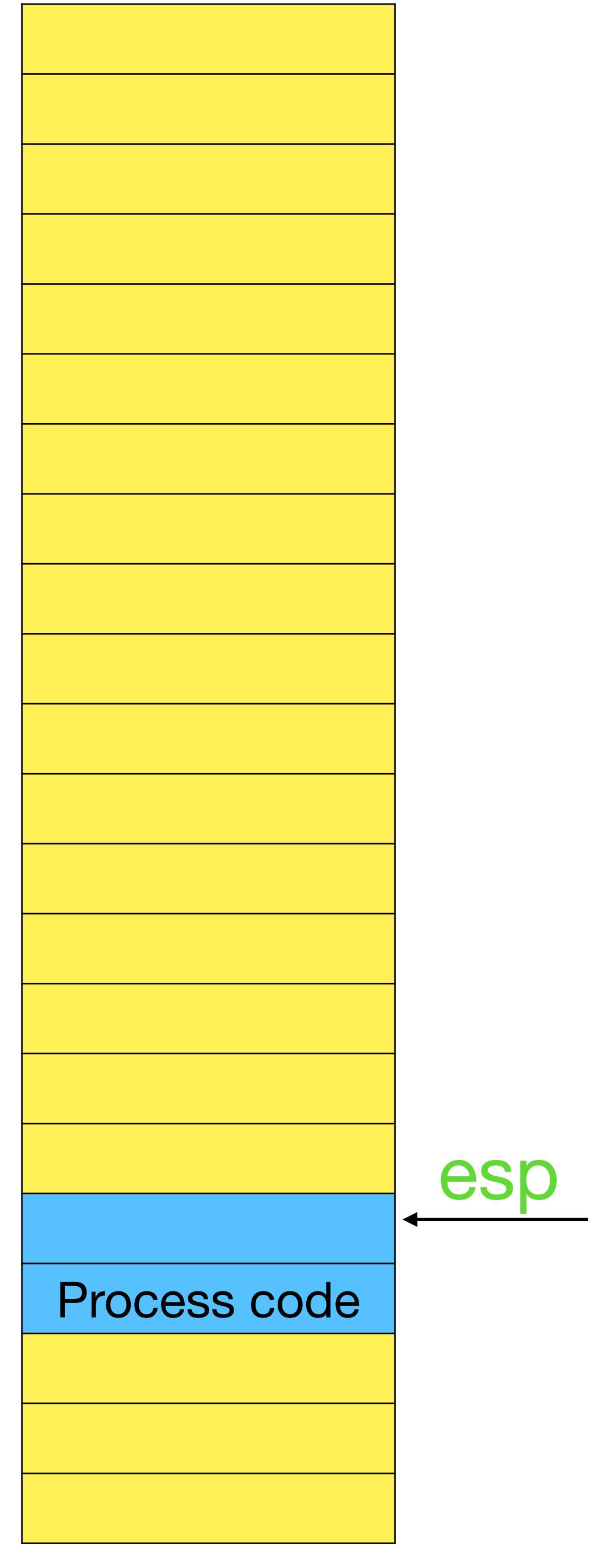
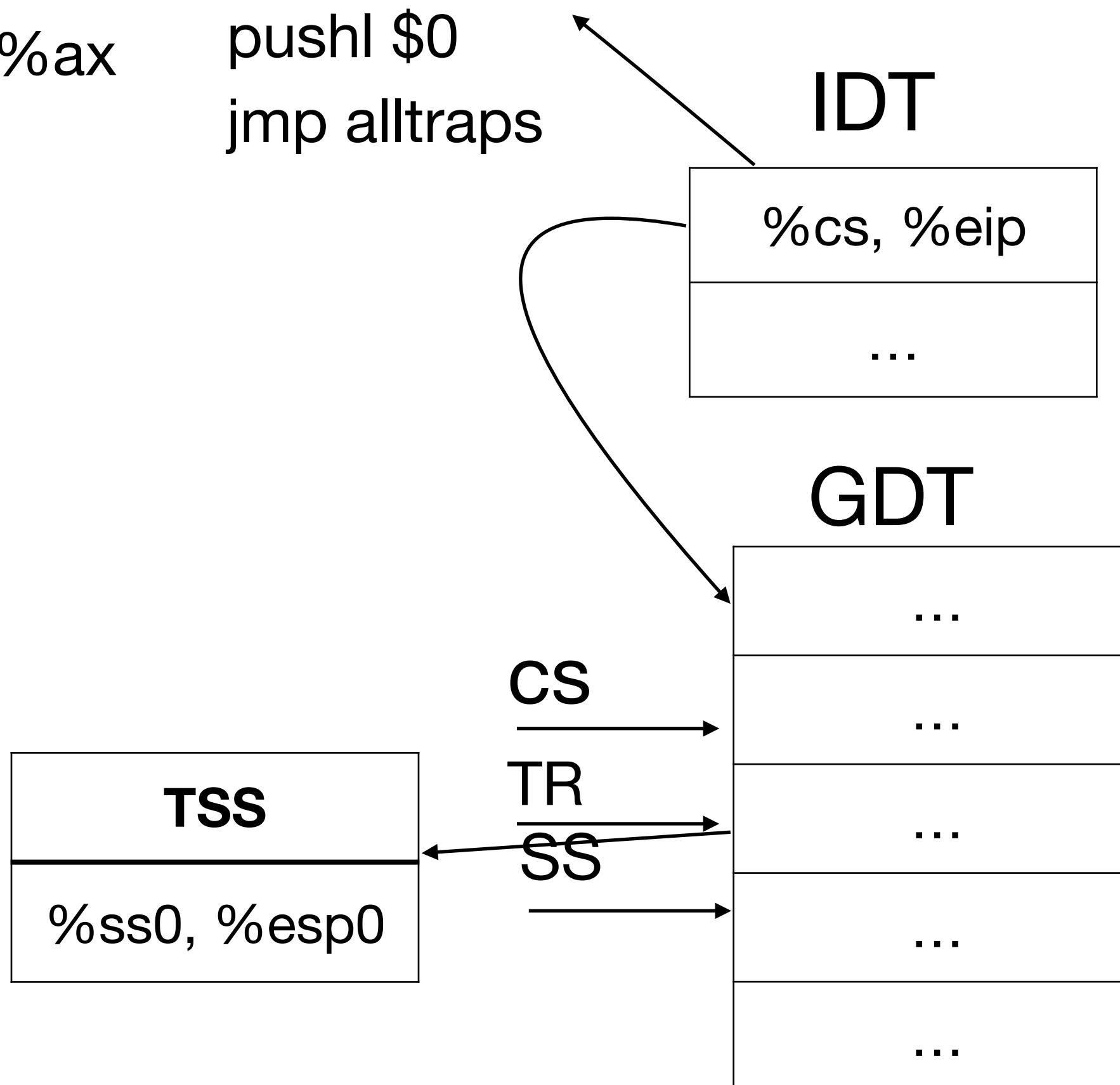


Interrupt handling with user process running

```
eip → for(;;)
      trapasm.S
      alltraps:
      pushl %ds..
      pushal
      movw $(SEG_KDATA<<3), %ax
      movw %ax, %ds..
      pushl %esp
      call trap
      addl $4, %esp
      popal
      popl %ds..
      addl $0x8, %esp
      iret
```

vectors.S

```
.globl vector0
vector0:
pushl $0
pushl $0
jmp alltraps
```

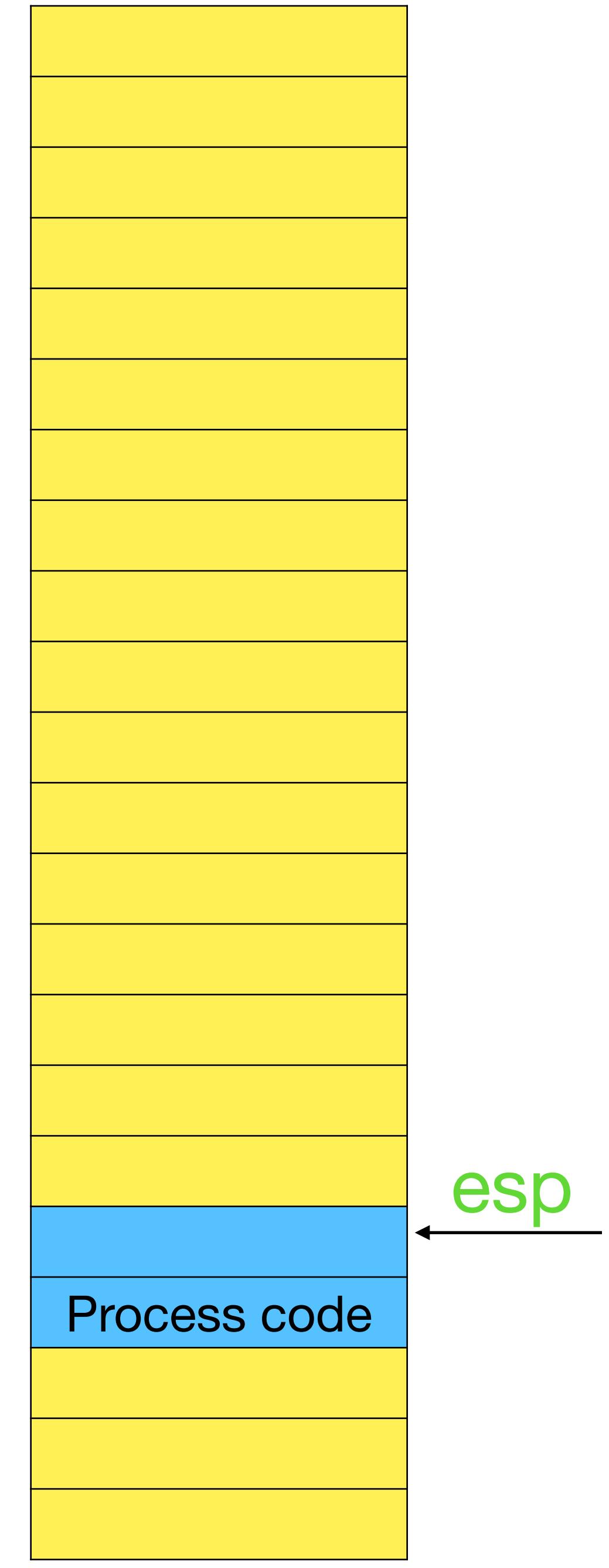
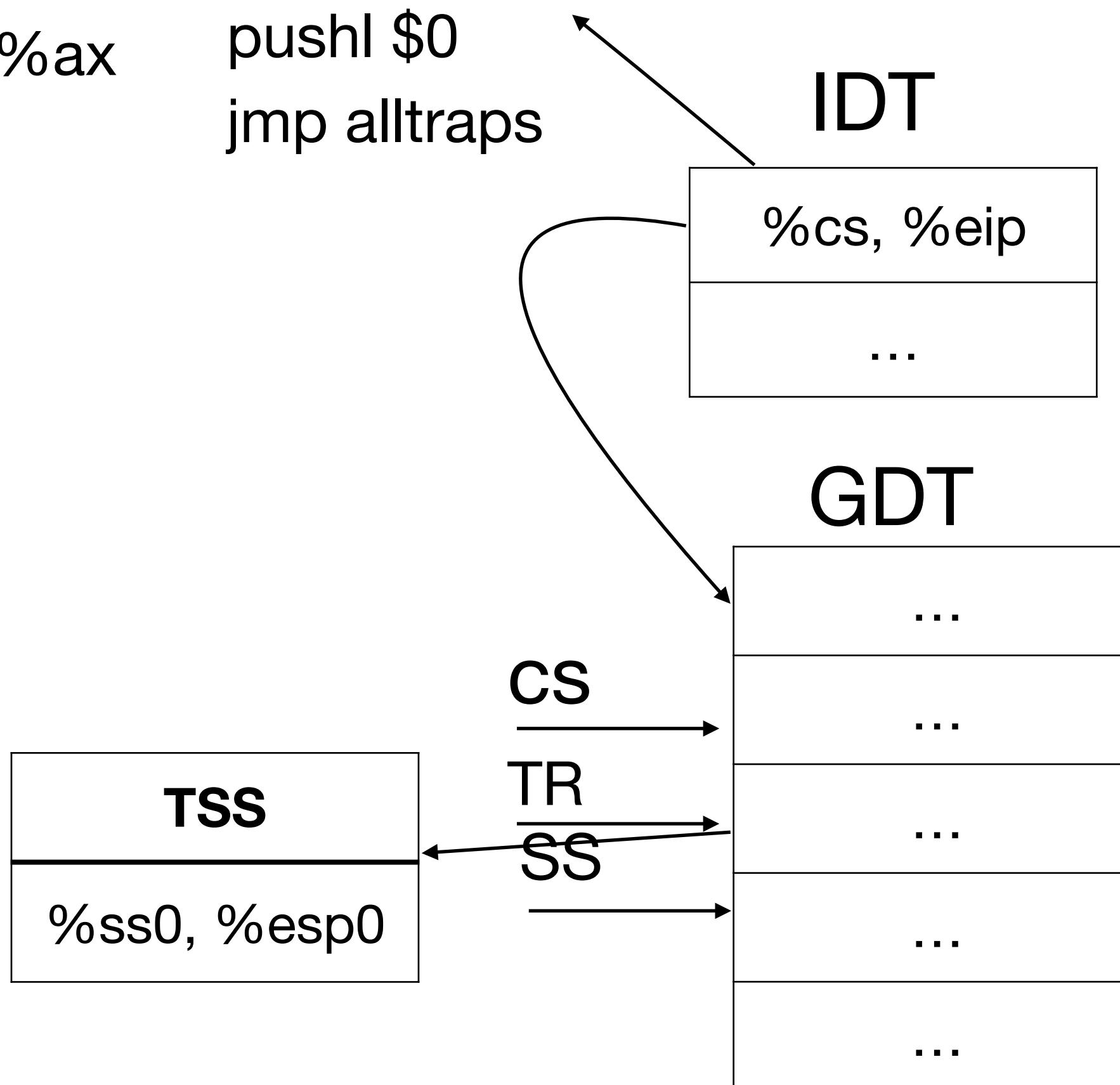


Interrupt handling with user process running

```
eip → for(;;)
      trapasm.S
      alltraps:
      pushl %ds..
      pushal
      movw $(SEG_KDATA<<3), %ax
      movw %ax, %ds..
      pushl %esp
      call trap
      addl $4, %esp
      popal
      popl %ds..
      addl $0x8, %esp
      iret
```

vectors.S

```
.globl vector0
vector0:
pushl $0
pushl $0
jmp alltraps
```

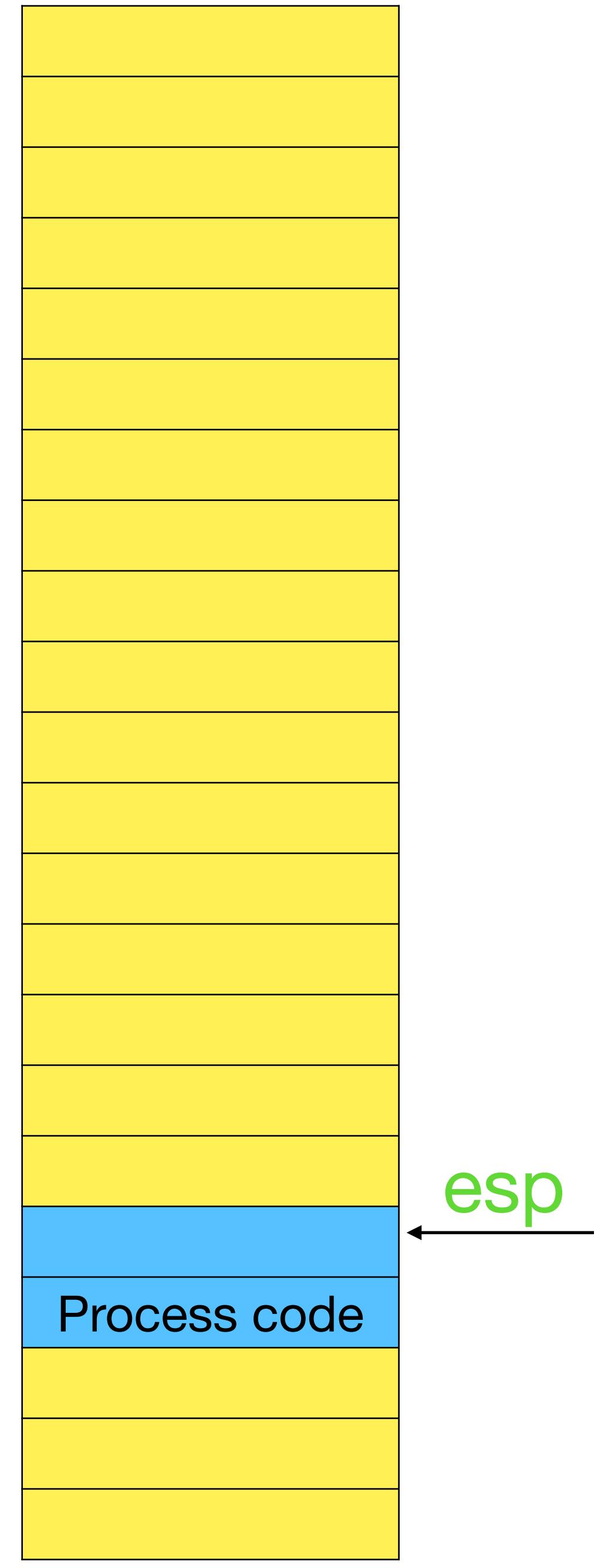
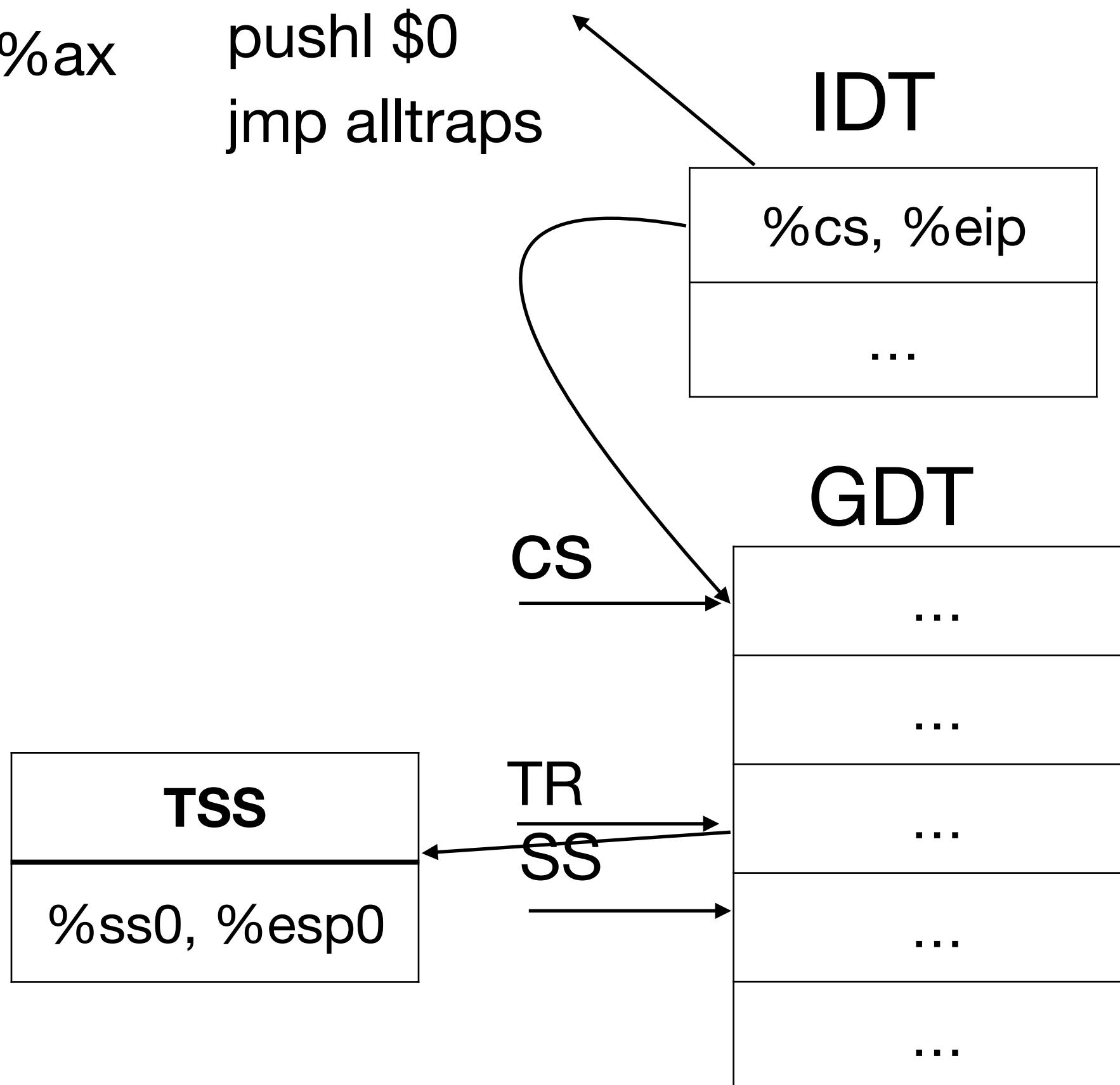


Interrupt handling with user process running

```
eip → for(;;)
      ;           trapasm.S
      alltraps:
      pushl %ds..
      pushal
      movw $(SEG_KDATA<<3), %ax
      movw %ax, %ds..
      pushl %esp
      call trap
      addl $4, %esp
      popal
      popl %ds..
      addl $0x8, %esp
      iret
```

vectors.S

```
.globl vector0
vector0:
pushl $0
pushl $0
jmp alltraps
```



Interrupt handling with user process running

eip

```
for(;;)
```

trapasm.S

```
alltraps:
```

```
    pushl %ds..
```

```
    pushal
```

```
    movw $(SEG_KDATA<<3), %ax
```

```
    movw %ax, %ds..
```

```
    pushl %esp
```

```
    call trap
```

```
    addl $4, %esp
```

```
    popal
```

```
    popl %ds..
```

```
    addl $0x8, %esp
```

```
    iret
```

vectors.S

```
.globl vector0
```

```
vector0:
```

```
    pushl $0
```

```
    pushl $0
```

```
    jmp alltraps
```

IDT

%cs, %eip

...

CS

GDT

...

...

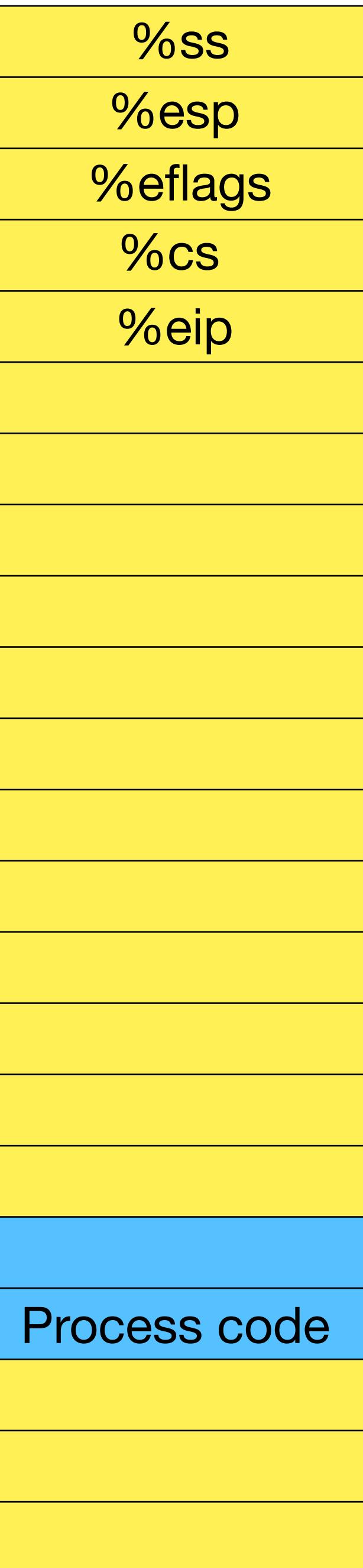
...

TR

SS

TSS

%ss0, %esp0

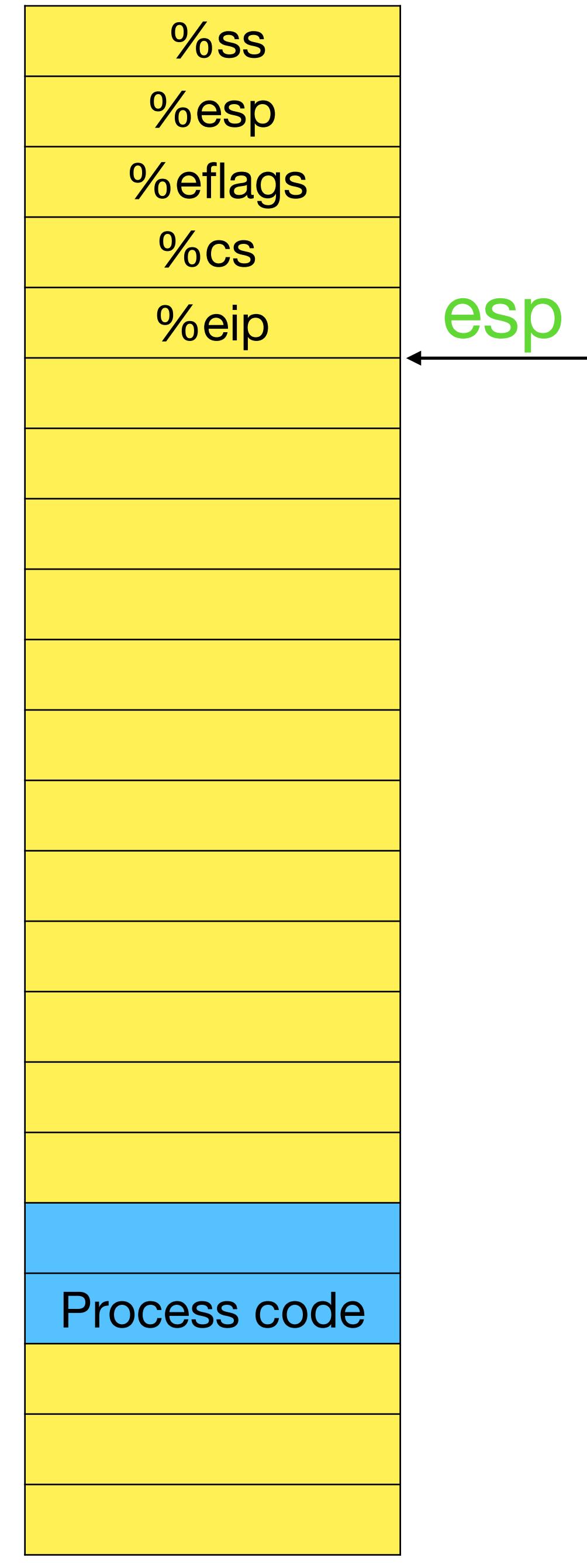
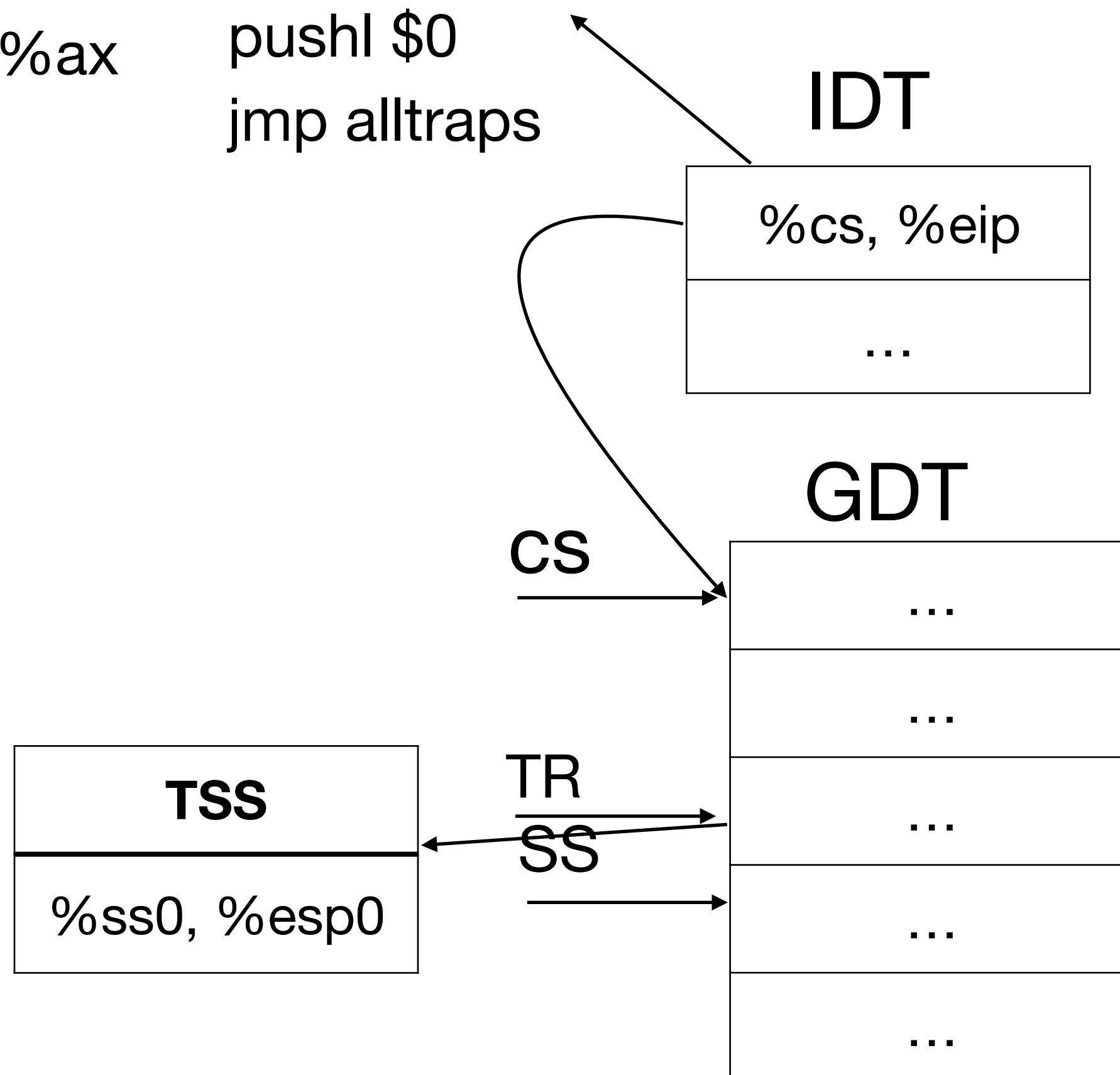


Interrupt handling with user process running

```
eip → for(;;)
;
trapasm.S
alltraps:
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
```

vectors.S

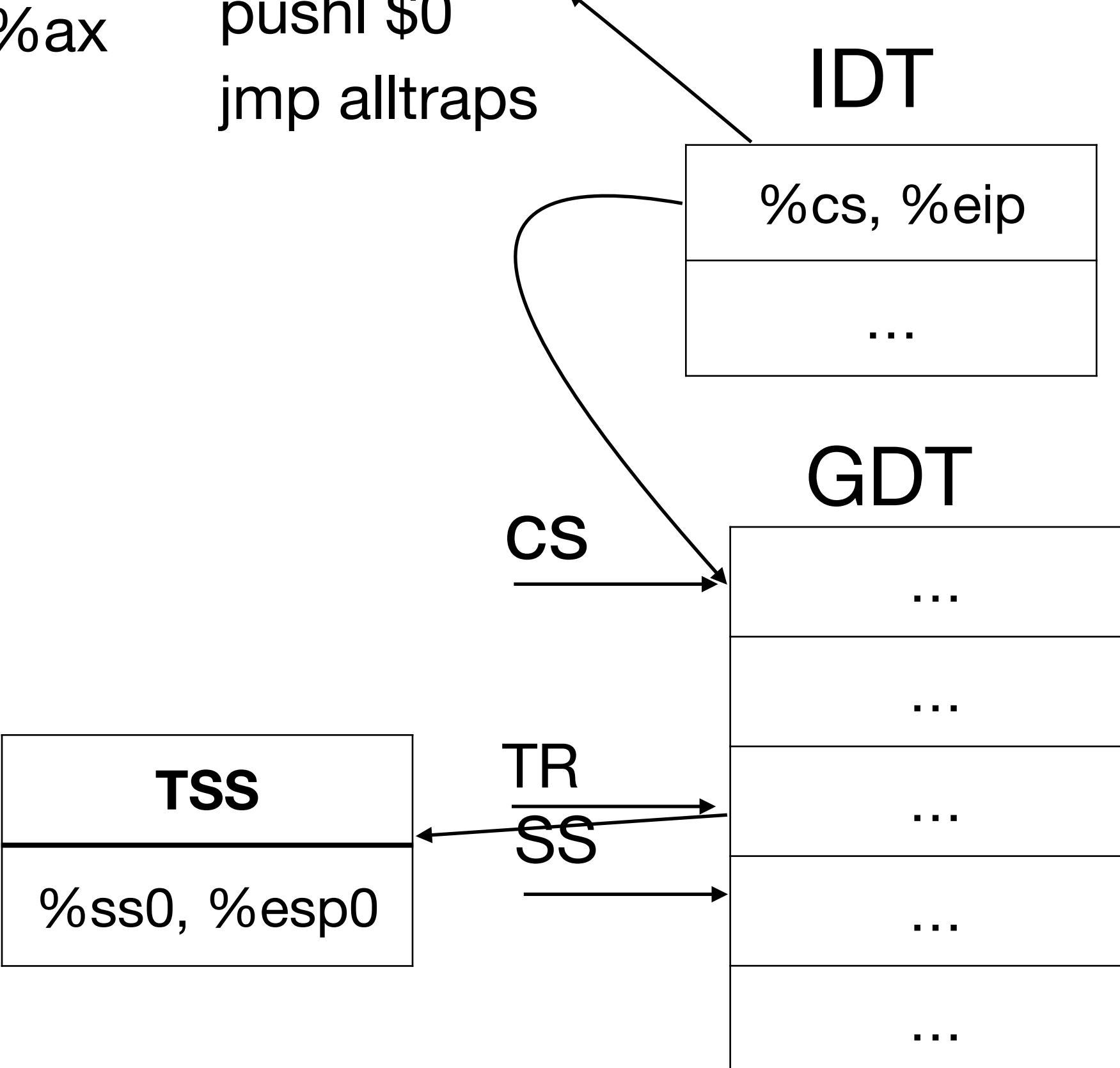
```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```



Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    alltraps:
        pushl %ds..
        pushal
        movw $(SEG_KDATA<<3), %ax
        movw %ax, %ds..
        pushl %esp
        call trap
        addl $4, %esp
        popal
        popl %ds..
        addl $0x8, %esp
        iret
```

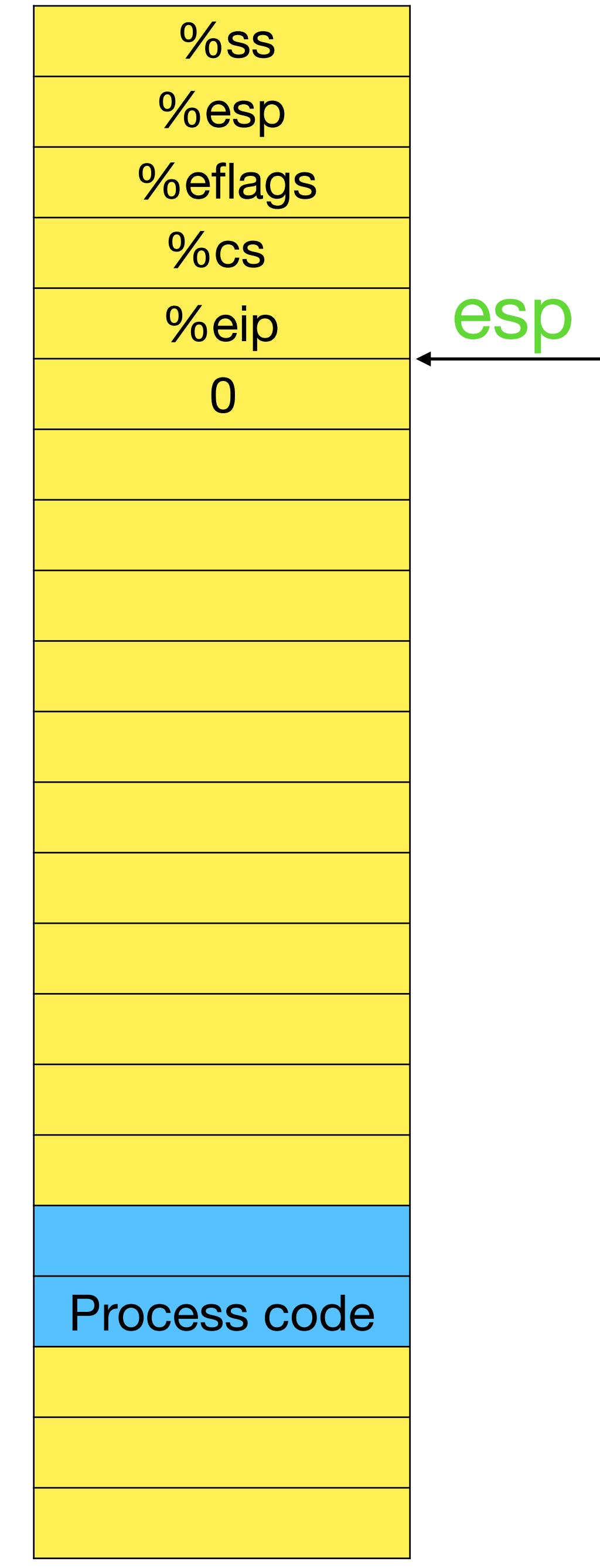
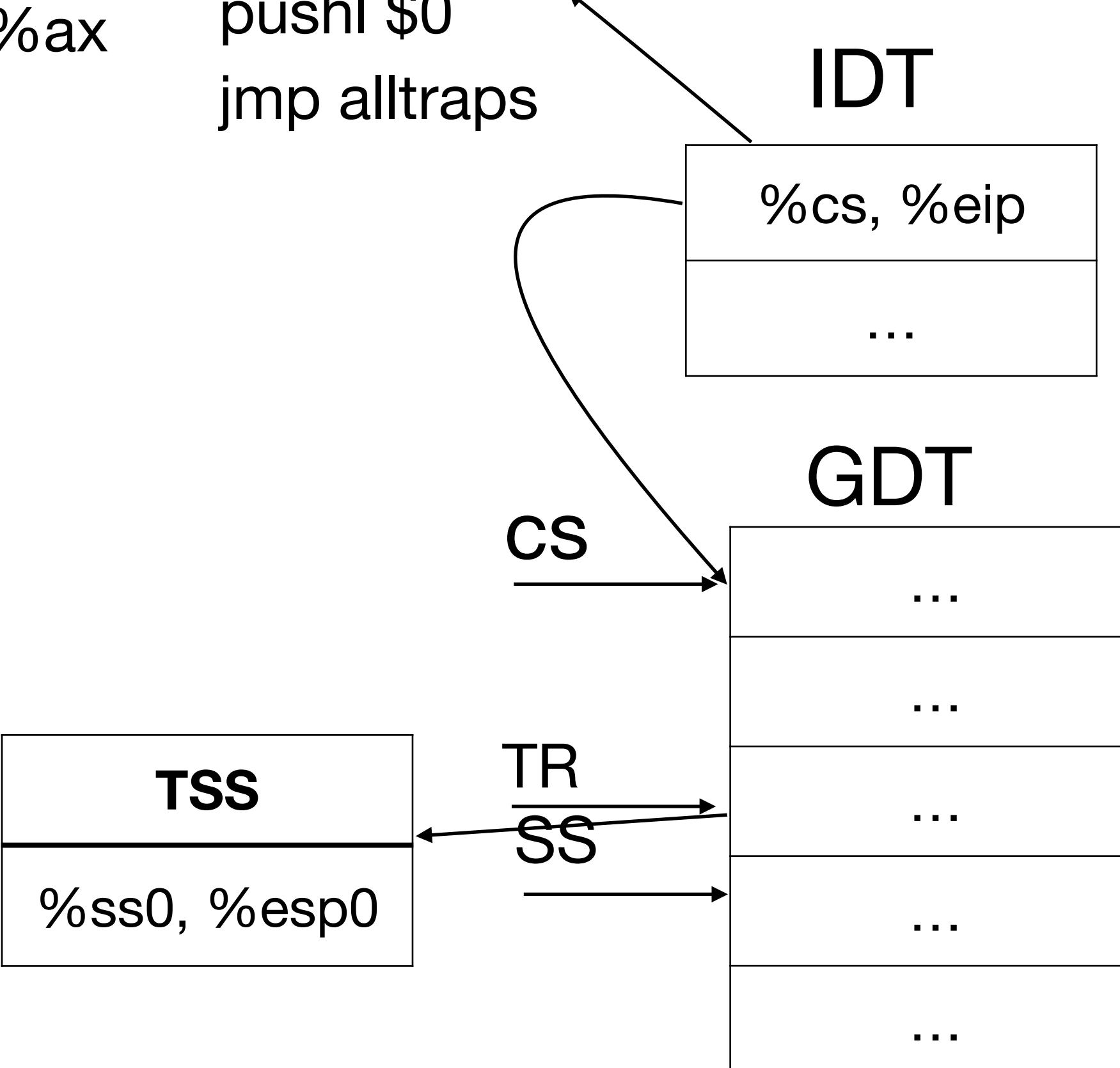
```
vectors.S
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```



Interrupt handling with user process running

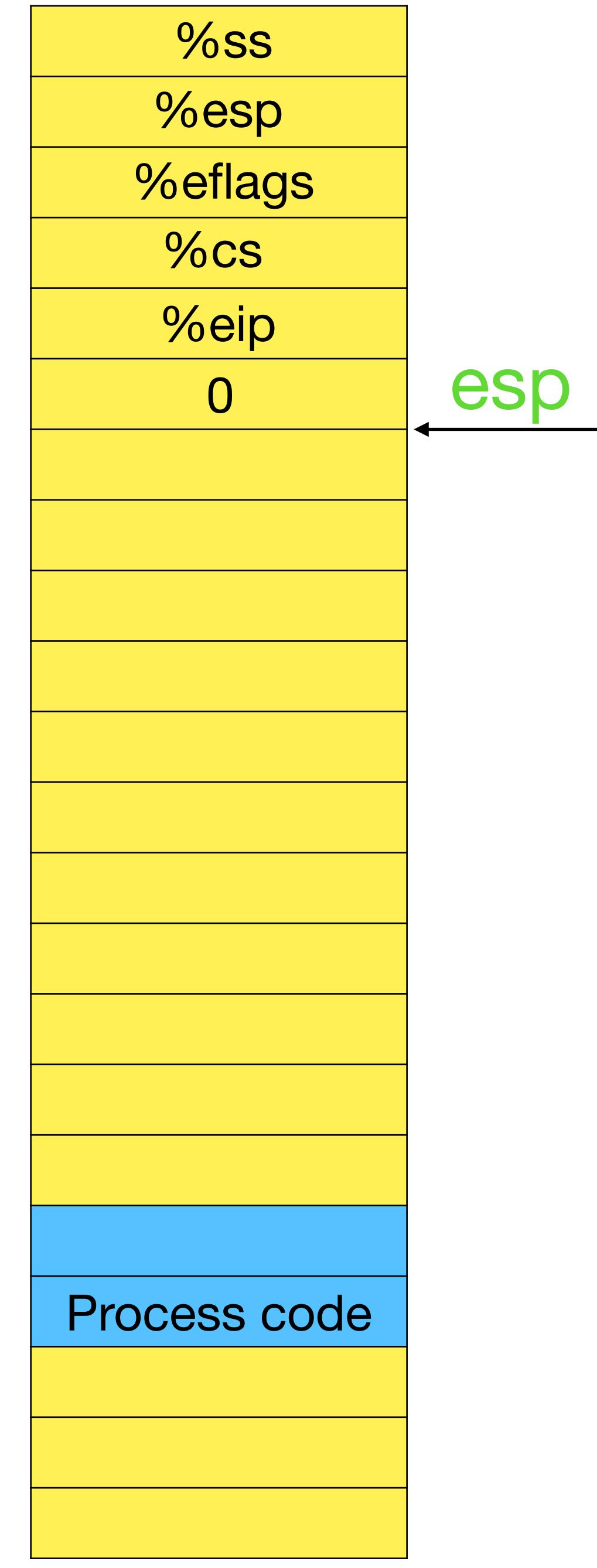
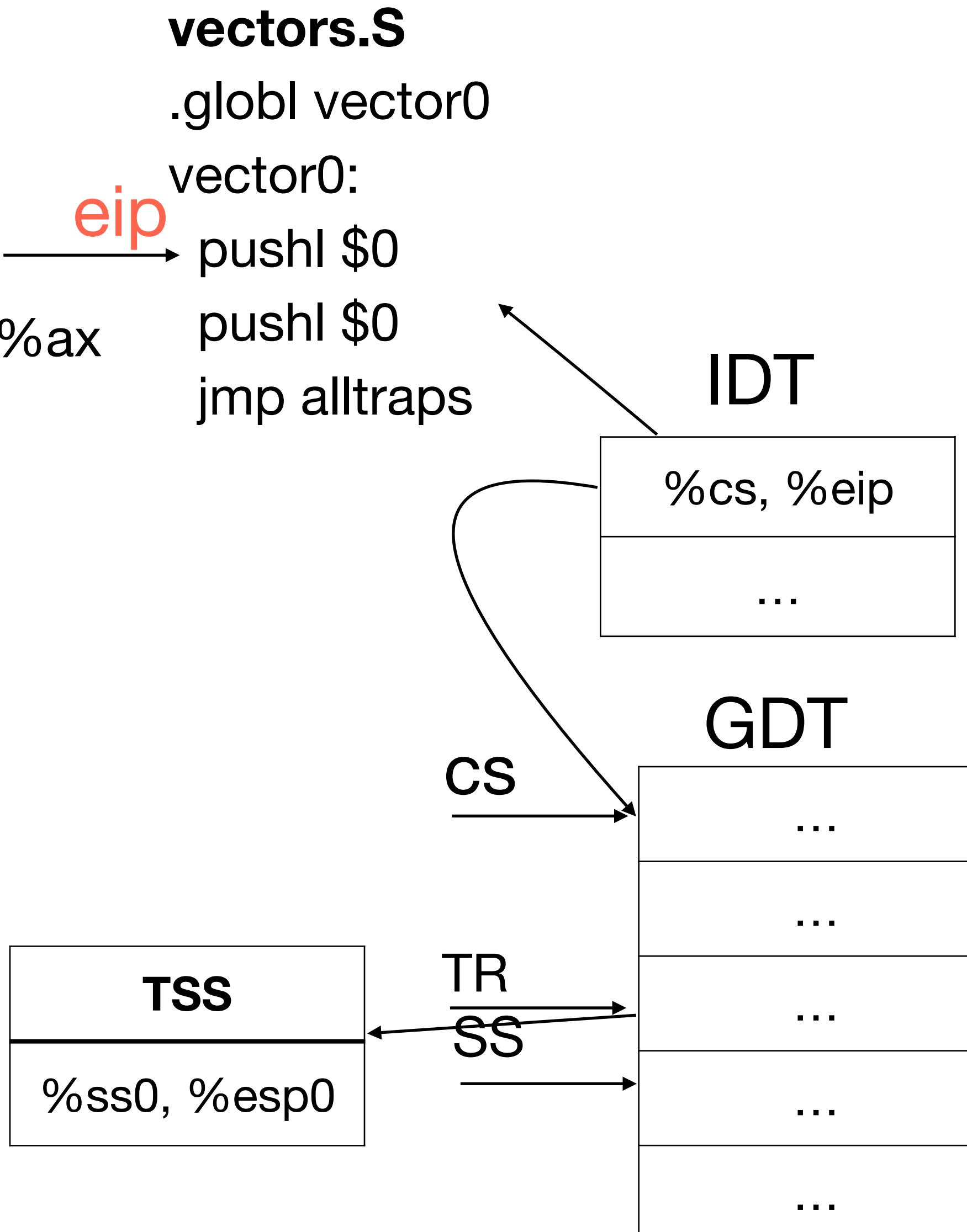
```
trapasm.S
for(;;)
;
    alltraps:
        pushl %ds..
        pushal
        movw $(SEG_KDATA<<3), %ax
        movw %ax, %ds..
        pushl %esp
        call trap
        addl $4, %esp
        popal
        popl %ds..
        addl $0x8, %esp
        iret
```

```
vectors.S
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```



Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    alltraps:
        pushl %ds..
        pushal
        movw $(SEG_KDATA<<3), %ax
        movw %ax, %ds..
        pushl %esp
        call trap
        addl $4, %esp
        popal
        popl %ds..
        addl $0x8, %esp
        iret
```



Interrupt handling with user process running

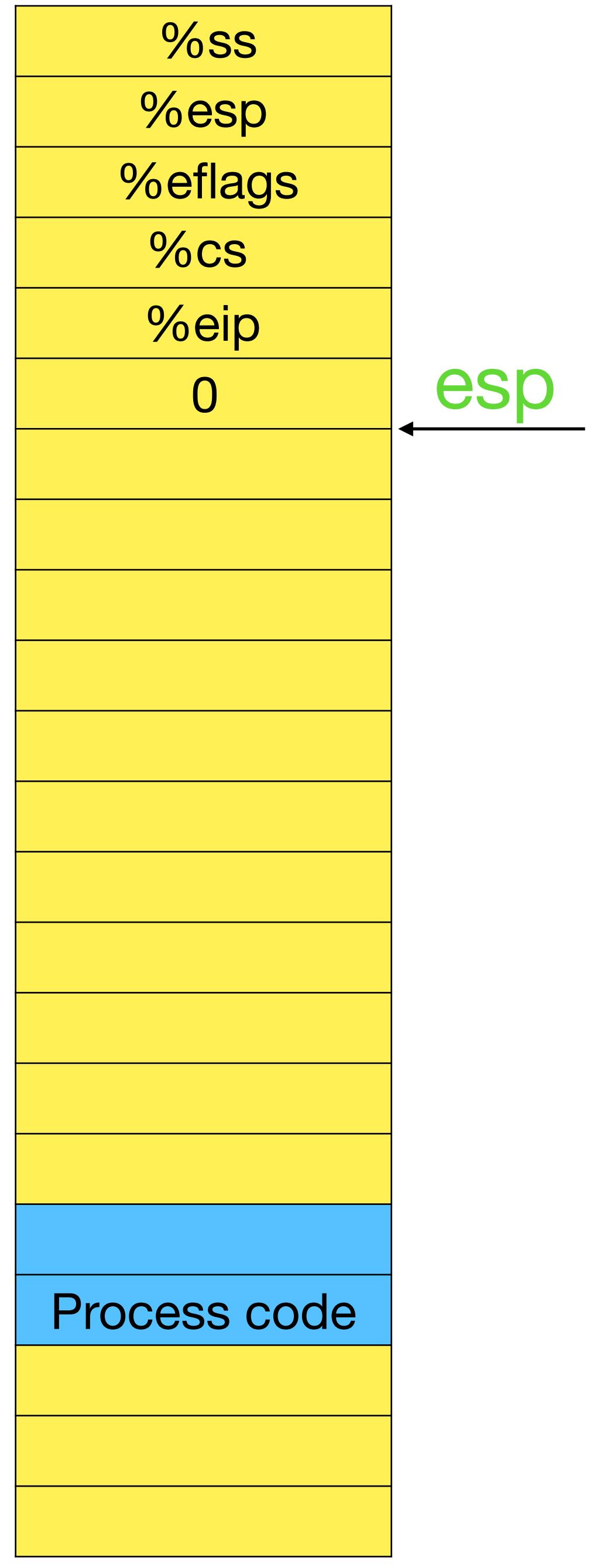
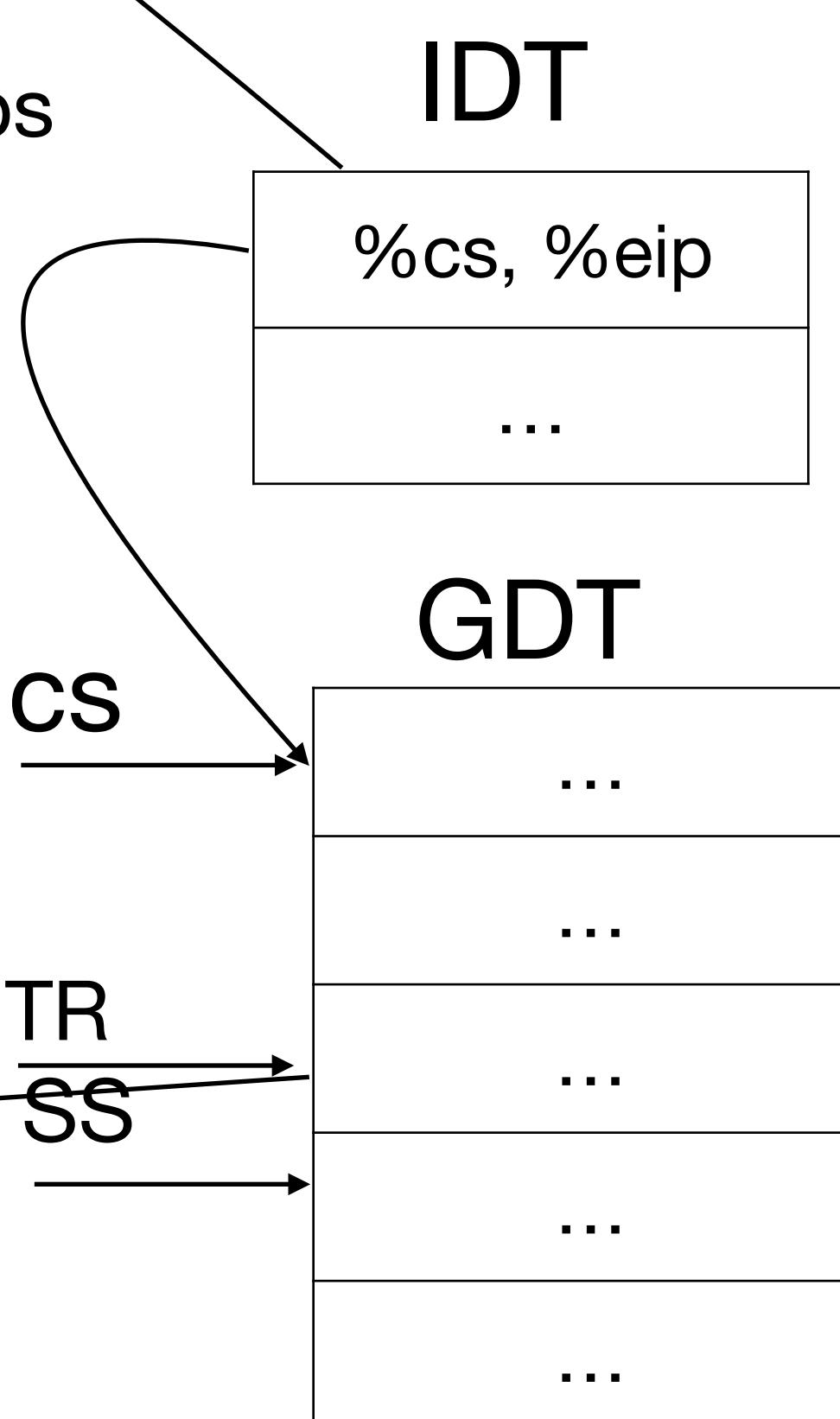
```
trapasm.S
for(;;)
;
    alltraps:
        pushl %ds..
        pushal
        movw $(SEG_KDATA<<3), %ax
        movw %ax, %ds..
        pushl %esp
        call trap
        addl $4, %esp
        popal
        popl %ds..
        addl $0x8, %esp
        iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

eip

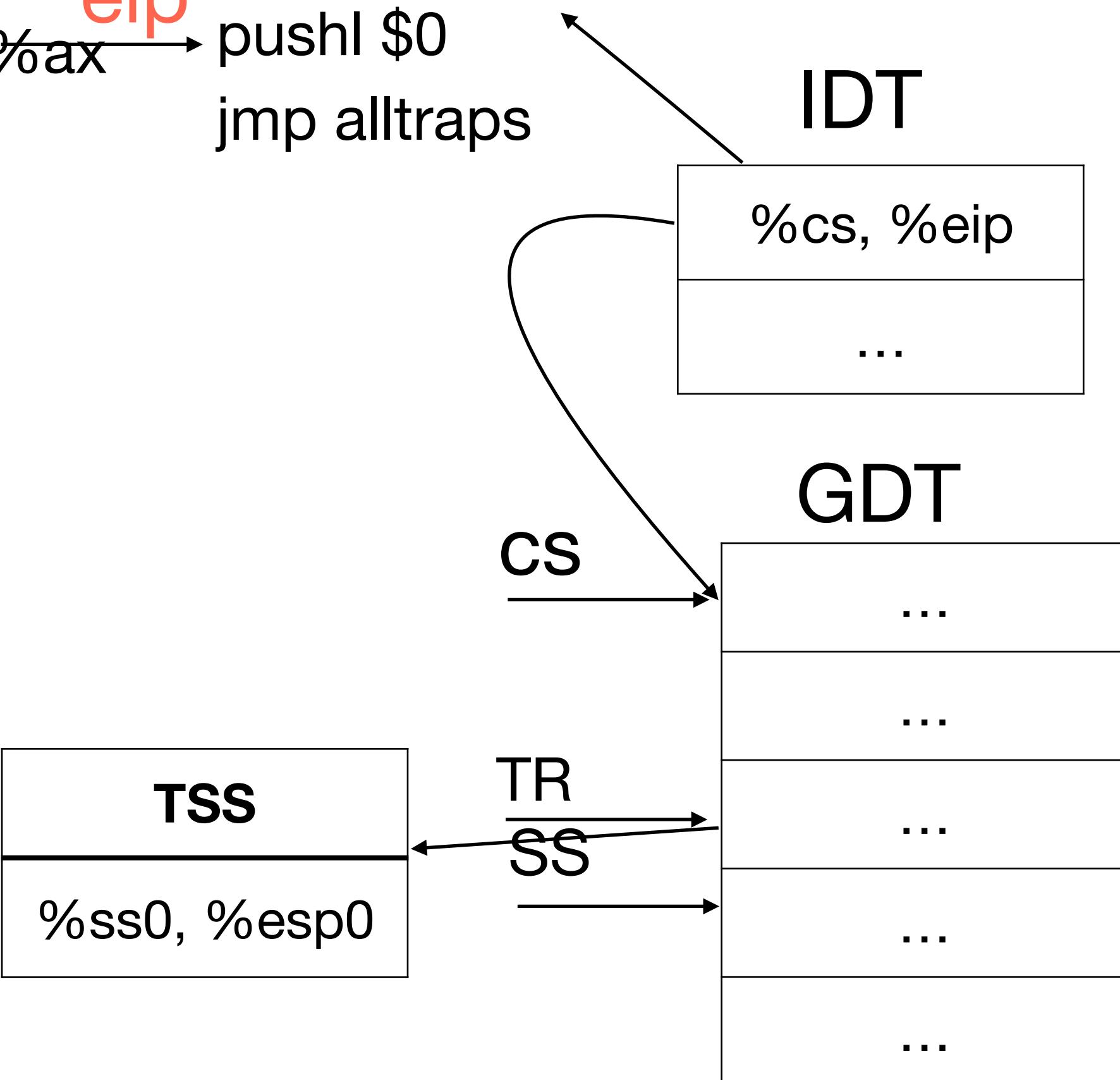
→



Interrupt handling with user process running

```
trapasm.S
for(;;)
;
alltraps:
    pushl %ds..
    pushal
    movw $(SEGMENT) %ax
    movw %ax, %ds
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
```

```
vectors.S  
.globl vector0  
vector0:  
    pushl $0  
    pushl $0  
    jmp alltraps
```



The diagram illustrates a memory stack structure. At the top, there are five yellow rectangular boxes containing the assembly register names: %ss, %esp, %eflags, %cs, and %eip. Below these are two more yellow boxes, both containing the value 0. A horizontal black arrow points from the right towards the bottom of the stack. To the right of this arrow, the text "esp" is written in green. At the very bottom of the stack, there is a blue rectangular box labeled "Process code". The entire stack is contained within a large black rectangular frame.

Interrupt handling with user process running

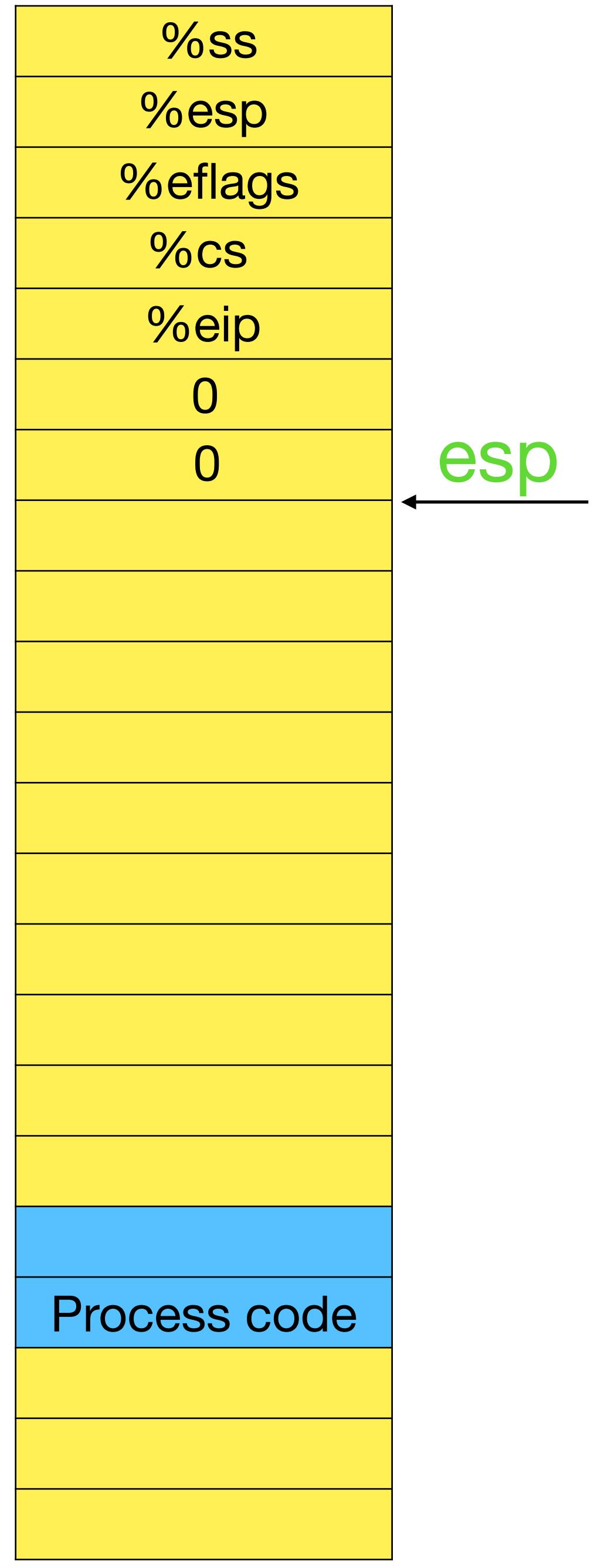
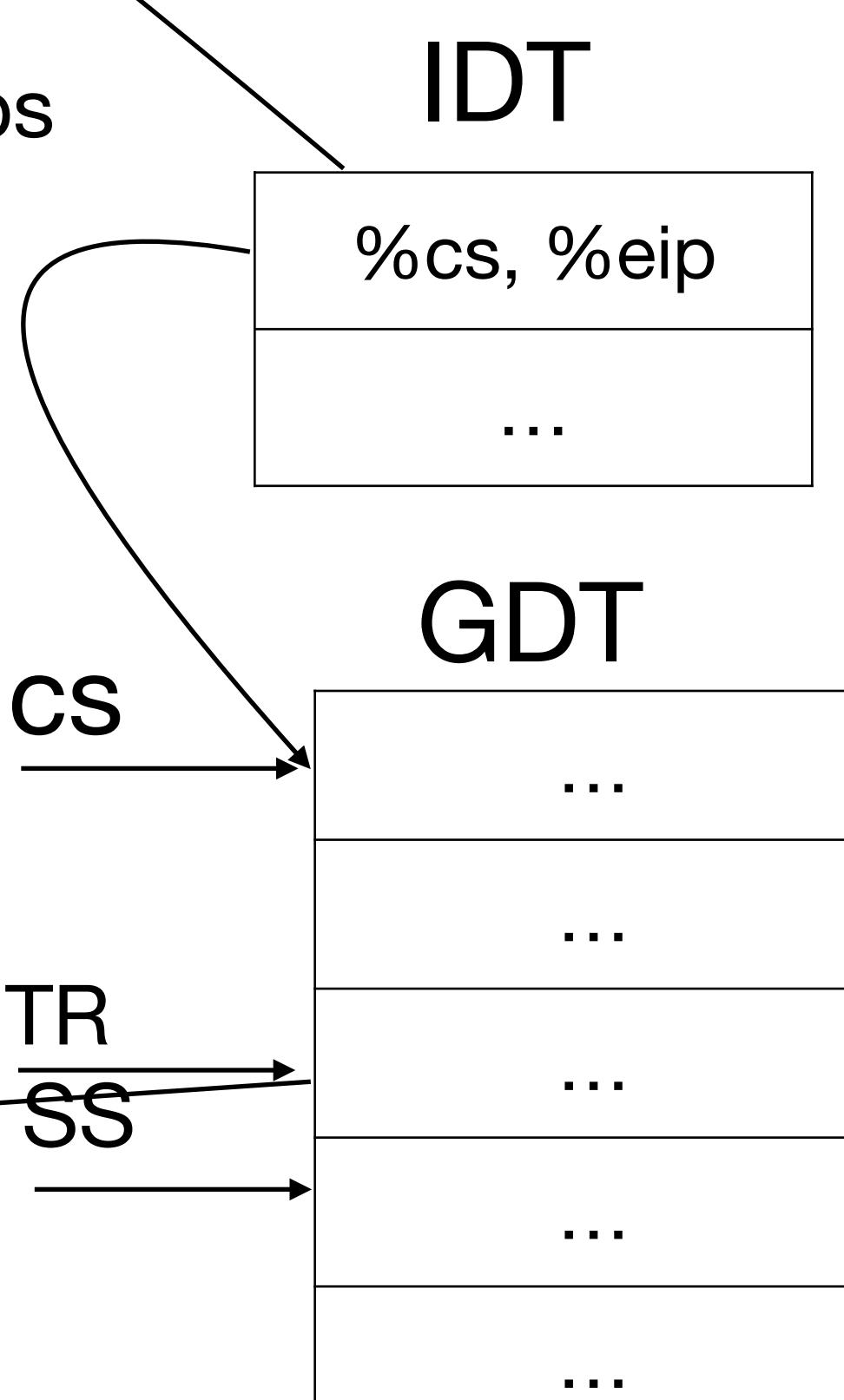
```
trapasm.S
for(;;)
;
    alltraps:
        pushl %ds..
        pushal
        movw $(SEG_KDATA<<3), %ax
        movw %ax, %ds..
        pushl %esp
        call trap
        addl $4, %esp
        popal
        popl %ds..
        addl $0x8, %esp
        iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

eip

→



Interrupt handling with user process running

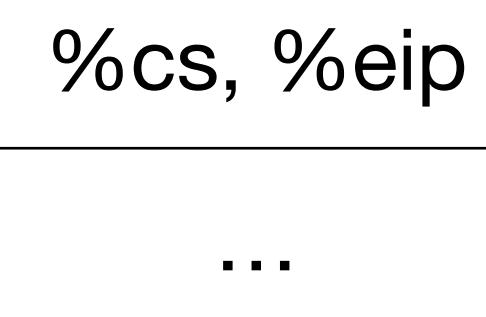
```
trapasm.S
for(;;)
;
    alltraps:
        pushl %ds..
        pushal
        movw $(SEG_KDATA<<3), %ax
        movw %ax, %ds..
        pushl %esp
        call trap
        addl $4, %esp
        popal
        popl %ds..
        addl $0x8, %esp
        iret
```

vectors.S

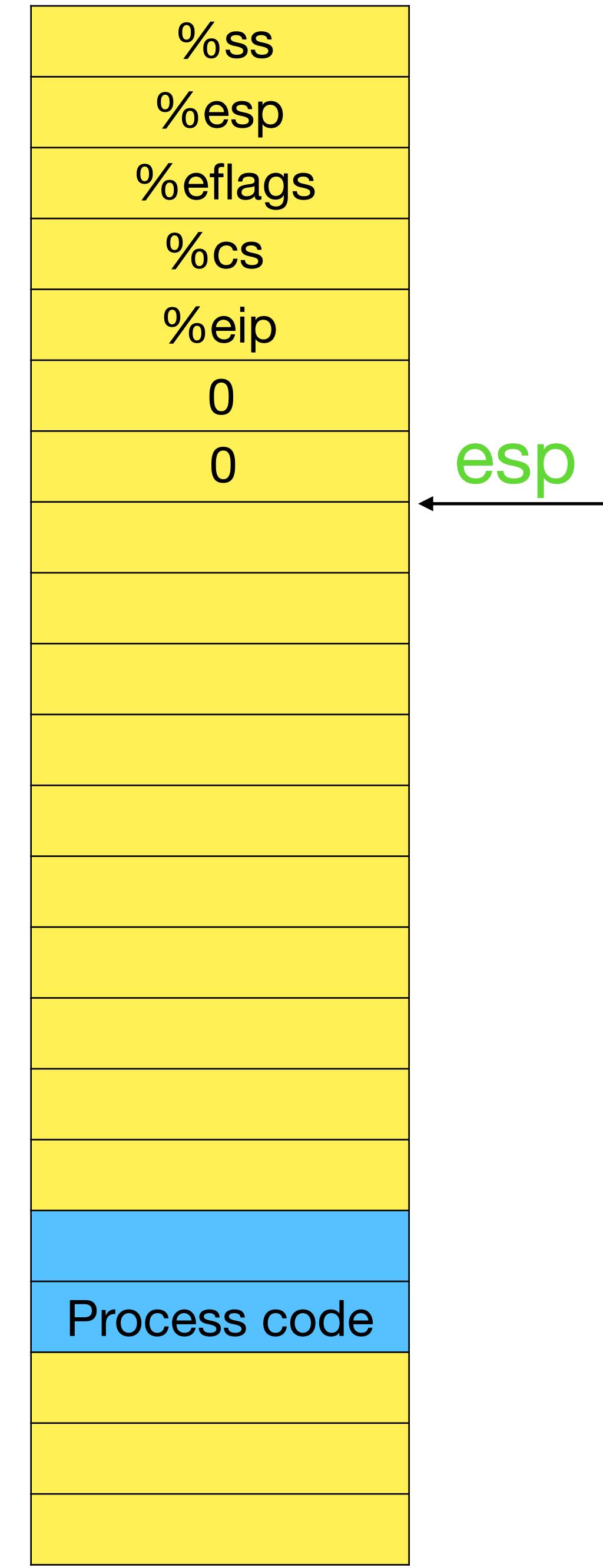
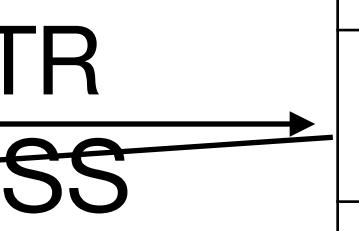
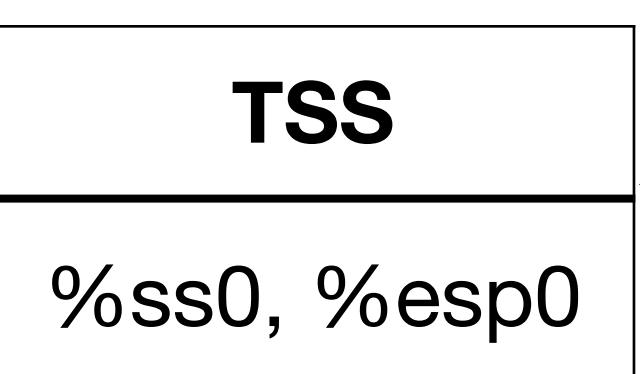
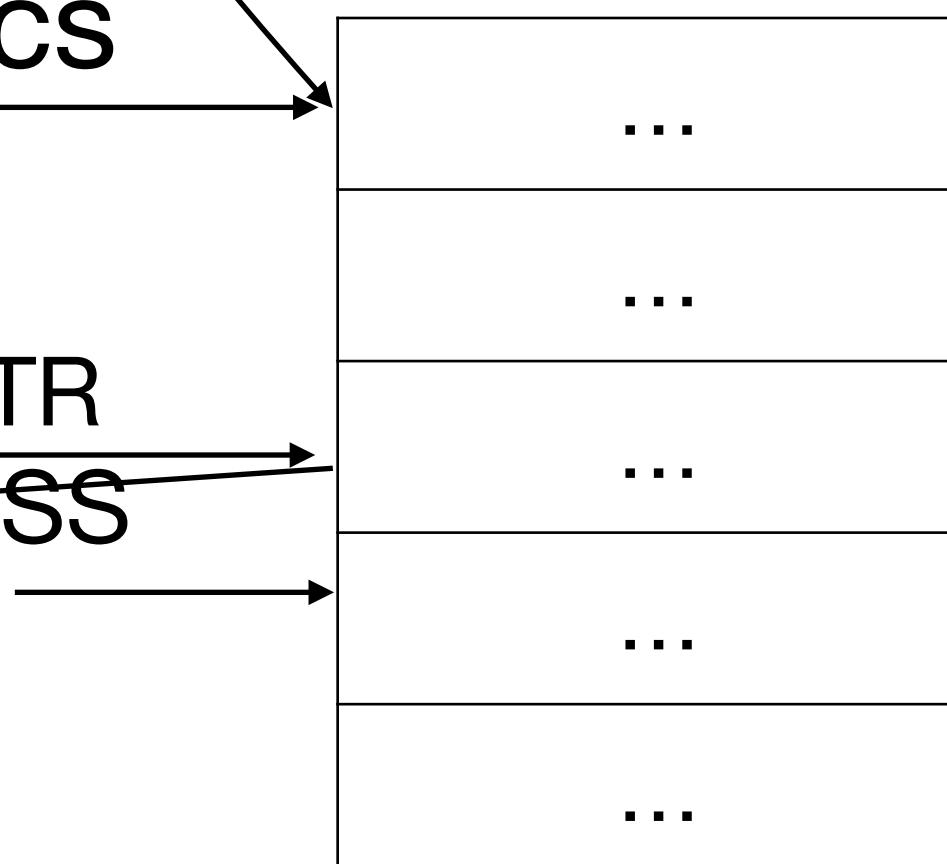
```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

eip

IDT



GDT

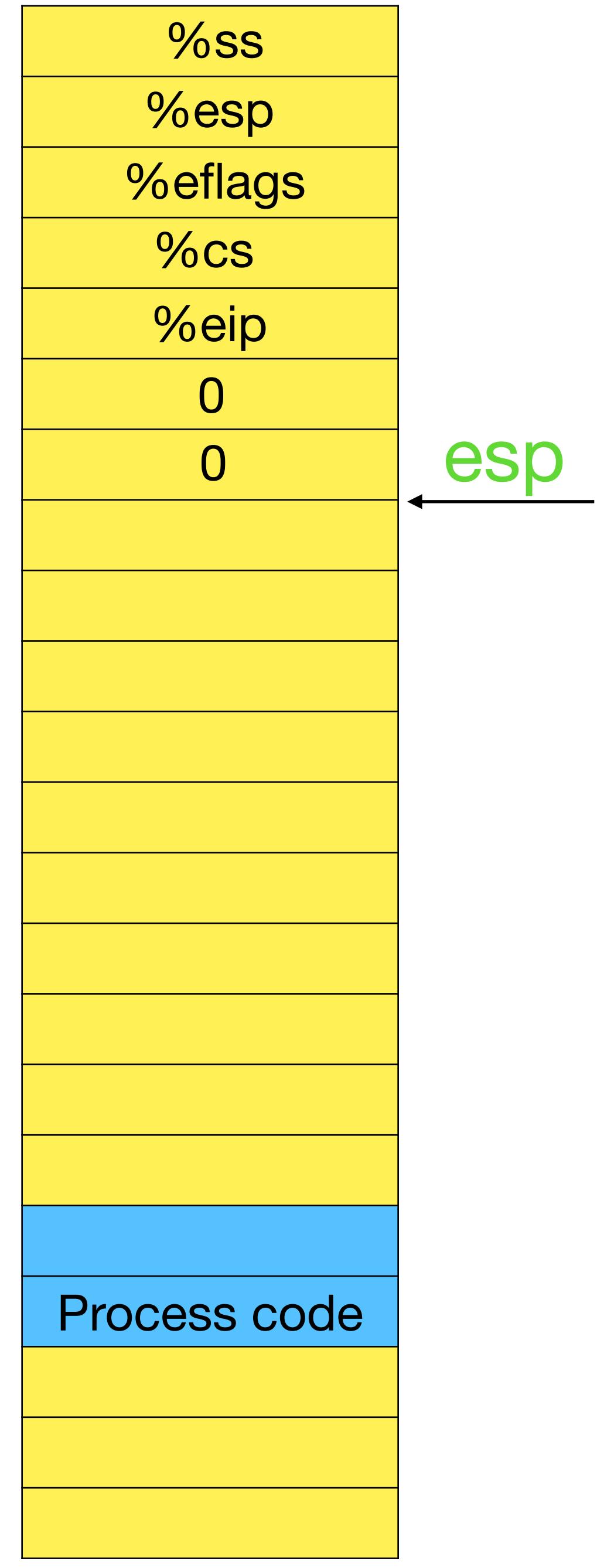
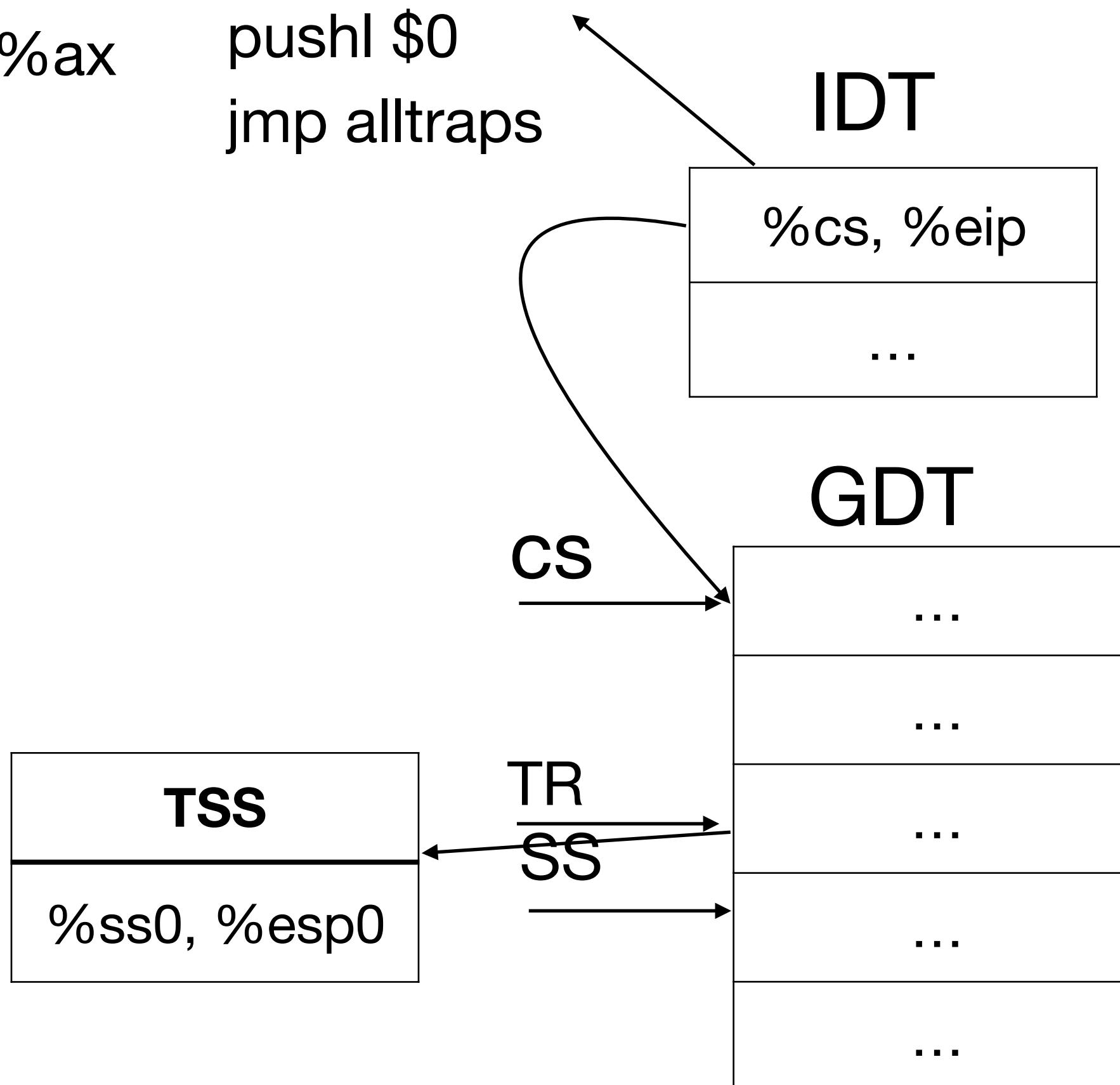


Interrupt handling with user process running

```
trapasm.S
for(;;)
;   eip    alltraps:
        pushl %ds..
        pushal
        movw $(SEG_KDATA<<3), %ax
        movw %ax, %ds..
        pushl %esp
        call trap
        addl $4, %esp
        popal
        popl %ds..
        addl $0x8, %esp
        iret
```

vectors.S

```
.globl vector0
vector0:
        pushl $0
        pushl $0
        jmp alltraps
```

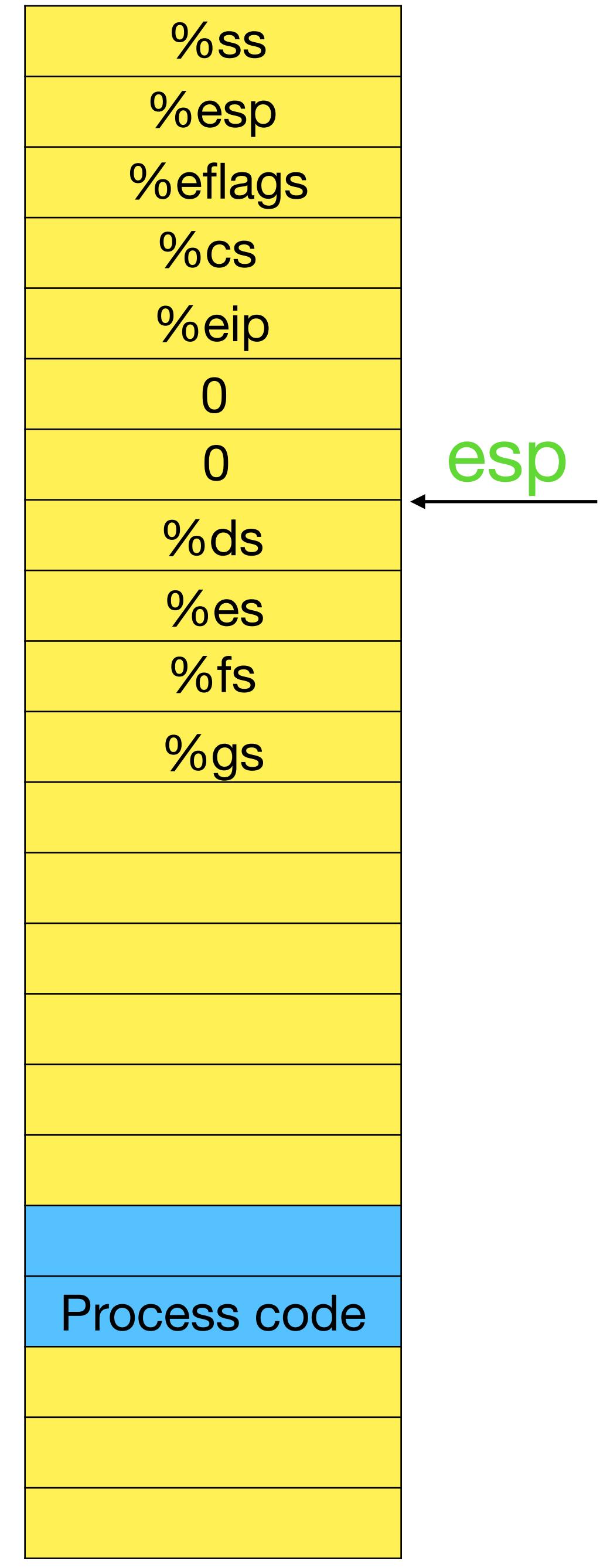
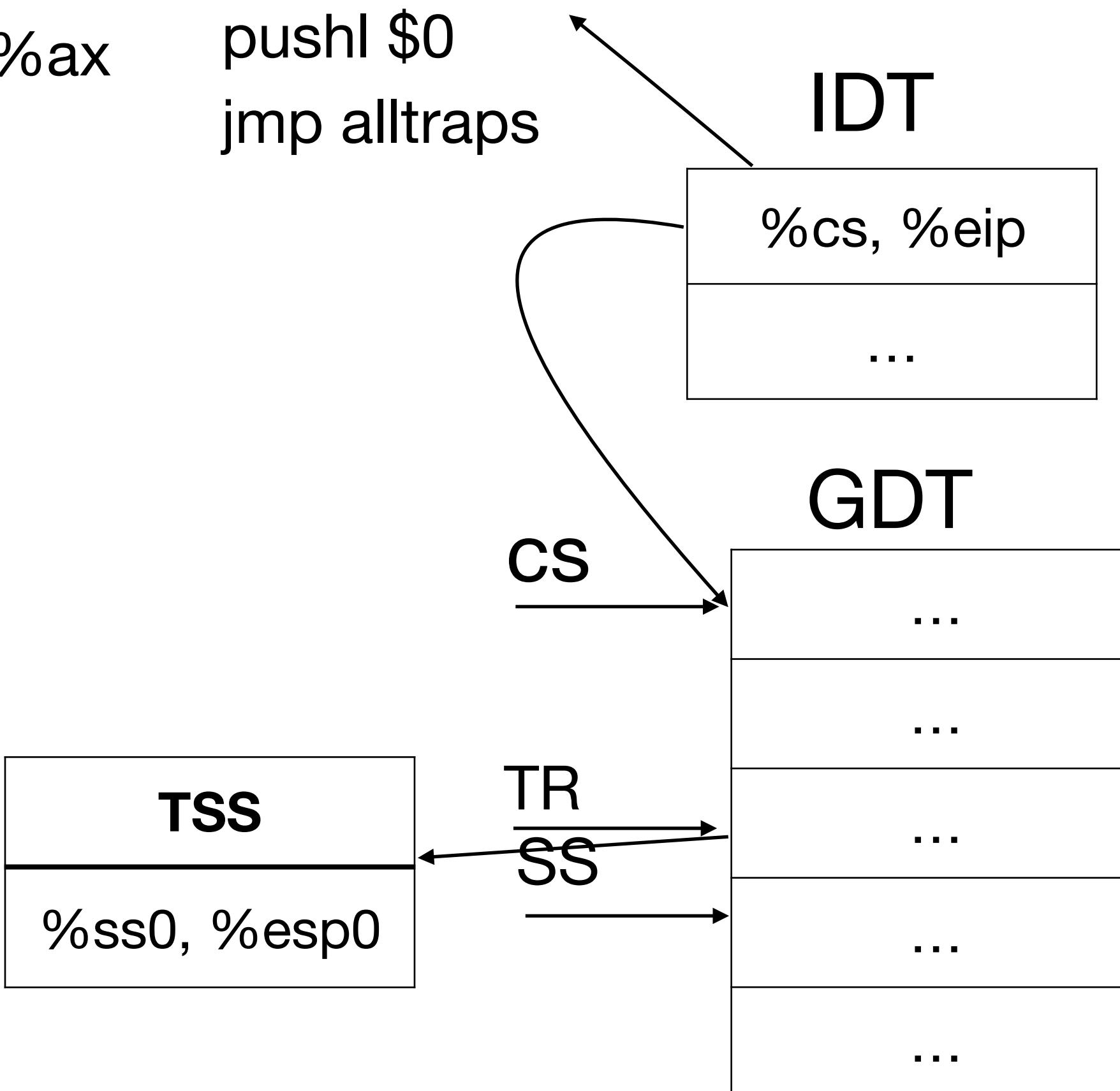


Interrupt handling with user process running

```
trapasm.S
for(;;)
;   eip    alltraps:
        pushl %ds..
        pushal
        movw $(SEG_KDATA<<3), %ax
        movw %ax, %ds..
        pushl %esp
        call trap
        addl $4, %esp
        popal
        popl %ds..
        addl $0x8, %esp
        iret
```

vectors.S

```
.globl vector0
vector0:
        pushl $0
        pushl $0
        jmp alltraps
```

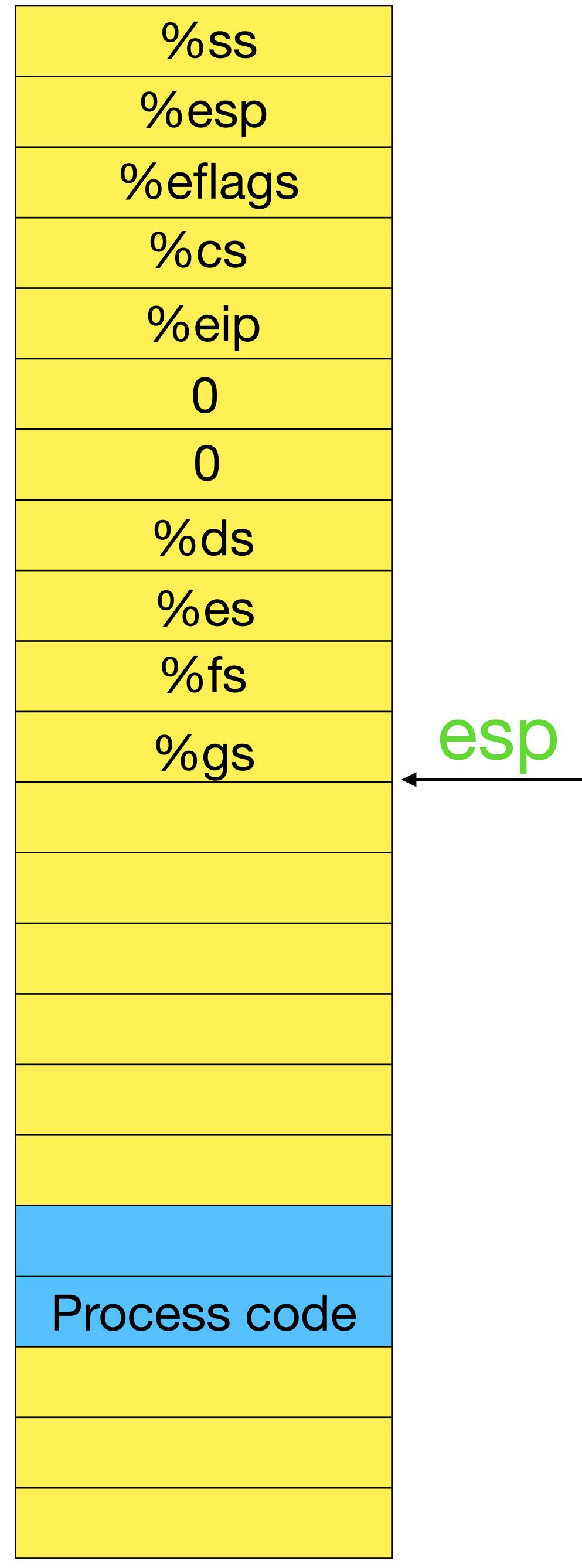
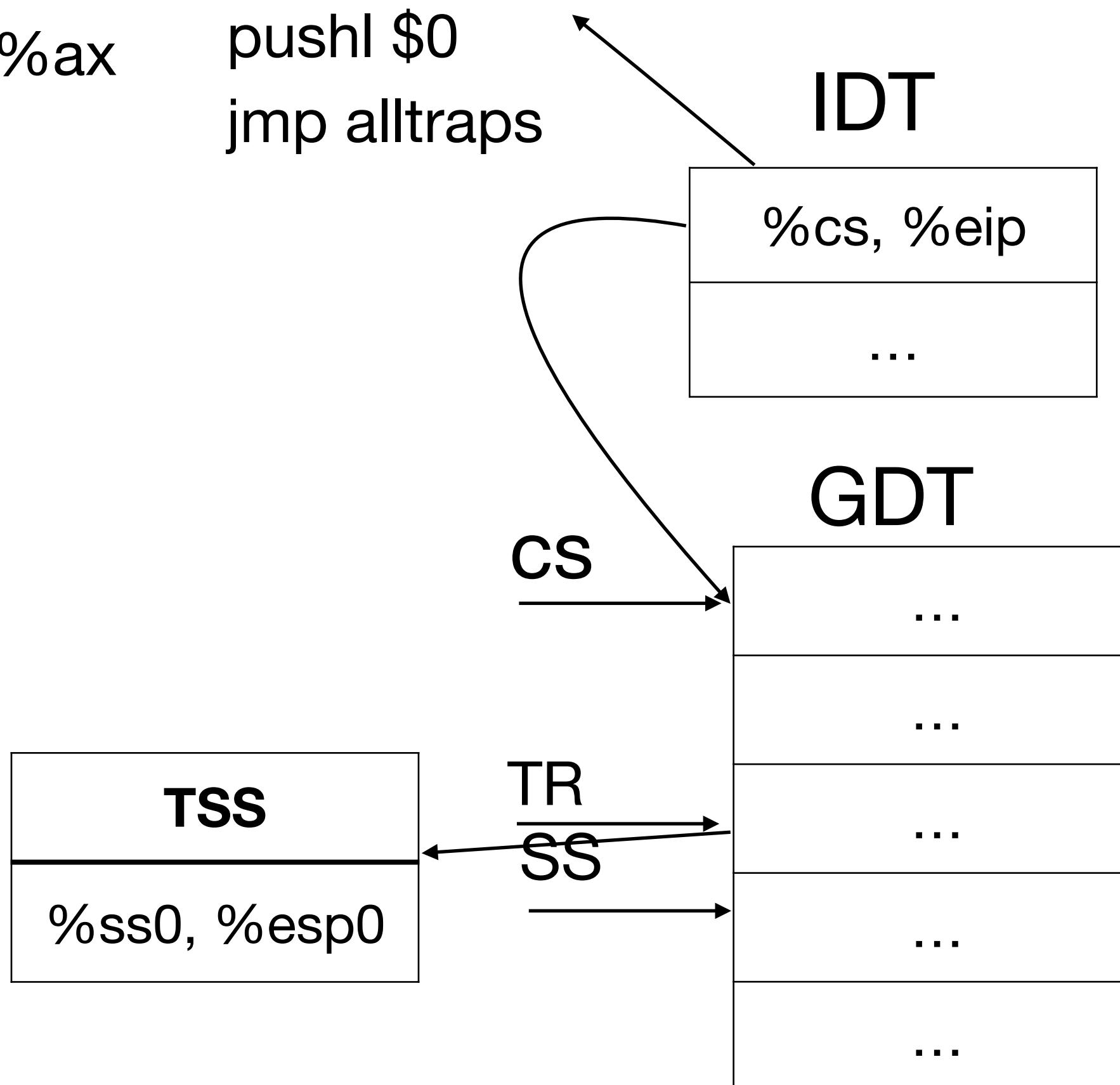


Interrupt handling with user process running

```
trapasm.S
for(;;)    eip    alltraps:
;   → pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
```

vectors.S

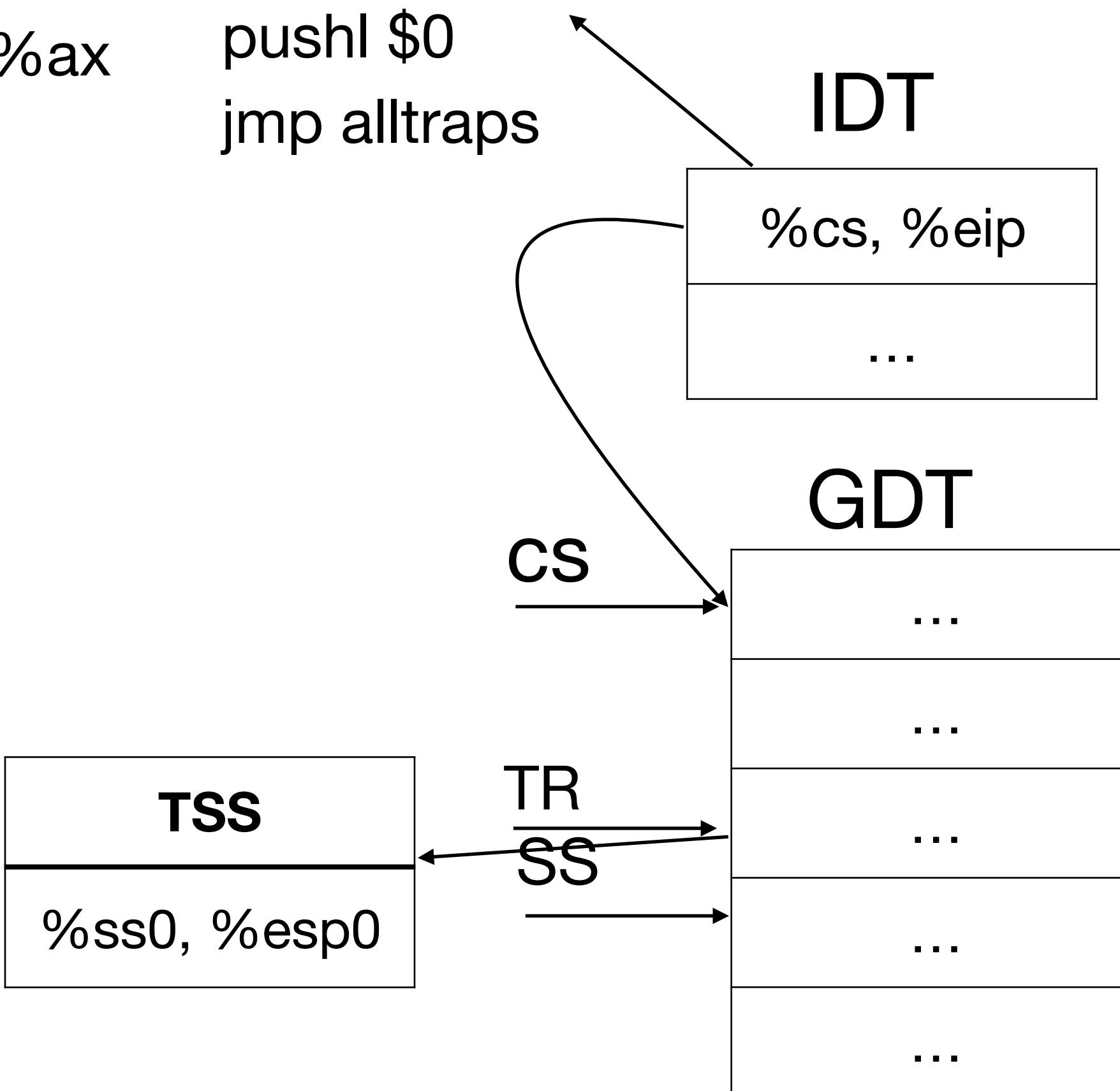
```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```



Interrupt handling with user process running

```
trapasm.S  
for(;;)    alltraps:  
;  
          pushl %ds..  
          eip → pushal  
          movw $(SEG_KDATA<<3), %ax  
          movw %ax, %ds..  
          pushl %esp  
          call trap  
          addl $4, %esp  
          popal  
          popl %ds..  
          addl $0x8, %esp  
          iret
```

```
vectors.S
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

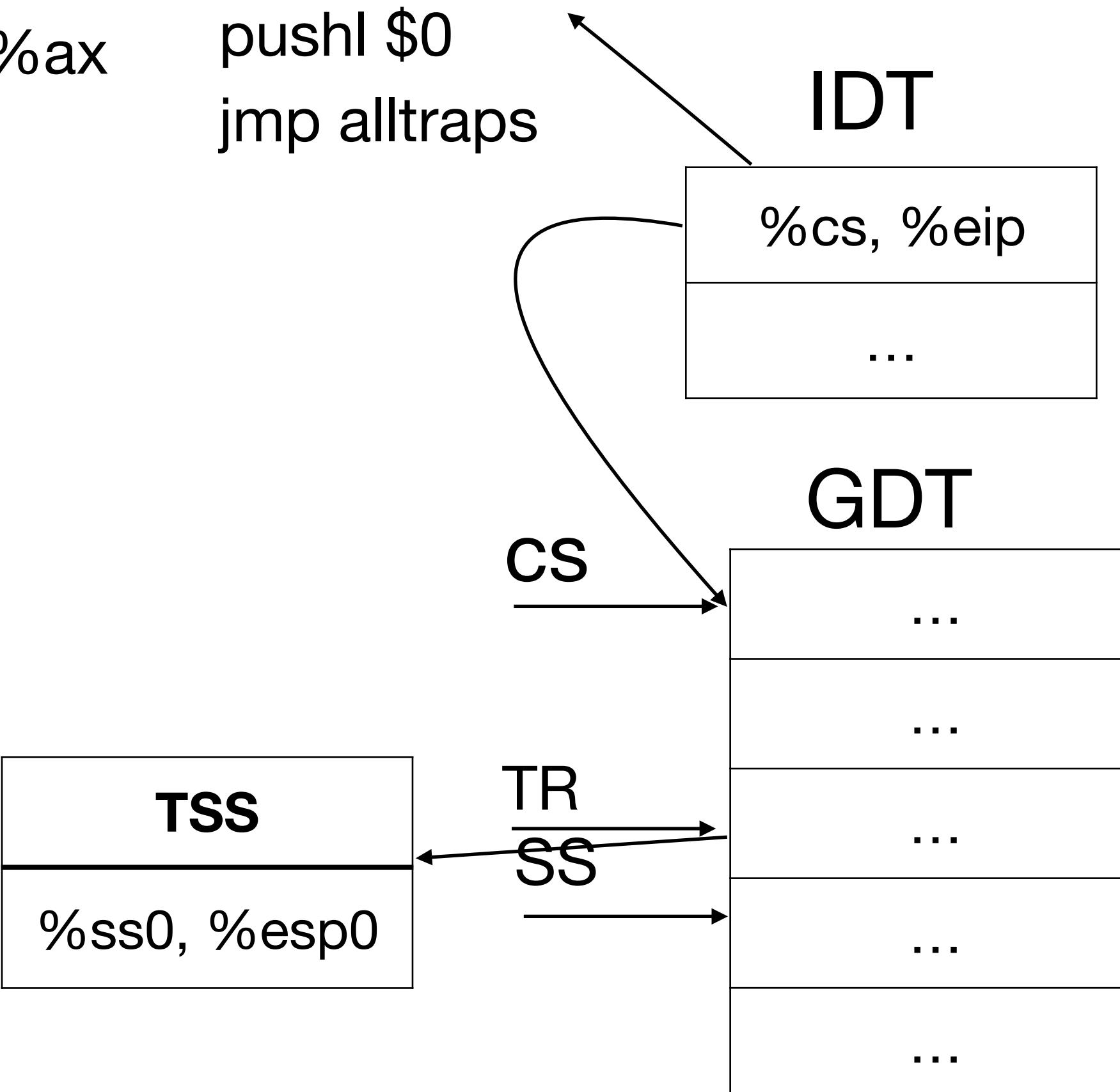


Interrupt handling with user process running

```
trapasm.S
for(;;)    alltraps:
;           eip    pushl %ds..
;           pushal
;           movw $(SEG_KDATA<<3), %ax
;           movw %ax, %ds..
;           pushl %esp
;           call trap
;           addl $4, %esp
;           popal
;           popl %ds..
;           addl $0x8, %esp
;           iret
```

vectors.S

```
.globl vector0
vector0:
pushl $0
pushl $0
jmp alltraps
```



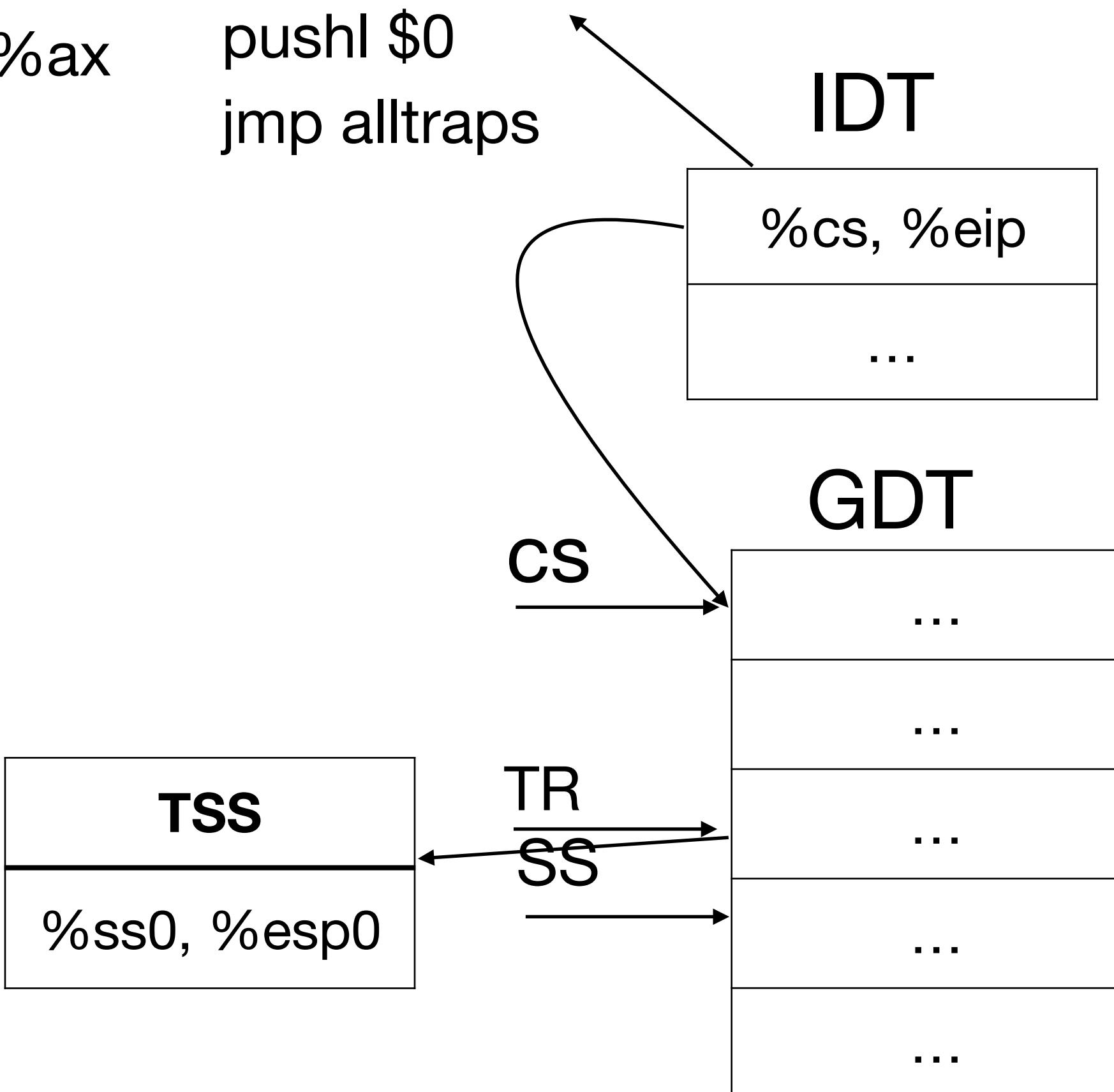
%ss
%esp
%eflags
%cs
%eip
0
0
%ds
%es
%fs
%gs
%eax
%ecx
...
%edi
...
Process code
...
...
...
...

Interrupt handling with user process running

```
trapasm.S
for(;;)    alltraps:
;           eip    pushl %ds..
;           pushal
;           movw $(SEG_KDATA<<3), %ax
;           movw %ax, %ds..
;           pushl %esp
;           call trap
;           addl $4, %esp
;           popal
;           popl %ds..
;           addl $0x8, %esp
;           iret
```

vectors.S

```
.globl vector0
vector0:
pushl $0
pushl $0
jmp alltraps
```



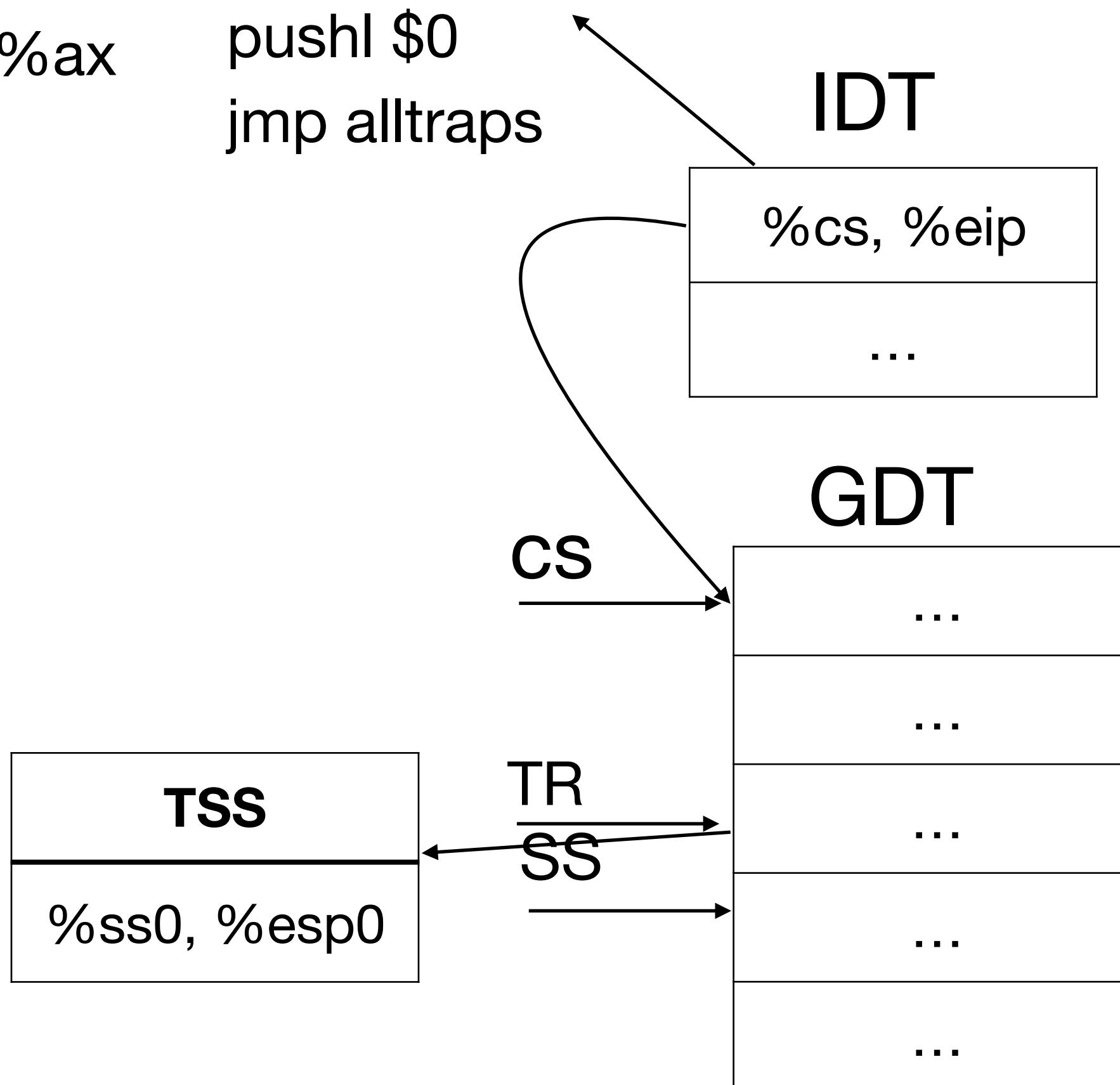
%ss
%esp
%eflags
%cs
%eip
0
0
%ds
%es
%fs
%gs
%eax
%ecx
...
%edi
...
esp

Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```



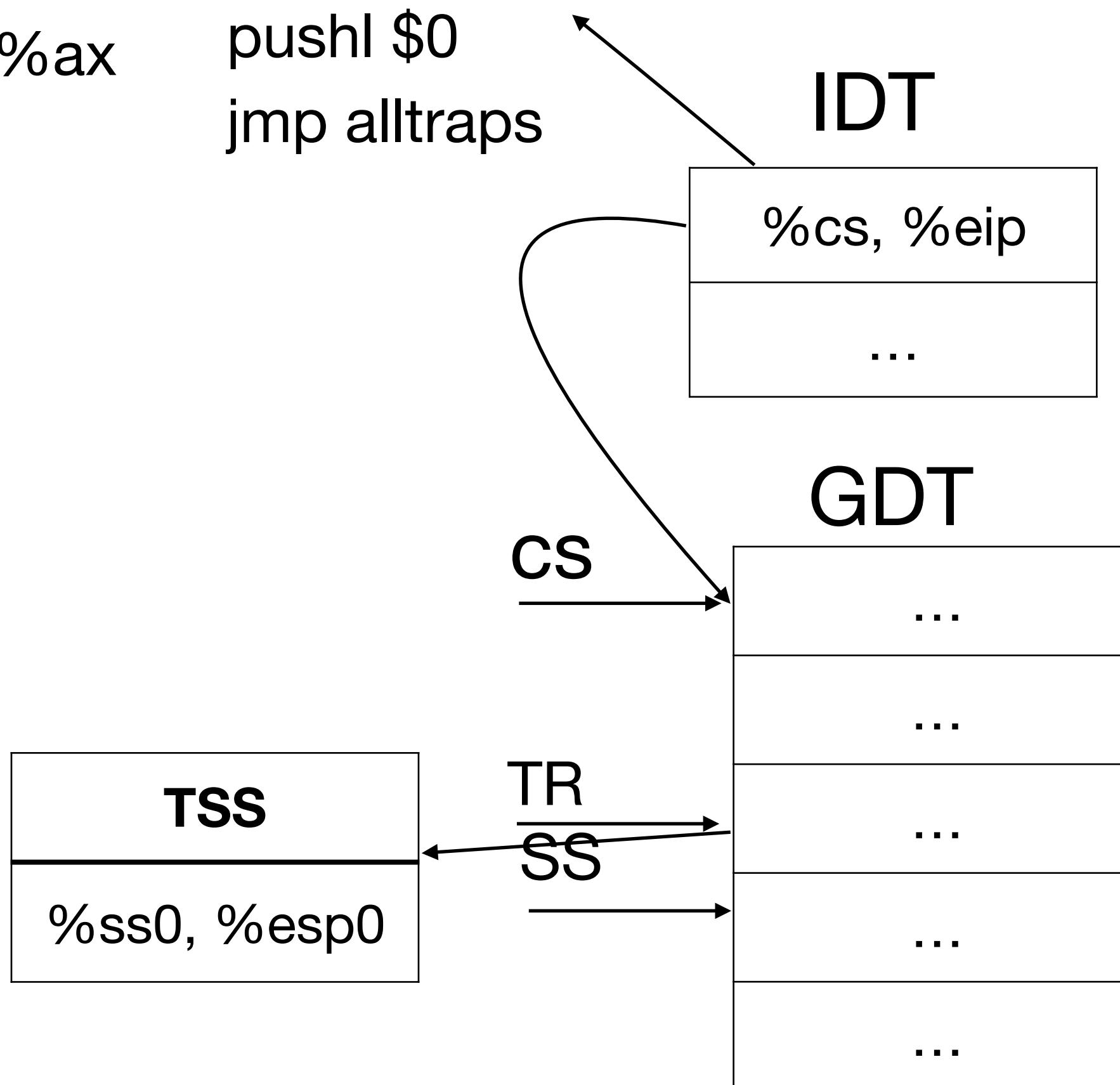
%ss
%esp
%eflags
%cs
%eip
0
0
%ds
%es
%fs
%gs
%eax
%ecx
...
%edi
Process code
...
...
...
...
...

Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```



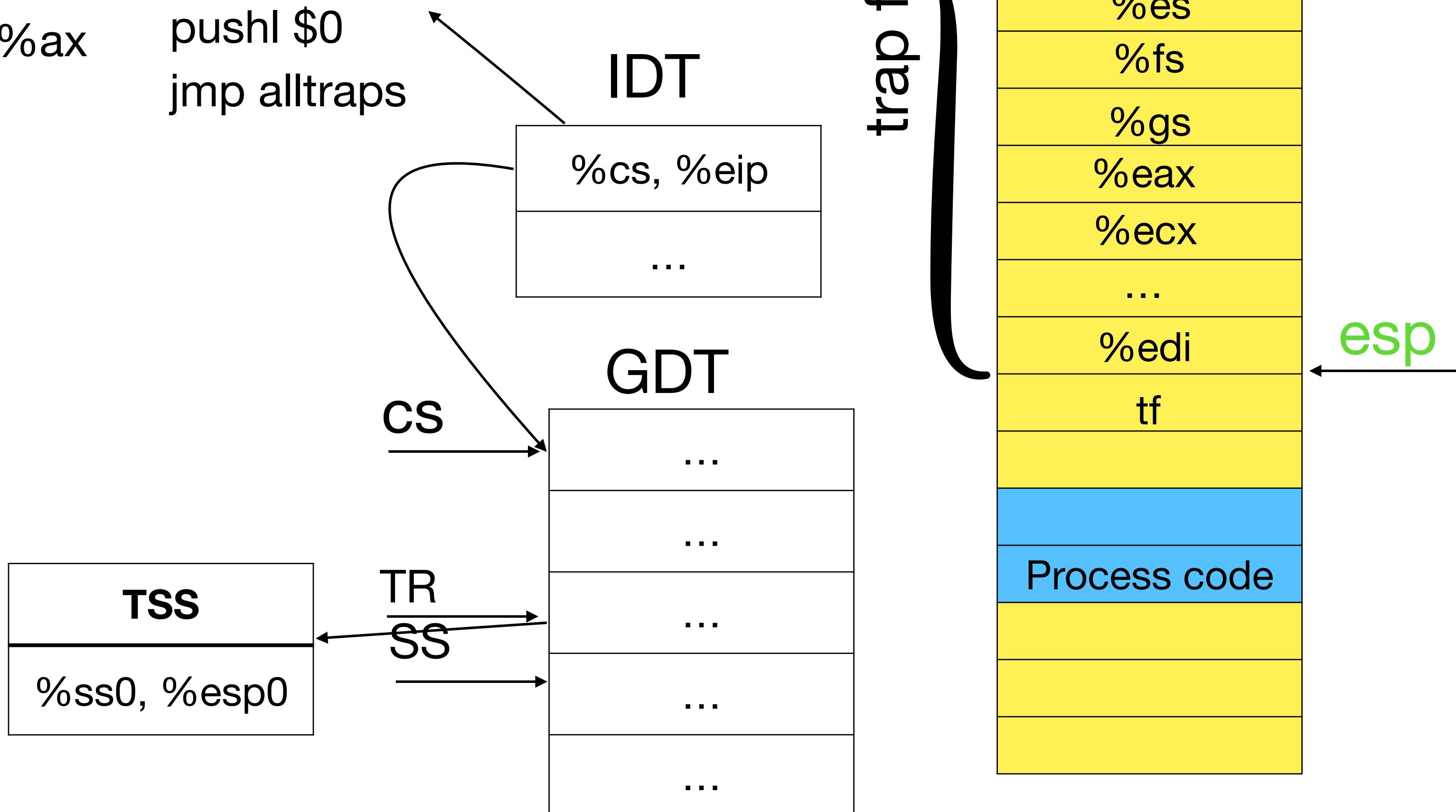
%ss
%esp
%eflags
%cs
%eip
0
0
%ds
%es
%fs
%gs
%eax
%ecx
...
%edi
tf
Process code
...
...
...
...
...

Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

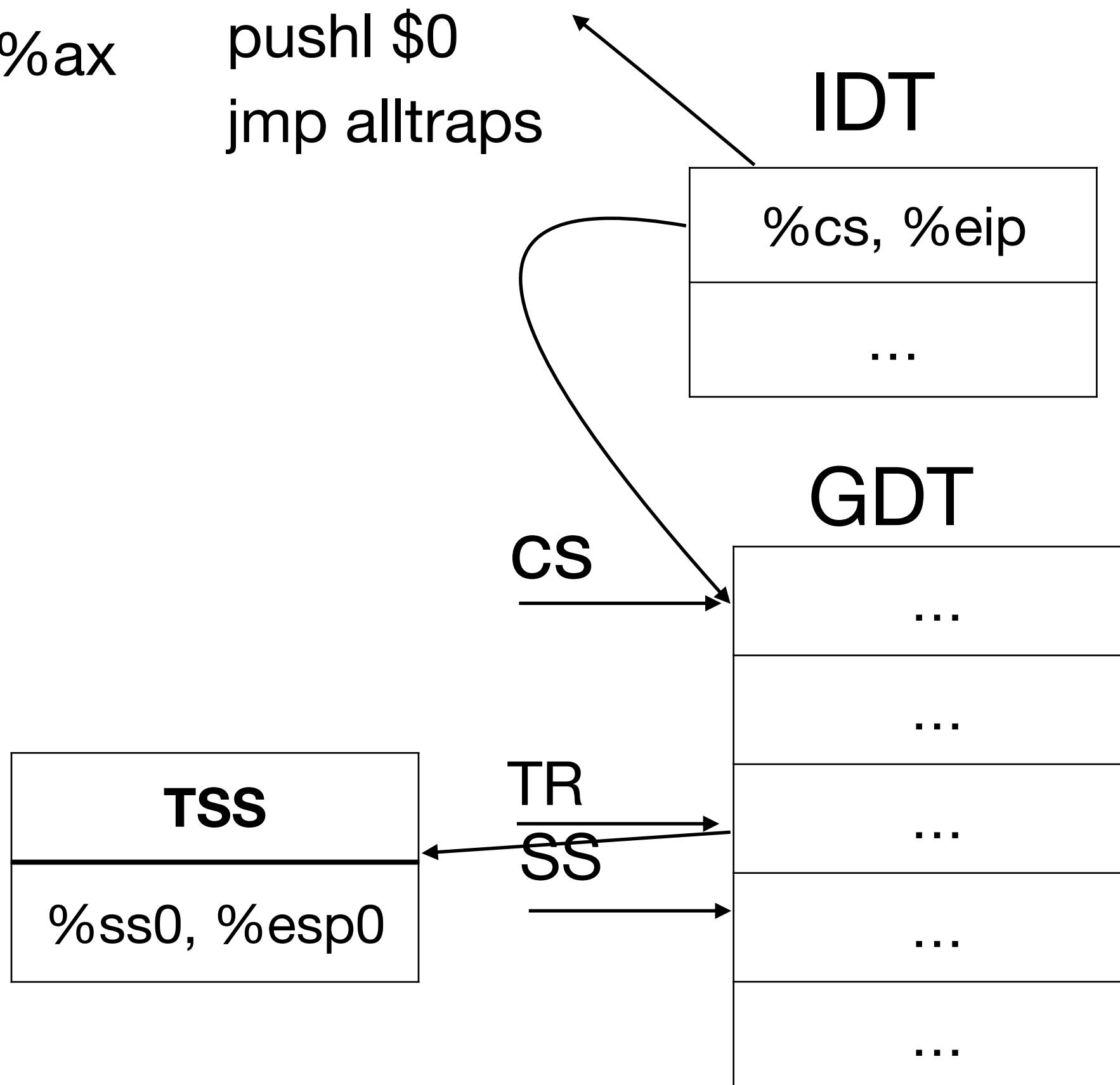


Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```



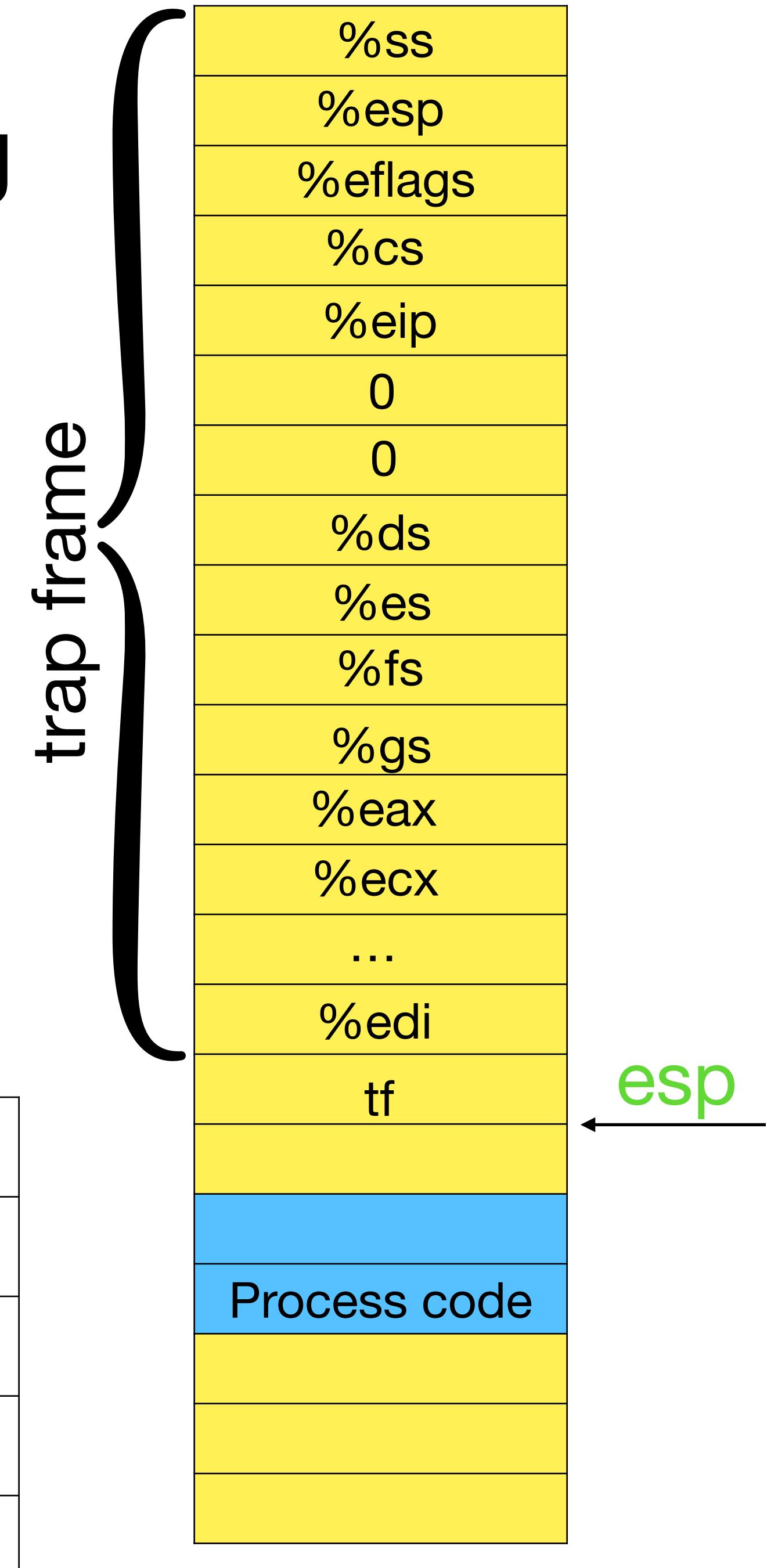
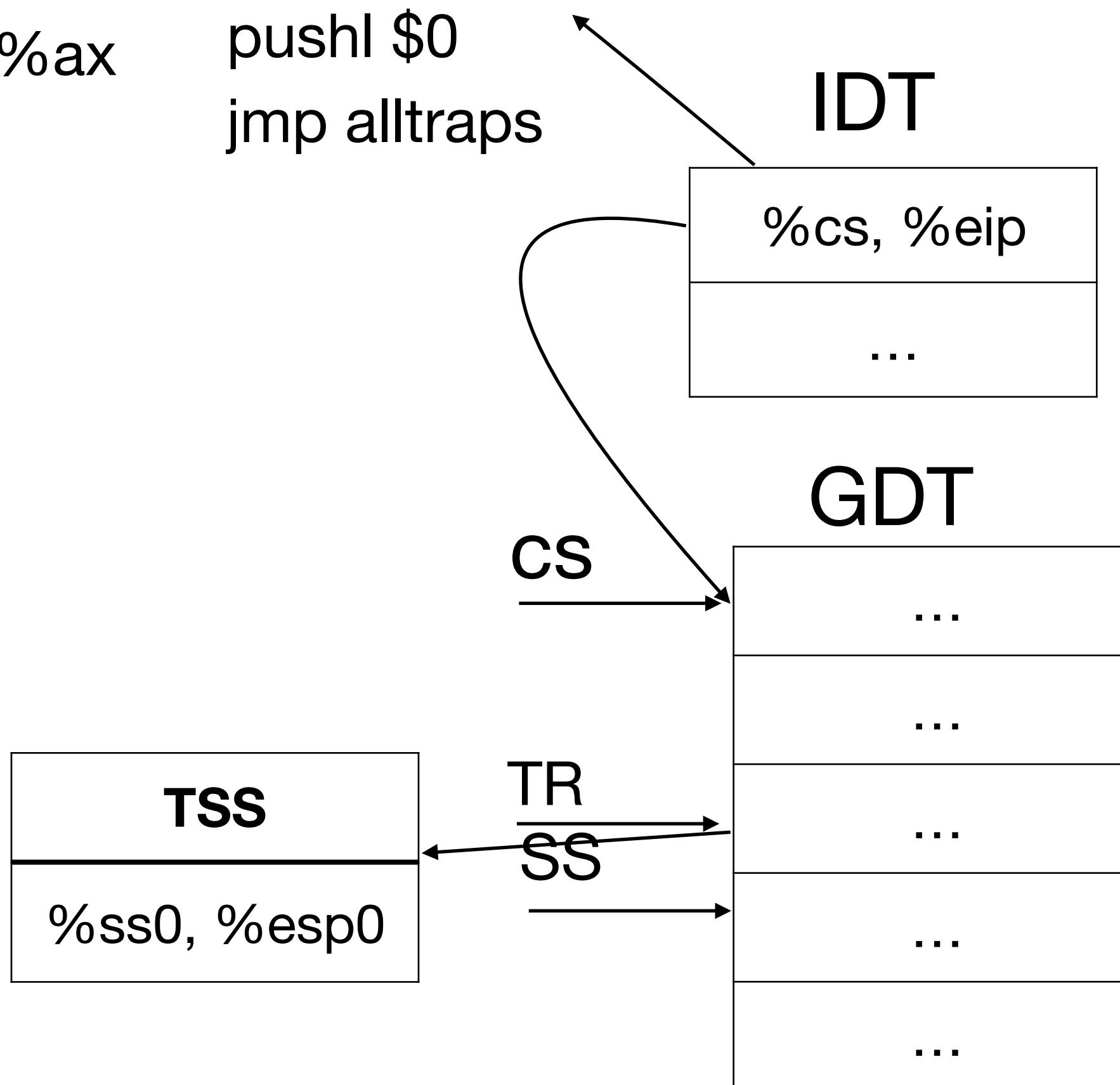
%ss
%esp
%eflags
%cs
%eip
0
0
%ds
%es
%fs
%gs
%eax
%ecx
...
%edi
tf
Process code
...
...
...
...
...

Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

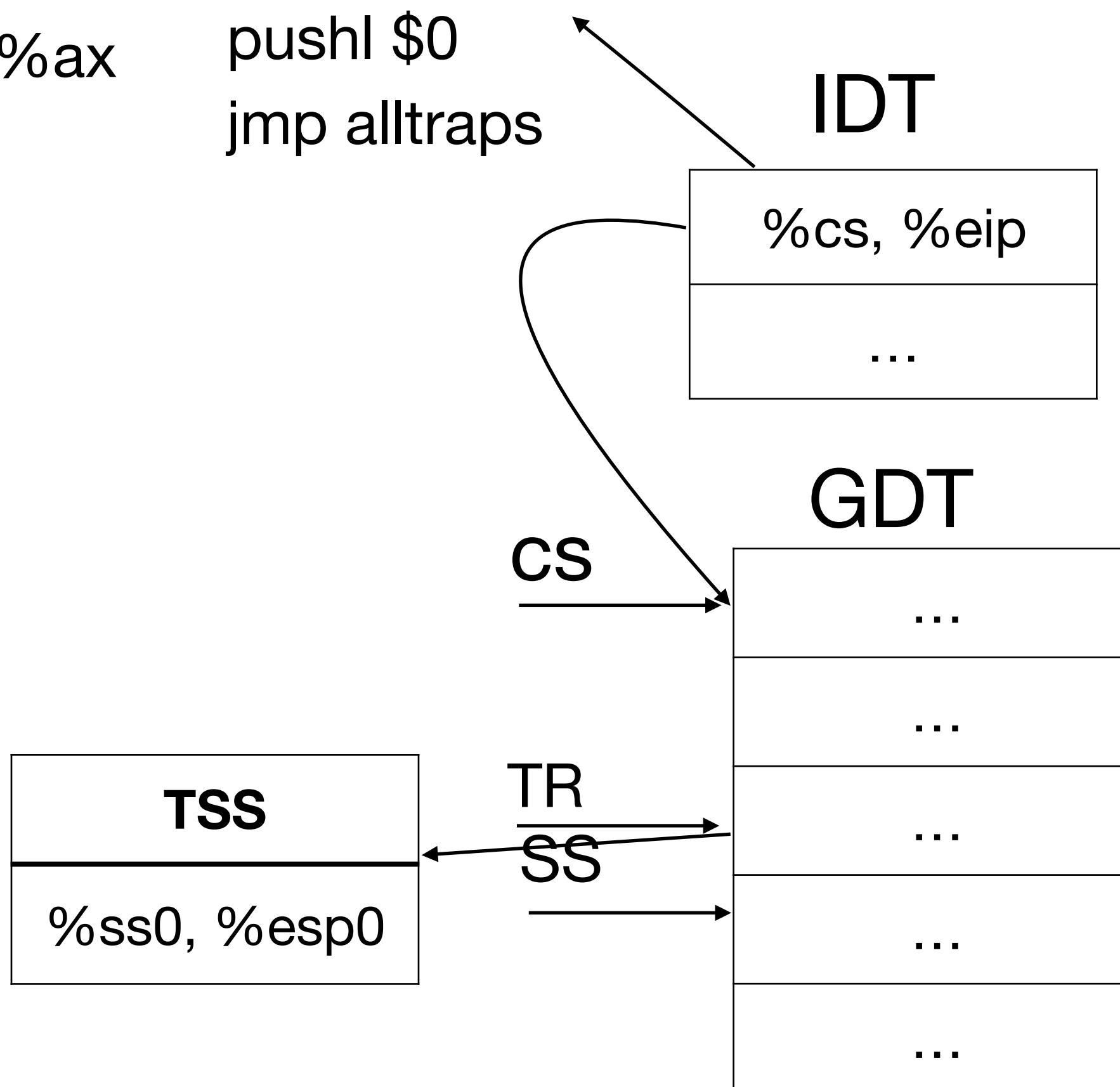


Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

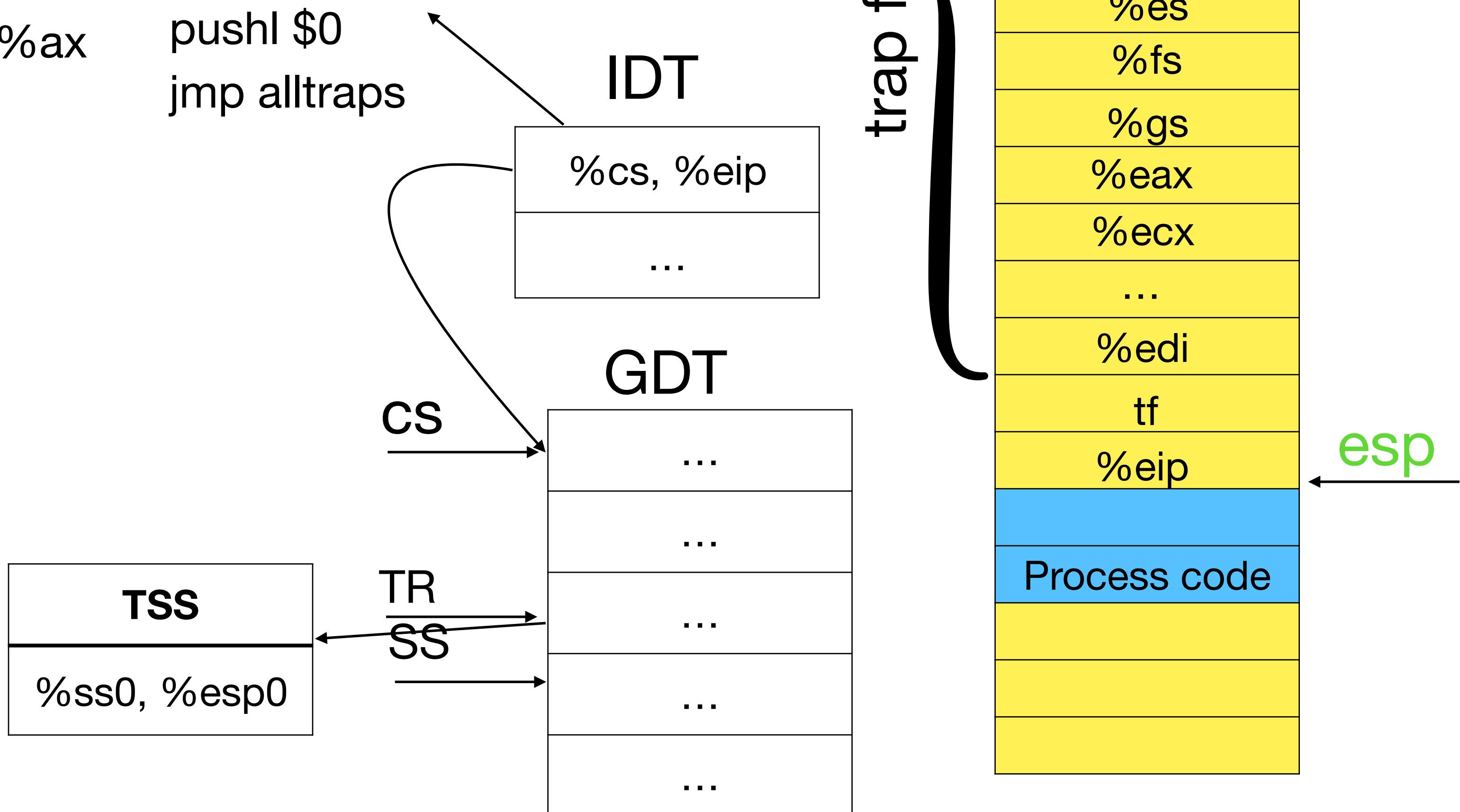


Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

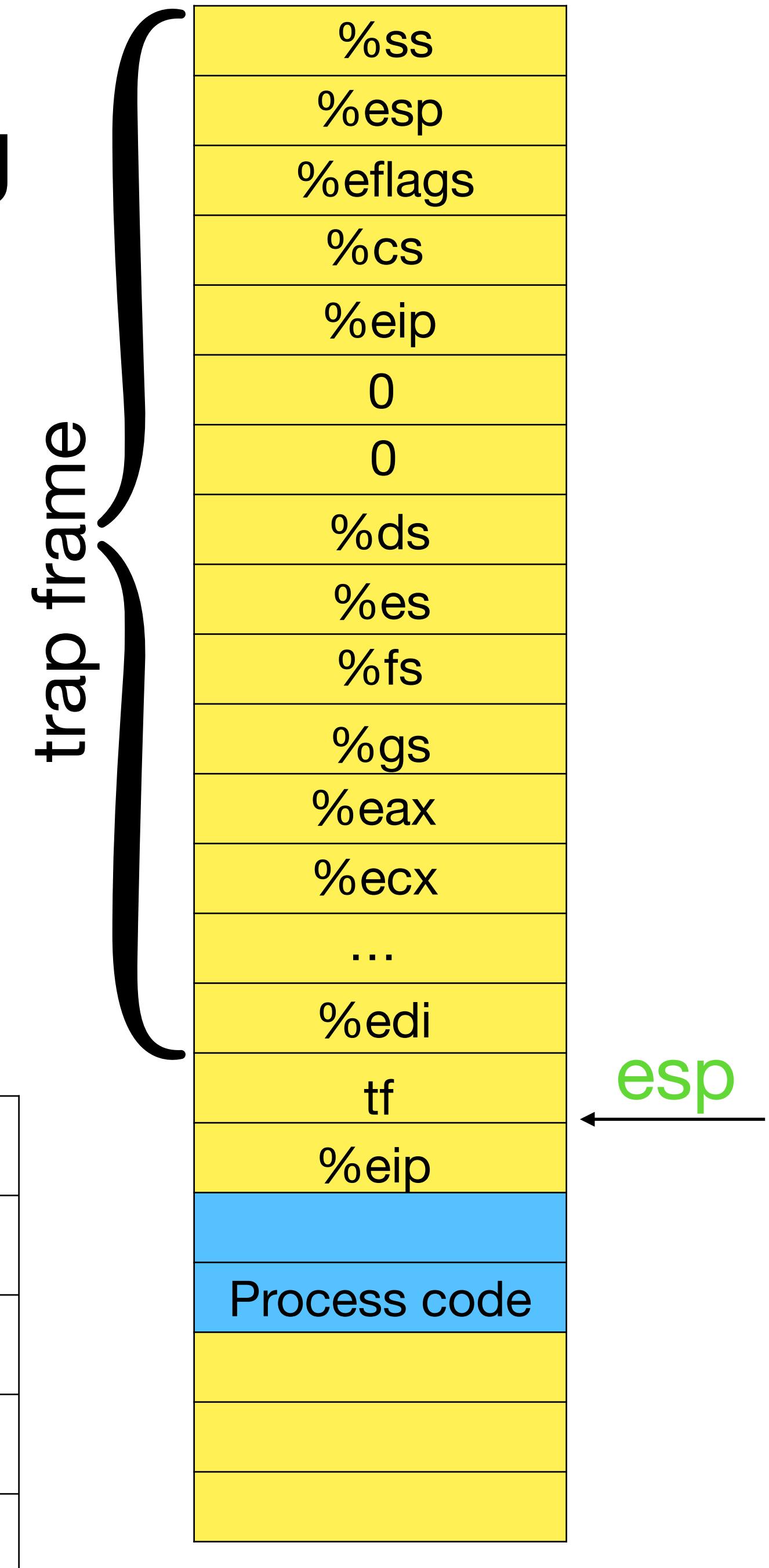
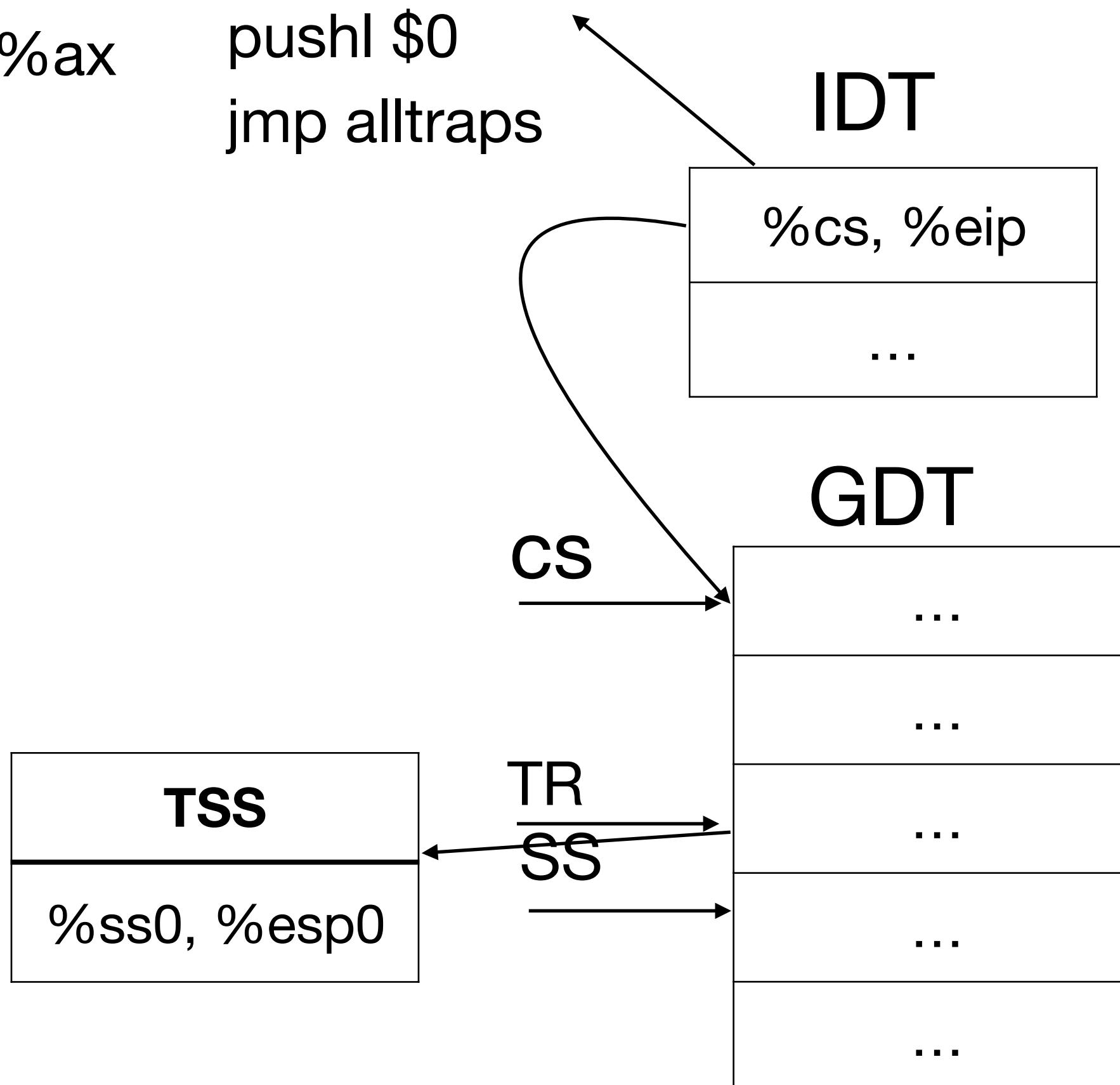


Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

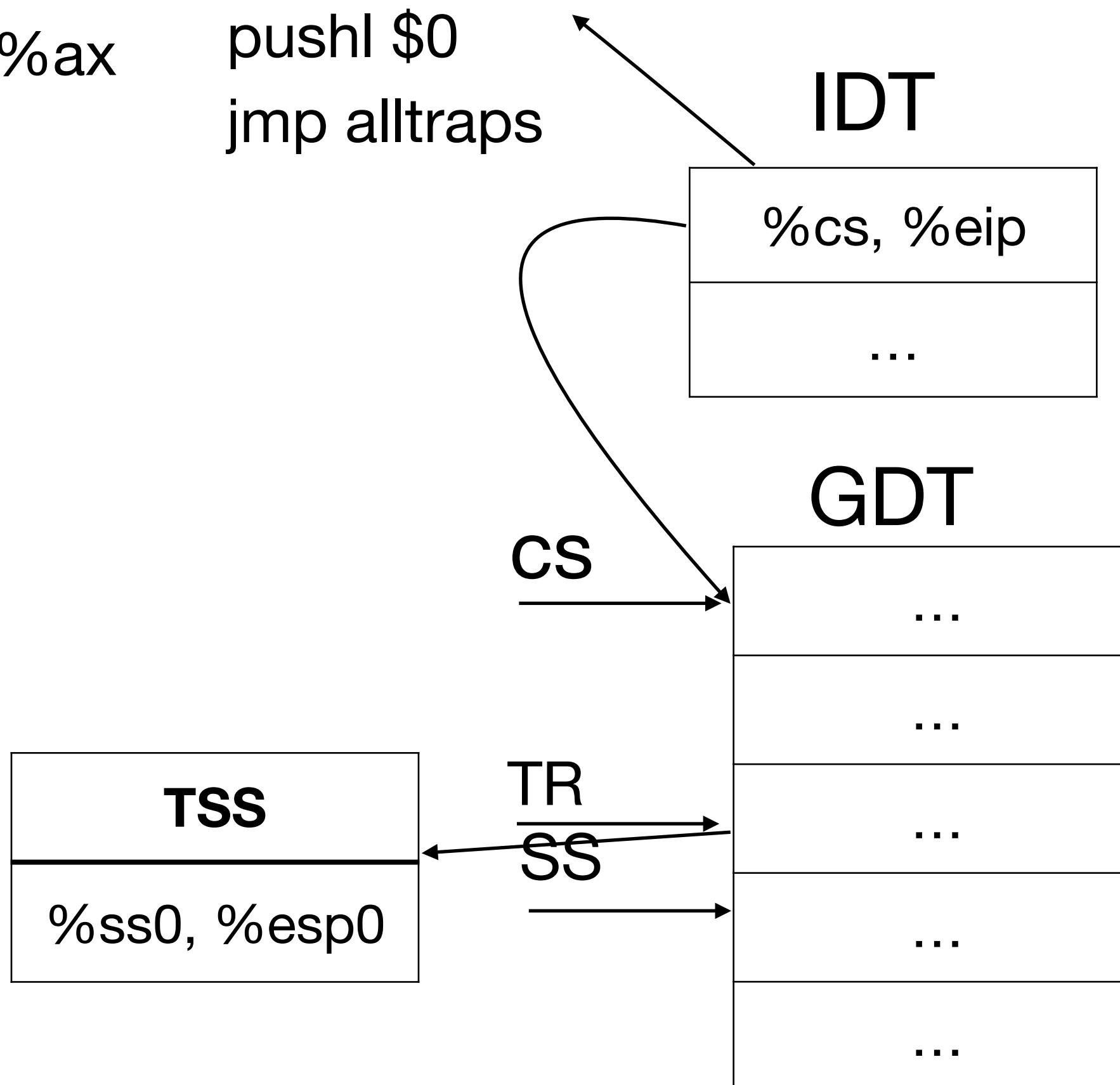


Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```



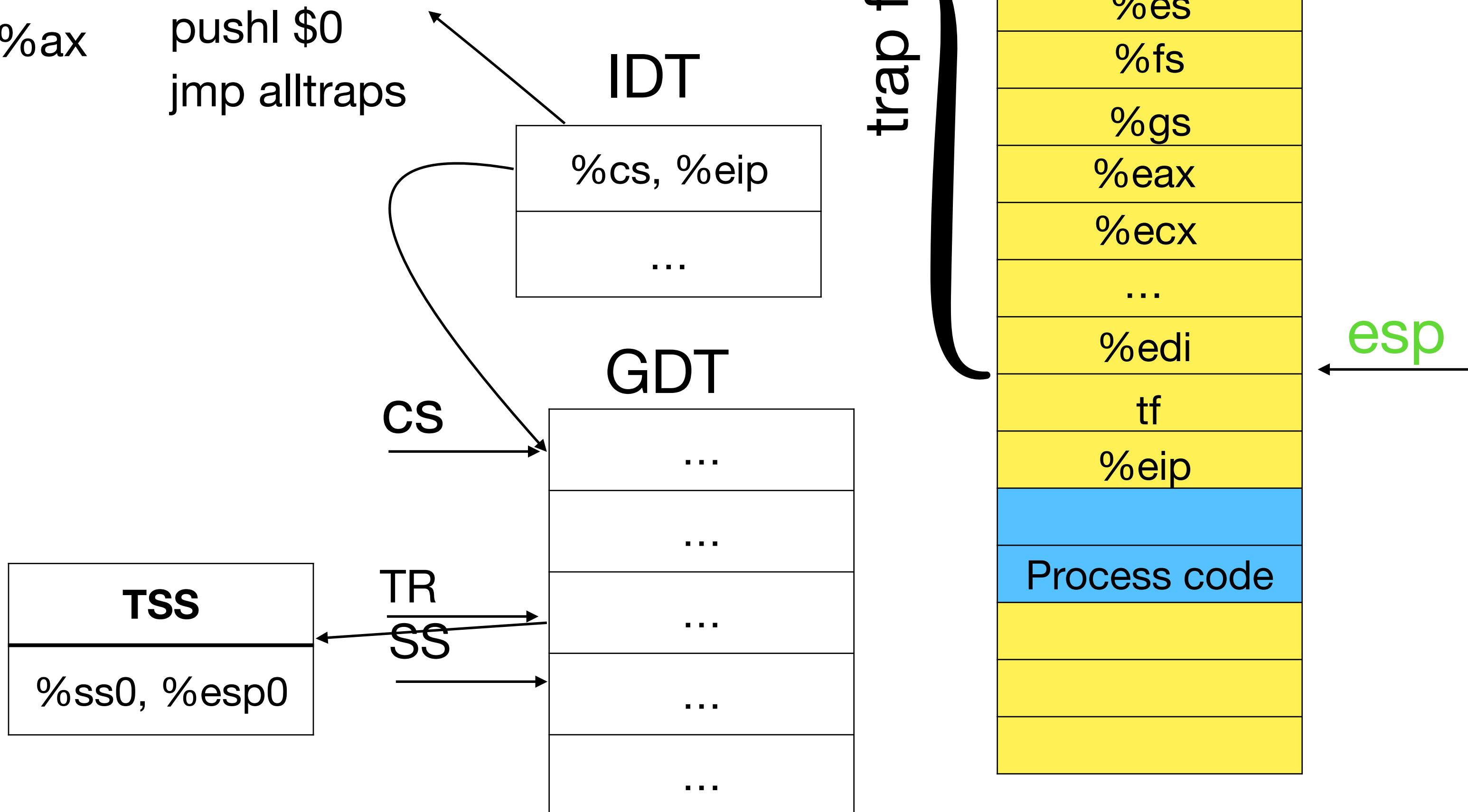
%ss
%esp
%eflags
%cs
%eip
0
0
%ds
%es
%fs
%gs
%eax
%ecx
...
%edi
tf
%eip
Process code

Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```



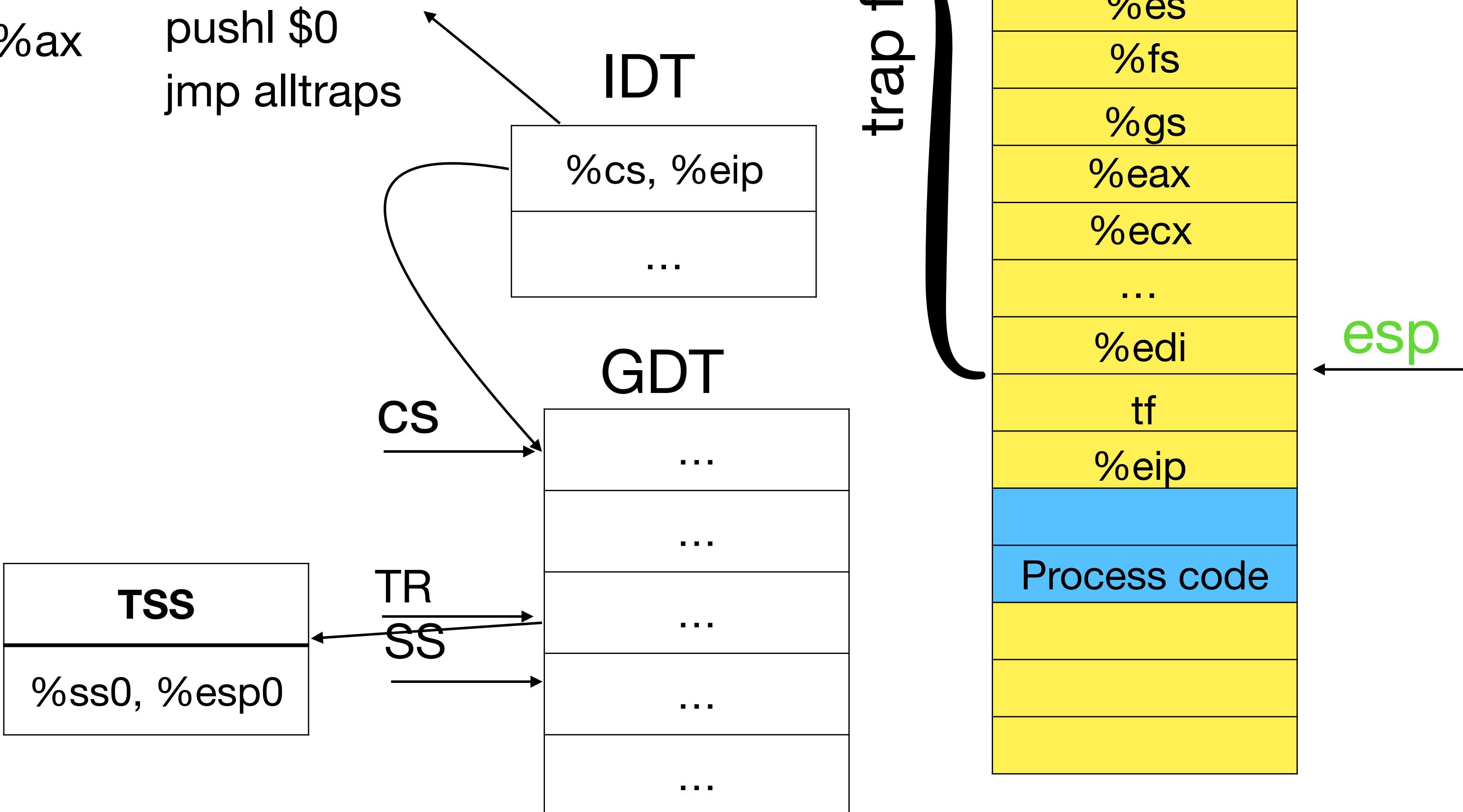
Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
```

eip

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

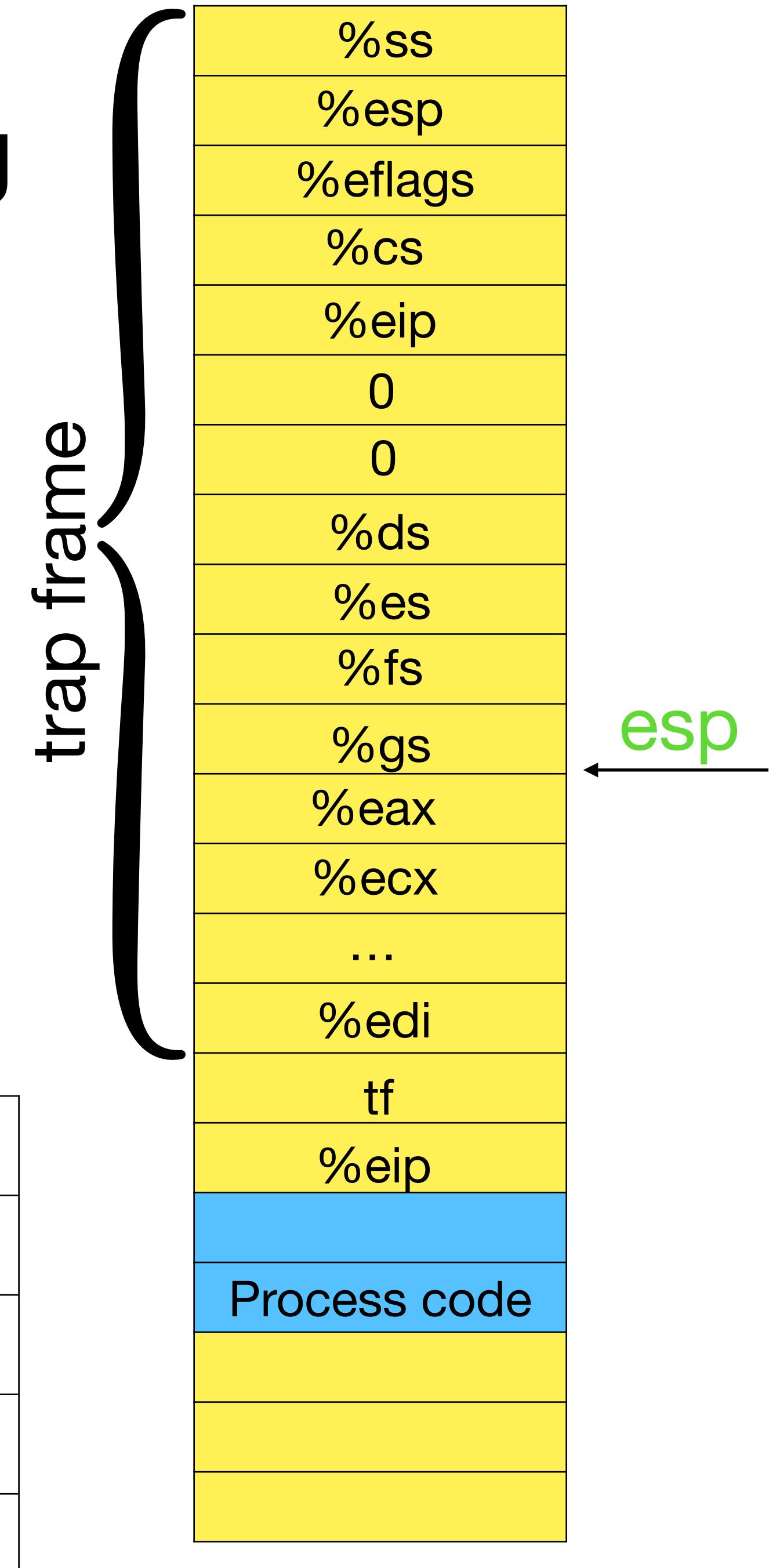
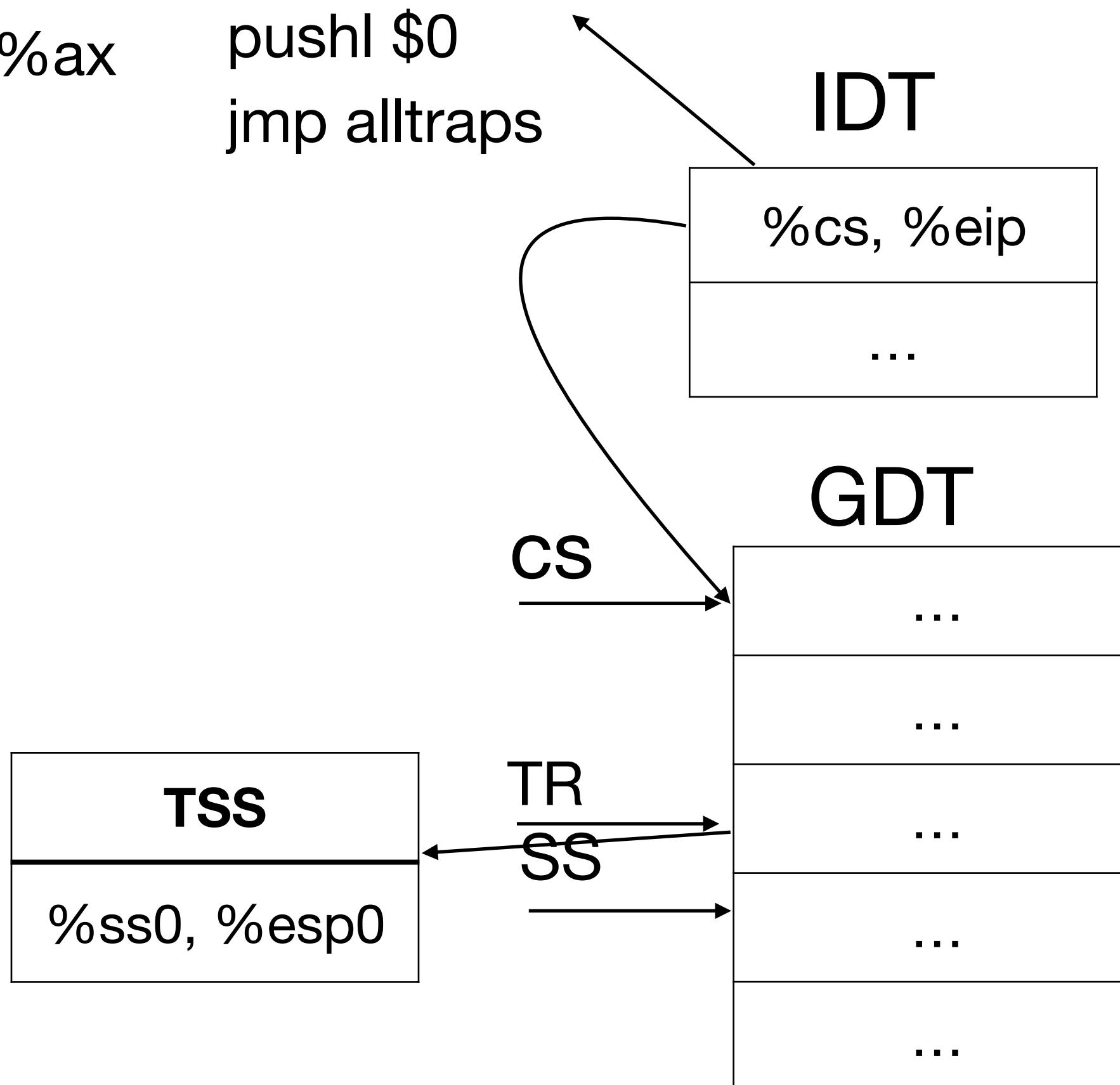


Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

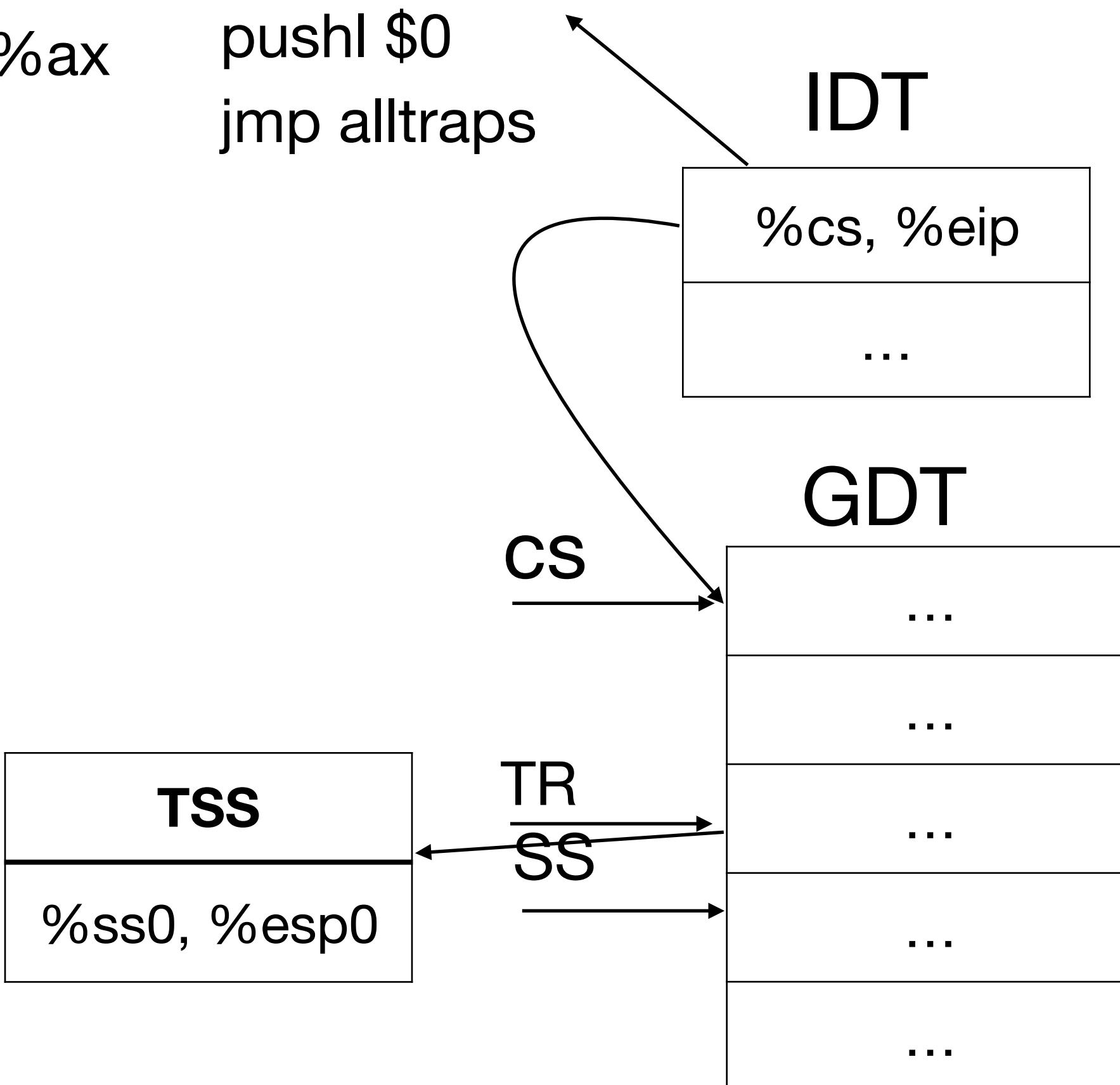


Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
    eip →
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```



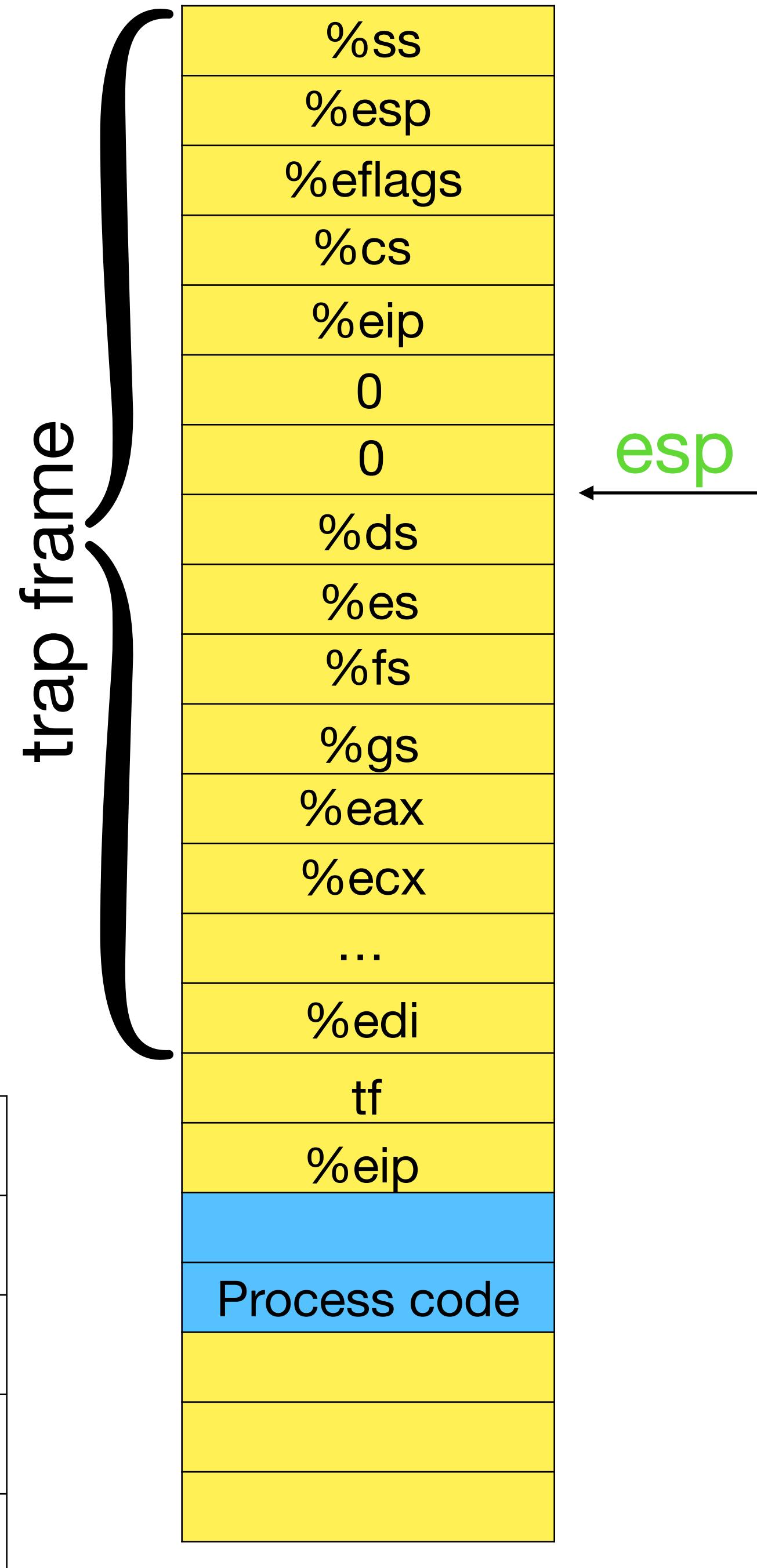
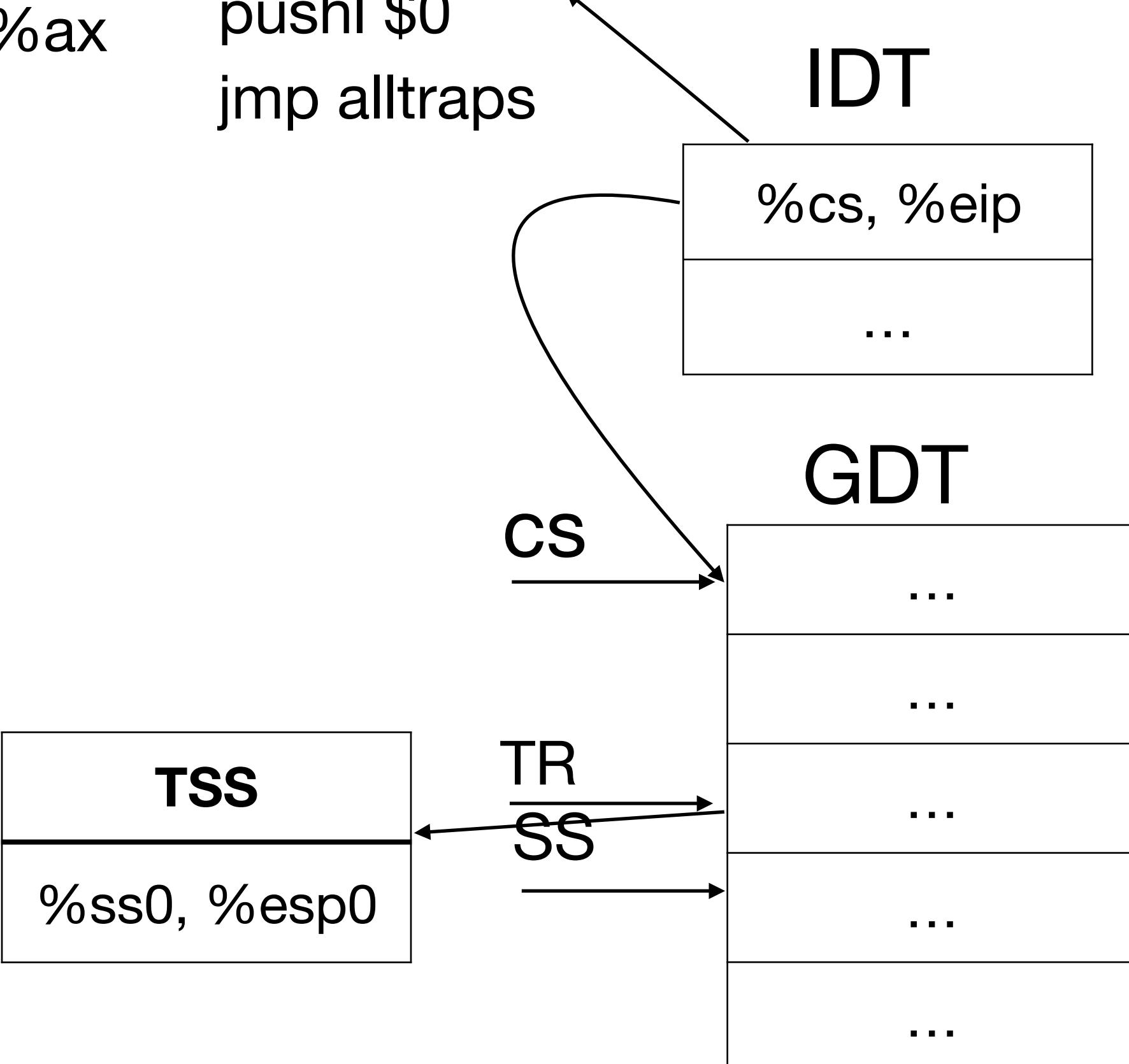
%ss
%esp
%eflags
%cs
%eip
0
0
%ds
%es
%fs
%gs
%eax
%ecx
...
%edi
tf
%eip
Process code

Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
    eip →
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

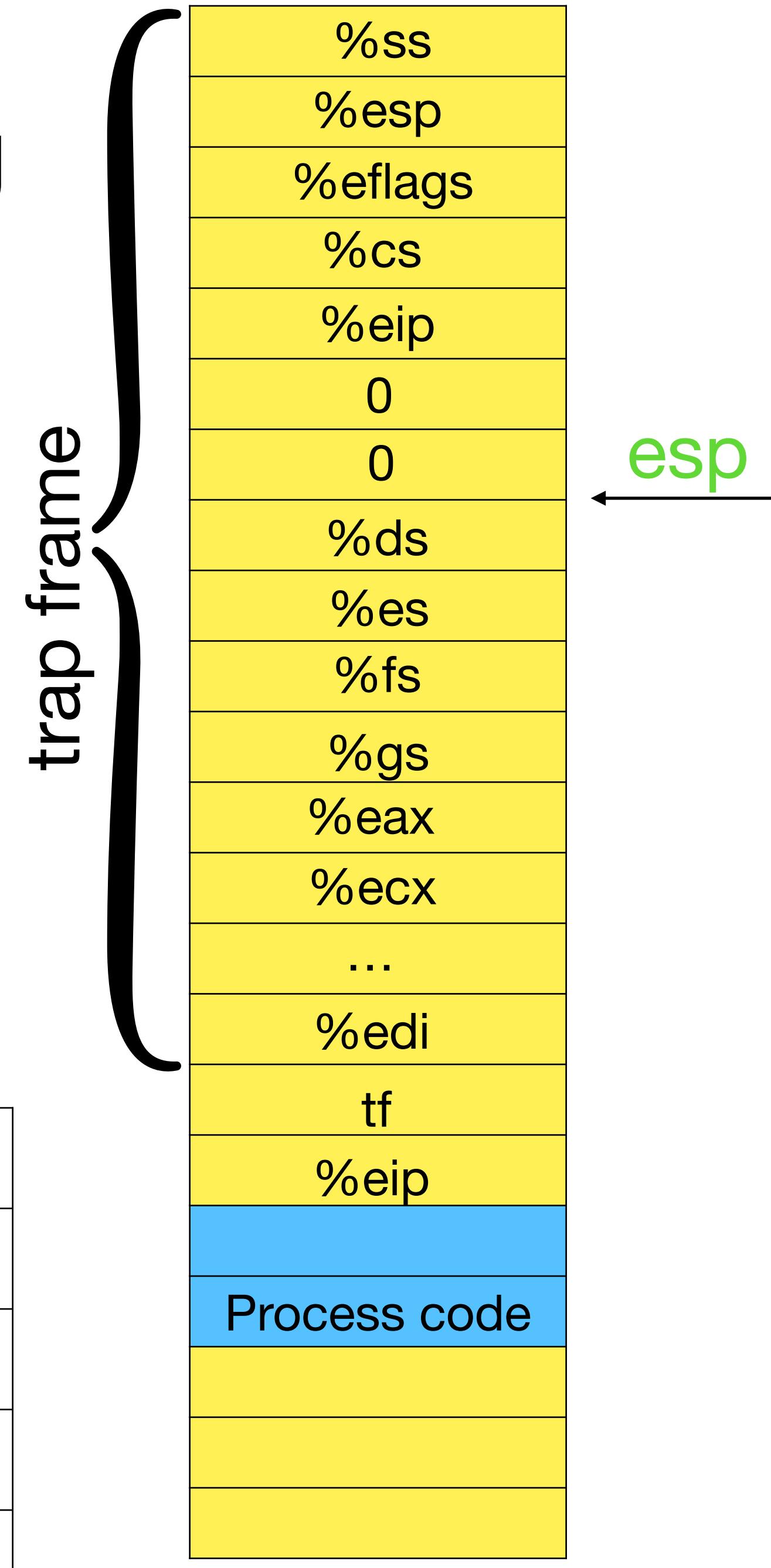
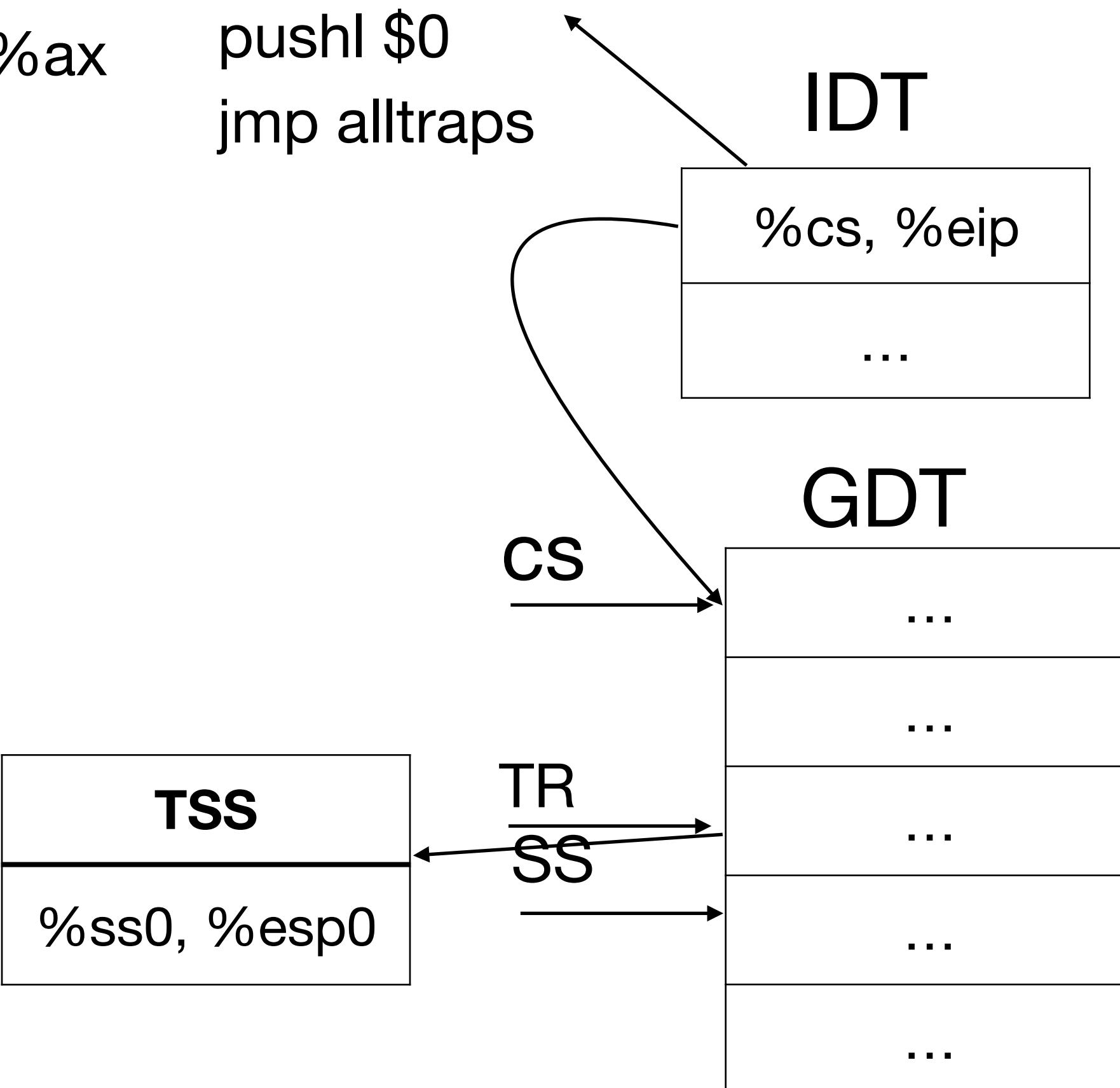


Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
    eip → addl $0x8, %esp
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

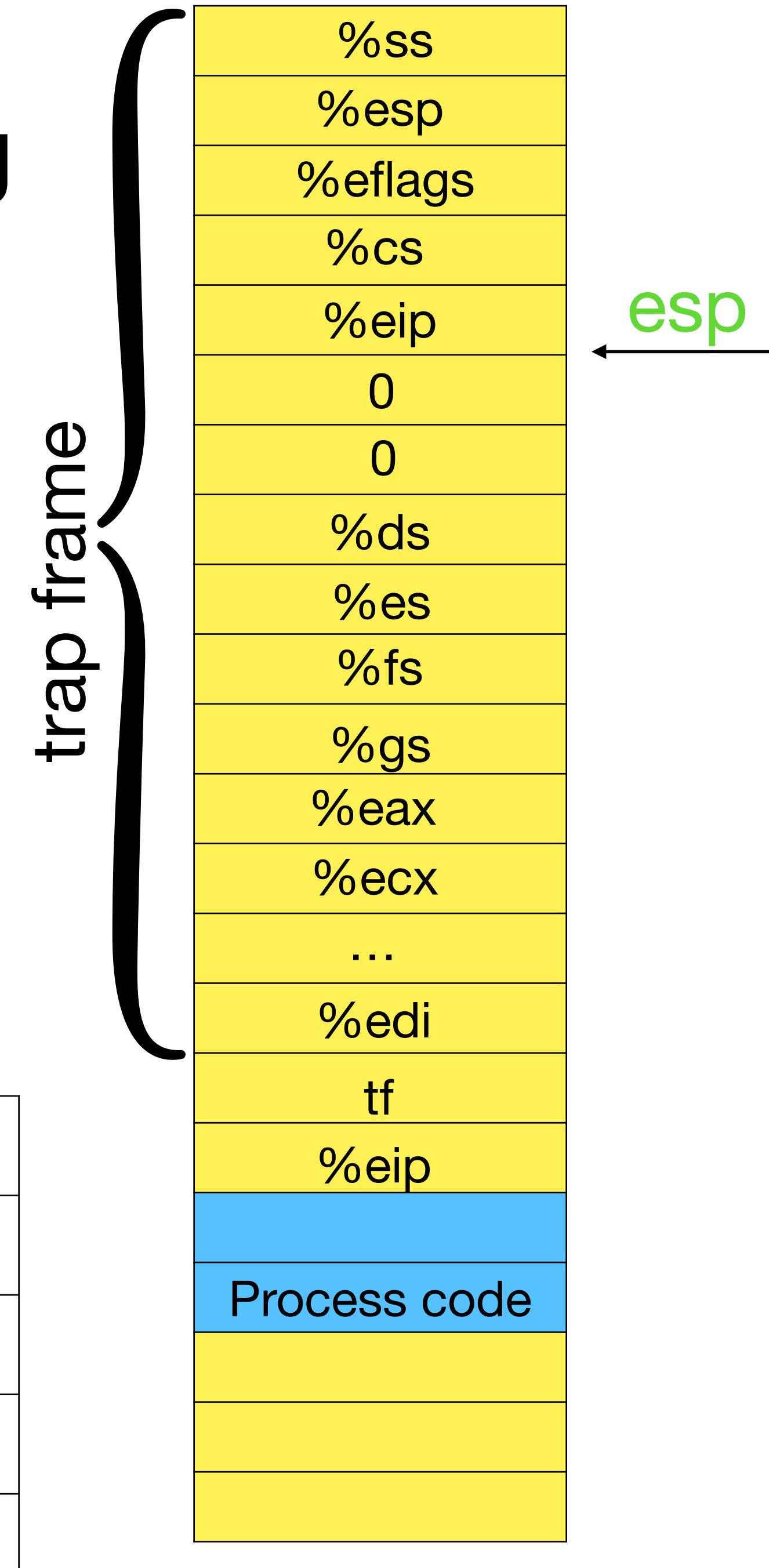
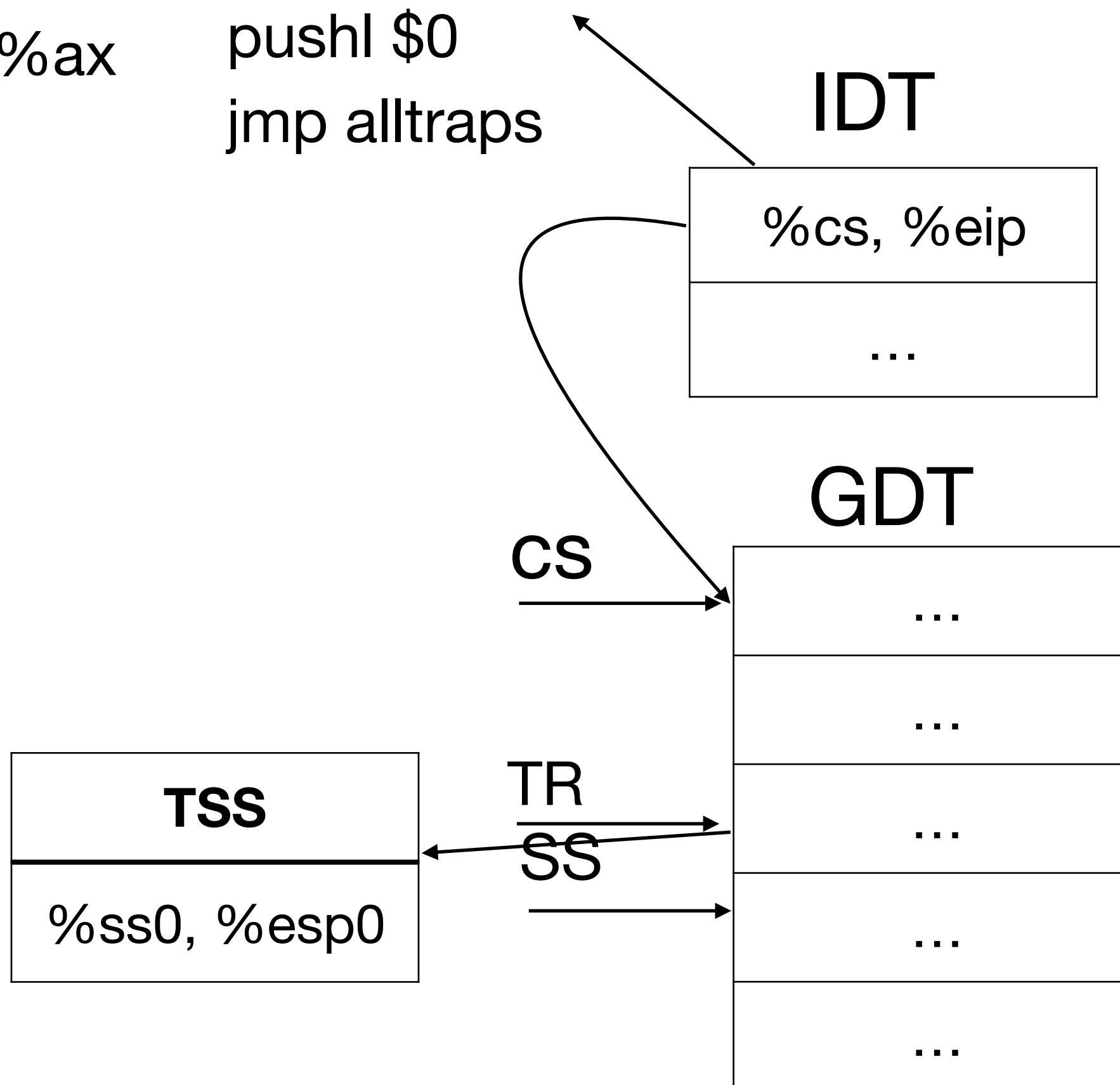


Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
    eip → addl $0x8, %esp
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

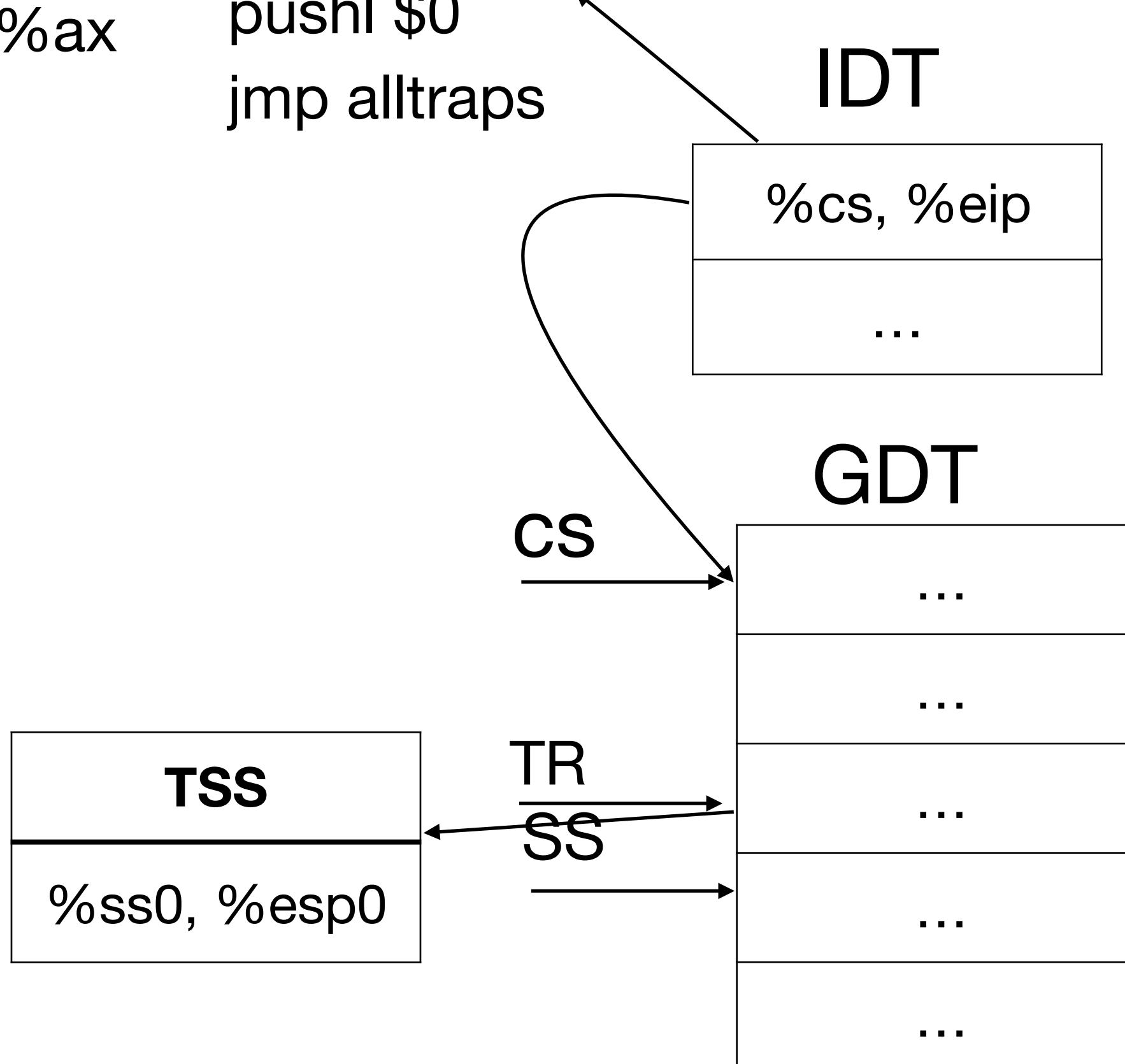


Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
    eip → iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```



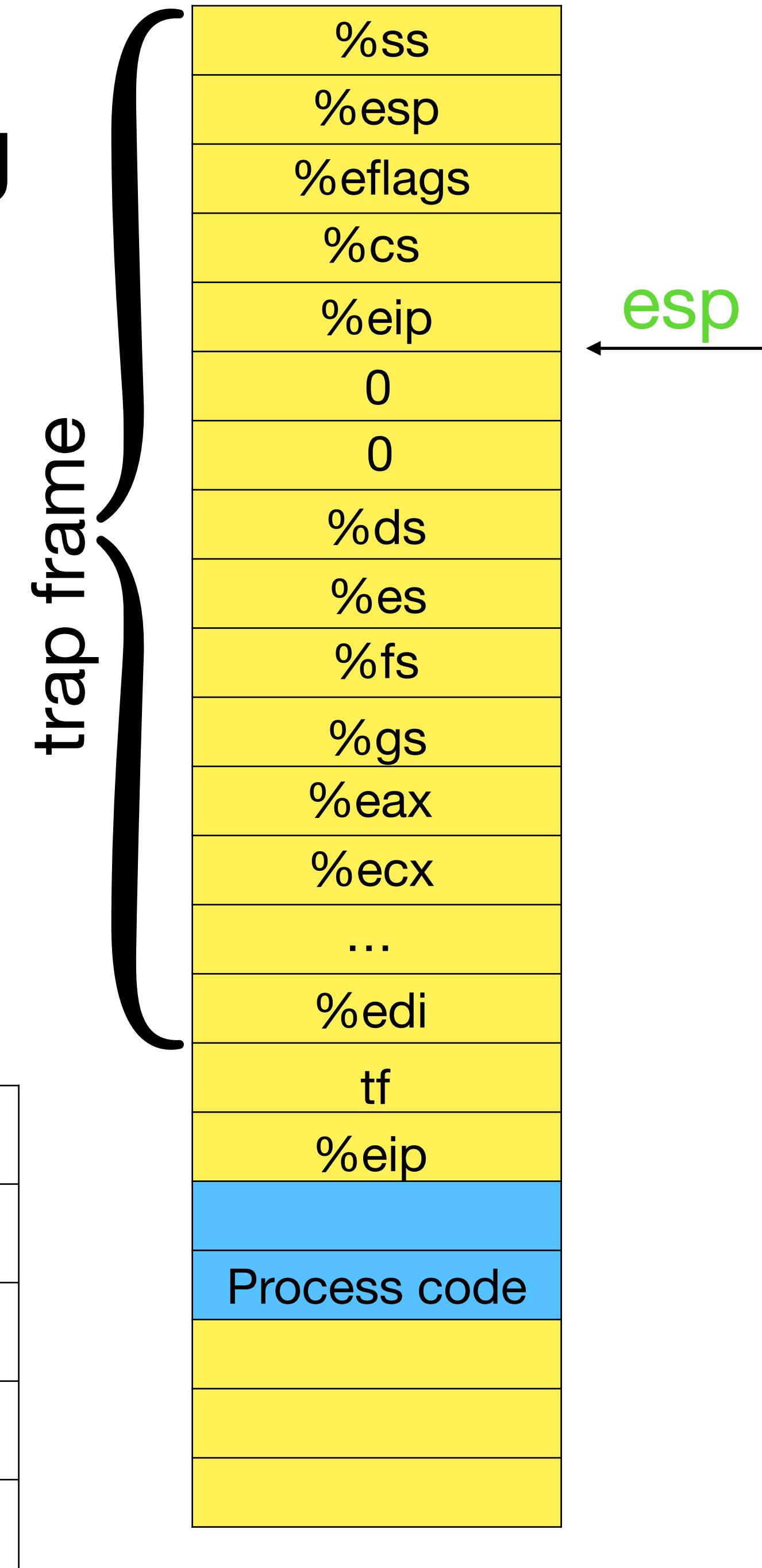
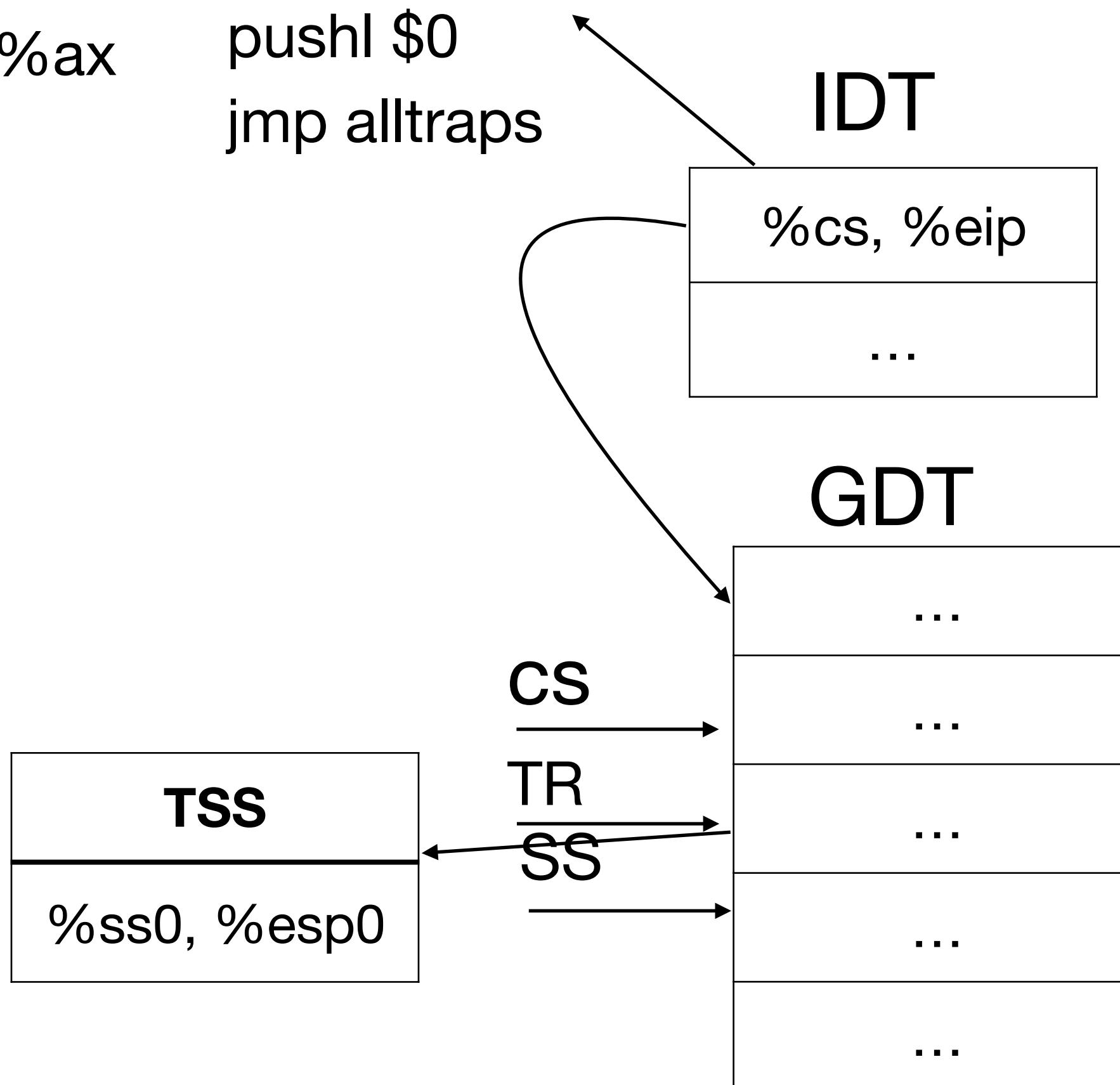
%ss
%esp
%eflags
%cs
%eip
0
0
%ds
%es
%fs
%gs
%eax
%ecx
...
%edi
tf
%eip
Process code
...
...

Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
    eip → iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

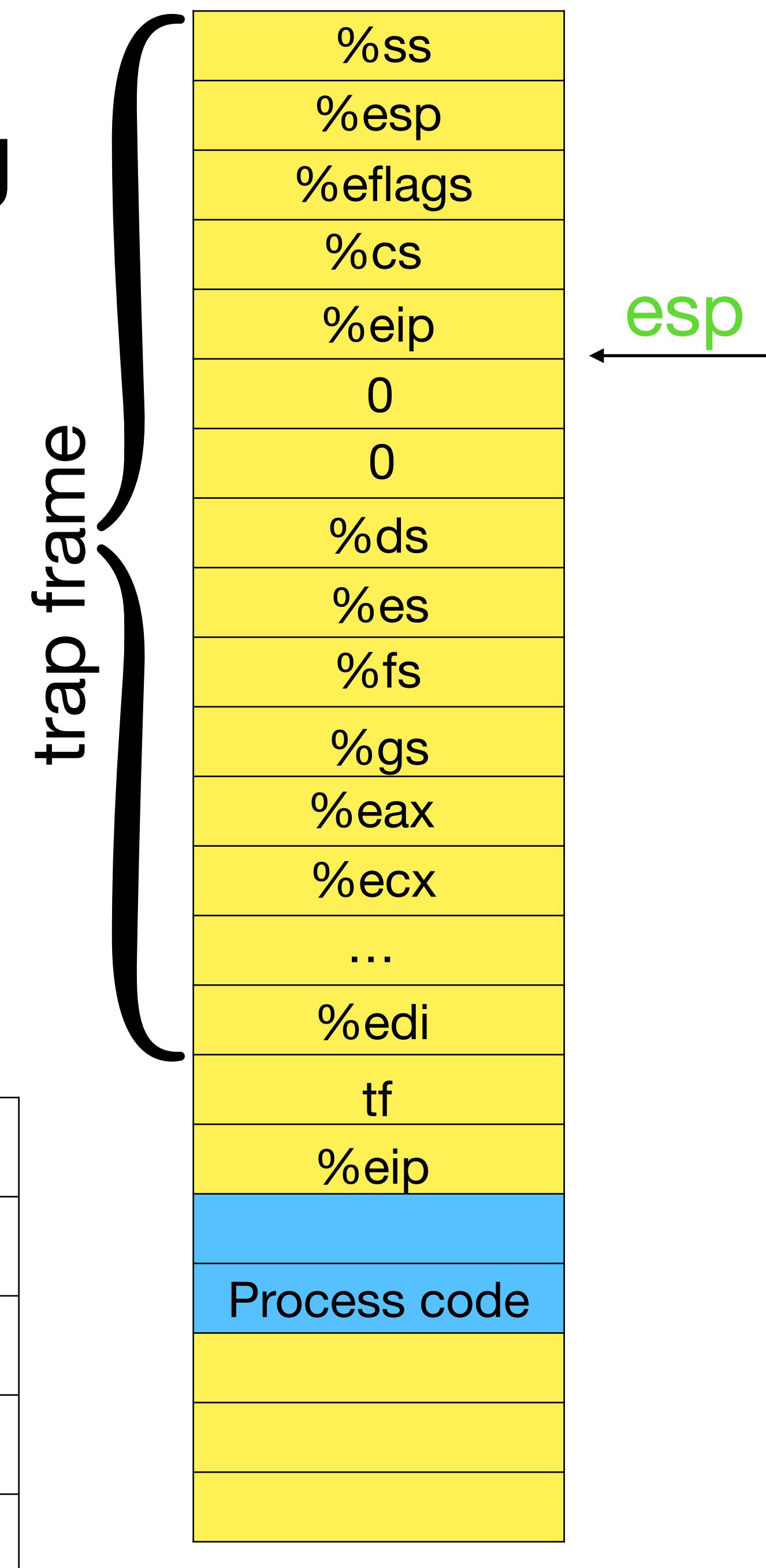
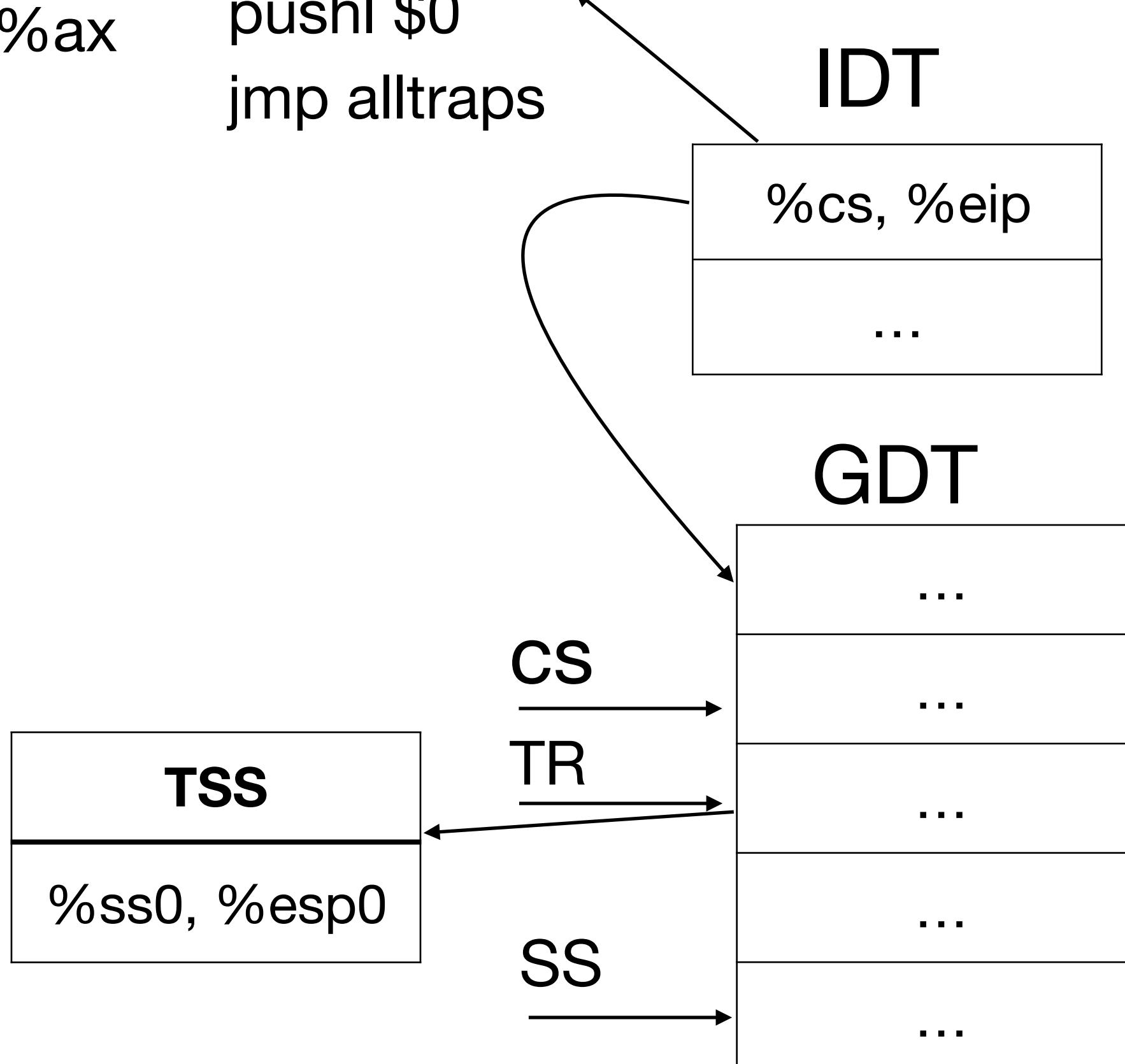


Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
    eip → iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```

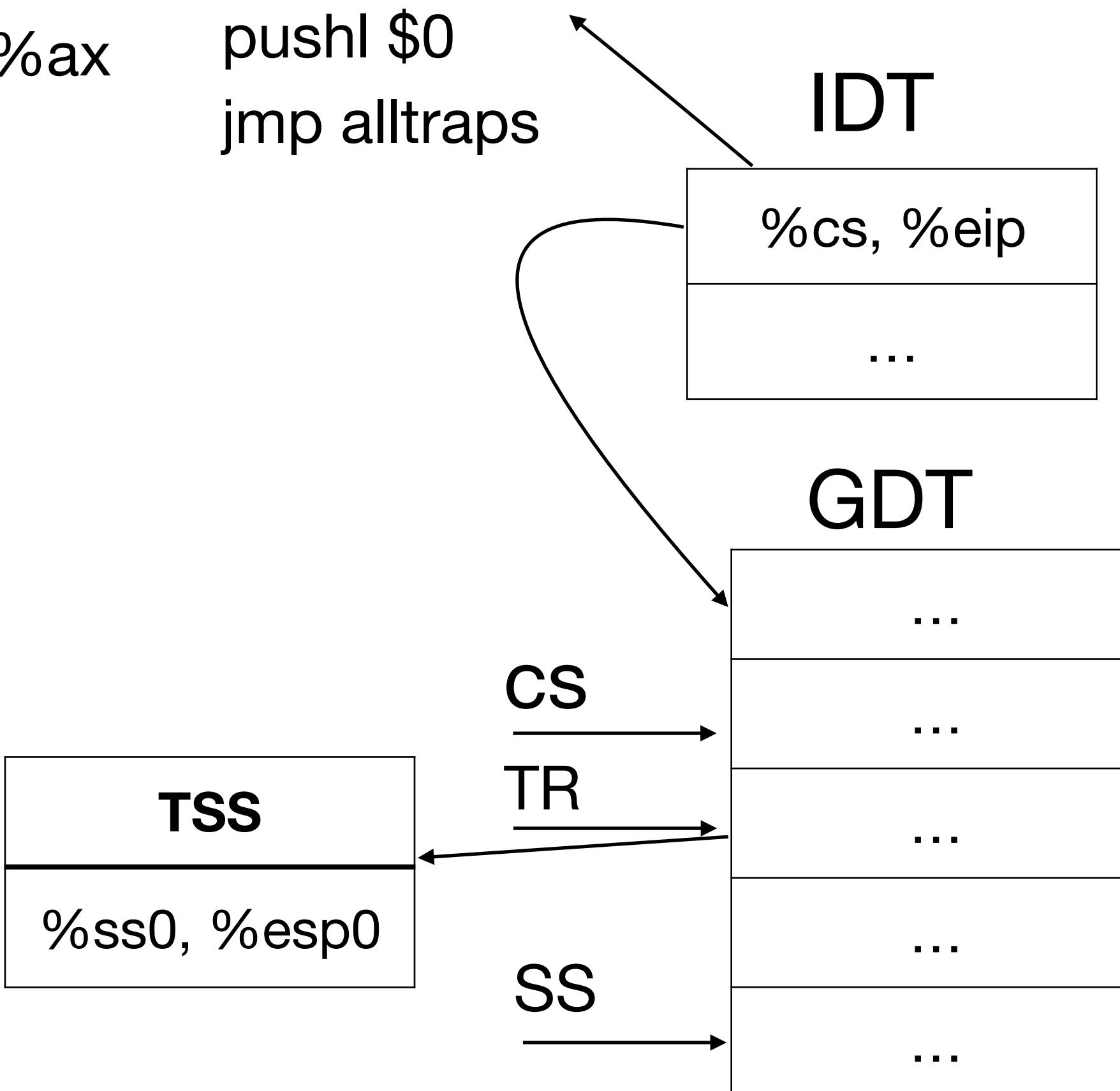


Interrupt handling with user process running

```
trapasm.S
for(;;)
;
    pushl %ds..
    pushal
    movw $(SEG_KDATA<<3), %ax
    movw %ax, %ds..
    pushl %esp
    call trap
    addl $4, %esp
    popal
    popl %ds..
    addl $0x8, %esp
    iret
    eip → iret
```

vectors.S

```
.globl vector0
vector0:
    pushl $0
    pushl $0
    jmp alltraps
```



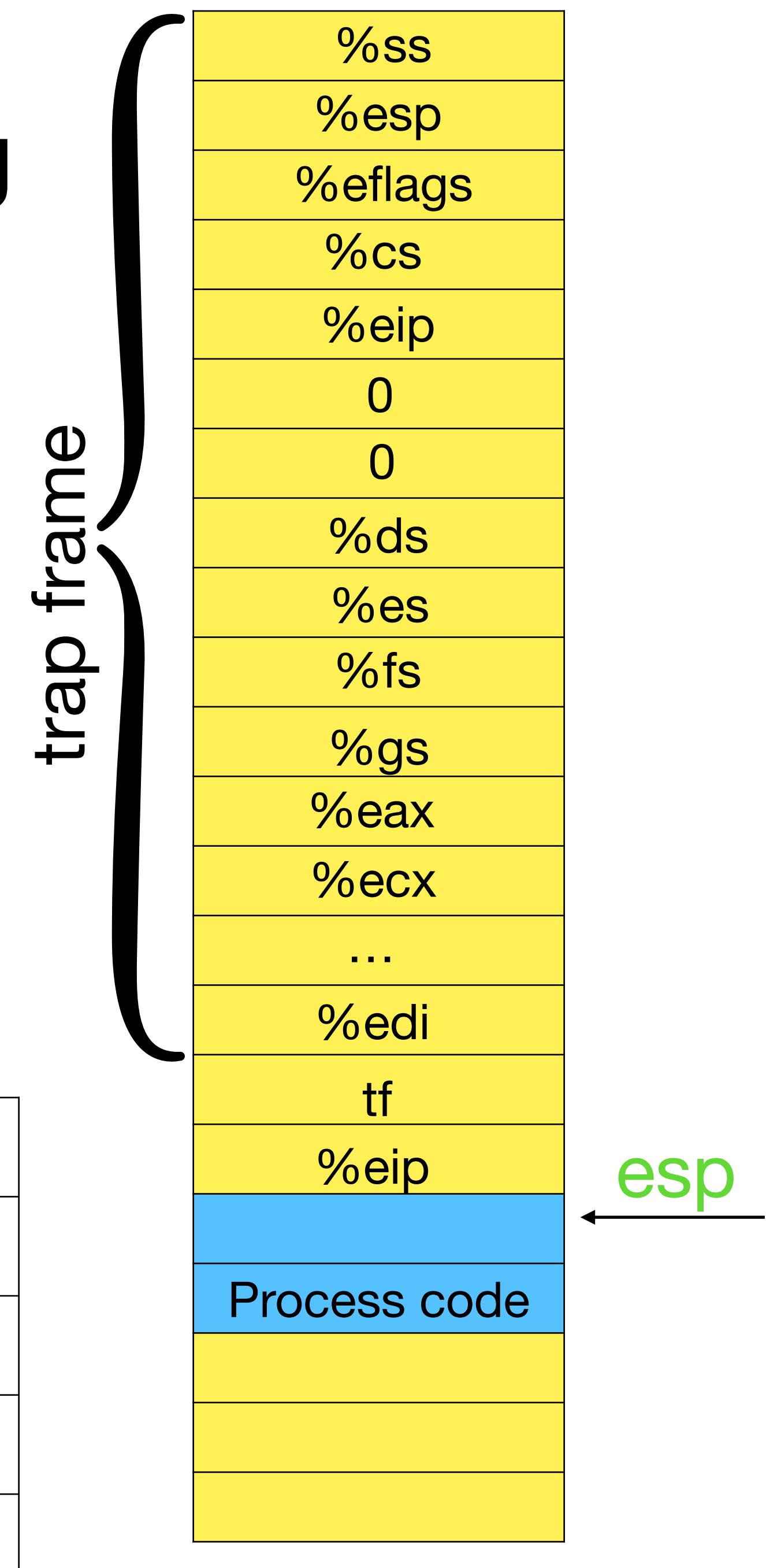
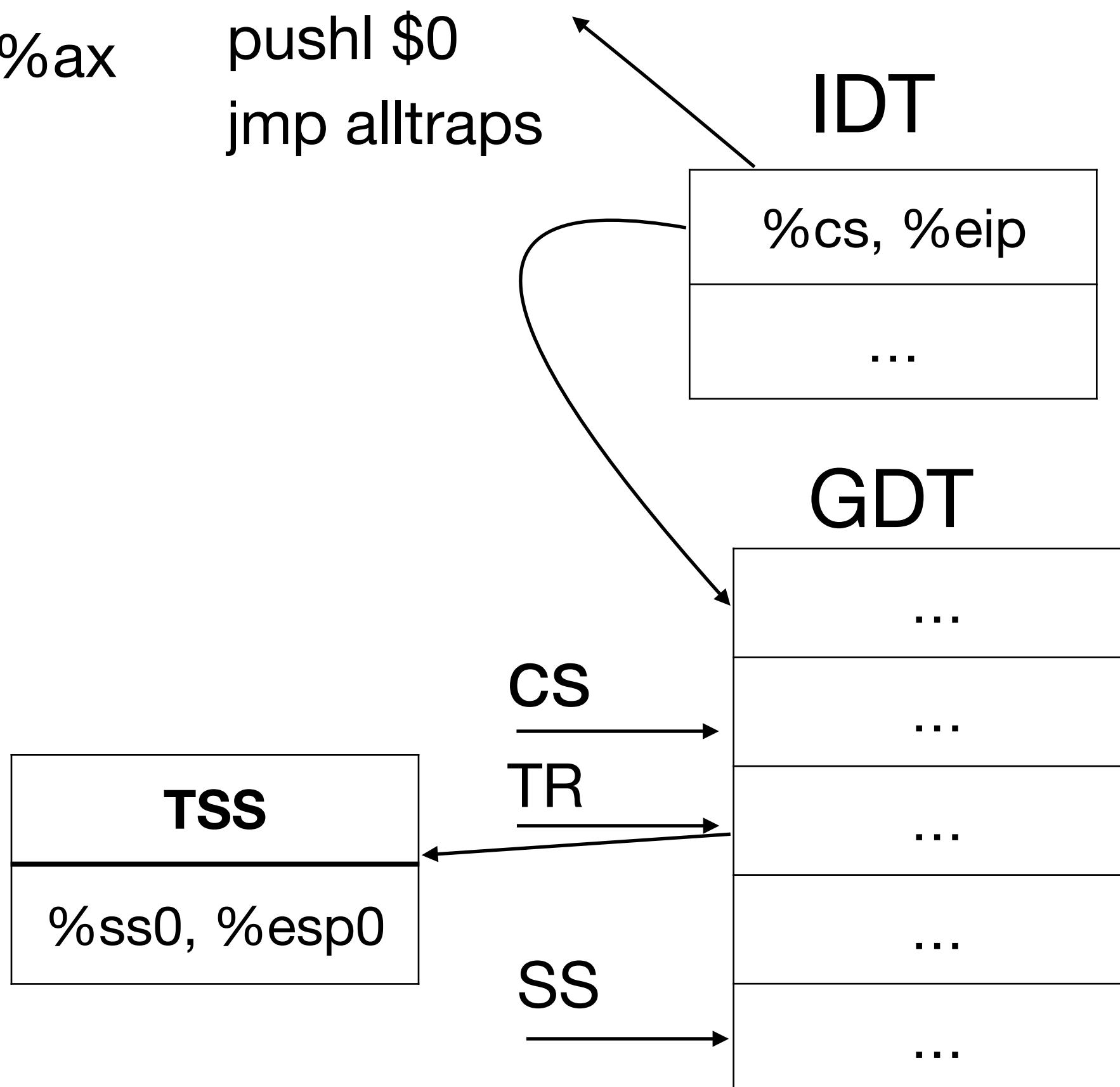
%ss
%esp
%eflags
%cs
%eip
0
0
%ds
%es
%fs
%gs
%eax
%ecx
...
%edi
tf
%eip
Process code
...
...
...
...
...

Interrupt handling with user process running

```
trapasm.S
for(;;)
    eip
    ;  
    alltraps:  
        pushl %ds..  
        pushal  
        movw $(SEG_KDATA<<3), %ax  
        movw %ax, %ds..  
        pushl %esp  
        call trap  
        addl $4, %esp  
        popal  
        popl %ds..  
        addl $0x8, %esp  
        iret
```

vectors.S

```
.globl vector0
vector0:  
    pushl $0
    pushl $0
    jmp alltraps
```



Protection so far

Protection so far

- Can processes read/write each other's memory?
 - No, since the other process' memory is not *addressable* by (not in the *address space* of) the process

Protection so far

- Can processes read/write each other's memory?
 - No, since the other process' memory is not *addressable* by (not in the *address space* of) the process
- Can process take over IDT, GDT, TSS?
 - IDT, GDT, TSS are not in address space. Cannot run LIDT, LGDT, LTR instructions since they are *privileged instructions* (only runnable from ring 0)

Protection so far

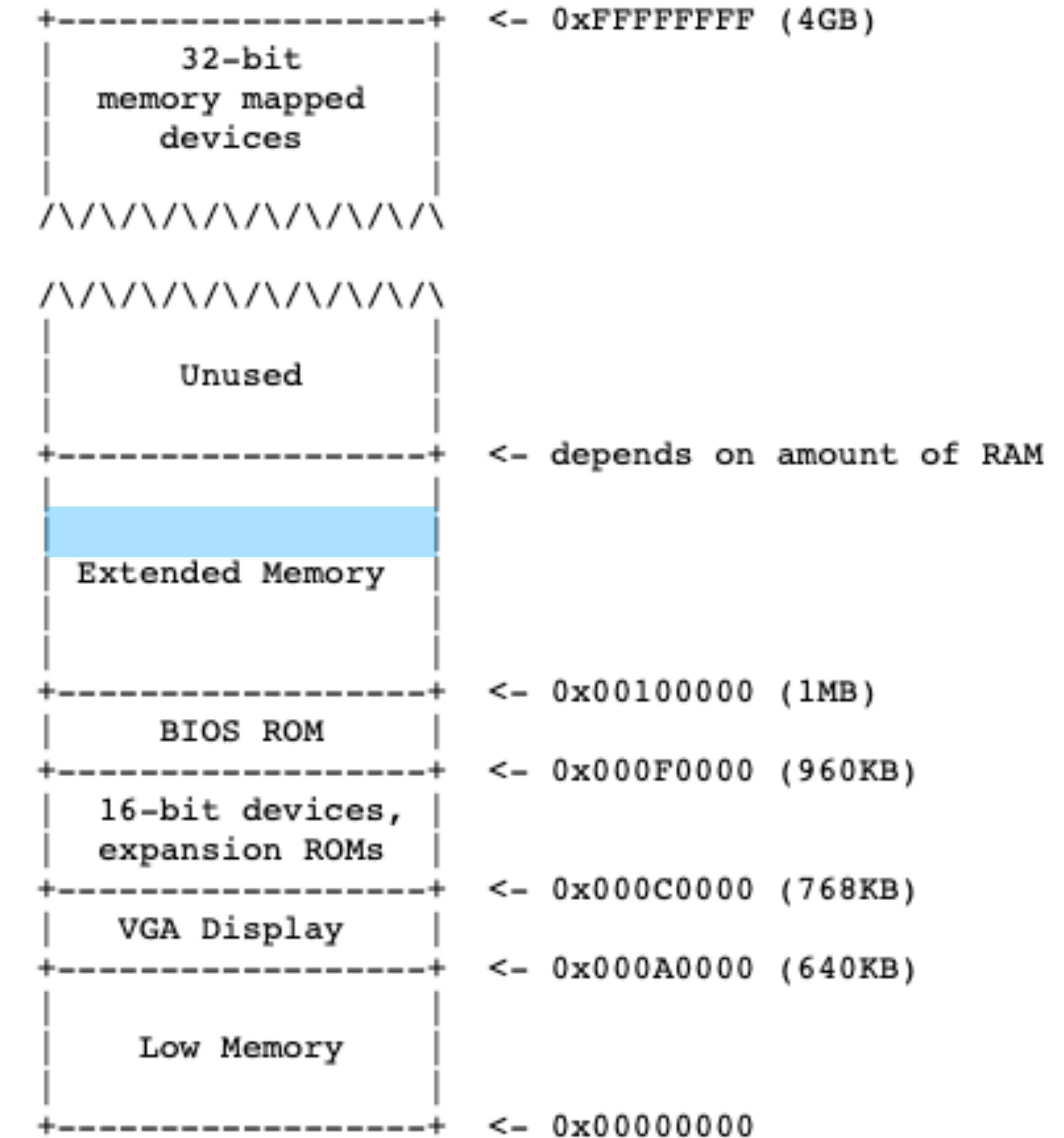
- Can processes read/write each other's memory?
 - No, since the other process' memory is not *addressable* by (not in the *address space* of) the process
- Can process take over IDT, GDT, TSS?
 - IDT, GDT, TSS are not in address space. Cannot run LIDT, LGDT, LTR instructions since they are *privileged instructions* (only runnable from ring 0)
- Can process change its privilege to ring 0?
 - Not directly. It can change only via trap handling mechanism. OS code runs upon a trap. Trap handling writes to kernel stack.

Protection so far

- Can processes read/write each other's memory?
 - No, since the other process' memory is not *addressable* by (not in the *address space* of) the process
- Can process take over IDT, GDT, TSS?
 - IDT, GDT, TSS are not in address space. Cannot run LIDT, LGDT, LTR instructions since they are *privileged instructions* (only runnable from ring 0)
- Can process change its privilege to ring 0?
 - Not directly. It can change only via trap handling mechanism. OS code runs upon a trap. Trap handling writes to kernel stack.
- Can process run away with the CPU?
 - CPU will be snatched at the time of timer interrupt

IO protection

- Memory mapped IO
 - Can directly read from / write to IO devices if OS keeps it in *process address space*



Port-mapped IO protection

inb(0x1F7), outb(0x1F2, 1), ..

Port-mapped IO protection

inb(0x1F7), outb(0x1F2, 1), ..

- Option 1: Make in and out instruction privileged. Processes must do IO via kernel.

Port-mapped IO protection

`inb(0x1F7), outb(0x1F2, 1), ..`

- Option 1: Make in and out instruction privileged. Processes must do IO via kernel.
 - Expensive user mode-kernel mode transitions for IO intensive processes.

Port-mapped IO protection

`inb(0x1F7), outb(0x1F2, 1), ..`

- Option 1: Make in and out instruction privileged. Processes must do IO via kernel.
 - Expensive user mode-kernel mode transitions for IO intensive processes.
- Option 2: Set EFLAGS.IOPL = 3

Port-mapped IO protection

inb(0x1F7), outb(0x1F2, 1), ...

- Option 1: Make in and out instruction privileged. Processes must do IO via kernel.
 - Expensive user mode-kernel mode transitions for IO intensive processes.
- Option 2: Set EFLAGS.IOPL = 3

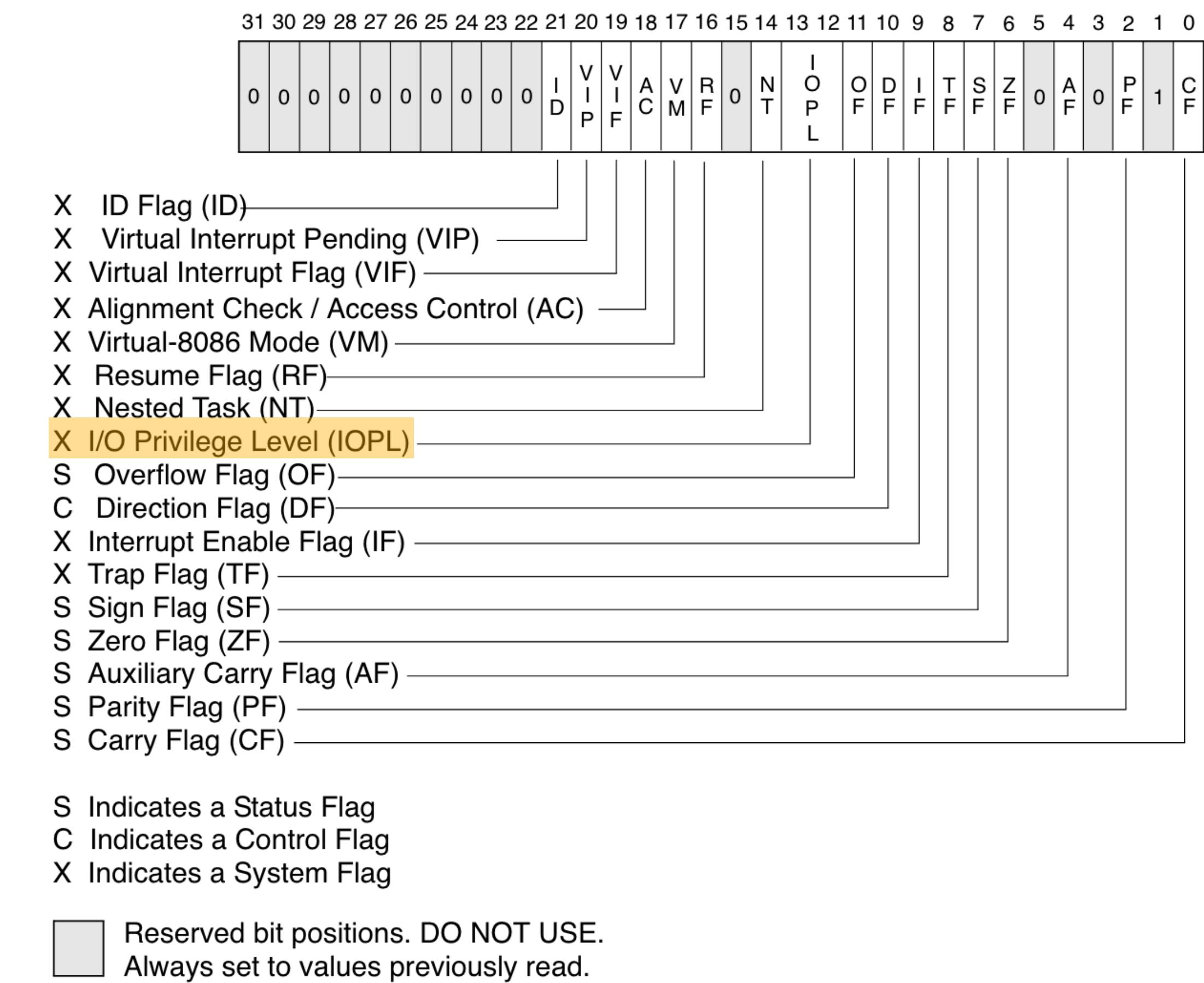


Figure 3-8. EFLAGS Register

Port-mapped IO protection

inb(0x1F7), outb(0x1F2, 1), ...

- Option 1: Make in and out instruction privileged. Processes must do IO via kernel.
 - Expensive user mode-kernel mode transitions for IO intensive processes.
- Option 2: Set EFLAGS.IOPL = 3
 - Processes (CPL=3) cannot modify EFLAGS.IOPL

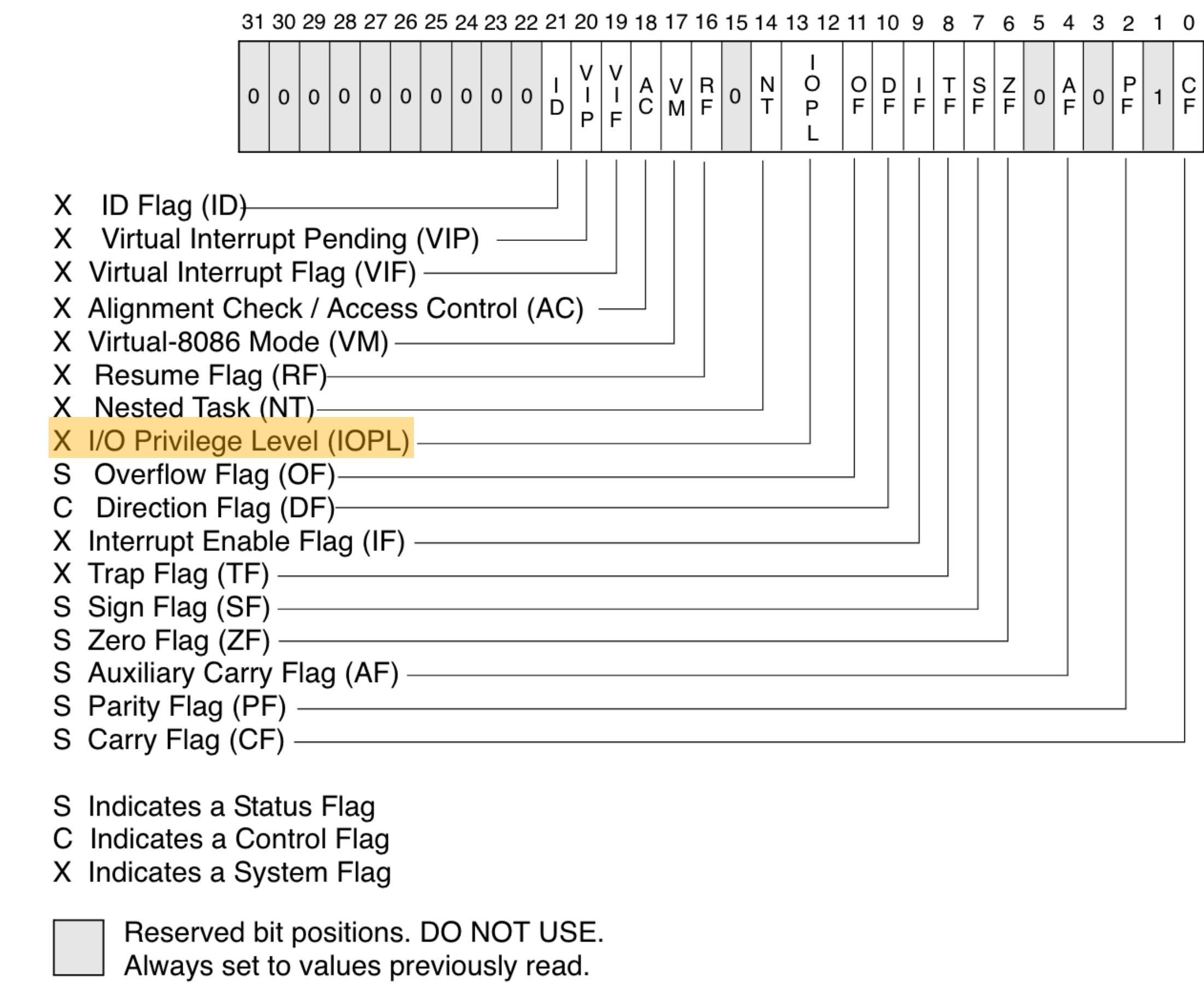


Figure 3-8. EFLAGS Register

Port-mapped IO protection

inb(0x1F7), outb(0x1F2, 1), ...

- Option 1: Make in and out instruction privileged. Processes must do IO via kernel.
 - Expensive user mode-kernel mode transitions for IO intensive processes.
- Option 2: Set EFLAGS.IOPL = 3
 - Processes (CPL=3) cannot modify EFLAGS.IOPL
 - Gives permission to access all IO ports

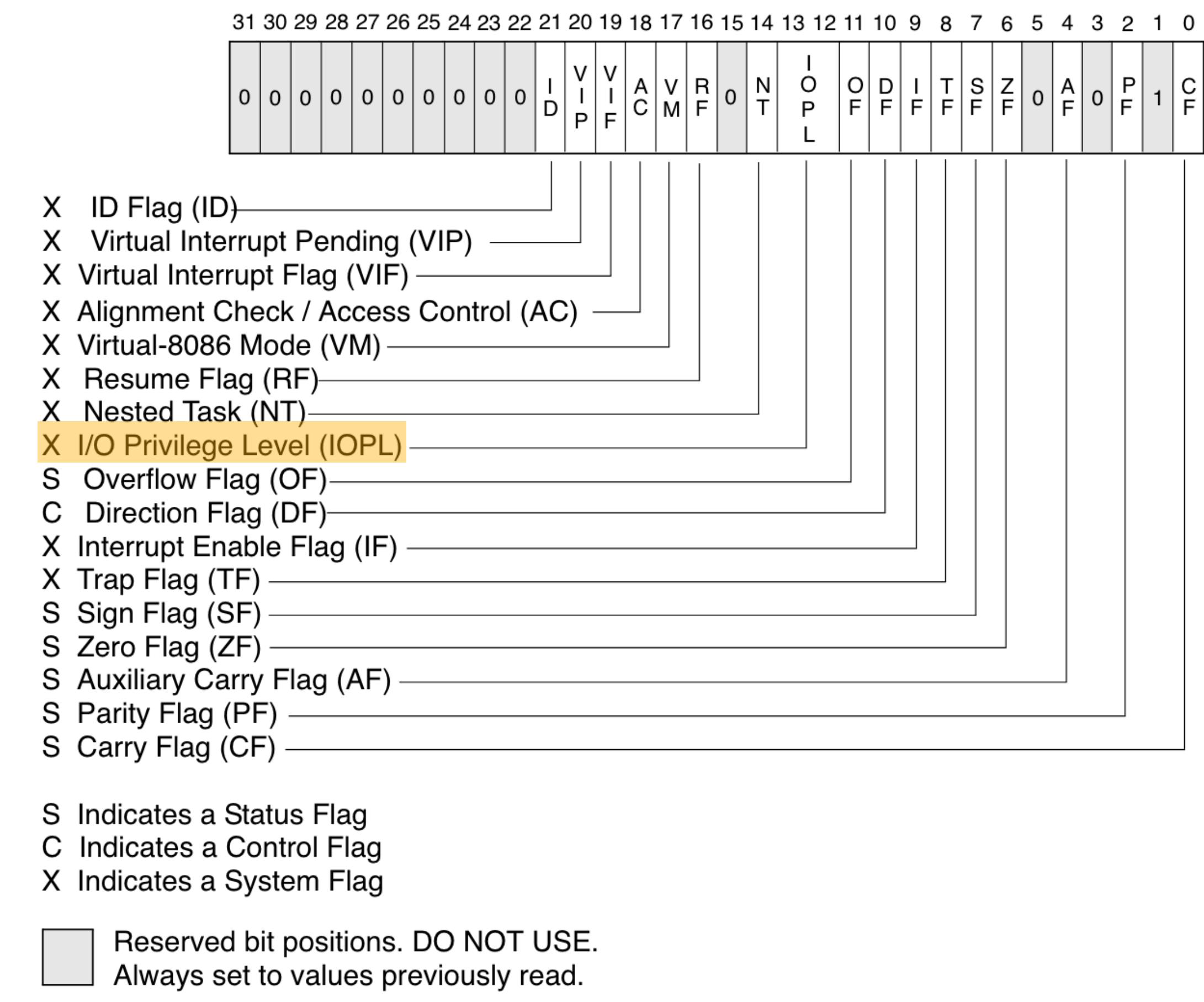


Figure 3-8. EFLAGS Register

Port-mapped IO protection (2)

Port-mapped IO protection (2)

- Option 3: Use a bitmap to decide which IO ports are directly accessible

Port-mapped IO protection (2)

- Option 3: Use a bitmap to decide which IO ports are directly accessible

31	15	0
I/O Map Base Address	Reserved	T 100
Reserved	LDT Segment Selector	96
Reserved	GS	92
Reserved	FS	88
Reserved	DS	84
Reserved	SS	80
Reserved	CS	76
Reserved	ES	72
	EDI	68
	ESI	64
	EBP	60
	ESP	56
	EBX	52
	EDX	48
	ECX	44
	EAX	40
	EFLAGS	36
	EIP	32
	CR3 (PDBR)	28
Reserved	SS2	24
	ESP2	20
Reserved	SS1	16
	ESP1	12
Reserved	SS0	8
	ESP0	4
Reserved	Previous Task Link	0

 Reserved bits. Set to 0.

Figure 7-2. 32-Bit Task-State Segment (TSS)

Port-mapped IO protection (2)

- Option 3: Use a bitmap to decide which IO ports are directly accessible

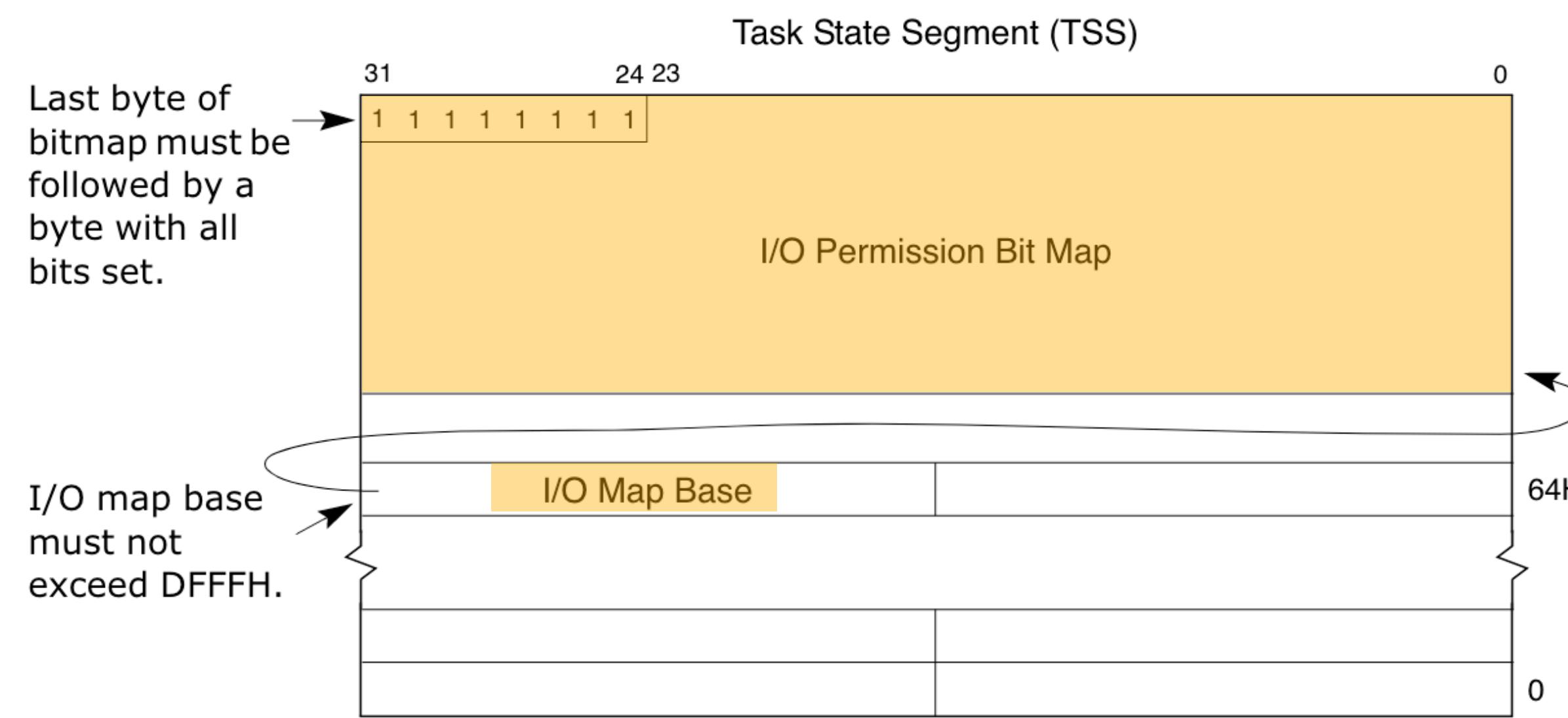


Figure 18-2. I/O Permission Bit Map

31	I/O Map Base Address	15	Reserved	0
	Reserved		LDT Segment Selector	100
	Reserved		GS	96
	Reserved		FS	92
	Reserved		DS	88
	Reserved		SS	84
	Reserved		CS	80
	Reserved		ES	76
			EDI	72
			ESI	68
			EBP	64
			ESP	60
			EBX	56
			EDX	52
			ECX	48
			EAX	44
	EFLAGS			40
	EIP			36
	CR3 (PDBR)			32
	Reserved		SS2	28
			ESP2	24
	Reserved		SS1	20
			ESP1	16
	Reserved		SS0	12
	ESP0			8
	Reserved		Previous Task Link	4
				0

Reserved bits. Set to 0.

Figure 7-2. 32-Bit Task-State Segment (TSS)

Port-mapped IO protection (2)

- Option 3: Use a bitmap to decide which IO ports are directly accessible
 - If CPL > EFLAGS.IOPL, use bitmap to decide if in, out instruction should generate a trap

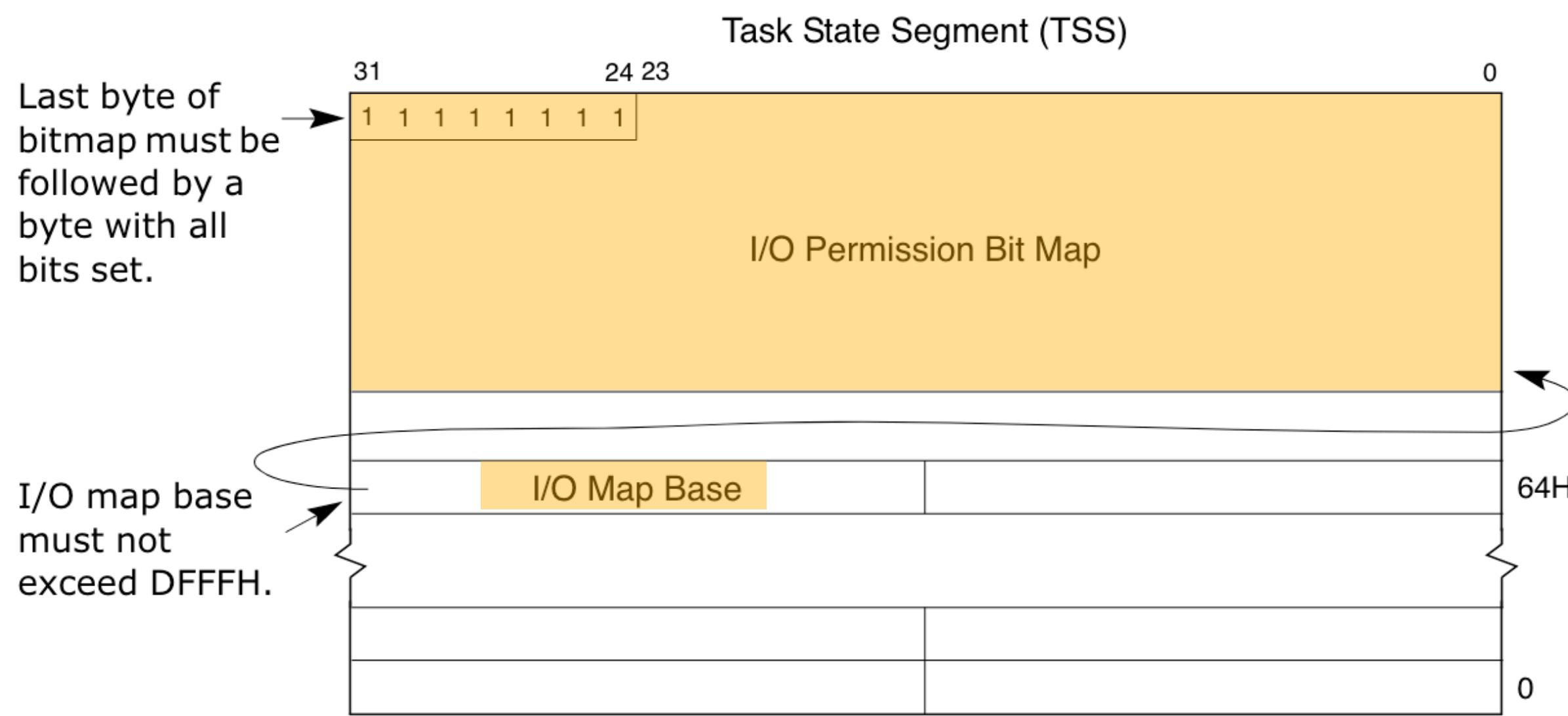


Figure 18-2. I/O Permission Bit Map

31	I/O Map Base Address	15	Reserved	0
	Reserved		LDT Segment Selector	100
	Reserved		GS	96
	Reserved		FS	92
	Reserved		DS	88
	Reserved		SS	84
	Reserved		CS	80
	Reserved		ES	76
			EDI	72
			ESI	68
			EBP	64
			ESP	60
			EBX	56
			EDX	52
			ECX	48
			EAX	44
			EFLAGS	40
			EIP	36
			CR3 (PDBR)	32
	Reserved		SS2	28
			ESP2	24
	Reserved		SS1	20
			ESP1	16
	Reserved		SS0	12
			ESP0	8
	Reserved		Previous Task Link	4
				0

Legend: Reserved bits. Set to 0.

Figure 7-2. 32-Bit Task-State Segment (TSS)

Port-mapped IO protection (2)

- Option 3: Use a bitmap to decide which IO ports are directly accessible
 - If CPL > EFLAGS.IOPL, use bitmap to decide if in, out instruction should generate a trap
 - Example: `outb(41, 1)` is allowed if 41st bit is set to 0

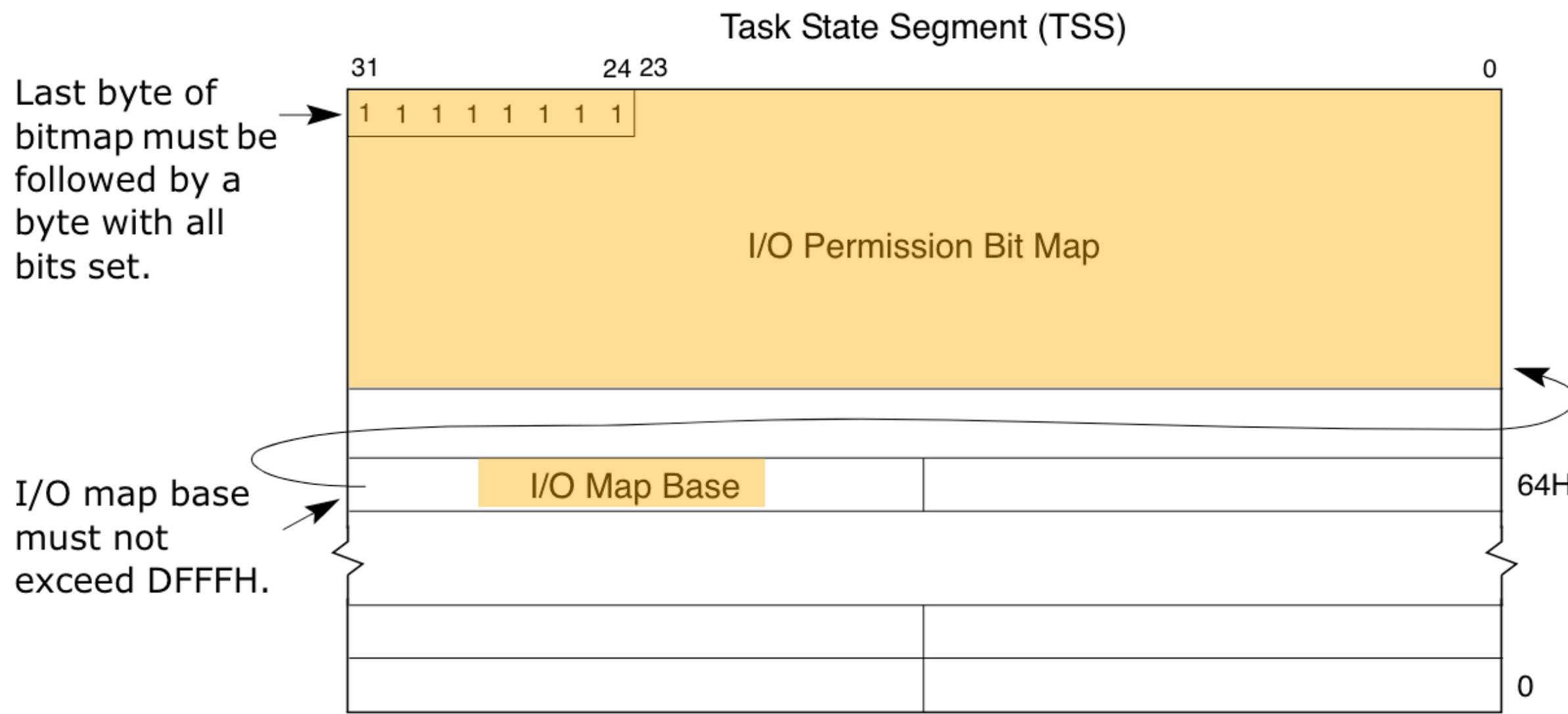


Figure 18-2. I/O Permission Bit Map

31	I/O Map Base Address	15	Reserved	0
	Reserved		LDT Segment Selector	100
	Reserved		GS	96
	Reserved		FS	92
	Reserved		DS	88
	Reserved		SS	84
	Reserved		CS	80
	Reserved		ES	76
			EDI	72
			ESI	68
			EBP	64
			ESP	60
			EBX	56
			EDX	52
			ECX	48
			EAX	44
			EFLAGS	40
			EIP	36
			CR3 (PDBR)	32
	Reserved		SS2	28
			ESP2	24
	Reserved		SS1	20
			ESP1	16
	Reserved		SS0	12
			ESP0	8
	Reserved		Previous Task Link	4
				0

Legend: Reserved bits. Set to 0.

Figure 7-2. 32-Bit Task-State Segment (TSS)

IO protection in xv6

- Processes cannot directly do IO
 - MMIO: Do not map IO addresses in process address space
 - PIO: Running in and out instructions will generate a trap (jump to OS)

IO protection in xv6

```
void pinit(void) {  
    ..  
    p->tf->eflags = FL_IF;  
}
```

```
void switchuvm(struct proc *p) {  
    mycpu()->gdt[SEG_TSS] = SEG16(STS_T32A, &mycpu()->ts,  
                                    sizeof(mycpu()->ts)-1, 0);  
    mycpu()->ts.ss0 = SEG_KDATA << 3;  
    mycpu()->ts.esp0 = (uint)p->kstack + KSTACKSIZE;  
    // setting IOPL=0 in eflags *and* iomb beyond the tss segment limit  
    // forbids I/O instructions (e.g., inb and outb) from user space  
    mycpu()->ts.iomb = (ushort) 0xFFFF;  
    ltr(SEG_TSS << 3);  
}
```

- Processes cannot directly do IO
 - MMIO: Do not map IO addresses in process address space
 - PIO: Running in and out instructions will generate a trap (jump to OS)

System calls

System calls

- Process wants OS to work on its behalf

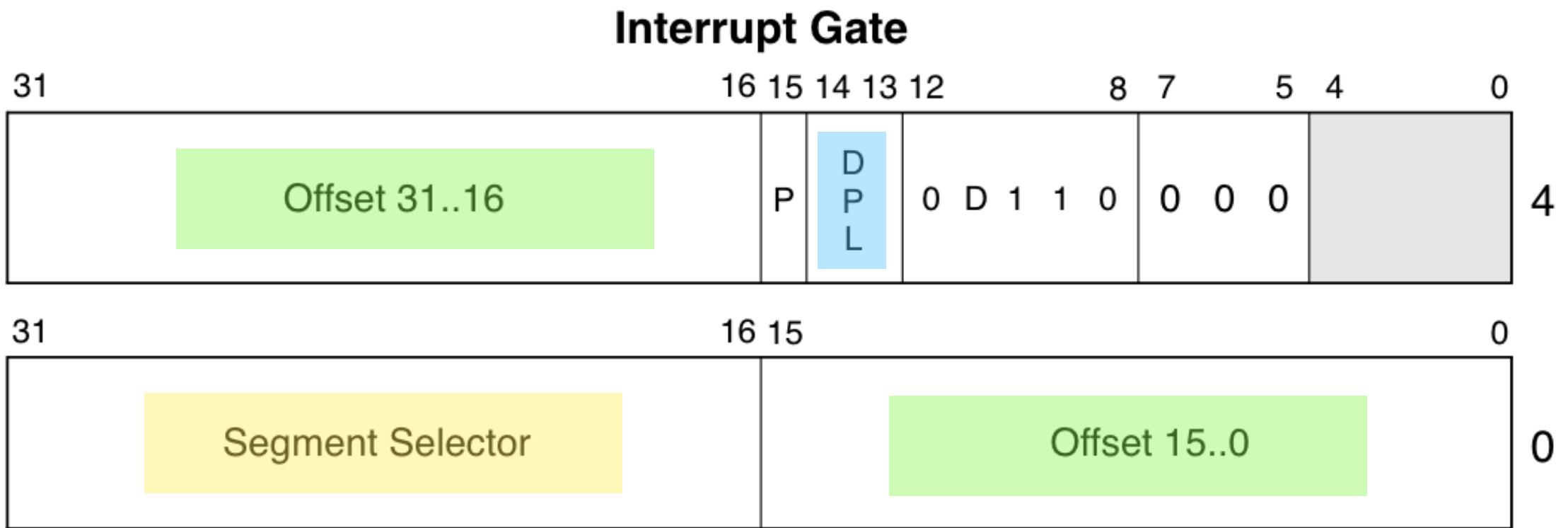


System calls

- Process wants OS to work on its behalf
 - Run INT xx instruction to trigger interrupt number xx



System calls

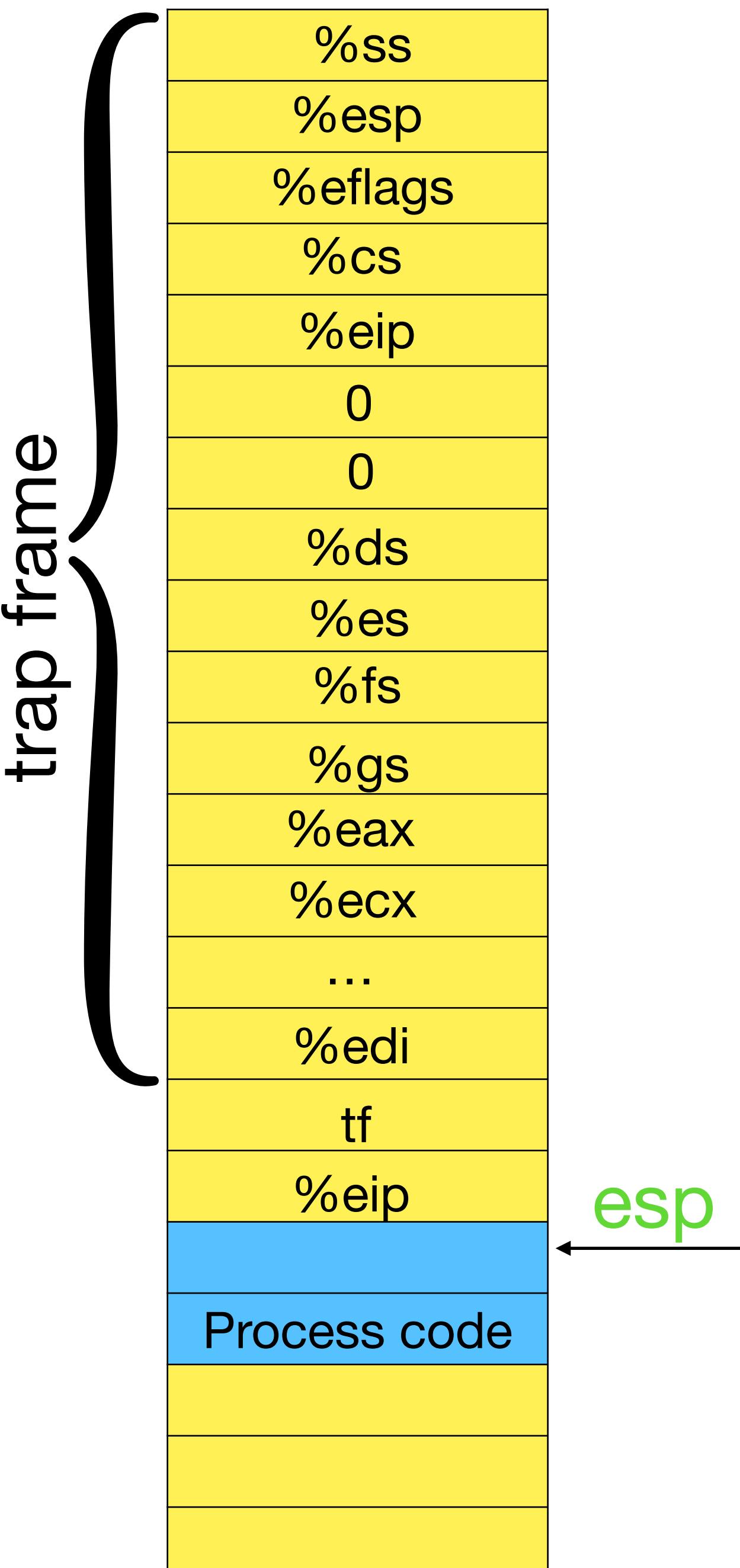


- Process wants OS to work on its behalf
 - Run INT xx instruction to trigger interrupt number xx
 - INT xx is allowed if CPL <= DPL for the xx interrupt vector

Code walkthrough

p19-syscall

- trap.c
 - tvinit sets DPL of IDT entry for T_SYSCALL to 3
 - trap calls syscall if interrupt vector is T_SYSCALL
- syscall.c
 - syscall reads eax from trap frame to find which syscall is made and calls that particular call
 - Return value of sys call is set in trap frame's eax
- initproc.S
 - Calls three system calls: open “console” file, write “hello world”, close



Visualising syscall handling p19-syscall

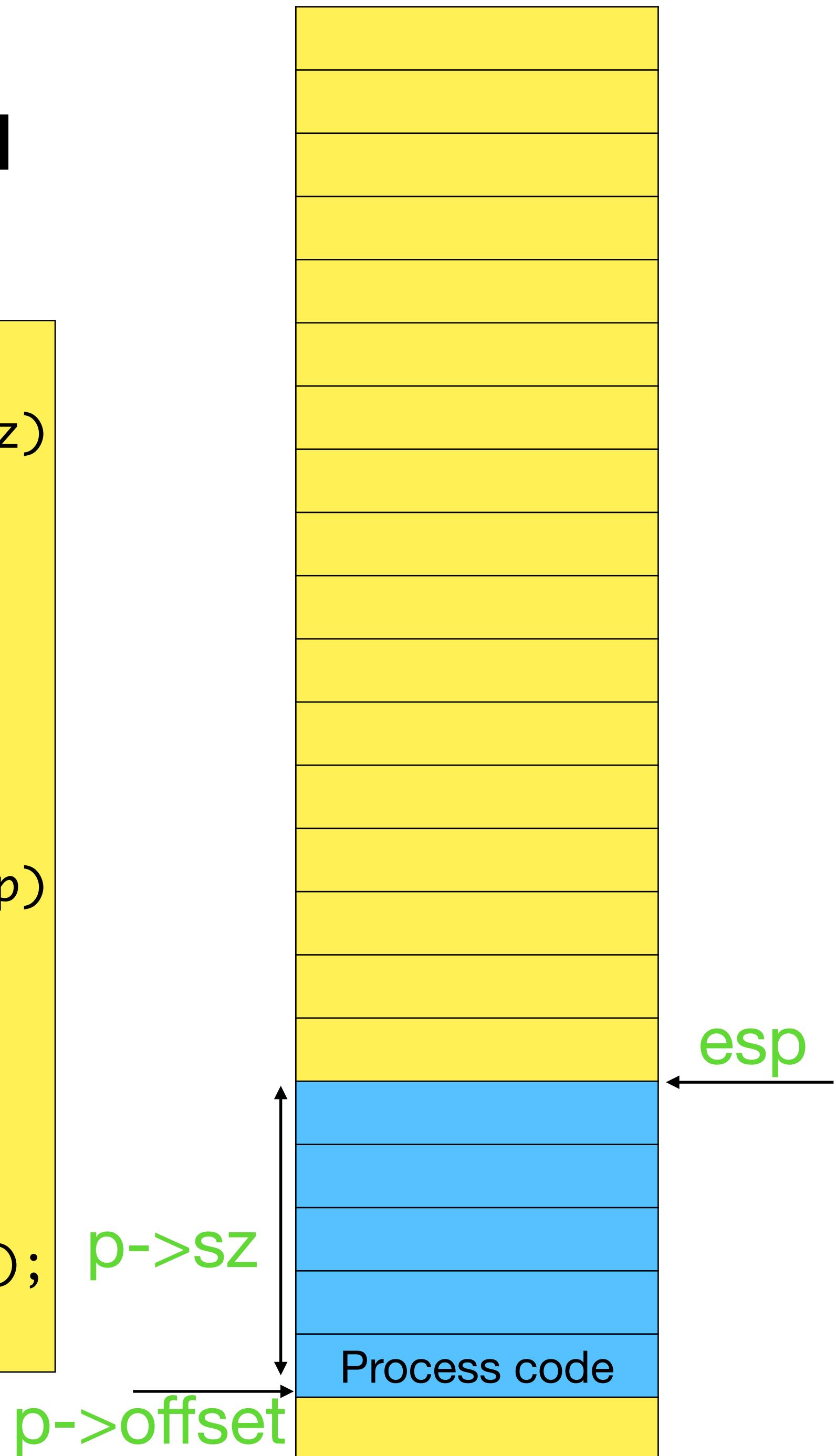
```
# sys_open("console", O_WRONLY)
eip    pushl $1
        pushl $console
        pushl $0
        movl $SYS_open, %eax
        int $T_SYSCALL
        pushl %eax
```

```
int sys_open(void) {
    int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```



Visualising syscall handling p19-syscall

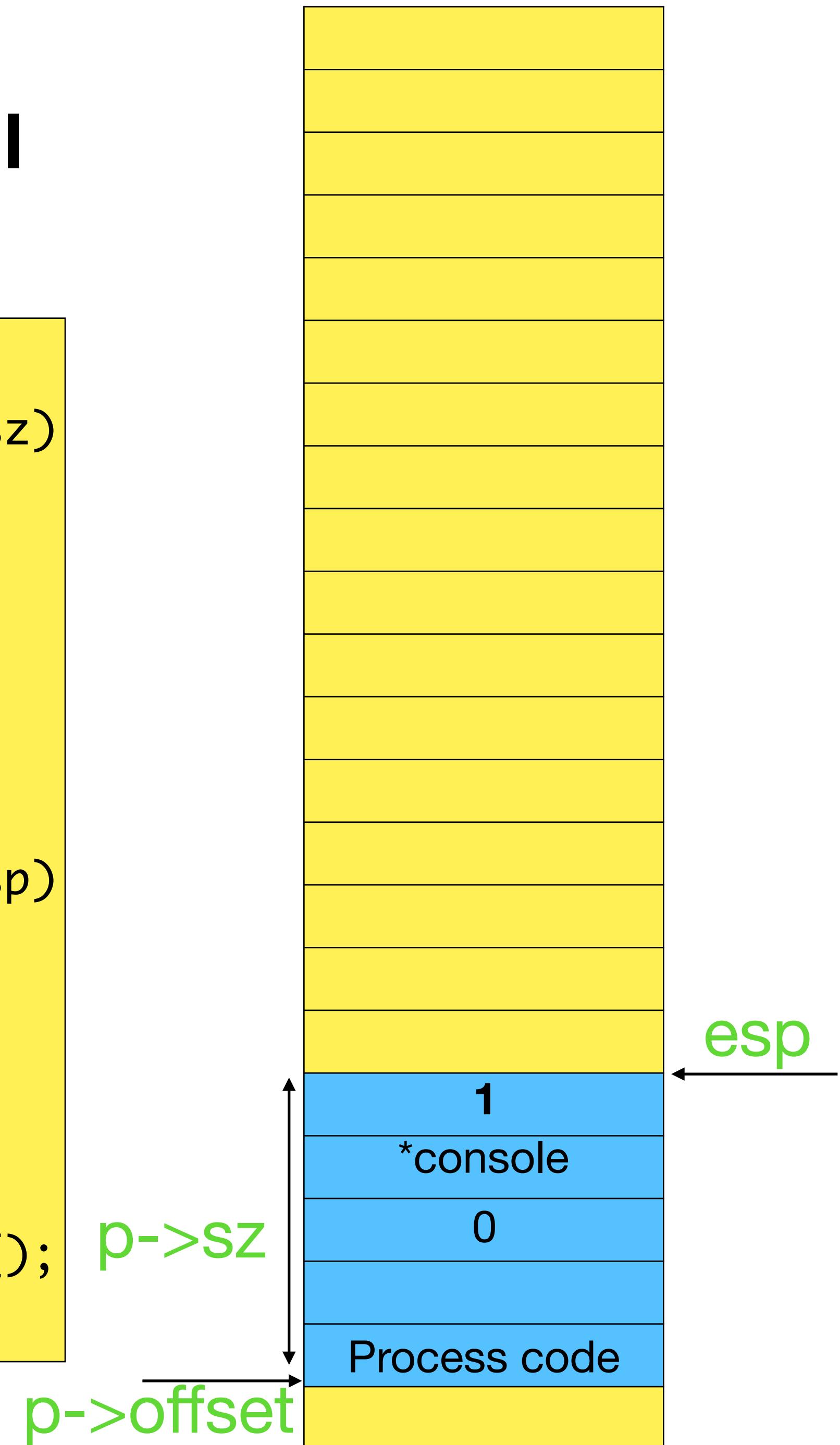
```
# sys_open("console", O_WRONLY)
eip    pushl $1
        pushl $console
        pushl $0
        movl $SYS_open, %eax
        int $T_SYSCALL
        pushl %eax
```

```
int sys_open(void) {
    int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```



Visualising syscall handling p19-syscall

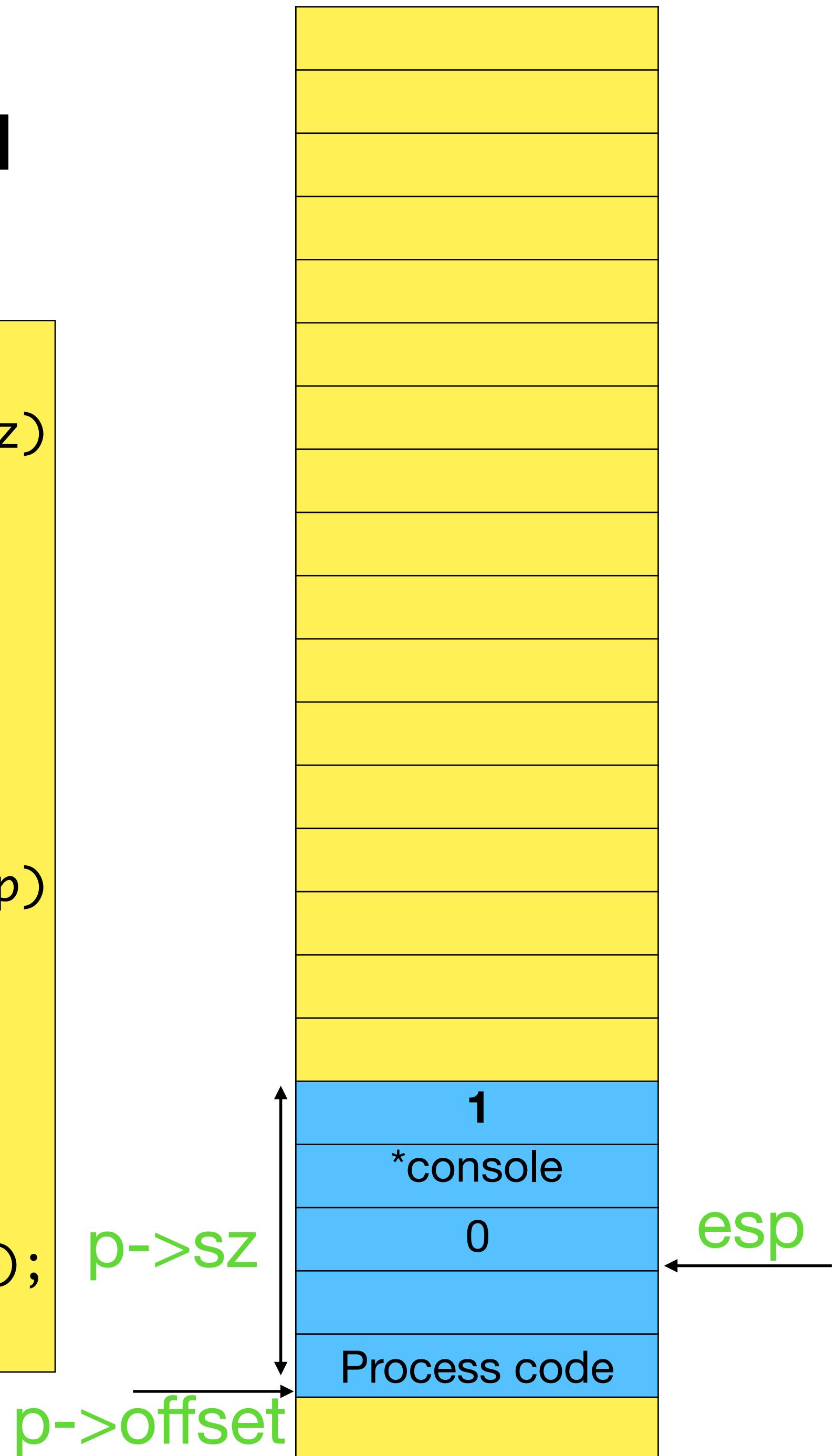
```
# sys_open("console", O_WRONLY)
eip    pushl $1
        pushl $console
        pushl $0
        movl $SYS_open, %eax
        int $T_SYSCALL
        pushl %eax
```

```
int sys_open(void) {
    int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```



Visualising syscall handling p19-syscall

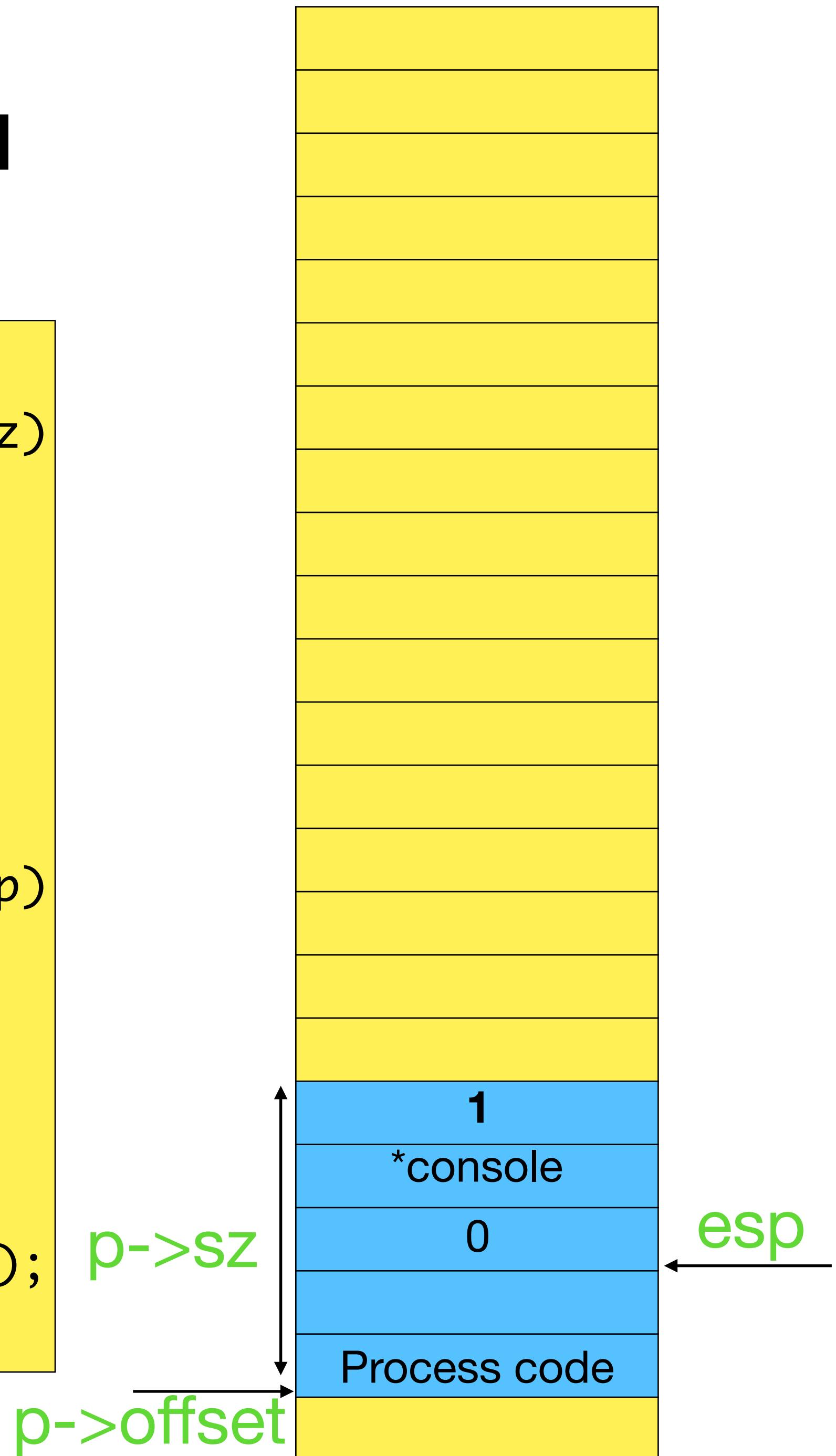
```
# sys_open("console", O_WRONLY)
    pushl $1
    pushl $console
    pushl $0
    eip → movl $SYS_open, %eax
    int $T_SYSCALL
    pushl %eax
```

```
int sys_open(void) {
    int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```



Visualising syscall handling p19-syscall

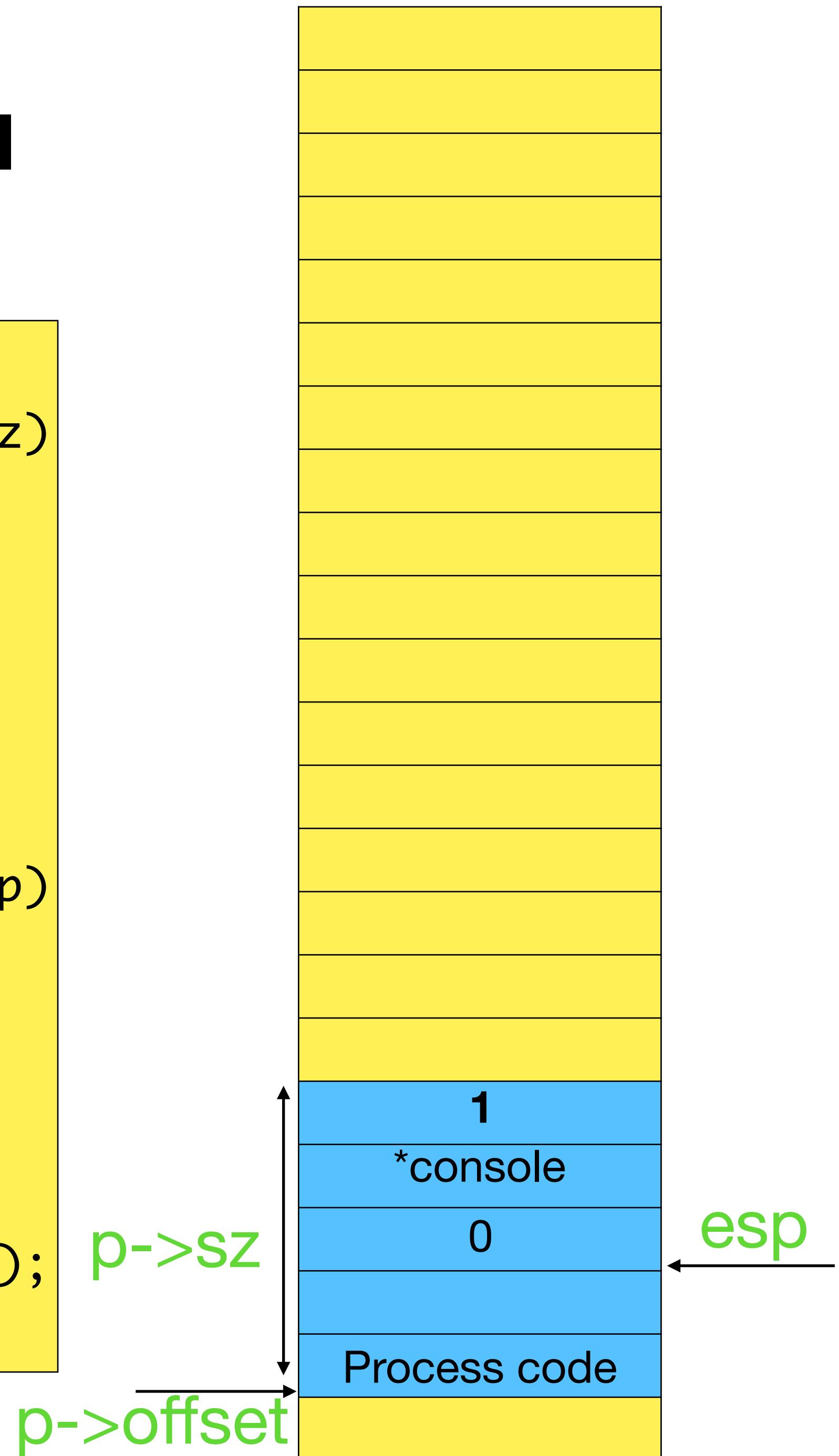
```
# sys_open("console", O_WRONLY)
    pushl $1
    pushl $console
    pushl $0
    movl $SYS_open, %eax
    eip → int $T_SYSCALL
    pushl %eax
```

```
int sys_open(void) {
    int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```



Visualising syscall handling p19-syscall

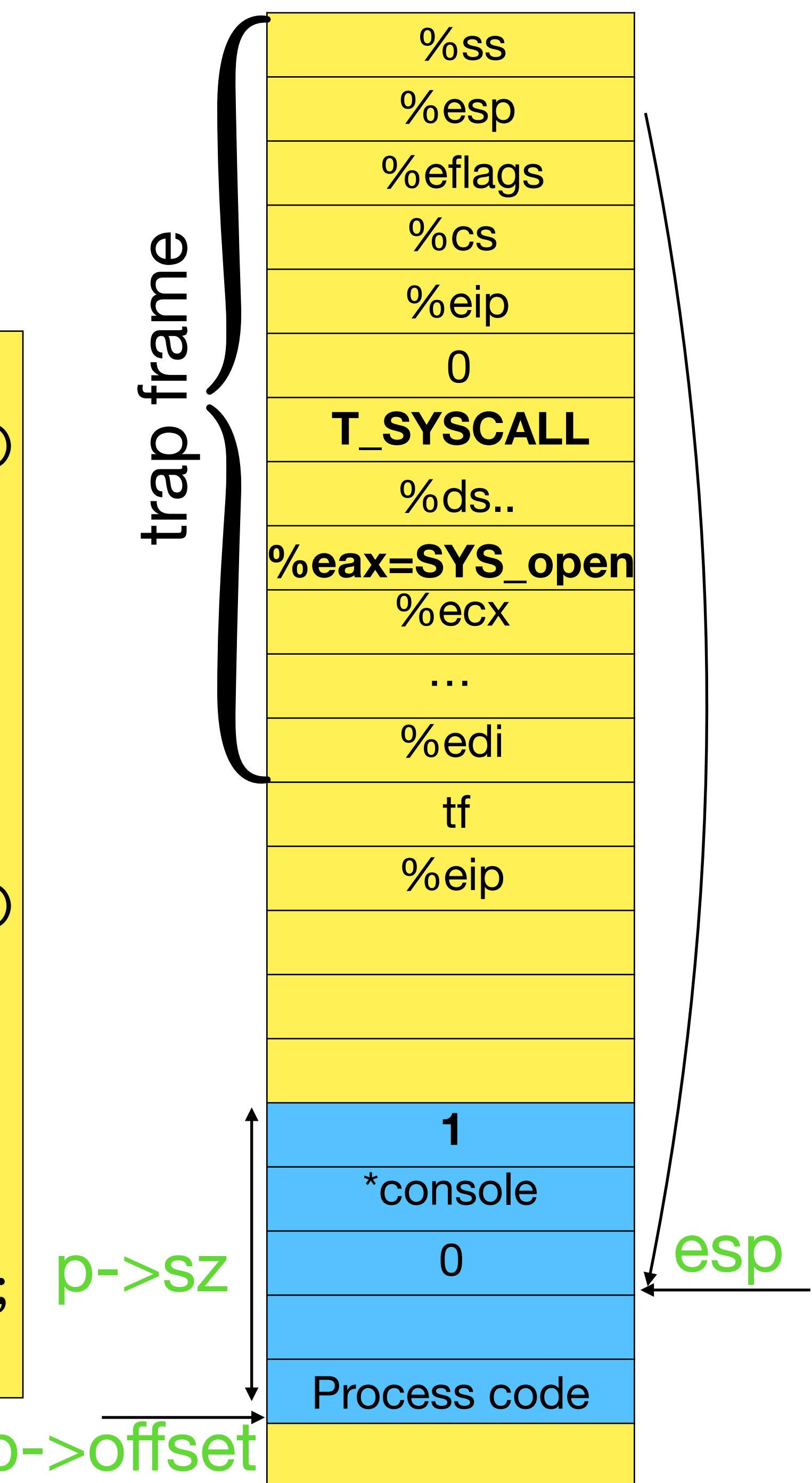
```
# sys_open("console", O_WRONLY)
    pushl $1
    pushl $console
    pushl $0
    movl $SYS_open, %eax
    eip → int $T_SYSCALL
    pushl %eax
```

```
int sys_open(void) {
    int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```



Visualising syscall handling

p19-syscall

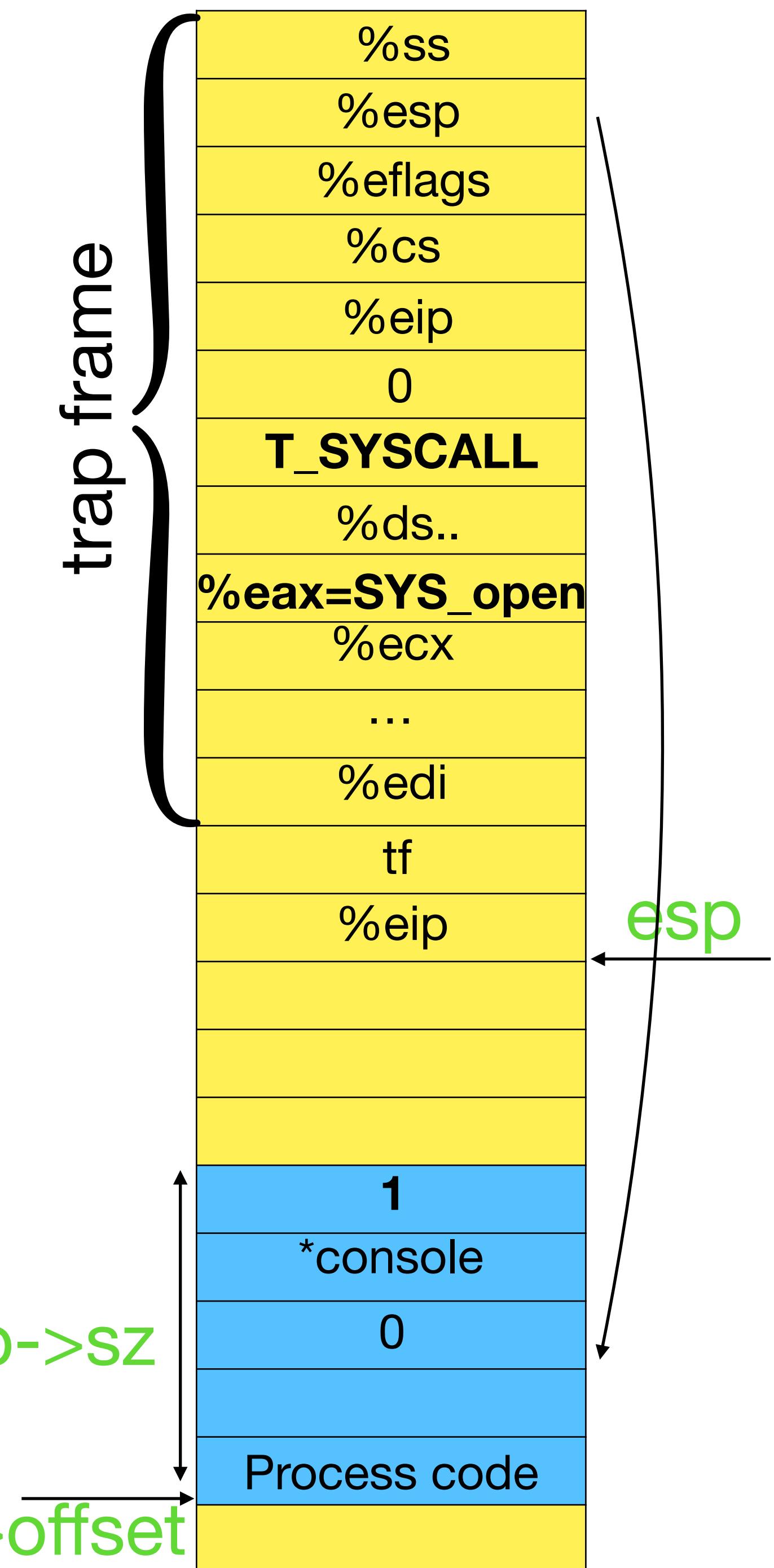
```
# sys_open("console", O_WRONLY)
    pushl $1
    pushl $console
    pushl $0
    movl $SYS_open, %eax
    eip → int $T_SYSCALL
    pushl %eax
```

```
int sys_open(void) {
    int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```



Visualising syscall handling

p19-syscall

```
# sys_open("console", O_WRONLY)
    pushl $1
    pushl $console
    pushl $0
    movl $SYS_open, %eax
    int $T_SYSCALL
    pushl %eax
```

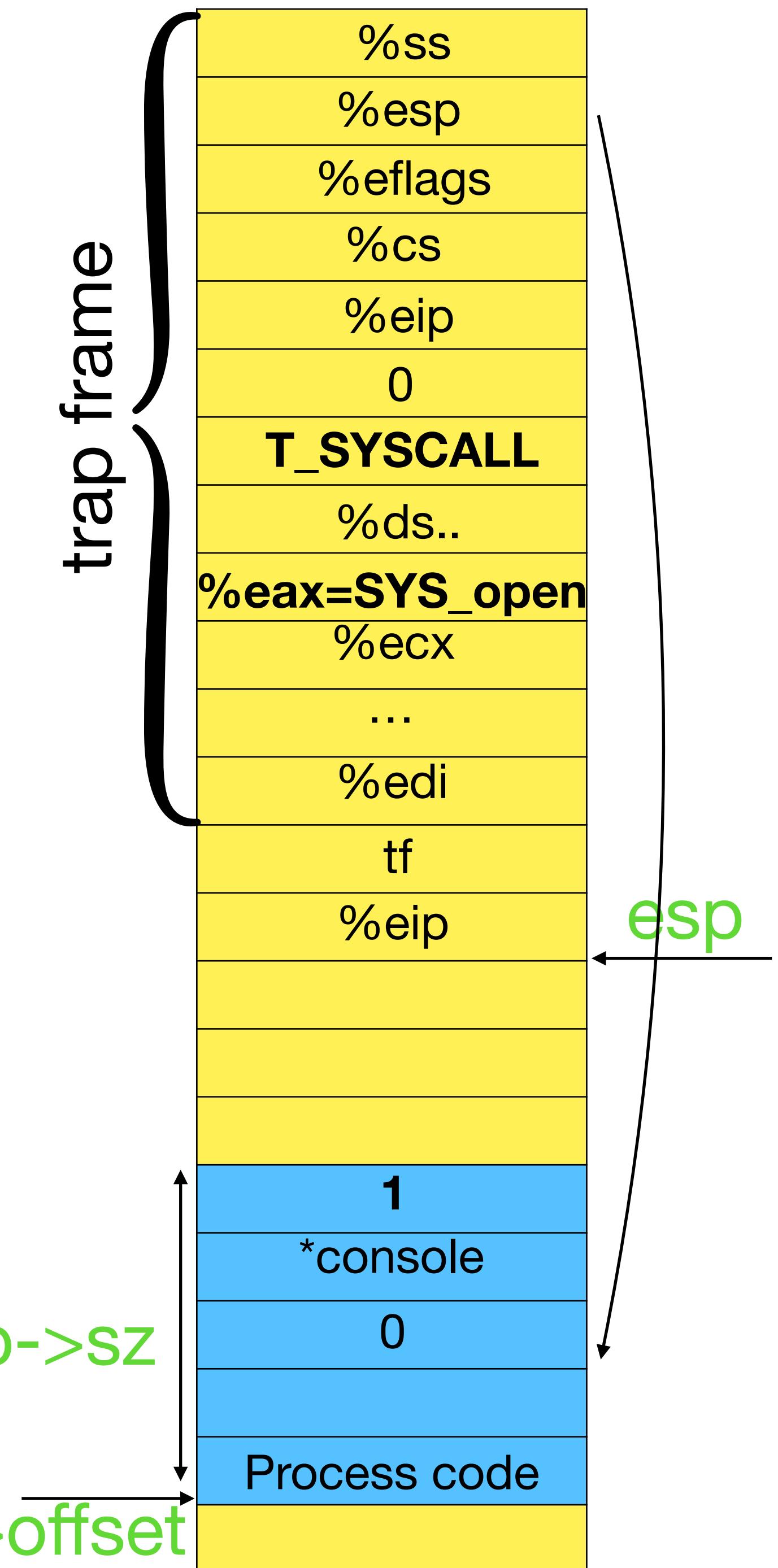
```
int sys_open(void) {
    int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```

eip



Visualising syscall handling

p19-syscall

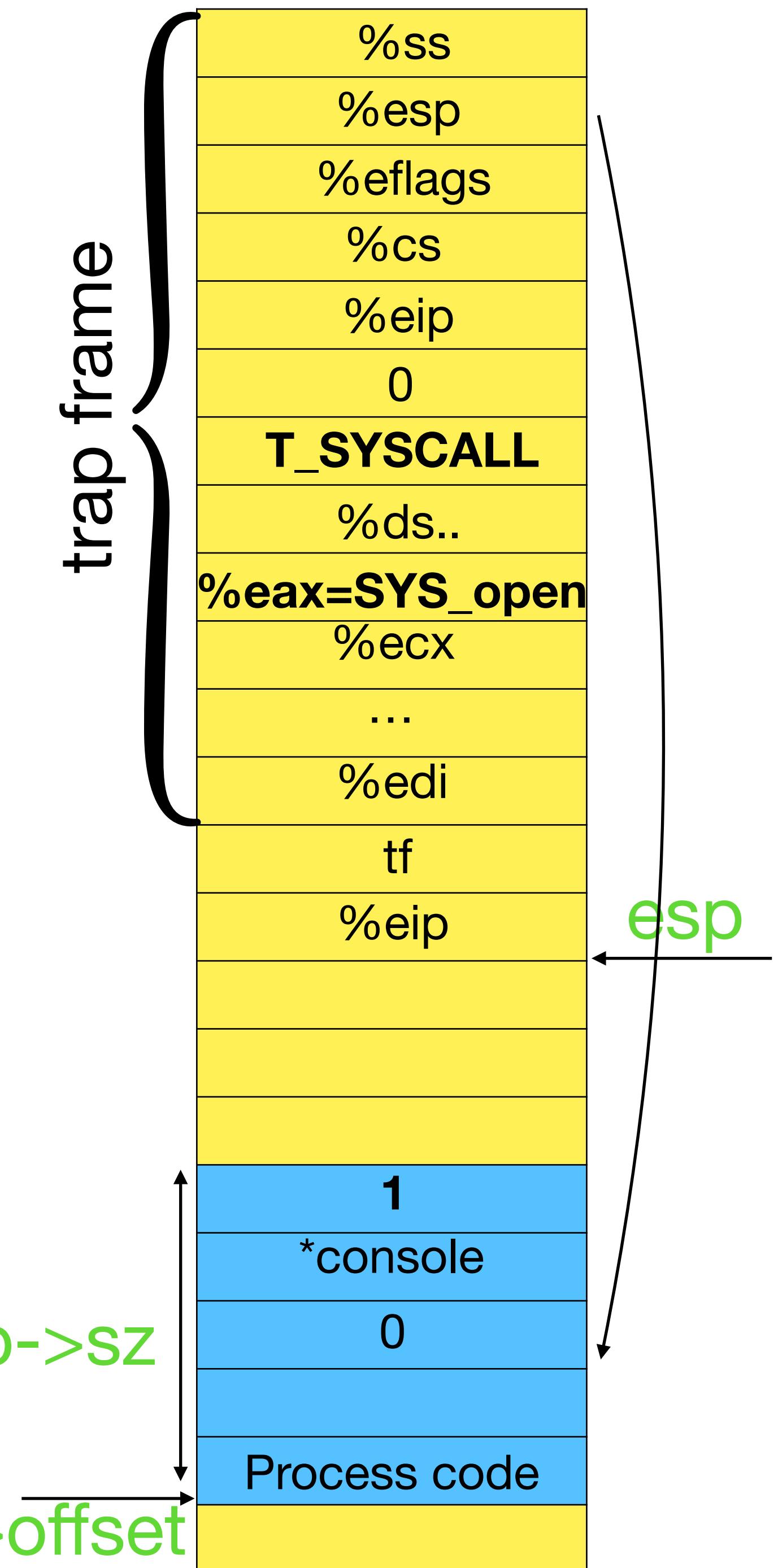
```
# sys_open("console", O_WRONLY)
    pushl $1
    pushl $console
    pushl $0
    movl $SYS_open, %eax
    int $T_SYSCALL
    pushl %eax
```

```
int sys_open(void) {
    eip → int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```



Visualising syscall handling

p19-syscall

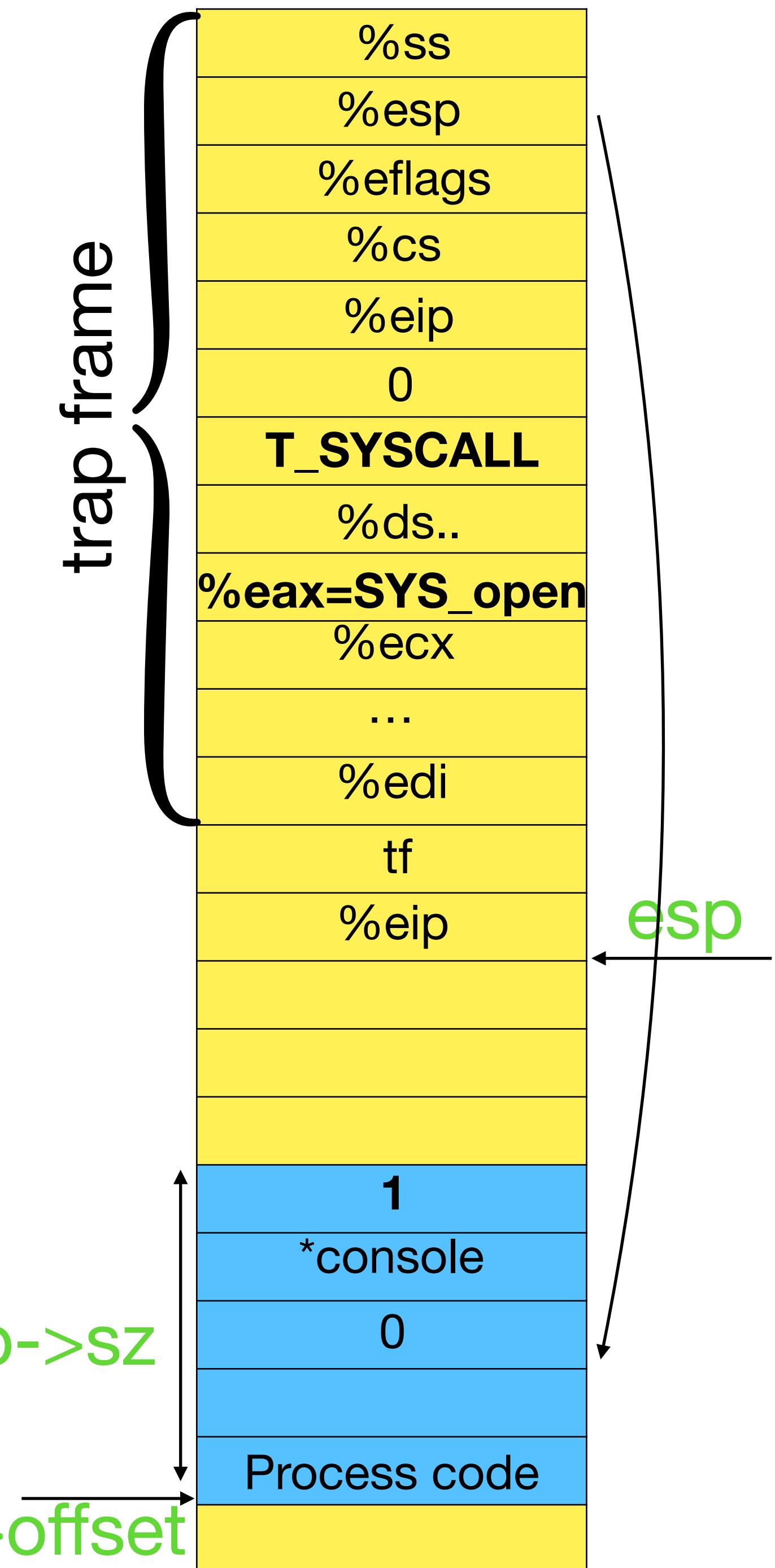
```
# sys_open("console", O_WRONLY)
    pushl $1
    pushl $console
    pushl $0
    movl $SYS_open, %eax
    int $T_SYSCALL
    pushl %eax
```

```
int sys_open(void) {
    eip int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```



Visualising syscall handling

p19-syscall

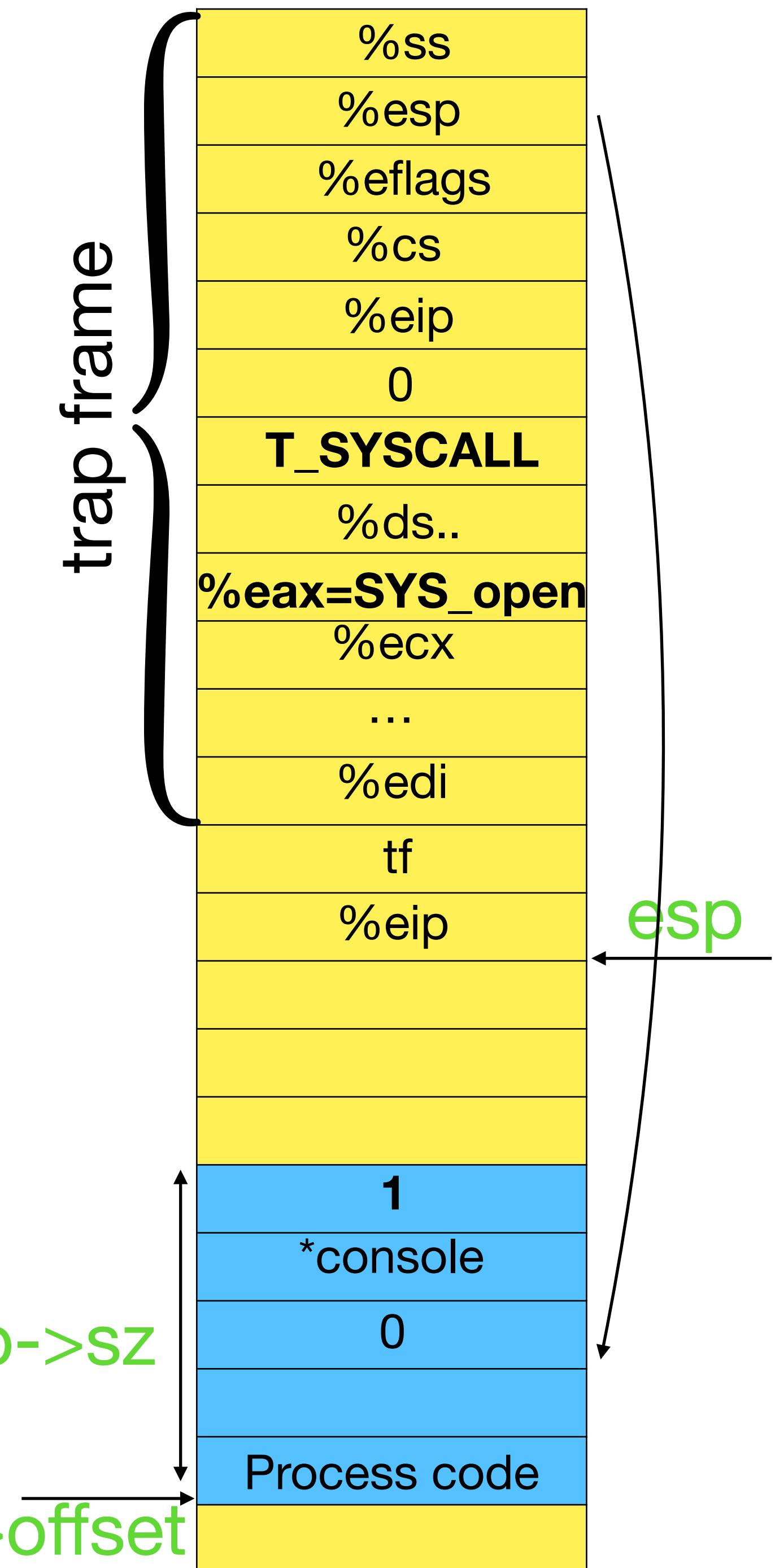
```
# sys_open("console", O_WRONLY)
    pushl $1
    pushl $console
    pushl $0
    movl $SYS_open, %eax
    int $T_SYSCALL
    pushl %eax
```

```
int sys_open(void) {
    int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```



Visualising syscall handling

p19-syscall

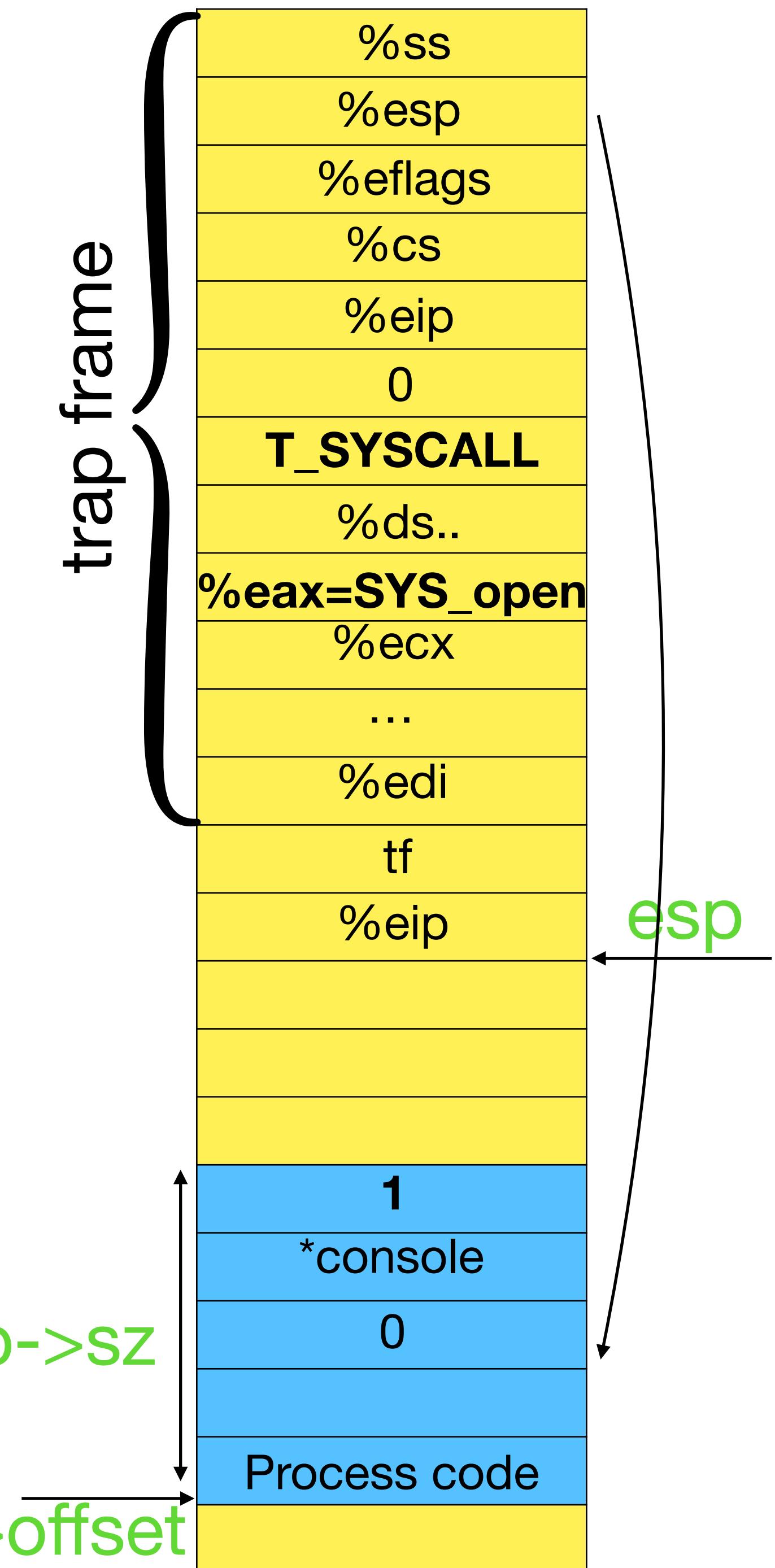
```
# sys_open("console", O_WRONLY)
    pushl $1
    pushl $console
    pushl $0
    movl $SYS_open, %eax
    int $T_SYSCALL
    pushl %eax
```

```
int sys_open(void) {
    int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```



Visualising syscall handling

p19-syscall

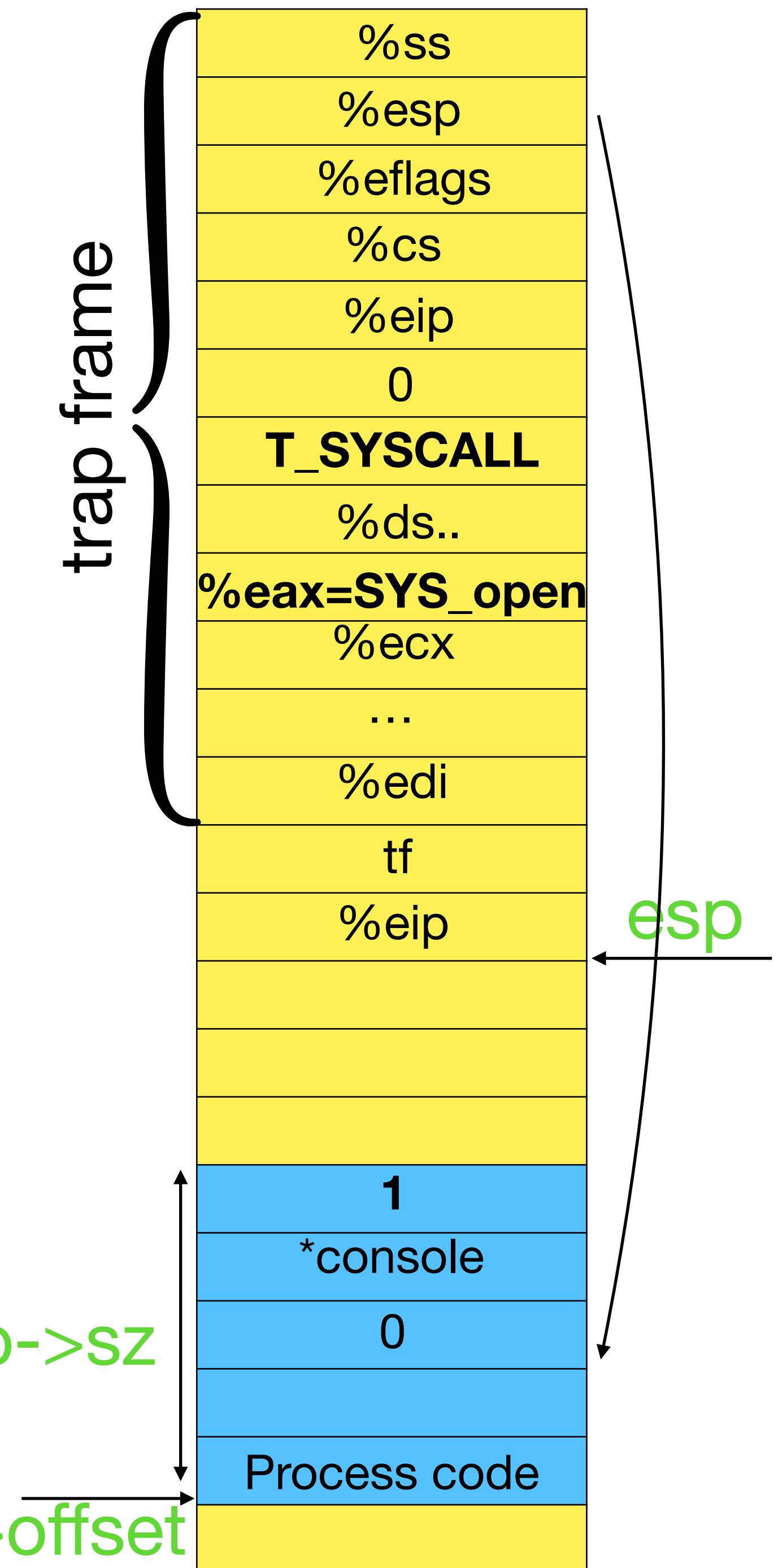
```
# sys_open("console", O_WRONLY)
    pushl $1
    pushl $console
    pushl $0
    movl $SYS_open, %eax
    int $T_SYSCALL
    pushl %eax
```

```
int sys_open(void) {
    int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```



Visualising syscall handling p19-syscall

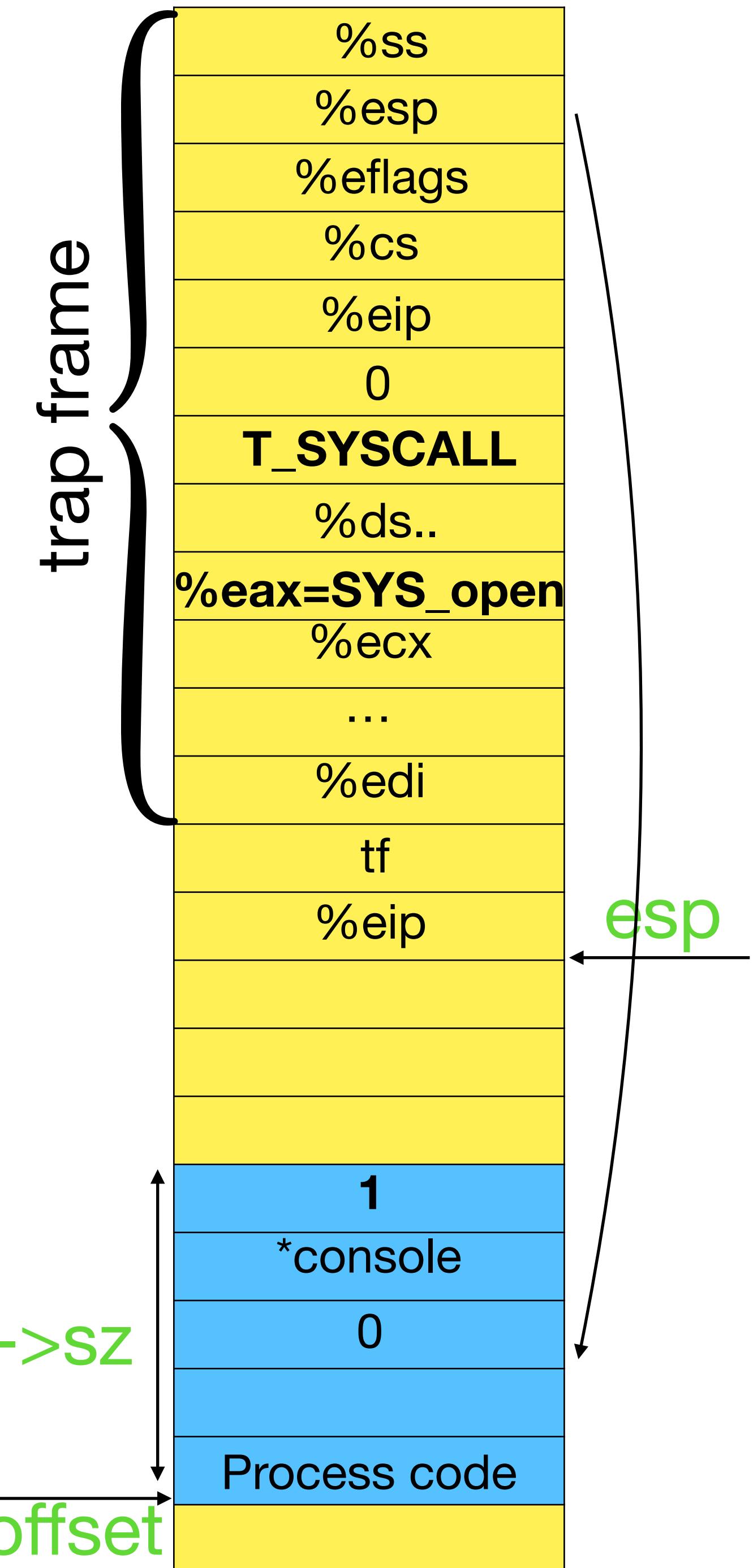
```
# sys_open("console", O_WRONLY)
    pushl $1
    pushl $console
    pushl $0
    movl $SYS_open, %eax
    int $T_SYSCALL
    pushl %eax
```

```
int sys_open(void) {
    int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```



Visualising syscall handling

p19-syscall

```
# sys_open("console", O_WRONLY)
    pushl $1
    pushl $console
    pushl $0
    movl $SYS_open, %eax
    int $T_SYSCALL
    pushl %eax
```

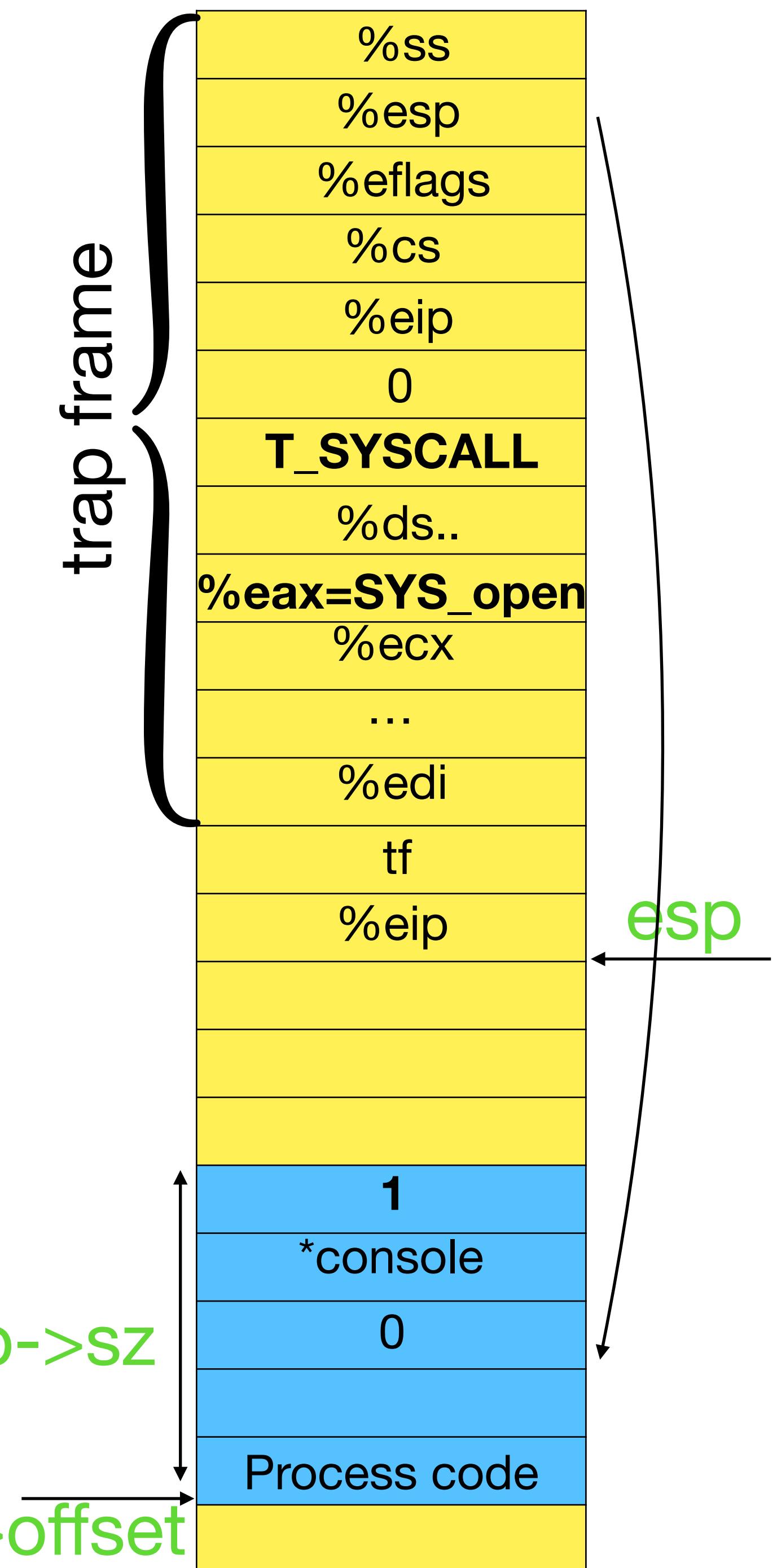
```
int sys_open(void) {
    int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```

eip



Visualising syscall handling

p19-syscall

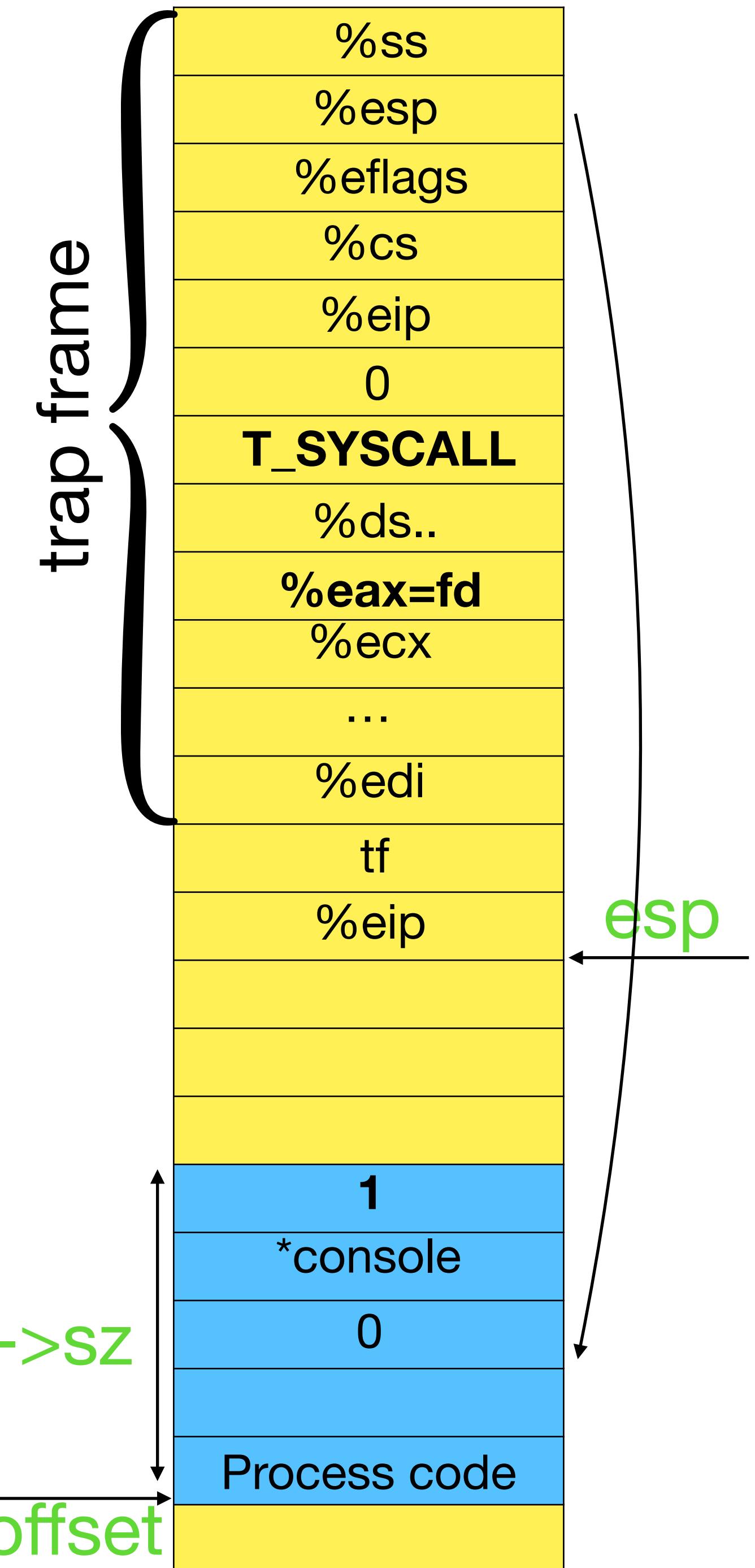
```
# sys_open("console", O_WRONLY)
    pushl $1
    pushl $console
    pushl $0
    movl $SYS_open, %eax
    int $T_SYSCALL
    pushl %eax
```

```
int sys_open(void) {
    int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```



Visualising syscall handling

p19-syscall

```
# sys_open("console", O_WRONLY)
    pushl $1
    pushl $console
    pushl $0
    movl $SYS_open, %eax
    int $T_SYSCALL
    pushl %eax
```

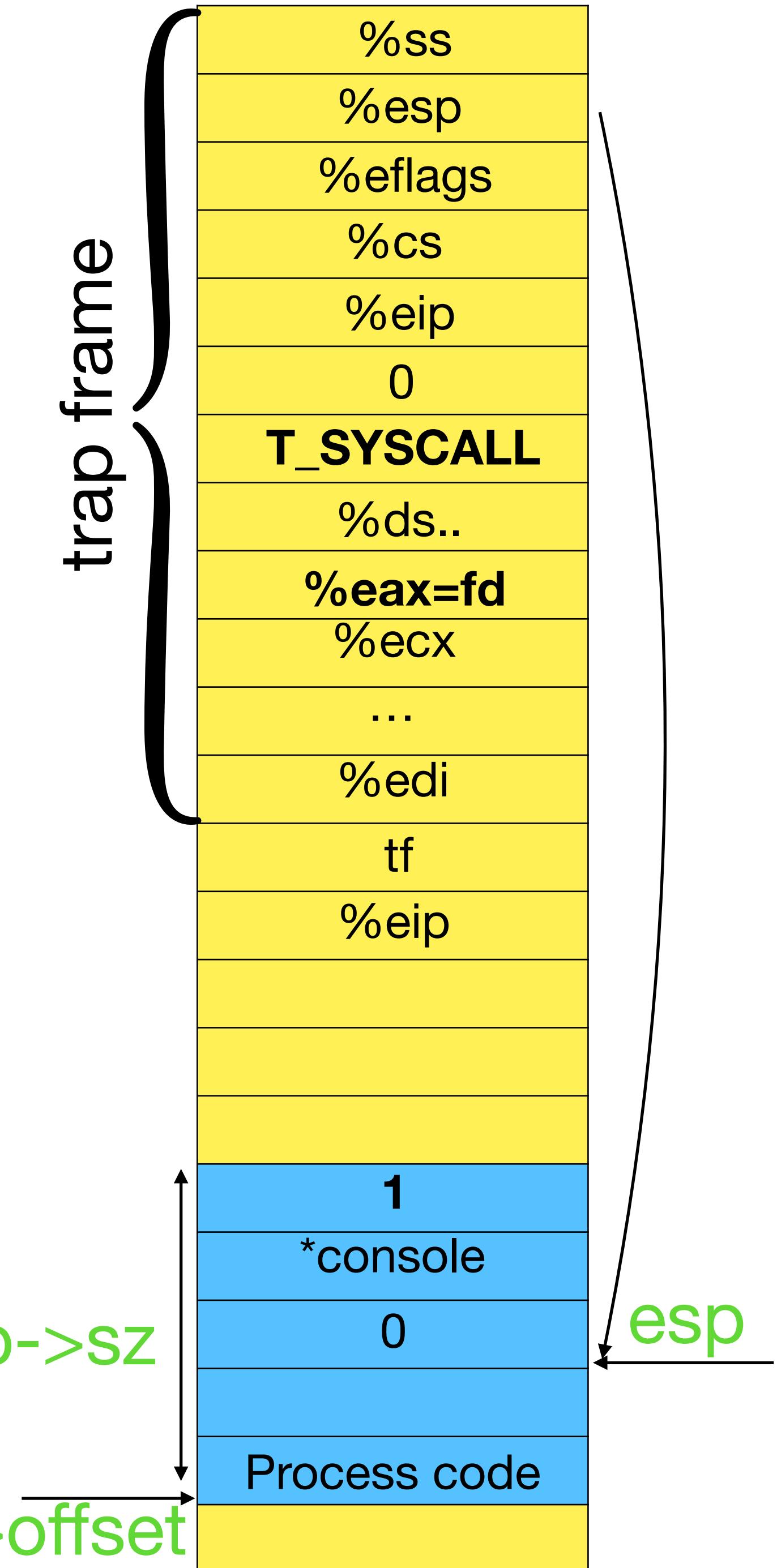
```
int sys_open(void) {
    int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```

eip



p->offset

esp

Visualising syscall handling p19-syscall

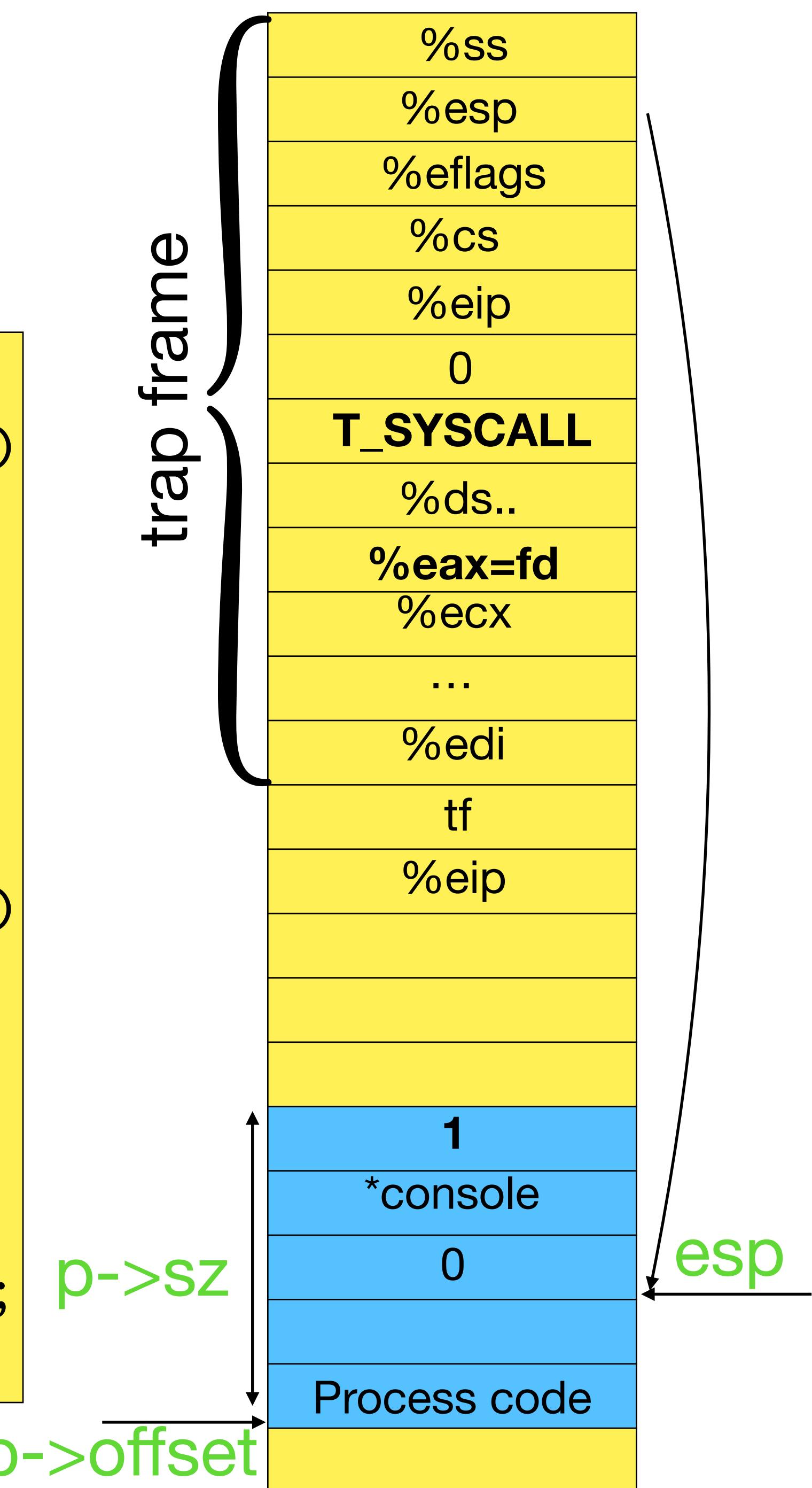
```
# sys_open("console", O_WRONLY)
    pushl $1
    pushl $console
    pushl $0
    movl $SYS_open, %eax
    int $T_SYSCALL
    eip → pushl %eax
```

```
int sys_open(void) {
    int fd, omode;
    if(argint(1, &omode) < 0) {
        return -1;
    }
    ...
    return fd;
}
```

```
int fetchint(uint addr, int *ip) {
    if(addr >= p->sz || addr+4 > p->sz)
        return -1;
    *ip = *(int*)(addr + p->offset);
}

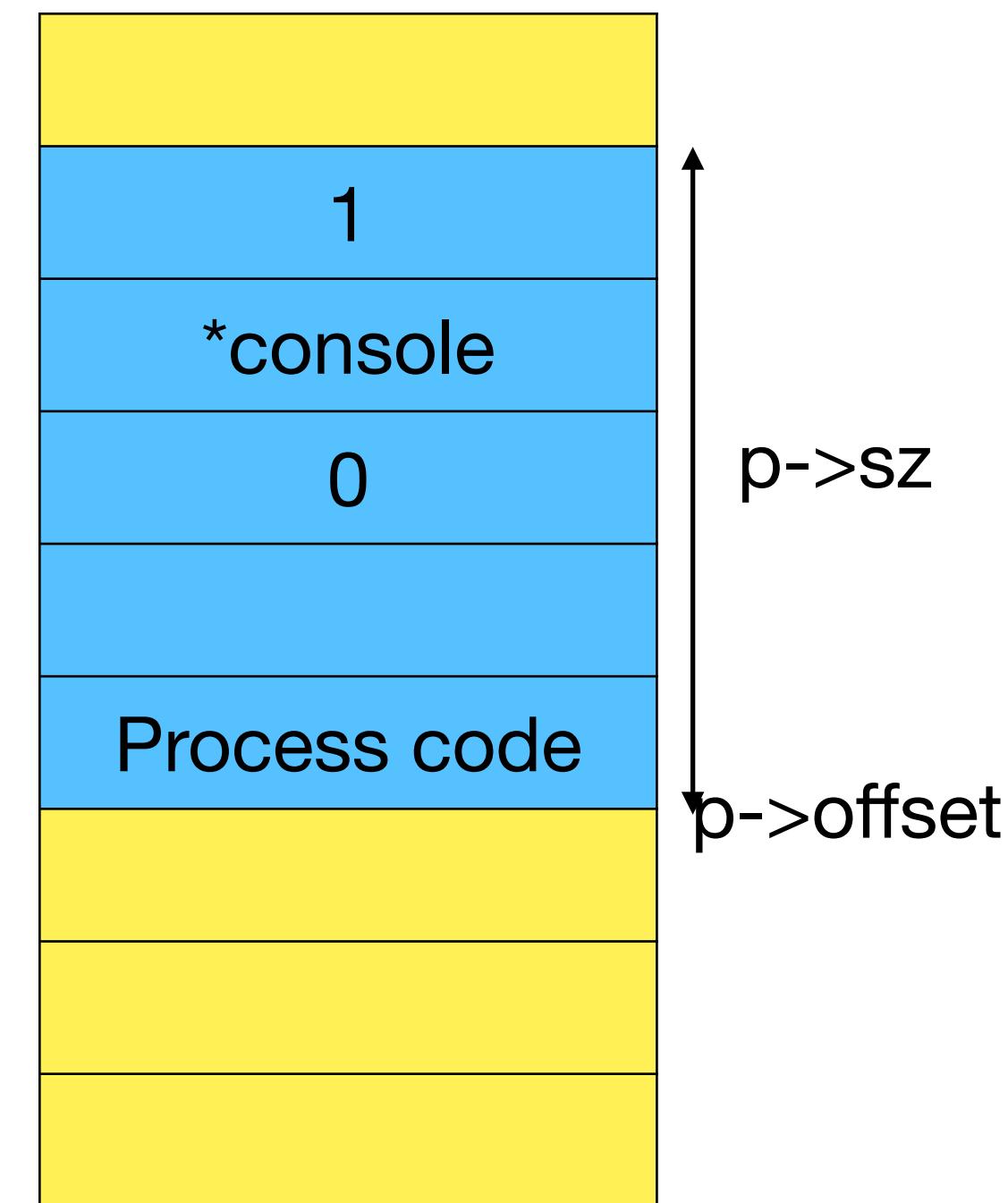
int argint(int n, int *ip) {
    return fetchint((myproc()->tf->esp)
                    + 4 + 4*n, ip);
}

void syscall(void) {
    int num = curproc->tf->eax;
    curproc->tf->eax = syscalls[num]();
}
```



Syscall parameter checking and translation

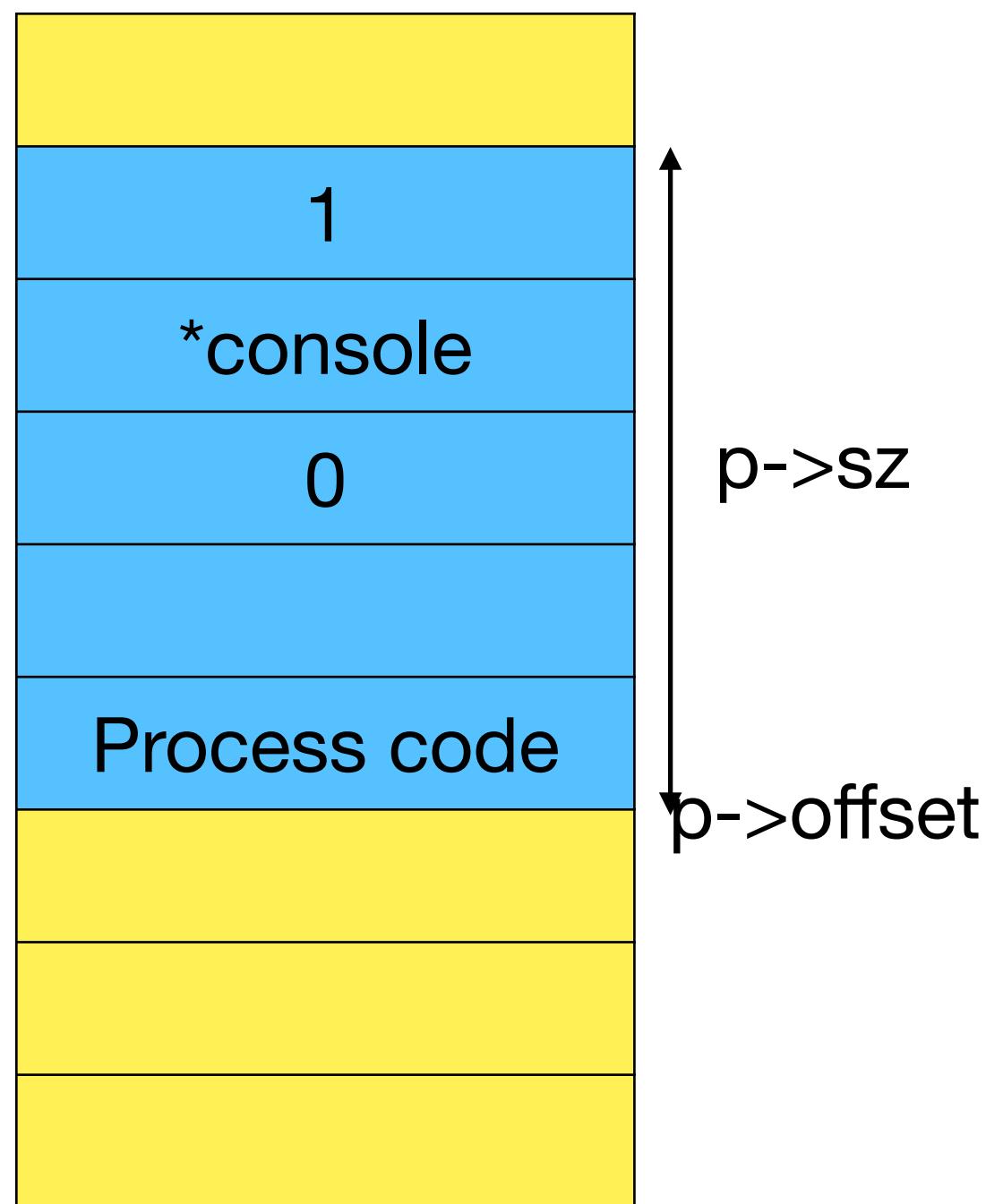
```
int fetchint(uint addr, int *ip) {  
    if(addr >= p->sz || addr+4 > p->sz)  
        return -1;  
    *ip = *(int*)(addr + p->offset);  
}
```



Syscall parameter checking and translation

```
int fetchint(uint addr, int *ip) {  
    if(addr >= p->sz || addr+4 > p->sz)  
        return -1;  
    *ip = *(int*)(addr + p->offset);  
}
```

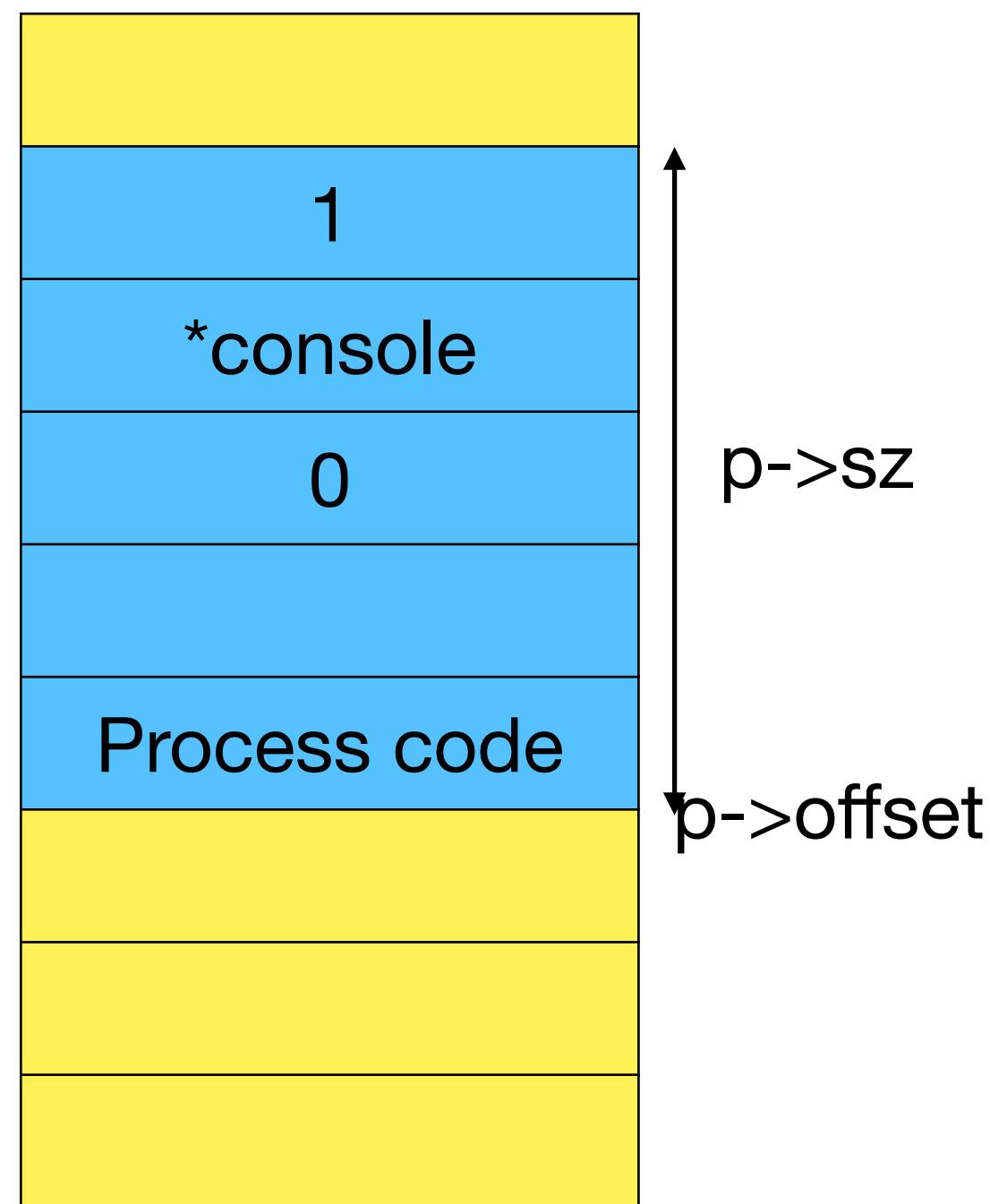
- Syscall parameters must be in process' address space:



Syscall parameter checking and translation

```
int fetchint(uint addr, int *ip) {  
    if(addr >= p->sz || addr+4 > p->sz)  
        return -1;  
    *ip = *(int*)(addr + p->offset);  
}
```

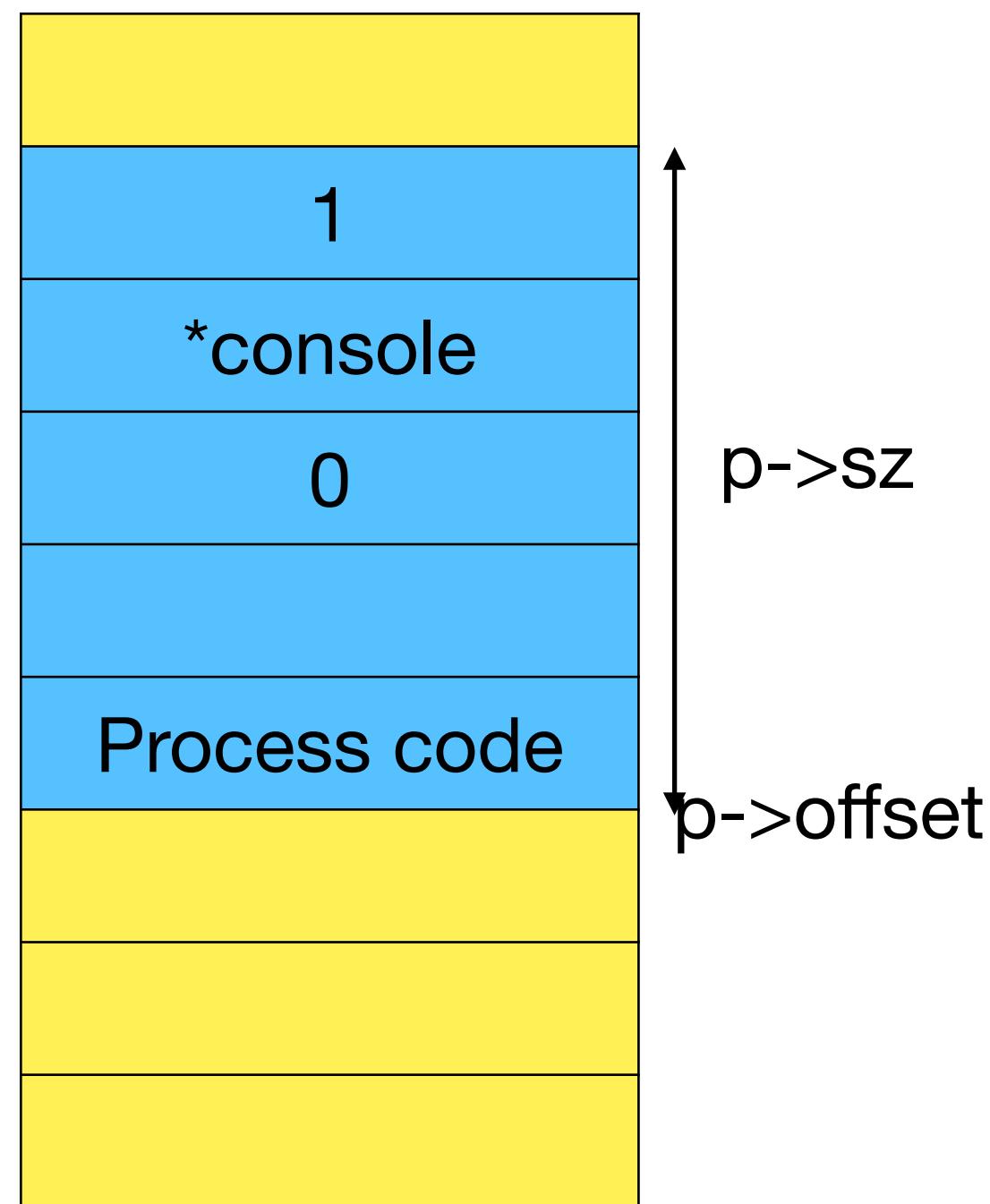
- Syscall parameters must be in process' address space:
 - Check virtual address (VA) is within limit



Syscall parameter checking and translation

```
int fetchint(uint addr, int *ip) {  
    if(addr >= p->sz || addr+4 > p->sz)  
        return -1;  
    *ip = *(int*)(addr + p->offset);  
}
```

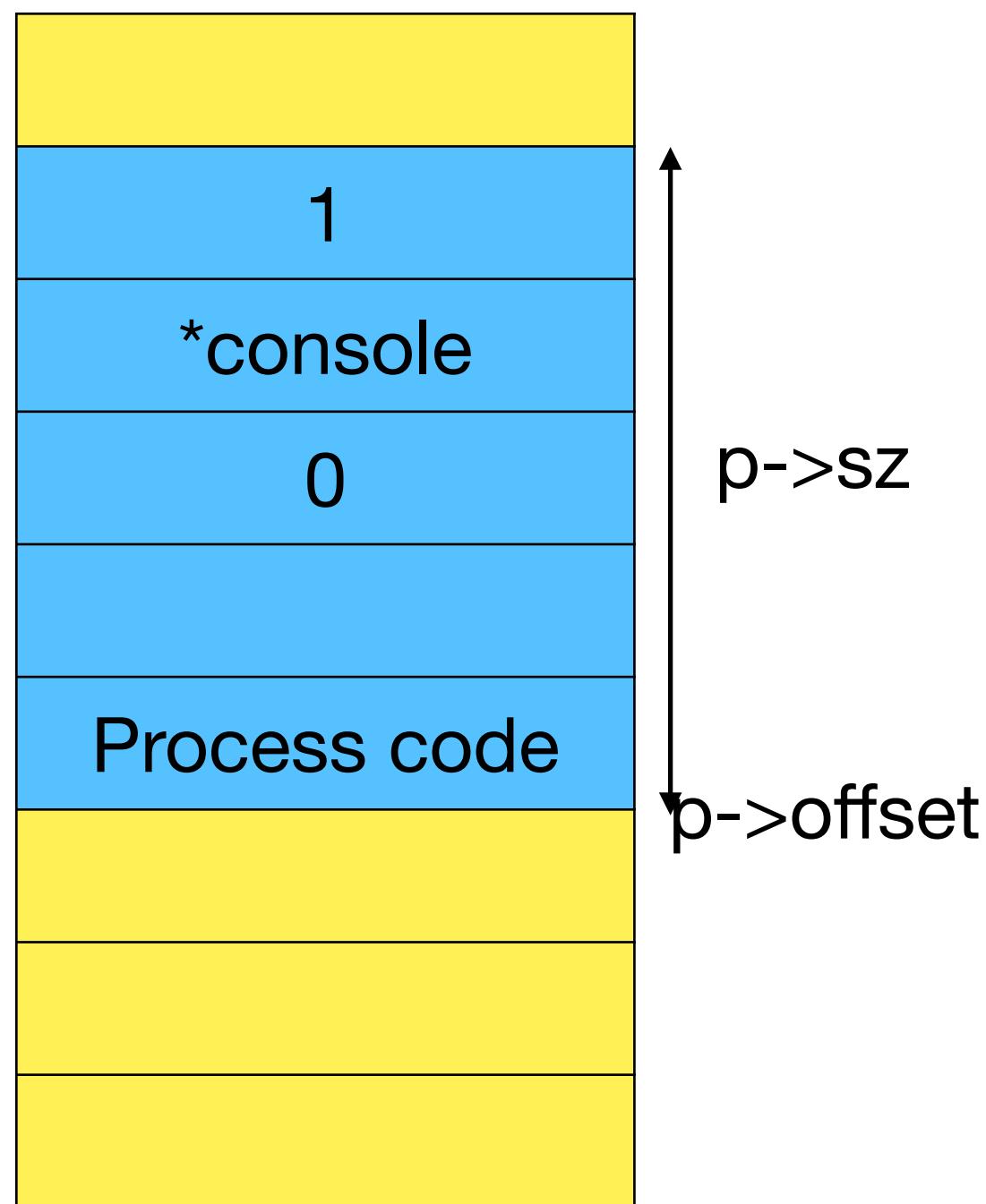
- Syscall parameters must be in process' address space:
 - Check virtual address (VA) is within limit
 - Add base to VA to get physical address (PA)



Syscall parameter checking and translation

```
int fetchint(uint addr, int *ip) {  
    if(addr >= p->sz || addr+4 > p->sz)  
        return -1;  
    *ip = *(int*)(addr + p->offset);  
}
```

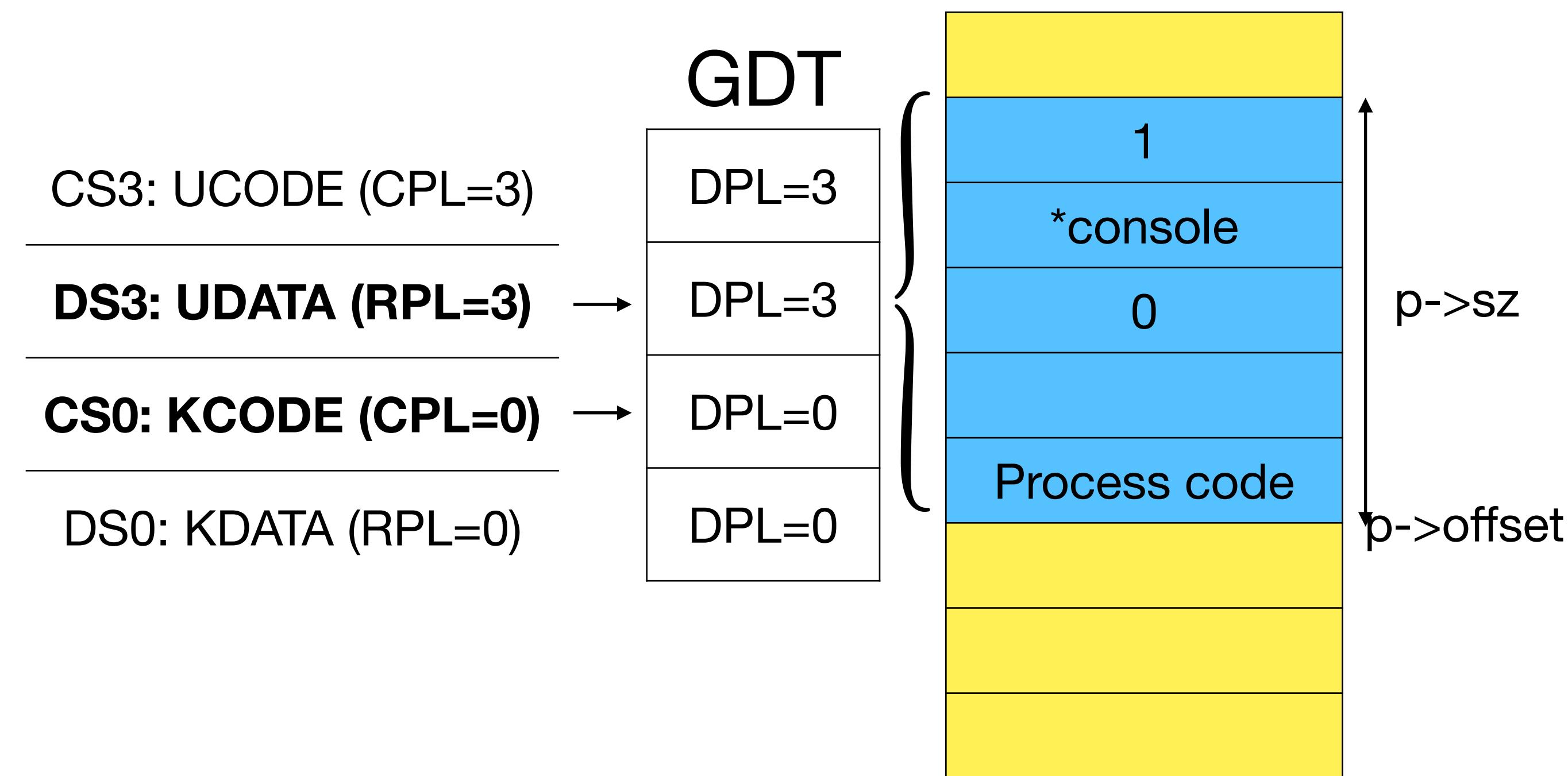
- Syscall parameters must be in process' address space:
 - Check virtual address (VA) is within limit
 - Add base to VA to get physical address (PA)
- Alternative design: use process' DS to read syscall parameters



Syscall parameter checking and translation

```
int fetchint(uint addr, int *ip) {  
    if(addr >= p->sz || addr+4 > p->sz)  
        return -1;  
  
    *ip = *(int*)(addr + p->offset);  
}
```

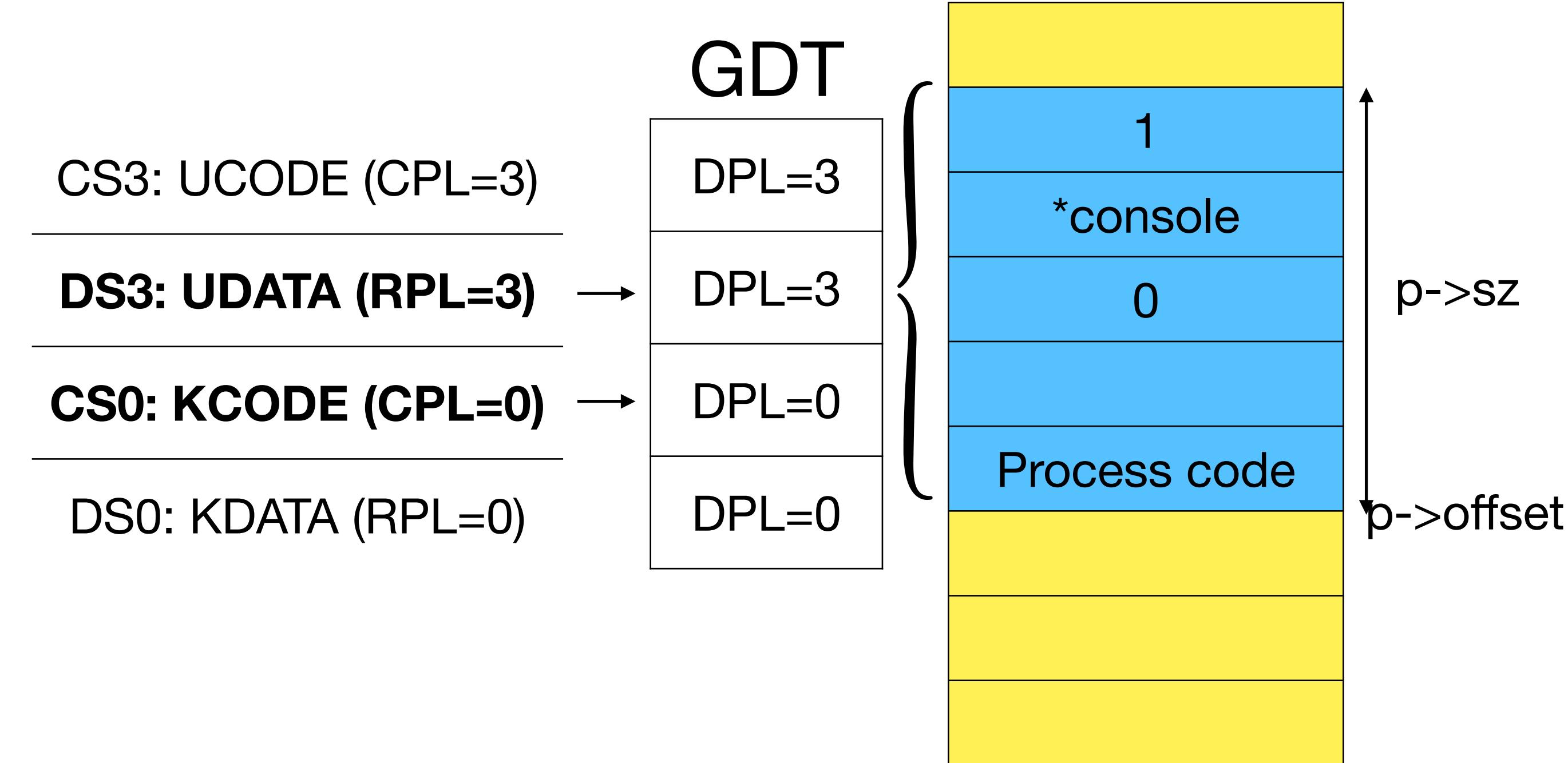
- Syscall parameters must be in process' address space:
 - Check virtual address (VA) is within limit
 - Add base to VA to get physical address (PA)
- Alternative design: use process' DS to read syscall parameters



Syscall parameter checking and translation

```
int fetchint(uint addr, int *ip) {  
    if(addr >= p->sz || addr+4 > p->sz)  
        return -1;  
  
    *ip = *(int*)(addr + p->offset);  
}
```

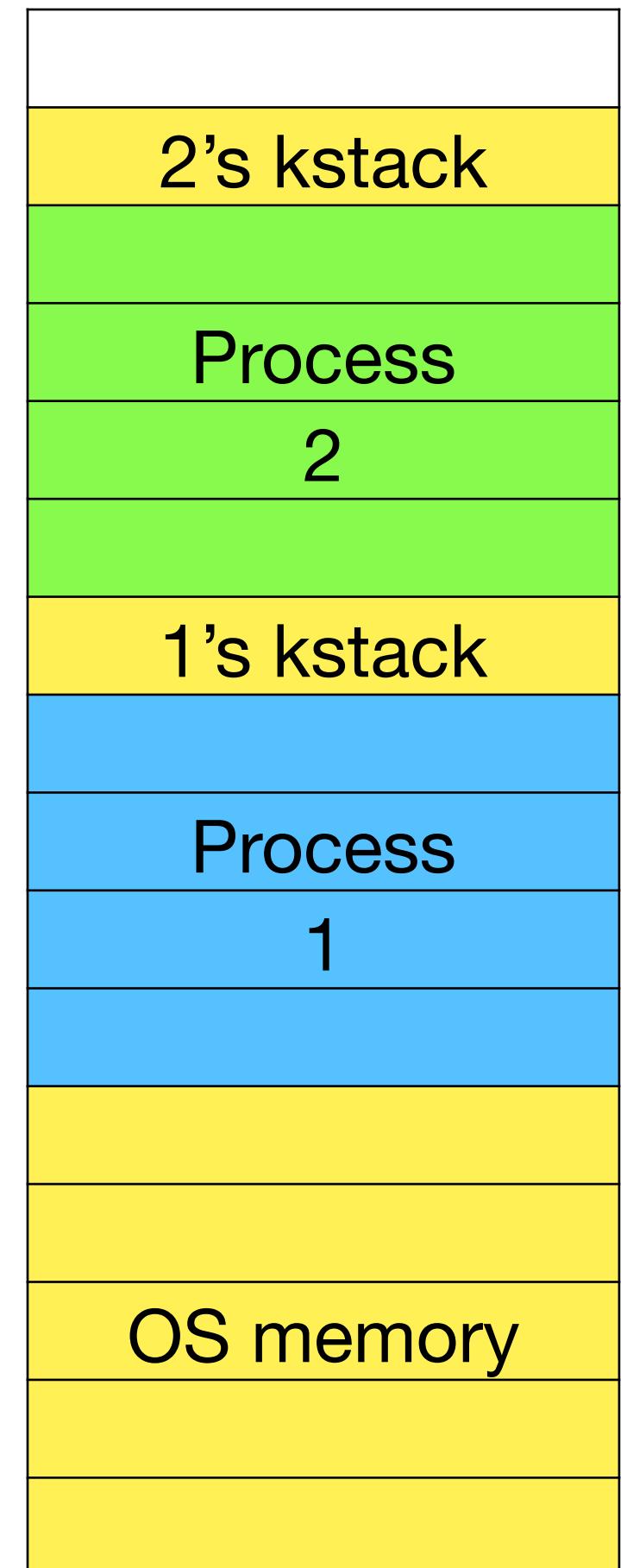
- Syscall parameters must be in process' address space:
 - Check virtual address (VA) is within limit
 - Add base to VA to get physical address (PA)
- Alternative design: use process' DS to read syscall parameters
 - CPL <= DPL and RPL <= DPL



Code walkthrough: Setting up multiple processes

p18-sched

- main.c
 - Calls pinit twice to start two processes and then calls scheduler
- proc.c
 - pinit calls allocproc which sets up processes like earlier with the added difference that it calls kalloc to find empty 1MB of space
- kalloc.c
 - Maintains free list in chunks of 1MB. kalloc returns the first element from free list. There is no coalescing and splitting since our OS always asks for 1MB.



Code walkthrough: Running multiple processes

p18-sched

- main.c
 - Calls scheduler to run the first RUNNABLE process. When giving control, we also remember “scheduler context” to come back to scheduler. Earlier, we were only running one process and were never coming back to the scheduler.
- trap.c
 - Calls yield when timer interrupt happens
- proc.c
 - yield calls sched which switches control from the process to the scheduler
 - scheduler looks for the next RUNNABLE process and gives control to it

Context switching in action: giving control

p18-sched

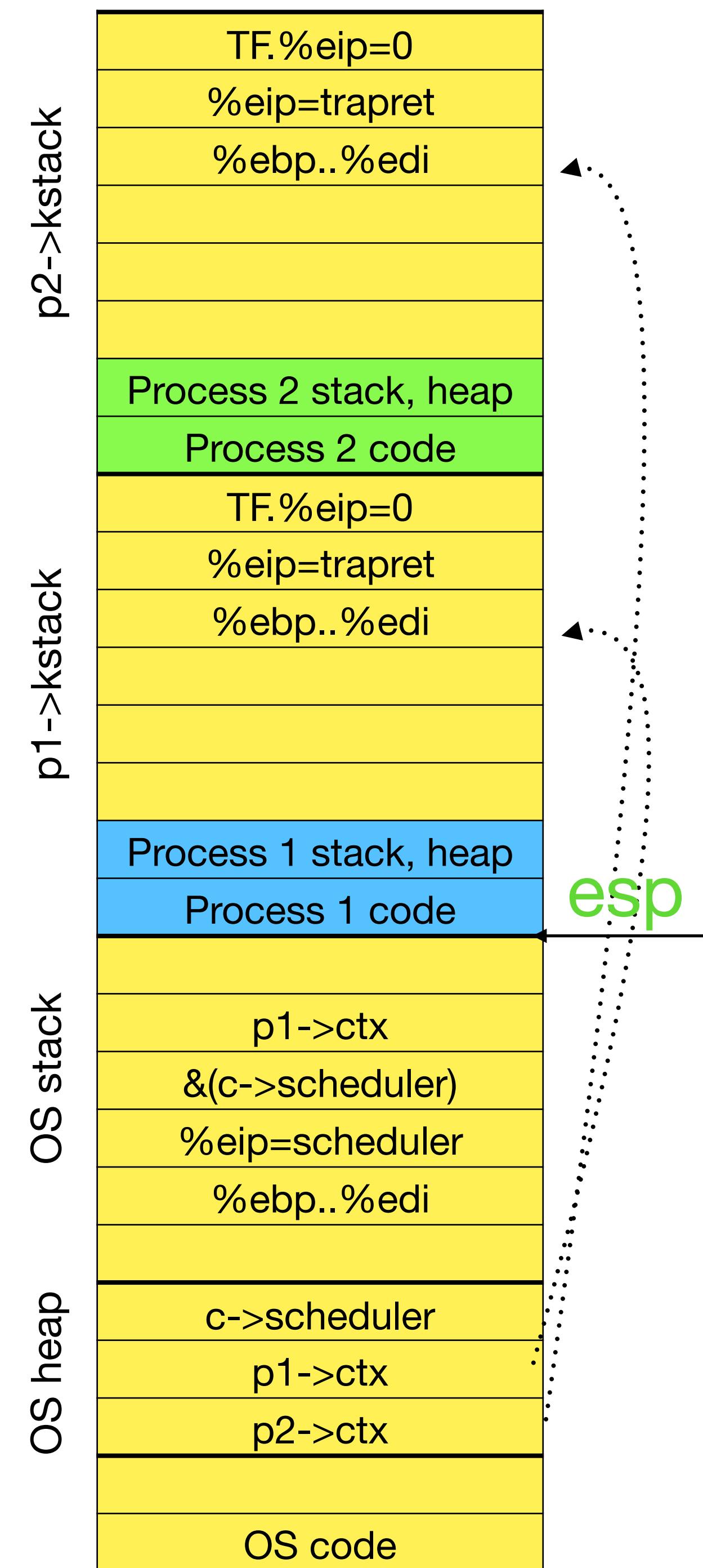
```

.globl swtch
swtch:           eip void scheduler(void) {
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret
                    }

void yield(void) {
    struct proc *p = myproc();
    p->state = RUNNABLE;
    swtch(&p->context, c->scheduler);
}

void trap(struct trapframe *tf) {
    ...
    if(tf->trapno == T_IRQ0+IRQ_TIMER)
        yield();
}

```



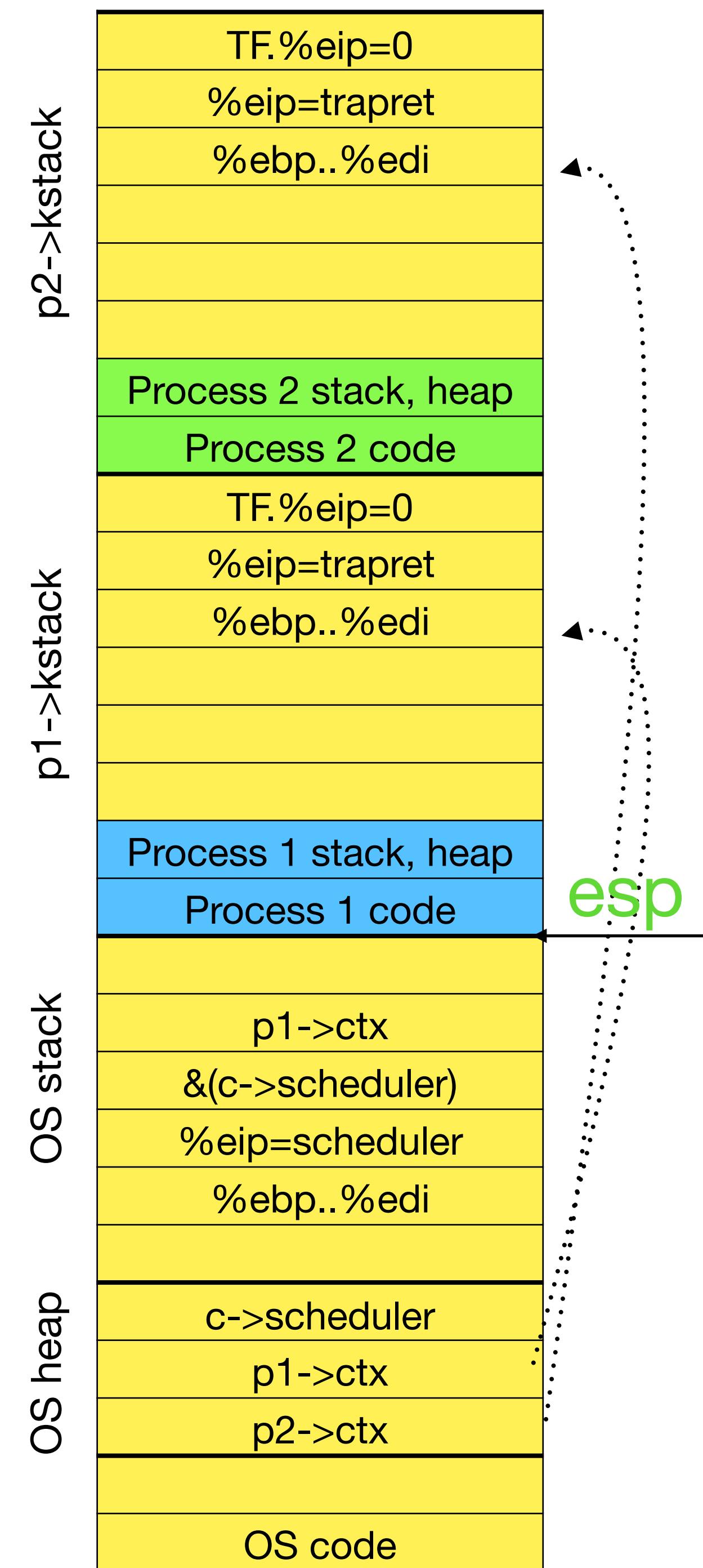
Context switching in action: giving control

p18-sched

```

.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret
        void scheduler(void) {
            struct proc *p; struct cpu *c = mycpu();
            for(;;){
                for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                    if(p->state != RUNNABLE)
                        continue;
                    p->state = RUNNING;
                    switchuvm(p);
                    swtch(&(c->scheduler), p->context);
                }
            }
        }
        void yield(void) {
            struct proc *p = myproc();
            p->state = RUNNABLE;
            swtch(&p->context, c->scheduler);
        }
        void trap(struct trapframe *tf) {
            ..
            if(tf->trapno == T_IRQ0+IRQ_TIMER)
                yield();
        }
    }
}

```



Context switching in action: giving control

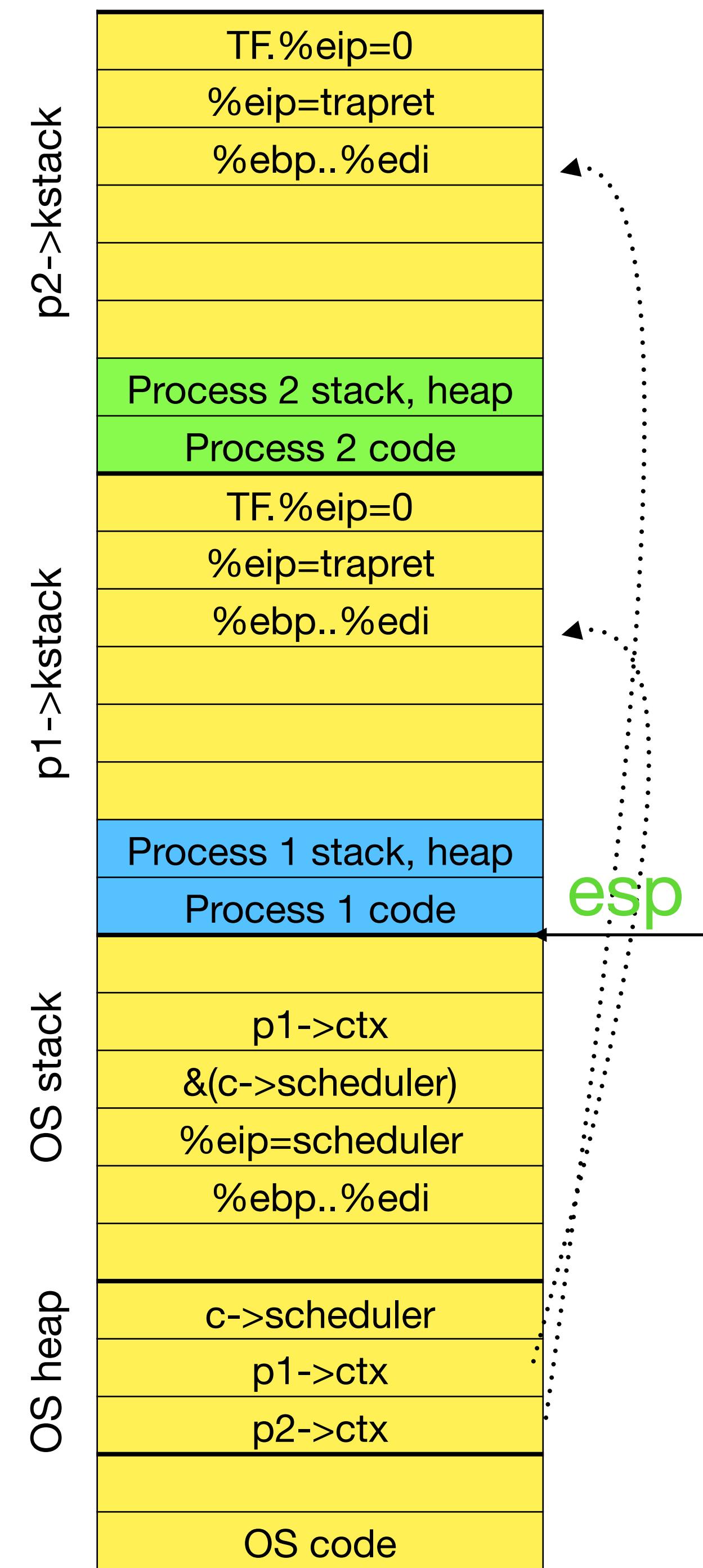
p18-sched

```

.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret
        void scheduler(void) {
            struct proc *p; struct cpu *c = mycpu();
            for(;;){
                for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                    if(p->state != RUNNABLE)
                        continue;
                    p->state = RUNNING;
                    switchuvm(p);
                    swtch(&(c->scheduler), p->context);
                }
            }
        }
        void yield(void) {
            struct proc *p = myproc();
            p->state = RUNNABLE;
            swtch(&p->context, c->scheduler);
        }
        void trap(struct trapframe *tf) {
            ..
            if(tf->trapno == T_IRQ0+IRQ_TIMER)
                yield();
        }
    }
}

```

eip →



Context switching in action: giving control

p18-sched

```

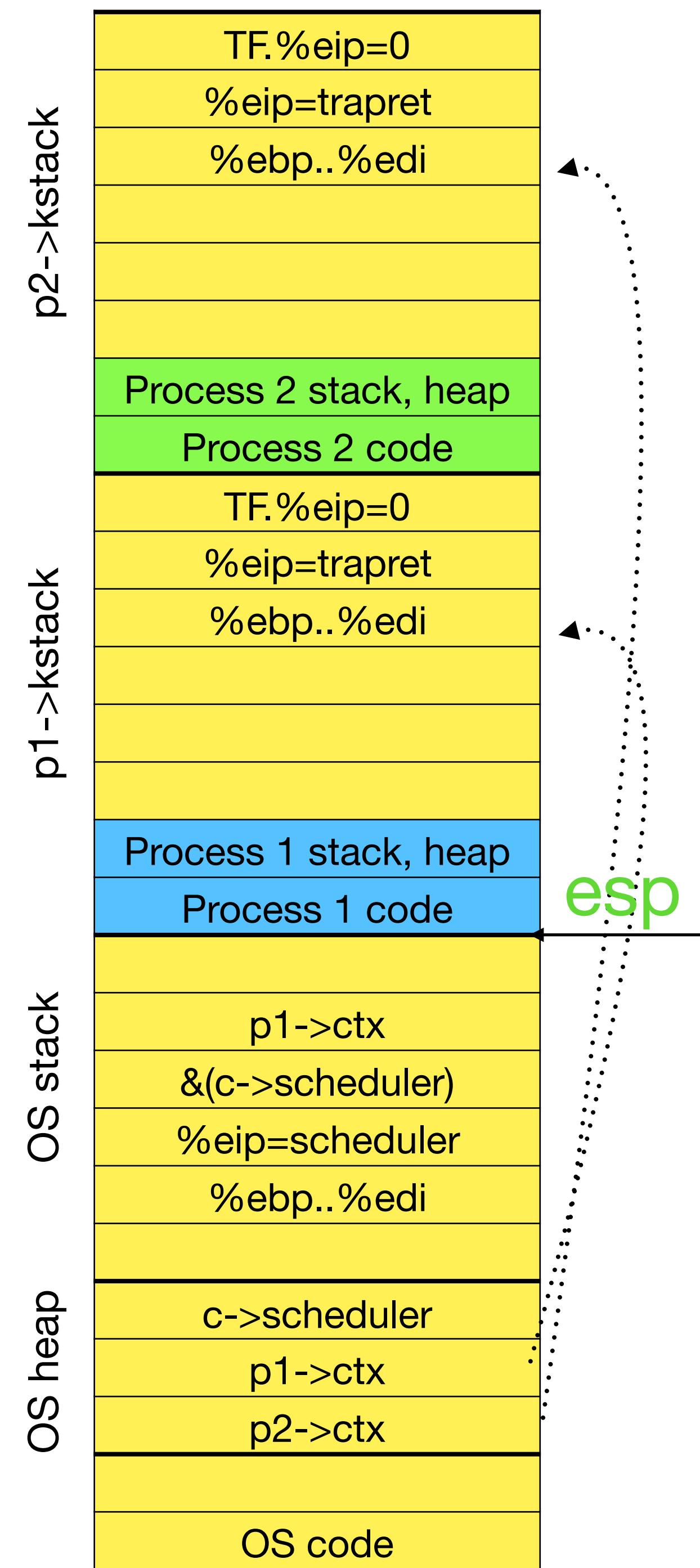
.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)      eip →
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret

    void scheduler(void) {
        struct proc *p; struct cpu *c = mycpu();
        for(;;){
            for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                if(p->state != RUNNABLE)
                    continue;
                p->state = RUNNING;
                switchuvm(p);
                swtch(&(c->scheduler), p->context);
            }
        }
    }

    void yield(void) {
        struct proc *p = myproc();
        p->state = RUNNABLE;
        swtch(&p->context, c->scheduler);
    }

    void trap(struct trapframe *tf) {
        ..
        if(tf->trapno == T_IRQ0+IRQ_TIMER)
            yield();
    }
}

```



Context switching in action: giving control

p18-sched

```

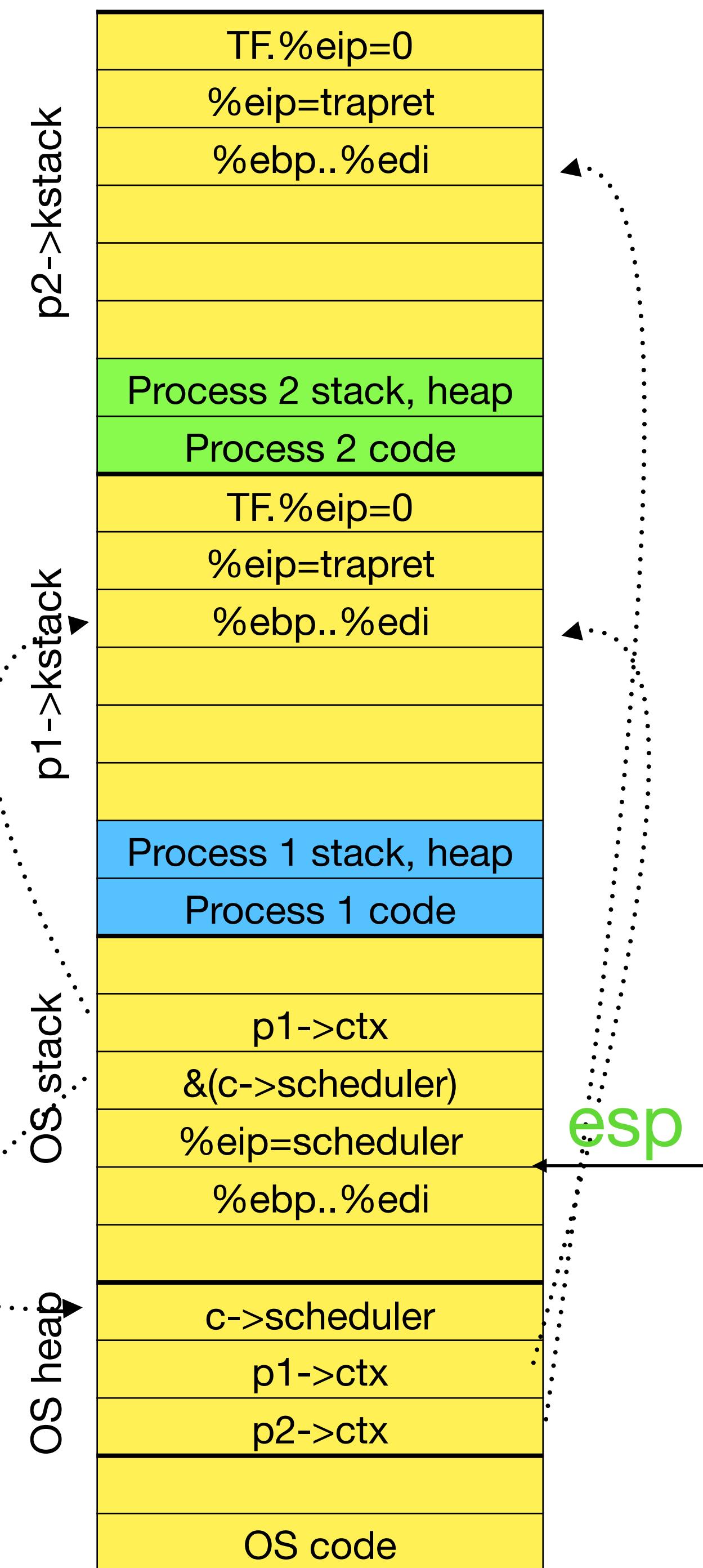
.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)      eip →
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret

    void scheduler(void) {
        struct proc *p; struct cpu *c = mycpu();
        for(;;){
            for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                if(p->state != RUNNABLE)
                    continue;
                p->state = RUNNING;
                switchuvm(p);
                swtch(&(c->scheduler), p->context);
            }
        }
    }

    void yield(void) {
        struct proc *p = myproc();
        p->state = RUNNABLE;
        swtch(&p->context, c->scheduler);
    }

    void trap(struct trapframe *tf) {
        ..
        if(tf->trapno == T_IRQ0+IRQ_TIMER)
            yield();
    }
}

```



Context switching in action: giving control

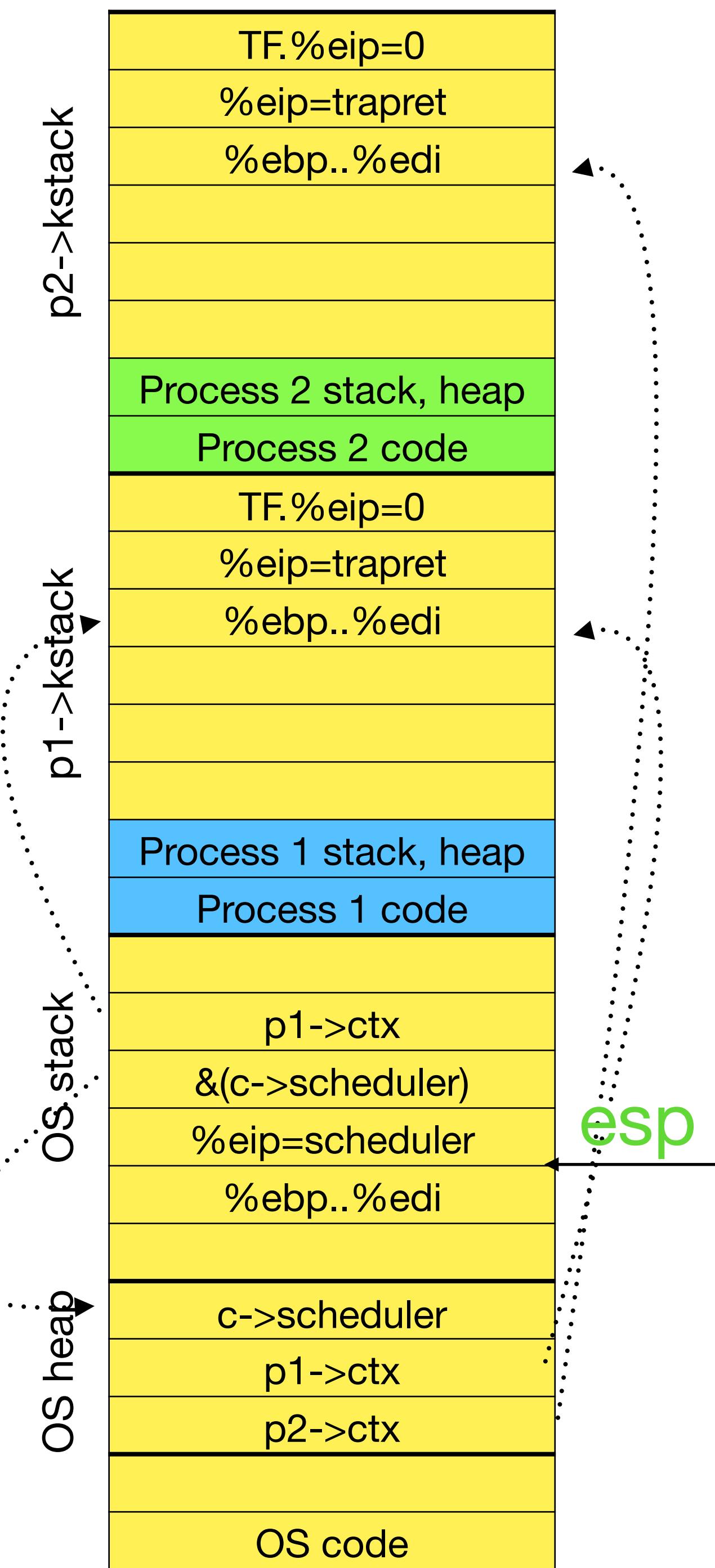
p18-sched

```
.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret

    void scheduler(void) {
        struct proc *p; struct cpu *c = mycpu();
        for(;;){
            for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                if(p->state != RUNNABLE)
                    continue;
                p->state = RUNNING;
                switchuvm(p);
                swtch(&(c->scheduler), p->context);
            }
        }
    }

    void yield(void) {
        struct proc *p = myproc();
        p->state = RUNNABLE;
        swtch(&p->context, c->scheduler);
    }

    void trap(struct trapframe *tf) {
        ..
        if(tf->trapno == T_IRQ0+IRQ_TIMER)
            yield();
    }
```



Context switching in action: giving control

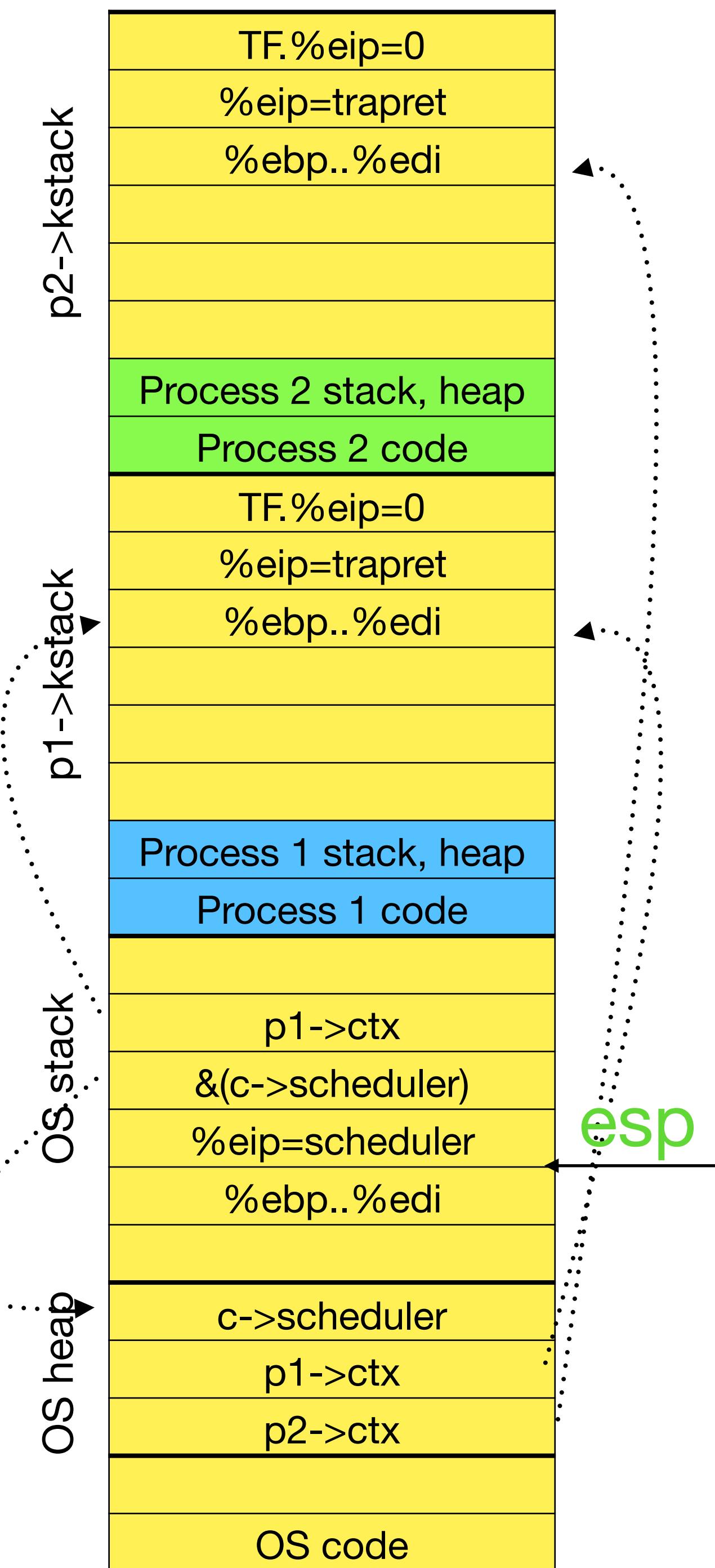
p18-sched

```
.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret
    
```

```
void scheduler(void) {
    struct proc *p; struct cpu *c = mycpu();
    for(;;){
        for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
            if(p->state != RUNNABLE)
                continue;
            p->state = RUNNING;
            switchuvm(p);
            swtch(&(c->scheduler), p->context);
        }
    }
}

void yield(void) {
    struct proc *p = myproc();
    p->state = RUNNABLE;
    swtch(&p->context, c->scheduler);
}

void trap(struct trapframe *tf) {
    ...
    if(tf->trapno == T_IRQ0+IRQ_TIMER)
        yield();
}
```



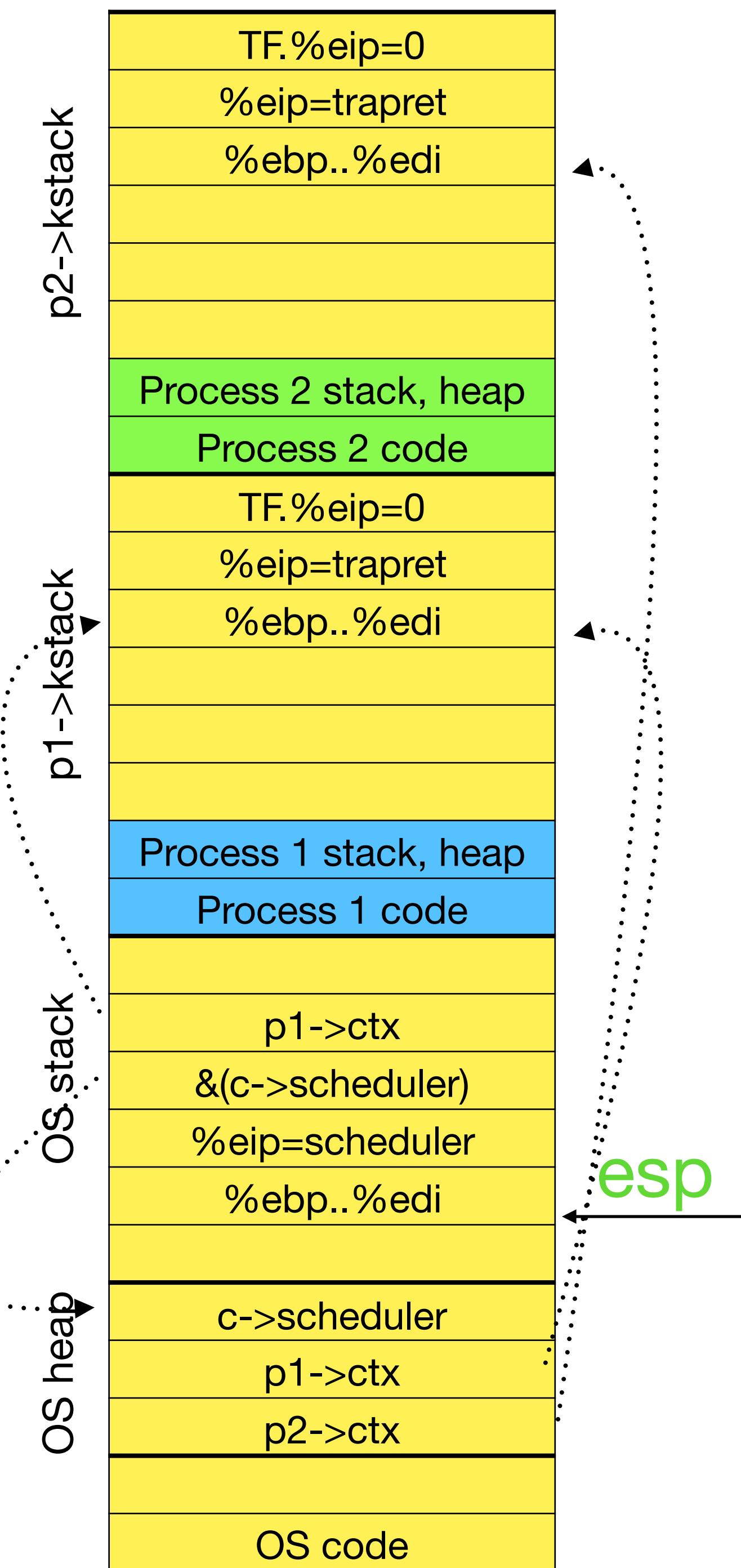
Context switching in action: giving control

p18-sched

```

.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret
        void scheduler(void) {
            struct proc *p; struct cpu *c = mycpu();
            for(;;){
                for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                    if(p->state != RUNNABLE)
                        continue;
                    p->state = RUNNING;
                    switchuvm(p);
                    swtch(&(c->scheduler), p->context);
                }
            }
        }
        void yield(void) {
            struct proc *p = myproc();
            p->state = RUNNABLE;
            swtch(&p->context, c->scheduler);
        }
        void trap(struct trapframe *tf) {
            ..
            if(tf->trapno == T_IRQ0+IRQ_TIMER)
                yield();
        }
    }
}

```



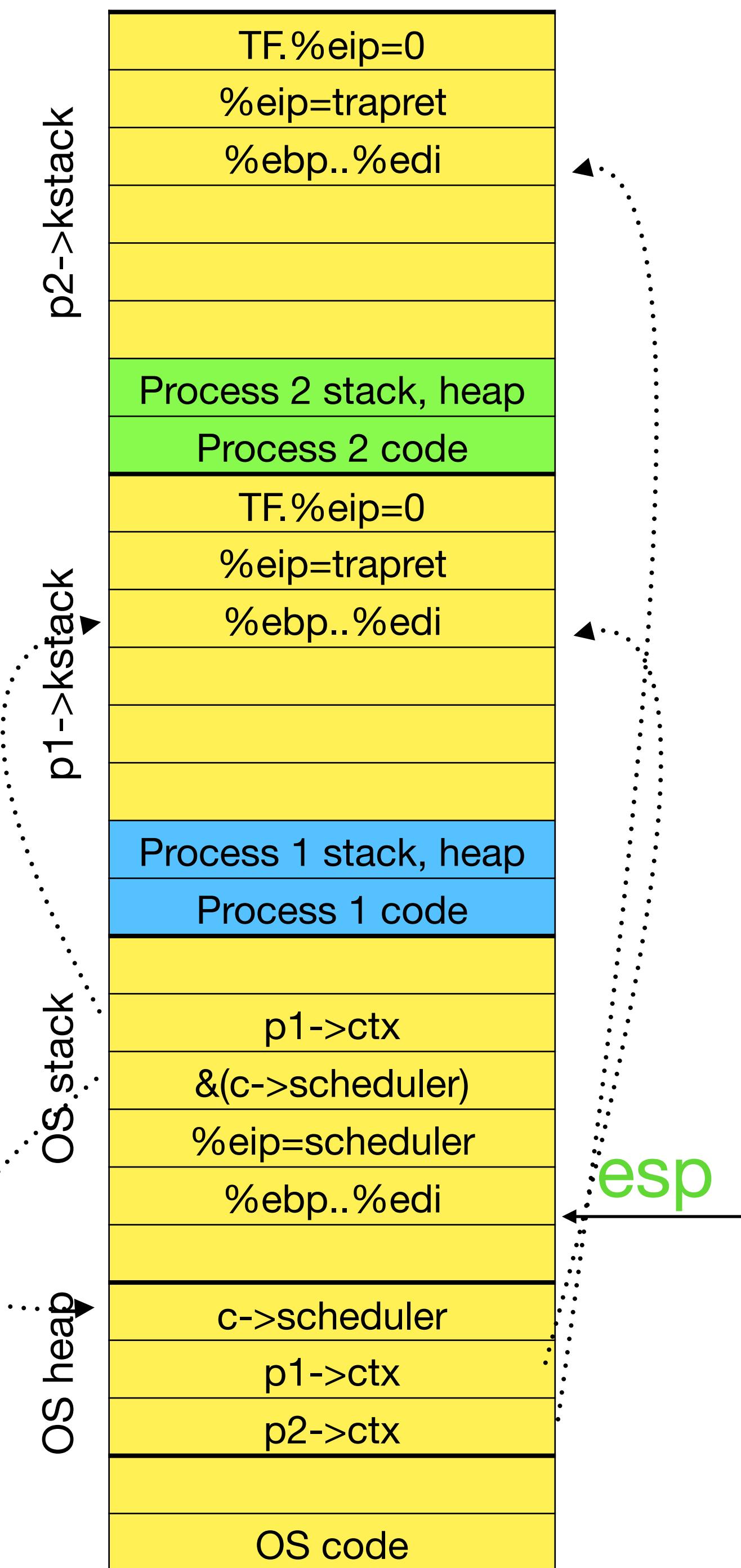
Context switching in action: giving control

p18-sched

```

.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret
    void scheduler(void) {
        struct proc *p; struct cpu *c = mycpu();
        for(;;){
            for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                if(p->state != RUNNABLE)
                    continue;
                p->state = RUNNING;
                switchuvm(p);
                swtch(&(c->scheduler), p->context);
            }
        }
    }
    void yield(void) {
        struct proc *p = myproc();
        p->state = RUNNABLE;
        swtch(&p->context, c->scheduler);
    }
    void trap(struct trapframe *tf) {
        ..
        if(tf->trapno == T_IRQ0+IRQ_TIMER)
            yield();
    }
}

```



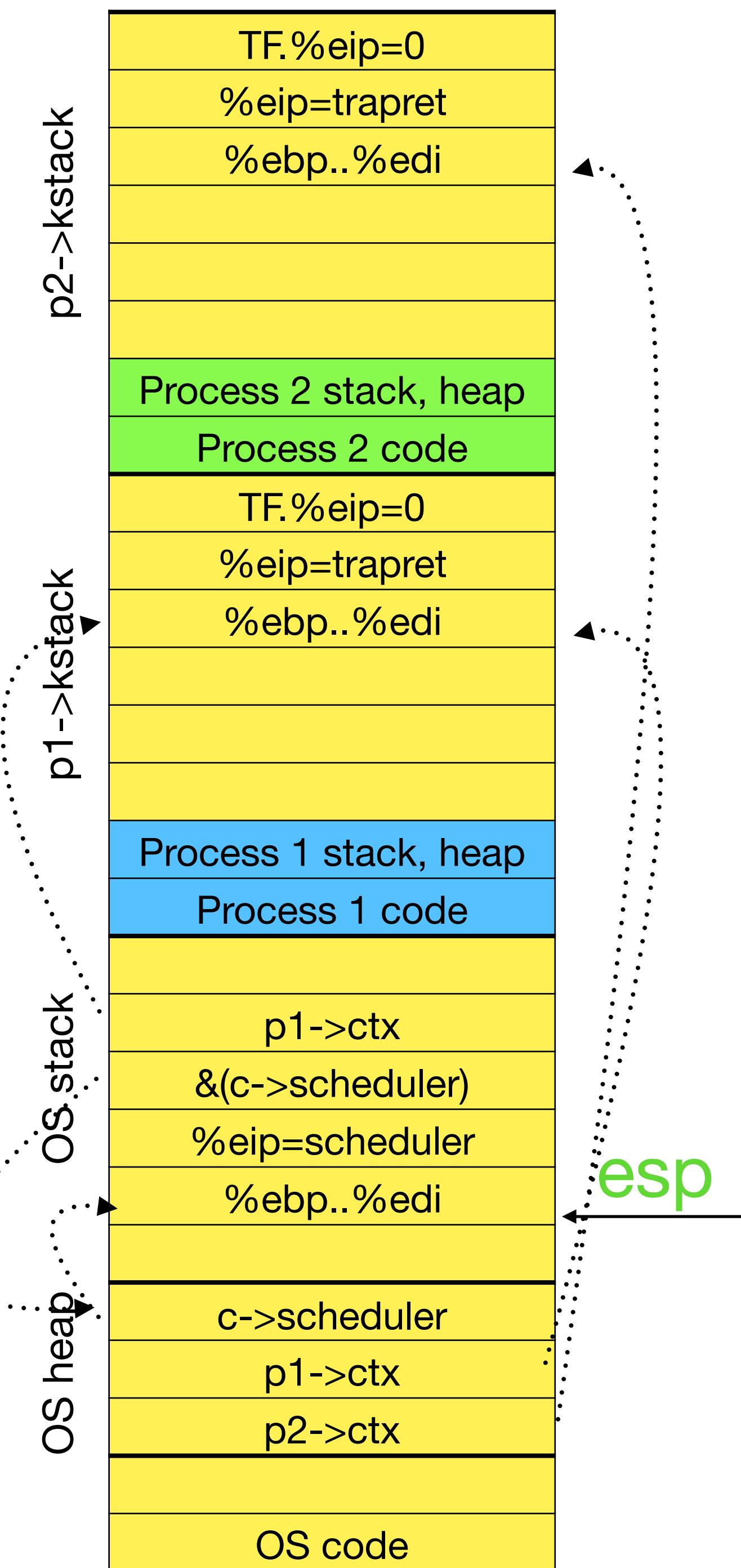
Context switching in action: giving control

p18-sched

```

.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret
    void scheduler(void) {
        struct proc *p; struct cpu *c = mycpu();
        for(;;){
            for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                if(p->state != RUNNABLE)
                    continue;
                p->state = RUNNING;
                switchuvm(p);
                swtch(&(c->scheduler), p->context);
            }
        }
    }
    void yield(void) {
        struct proc *p = myproc();
        p->state = RUNNABLE;
        swtch(&p->context, c->scheduler);
    }
    void trap(struct trapframe *tf) {
        ..
        if(tf->trapno == T_IRQ0+IRQ_TIMER)
            yield();
    }
}

```



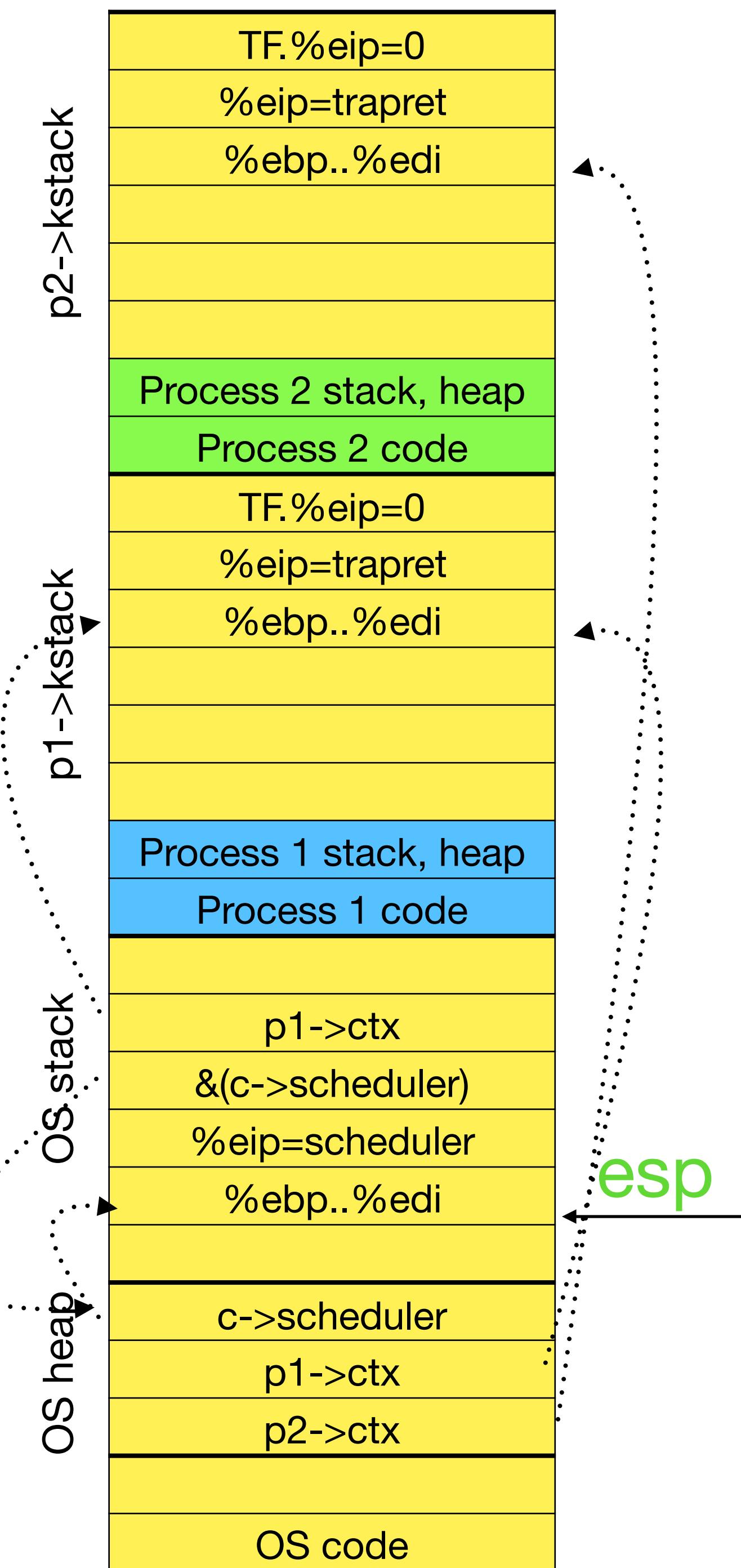
Context switching in action: giving control

p18-sched

```

.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret
        void scheduler(void) {
            struct proc *p; struct cpu *c = mycpu();
            for(;;){
                for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                    if(p->state != RUNNABLE)
                        continue;
                    p->state = RUNNING;
                    switchuvm(p);
                    swtch(&(c->scheduler), p->context);
                }
            }
        }
        void yield(void) {
            struct proc *p = myproc();
            p->state = RUNNABLE;
            swtch(&p->context, c->scheduler);
        }
        void trap(struct trapframe *tf) {
            ..
            if(tf->trapno == T_IRQ0+IRQ_TIMER)
                yield();
        }
    }
}

```



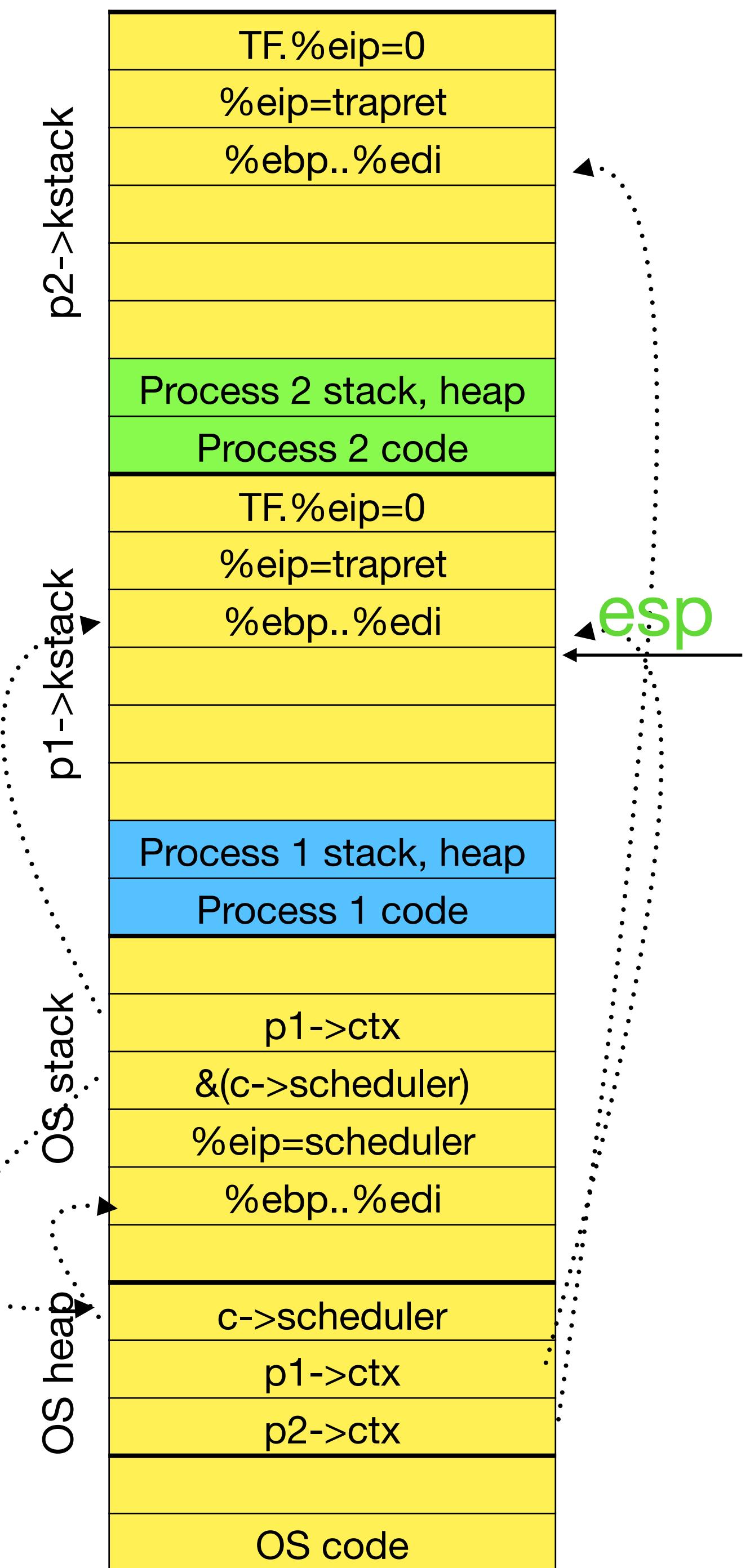
Context switching in action: giving control

p18-sched

```

.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret
        void scheduler(void) {
            struct proc *p; struct cpu *c = mycpu();
            for(;;){
                for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                    if(p->state != RUNNABLE)
                        continue;
                    p->state = RUNNING;
                    switchuvm(p);
                    swtch(&(c->scheduler), p->context);
                }
            }
        }
        void yield(void) {
            struct proc *p = myproc();
            p->state = RUNNABLE;
            swtch(&p->context, c->scheduler);
        }
        void trap(struct trapframe *tf) {
            ..
            if(tf->trapno == T_IRQ0+IRQ_TIMER)
                yield();
        }
    }
}

```



Context switching in action: giving control

p18-sched

```

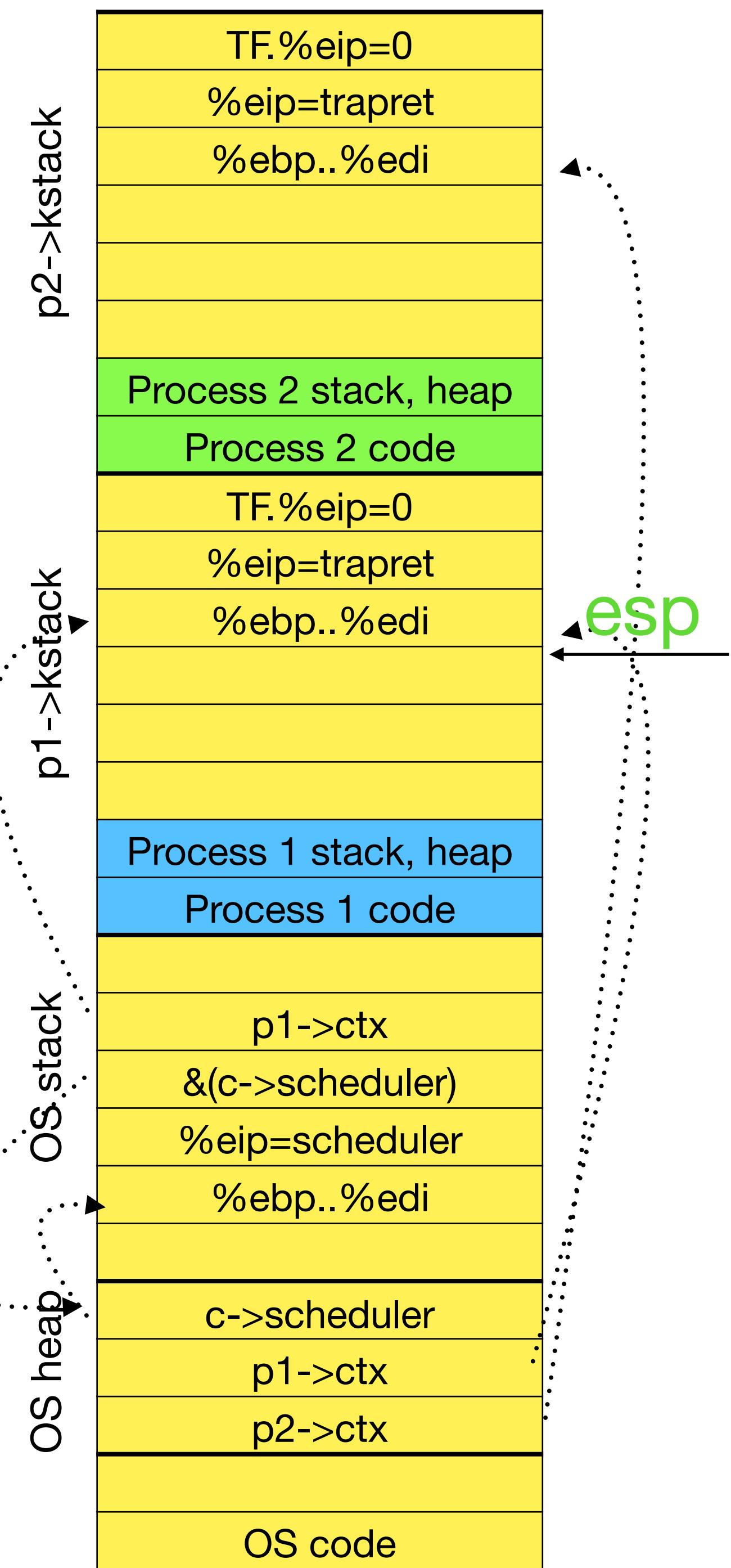
.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret
    eip →

        void scheduler(void) {
            struct proc *p; struct cpu *c = mycpu();
            for(;;){
                for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                    if(p->state != RUNNABLE)
                        continue;
                    p->state = RUNNING;
                    switchuvm(p);
                    swtch(&(c->scheduler), p->context);
                }
            }
        }

        void yield(void) {
            struct proc *p = myproc();
            p->state = RUNNABLE;
            swtch(&p->context, c->scheduler);
        }

        void trap(struct trapframe *tf) {
            ..
            if(tf->trapno == T_IRQ0+IRQ_TIMER)
                yield();
        }
    }
}

```



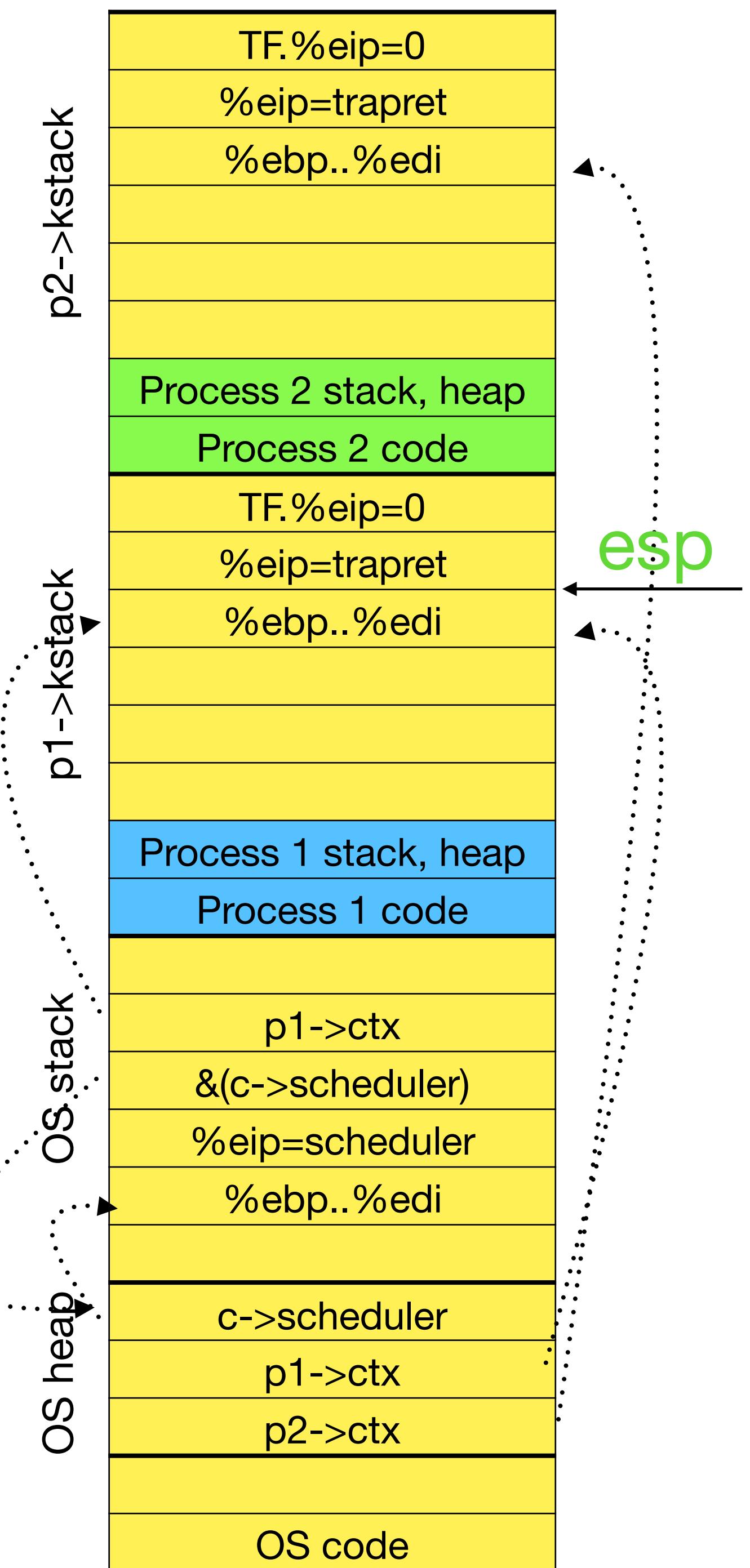
Context switching in action: giving control

p18-sched

```

.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    eip → popl %ebp
    ret
        void scheduler(void) {
            struct proc *p; struct cpu *c = mycpu();
            for(;;){
                for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                    if(p->state != RUNNABLE)
                        continue;
                    p->state = RUNNING;
                    switchuvm(p);
                    swtch(&(c->scheduler), p->context);
                }
            }
        }
        void yield(void) {
            struct proc *p = myproc();
            p->state = RUNNABLE;
            swtch(&p->context, c->scheduler);
        }
        void trap(struct trapframe *tf) {
            ..
            if(tf->trapno == T_IRQ0+IRQ_TIMER)
                yield();
        }
    }
}

```



Context switching in action: giving control

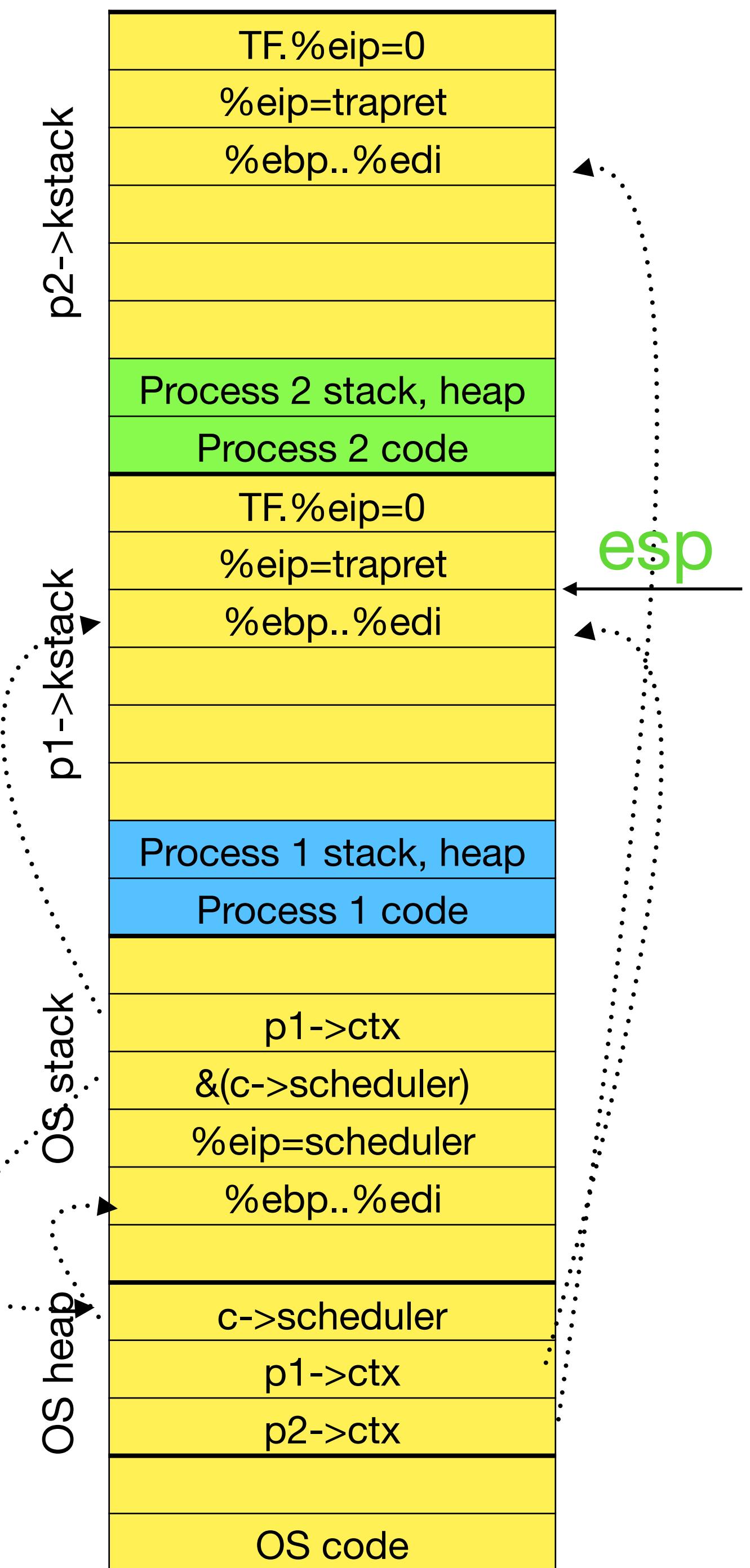
p18-sched

```

.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    void yield(void) {
        struct proc *p = myproc();
        p->state = RUNNABLE;
        swtch(&p->context, c->scheduler);
    }
    void scheduler(void) {
        struct proc *p; struct cpu *c = mycpu();
        for(;;){
            for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                if(p->state != RUNNABLE)
                    continue;
                p->state = RUNNING;
                switchuvm(p);
                swtch(&(c->scheduler), p->context);
            }
        }
    }
    void trap(struct trapframe *tf) {
        ..
        if(tf->trapno == T_IRQ0+IRQ_TIMER)
            yield();
    }
}

```

eip → ret



Context switching in action: giving control

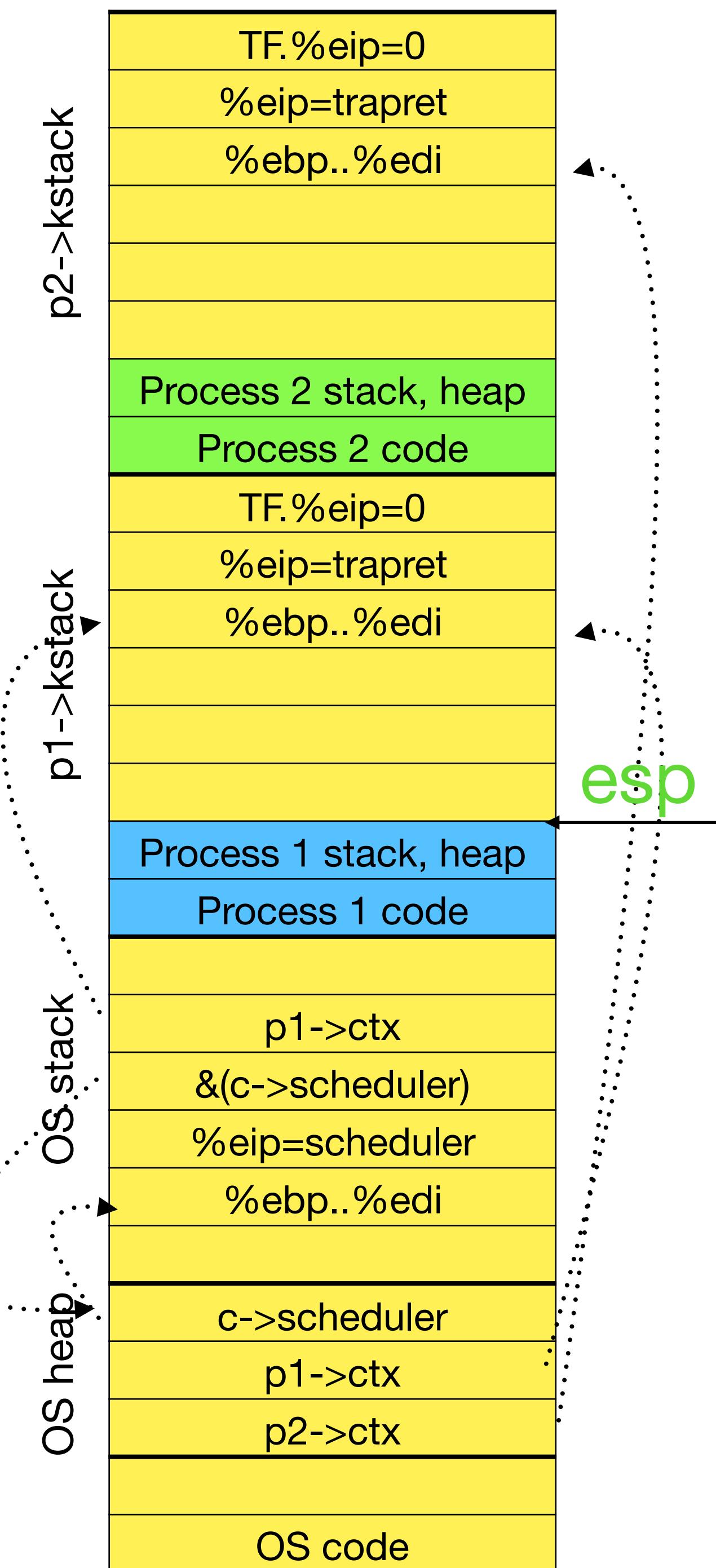
p18-sched

```

.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    void yield(void) {
        struct proc *p = myproc();
        p->state = RUNNABLE;
        swtch(&p->context, c->scheduler);
    }
    void scheduler(void) {
        struct proc *p; struct cpu *c = mycpu();
        for(;;){
            for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                if(p->state != RUNNABLE)
                    continue;
                p->state = RUNNING;
                switchuvm(p);
                swtch(&(c->scheduler), p->context);
            }
        }
    }
    void trap(struct trapframe *tf) {
        ..
        if(tf->trapno == T_IRQ0+IRQ_TIMER)
            yield();
    }
}

```

eip → ret



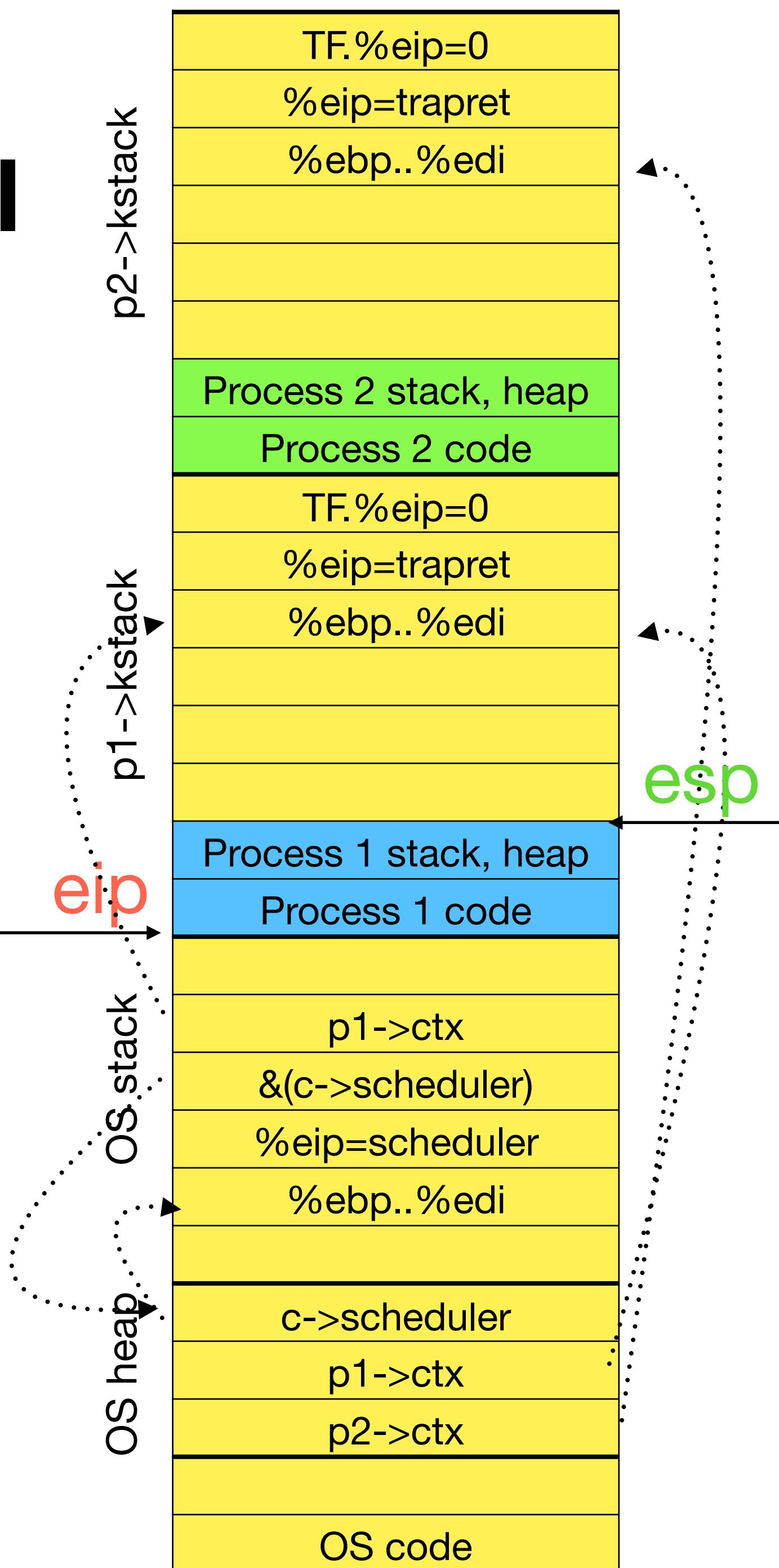
Context switching in action: giving control

p18-sched

```

.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret
    void scheduler(void) {
        struct proc *p; struct cpu *c = mycpu();
        for(;;){
            for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                if(p->state != RUNNABLE)
                    continue;
                p->state = RUNNING;
                switchuvm(p);
                swtch(&(c->scheduler), p->context);
            }
        }
    }
    void yield(void) {
        struct proc *p = myproc();
        p->state = RUNNABLE;
        swtch(&p->context, c->scheduler);
    }
    void trap(struct trapframe *tf) {
        ..
        if(tf->trapno == T_IRQ0+IRQ_TIMER)
            yield();
    }
}

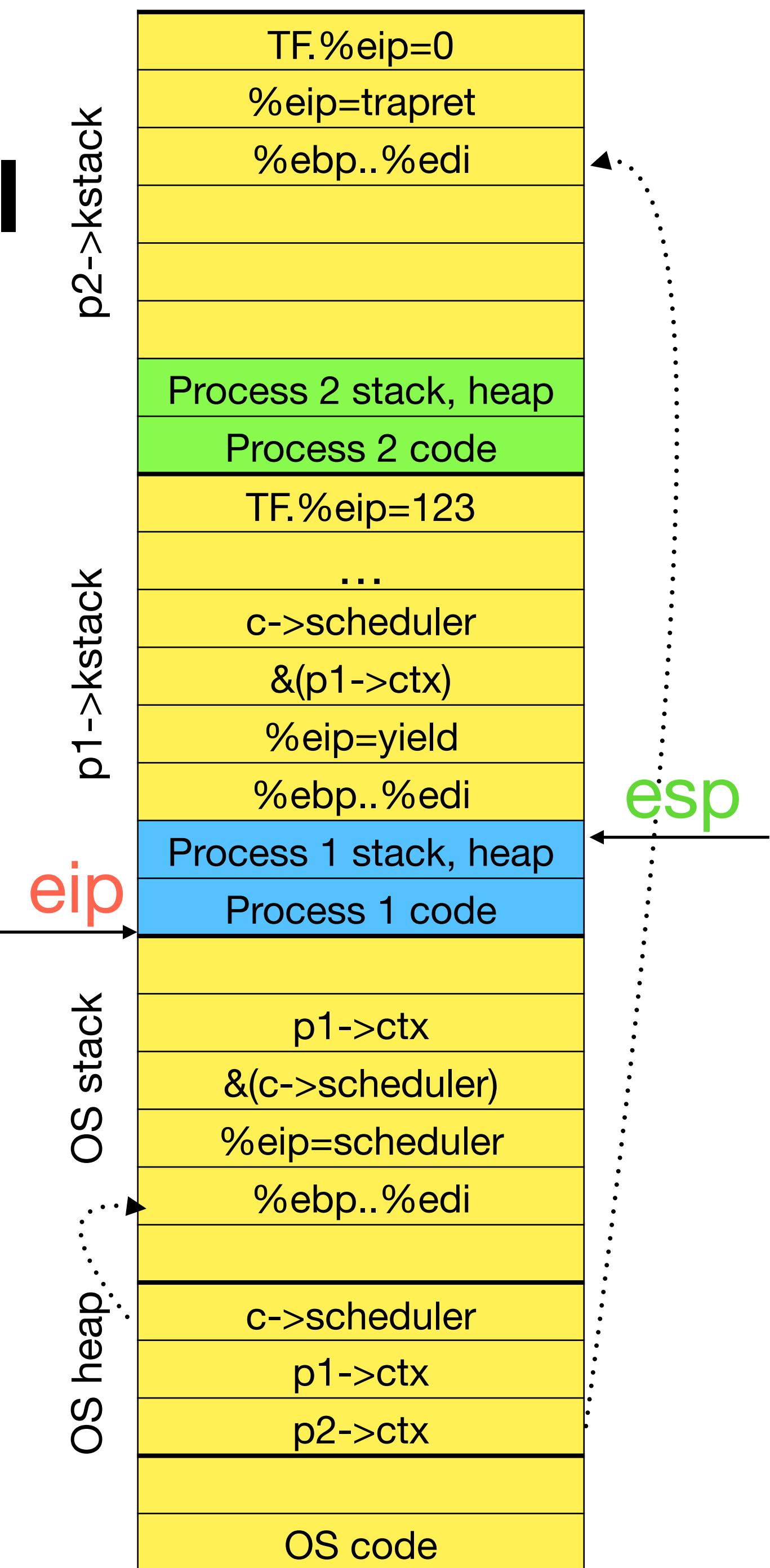
```



Context switching in action: taking control

p18-sched

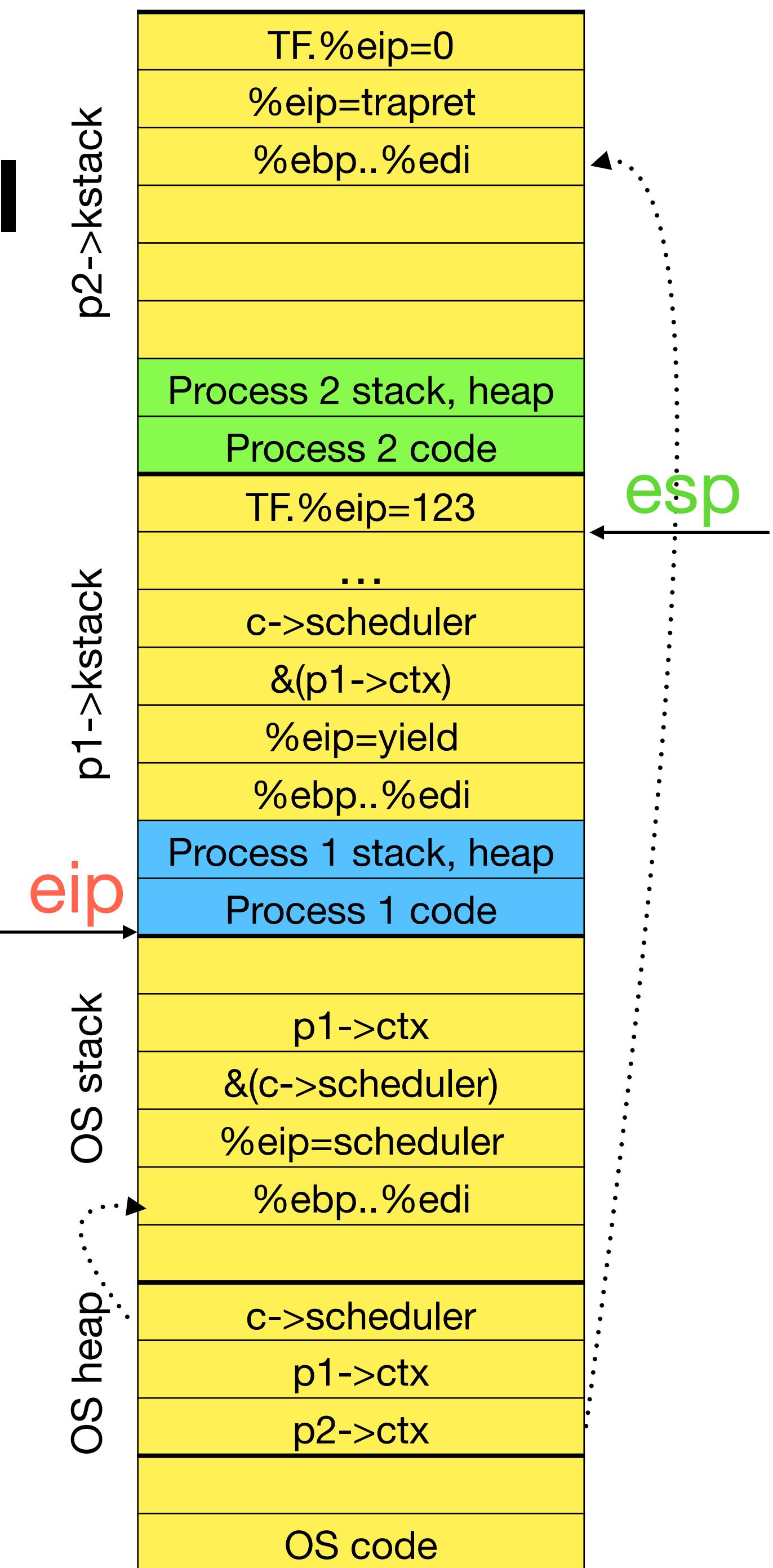
```
.globl swtch          void scheduler(void) {  
swtch:  
    movl 4(%esp), %eax  
    movl 8(%esp), %edx  
    pushl %ebp  
    pushl %ebx  
    pushl %esi  
    pushl %edi  
    movl %esp, (%eax)  
    movl %edx, %esp  
    popl %edi  
    popl %esi  
    popl %ebx  
    popl %ebp  
    ret  
    }  
    void yield(void) {  
        struct proc *p = myproc();  
        p->state = RUNNABLE;  
        swtch(&p->context, c->scheduler);  
    }  
    void trap(struct trapframe *tf) {  
        ..  
        if(tf->trapno == T_IRQ0+IRQ_TIMER)  
            yield();  
    }
```



Context switching in action: taking control

p18-sched

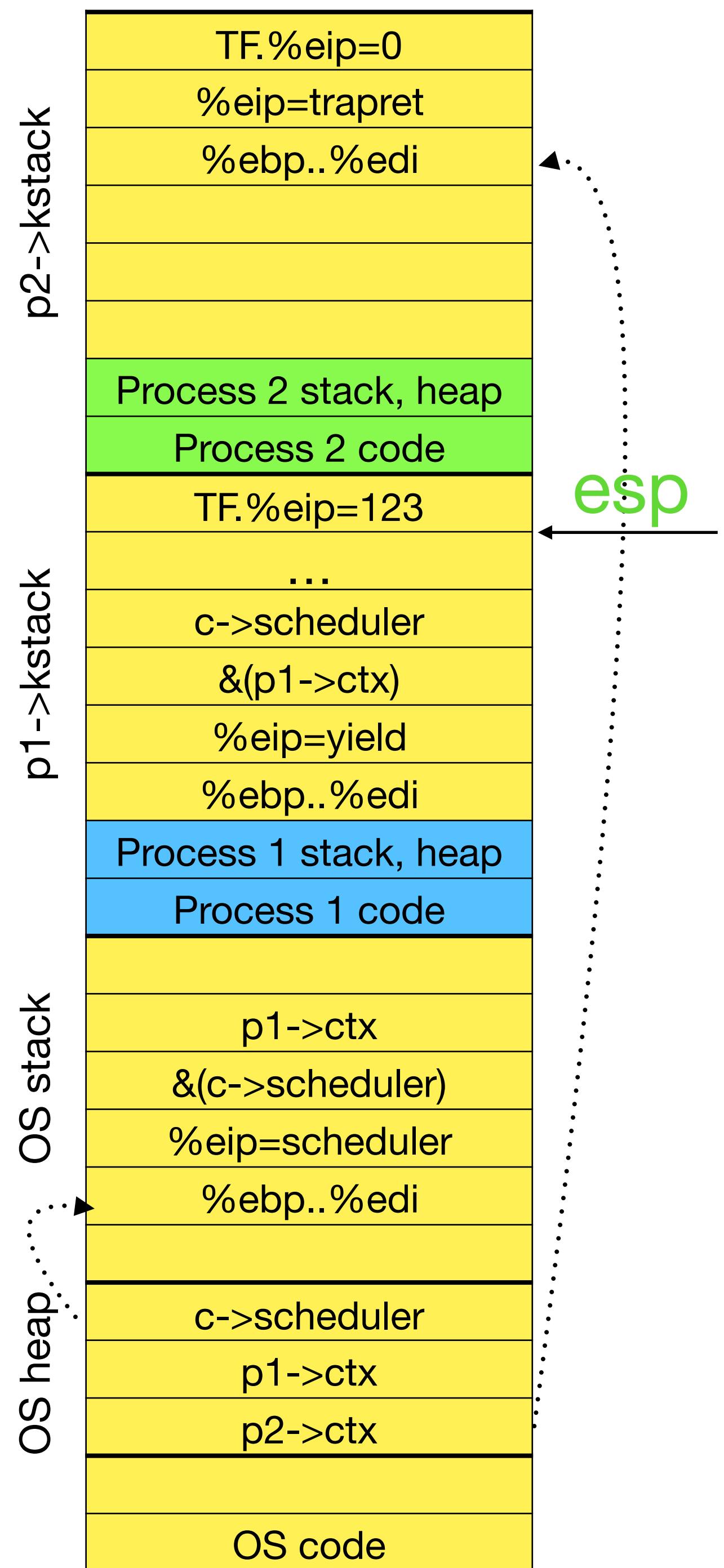
```
.globl swtch          void scheduler(void) {  
swtch:  
    movl 4(%esp), %eax  
    movl 8(%esp), %edx  
    pushl %ebp  
    pushl %ebx  
    pushl %esi  
    pushl %edi  
    movl %esp, (%eax)  
    movl %edx, %esp  
    popl %edi  
    popl %esi  
    popl %ebx  
    popl %ebp  
    void yield(void) {  
        struct proc *p = myproc();  
        p->state = RUNNABLE;  
        swtch(&p->context, c->scheduler);  
    }  
    ..  
    if(tf->trapno == T_IRQ0+IRQ_TIMER)  
        yield();  
}
```



Context switching in action: taking control

p18-sched

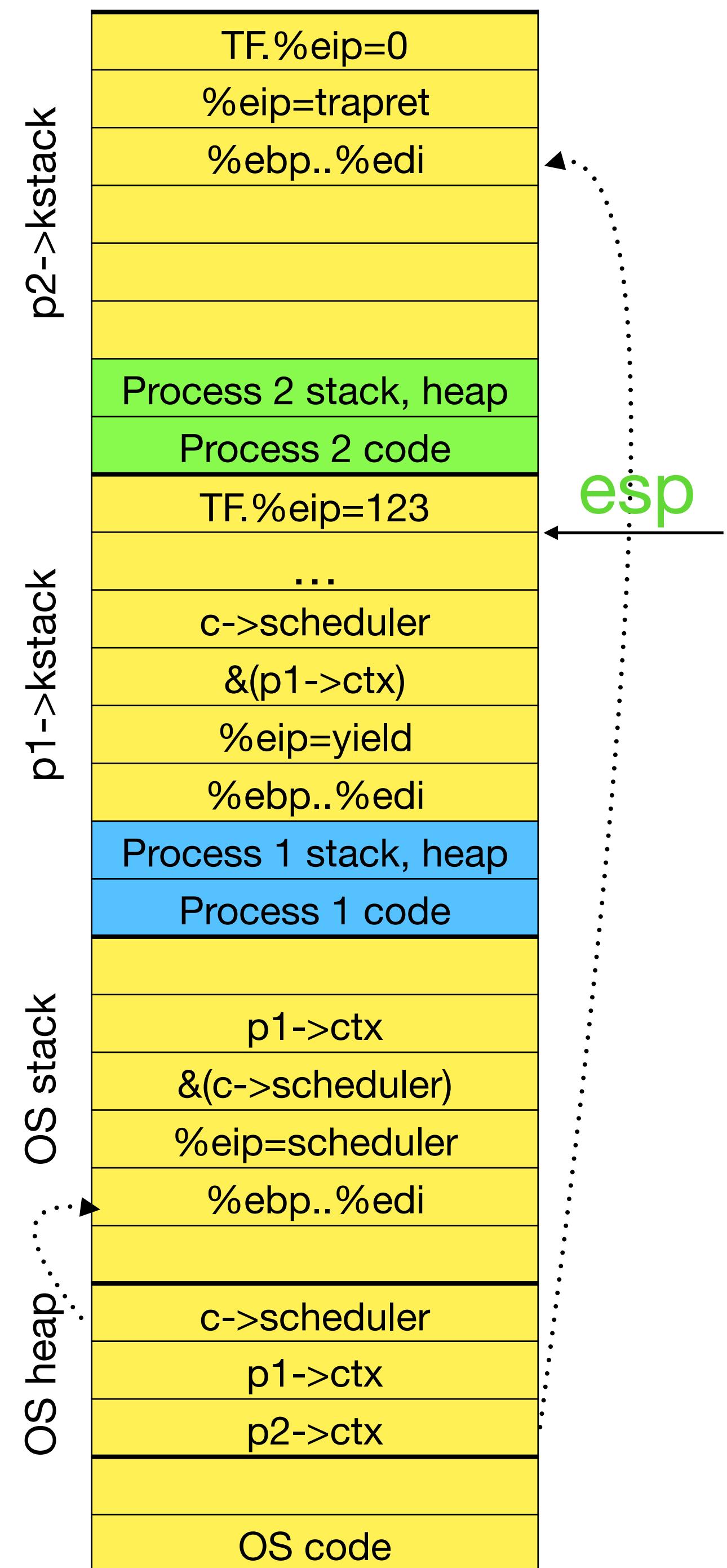
```
.globl swtch          void scheduler(void) {  
swtch:  
    movl 4(%esp), %eax  
    movl 8(%esp), %edx  
    pushl %ebp  
    pushl %ebx  
    pushl %esi  
    pushl %edi  
    movl %esp, (%eax)  
    movl %edx, %esp  
    popl %edi  
    popl %esi  
    popl %ebx  
    popl %ebp  
    ret  
    }  
    void yield(void) {  
        struct proc *p = myproc();  
        p->state = RUNNABLE;  
        swtch(&p->context, c->scheduler);  
    }  
    void trap(struct trapframe *tf) {  
        ..  
        if(tf->trapno == T_IRQ0+IRQ_TIMER)  
            eip  
            if(tf->trapno == T_IRQ0+IRQ_TIMER)  
                yield();  
    }
```



Context switching in action: taking control

p18-sched

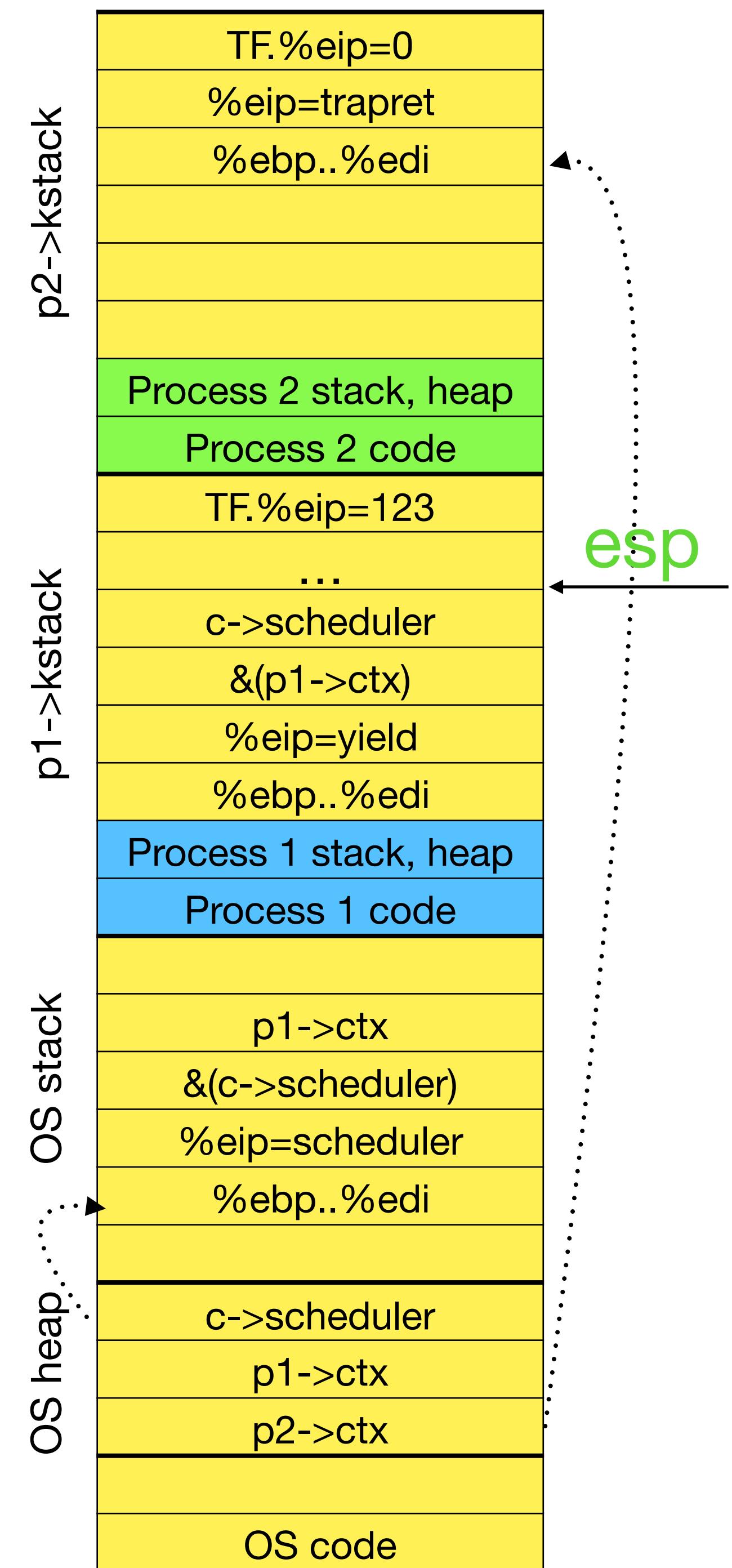
```
.globl swtch          void scheduler(void) {  
swtch:  
    movl 4(%esp), %eax  
    movl 8(%esp), %edx  
    pushl %ebp  
    pushl %ebx  
    pushl %esi  
    pushl %edi  
    movl %esp, (%eax)  
    movl %edx, %esp  
    popl %edi  
    popl %esi  
    popl %ebx  
    popl %ebp eip  
    void yield(void) {  
        struct proc *p = myproc();  
        p->state = RUNNABLE;  
        swtch(&p->context, c->scheduler);  
    }  
    ...  
    if(tf->trapno == T_IRQ0+IRQ_TIMER)  
        yield();  
}
```



Context switching in action: taking control

p18-sched

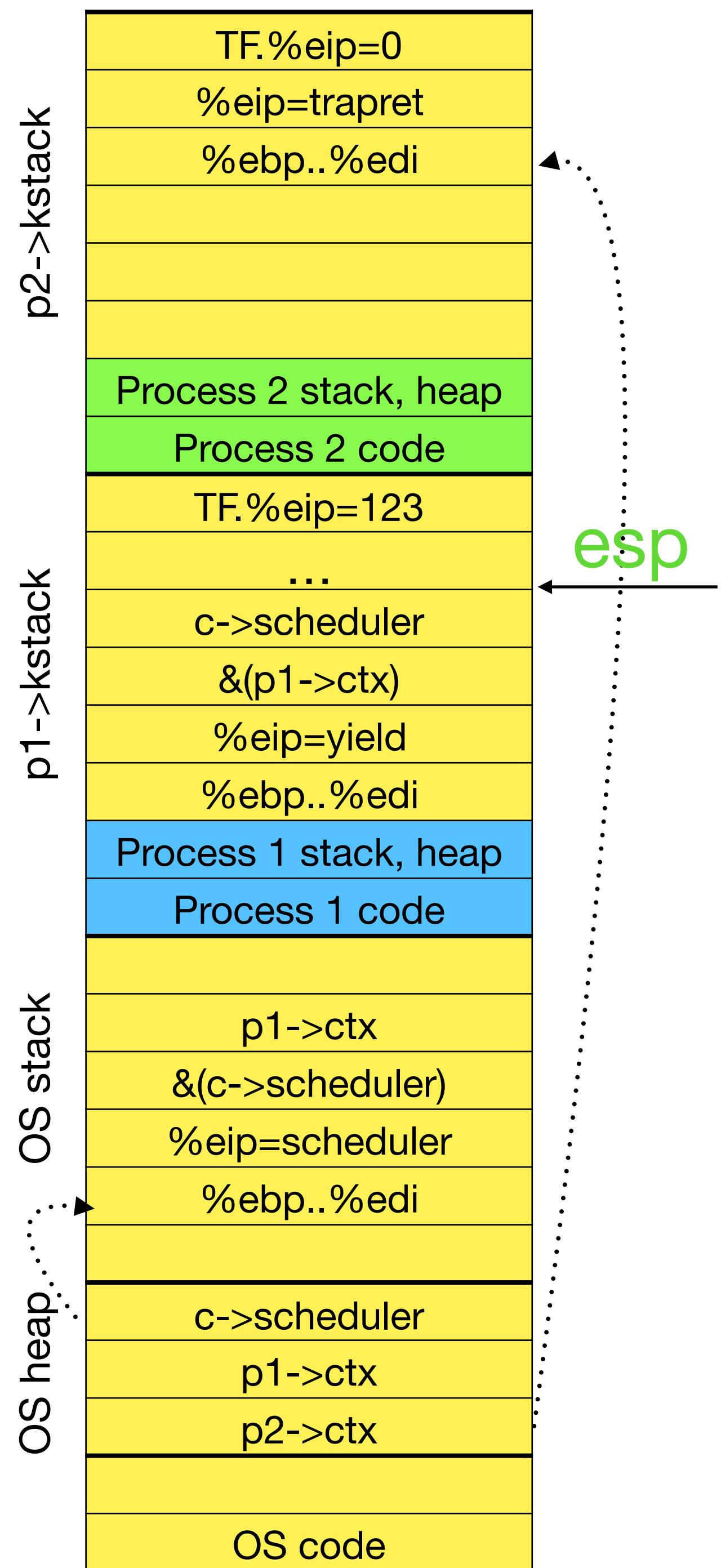
```
.globl swtch          void scheduler(void) {  
swtch:  
    movl 4(%esp), %eax  
    movl 8(%esp), %edx  
    pushl %ebp  
    pushl %ebx  
    pushl %esi  
    pushl %edi  
    movl %esp, (%eax)  
    movl %edx, %esp  
    popl %edi  
    popl %esi  
    popl %ebx  
    popl %ebp eip  
    void yield(void) {  
        struct proc *p = myproc();  
        p->state = RUNNABLE;  
        swtch(&p->context, c->scheduler);  
    }  
    ..  
    if(tf->trapno == T_IRQ0+IRQ_TIMER)  
        yield();  
}
```



Context switching in action: taking control

p18-sched

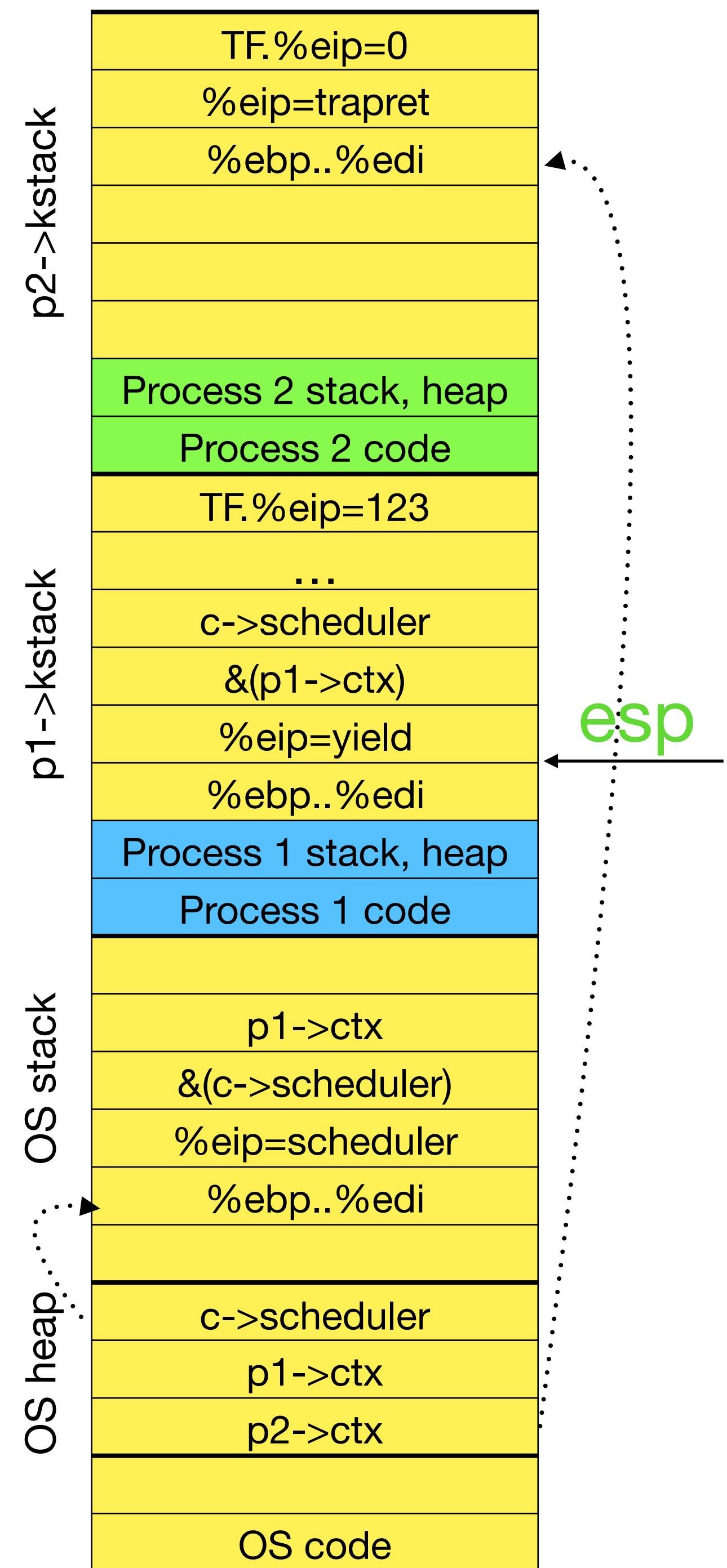
```
.globl swtch          void scheduler(void) {  
swtch:  
    movl 4(%esp), %eax  
    movl 8(%esp), %edx  
    pushl %ebp  
    pushl %ebx  
    pushl %esi  
    pushl %edi  
    movl %esp, (%eax)  
    movl %edx, %esp  
    popl %edi  
    popl %esi  
    popl %ebx  
    popl %ebp  
    void yield(void) {  
        struct proc *p = myproc();  
        p->state = RUNNABLE;  
        ret  
        eip → swtch(&p->context, c->scheduler);  
    }  
    struct proc *p;  struct cpu *c = mycpu();  
    for(;;){  
        for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){  
            if(p->state != RUNNABLE)  
                continue;  
            p->state = RUNNING;  
            switchuvm(p);  
            swtch(&(c->scheduler), p->context);  
        }  
    }  
}  
void trap(struct trapframe *tf) {  
    ..  
    if(tf->trapno == T_IRQ0+IRQ_TIMER)  
        yield();  
}
```



Context switching in action: taking control

p18-sched

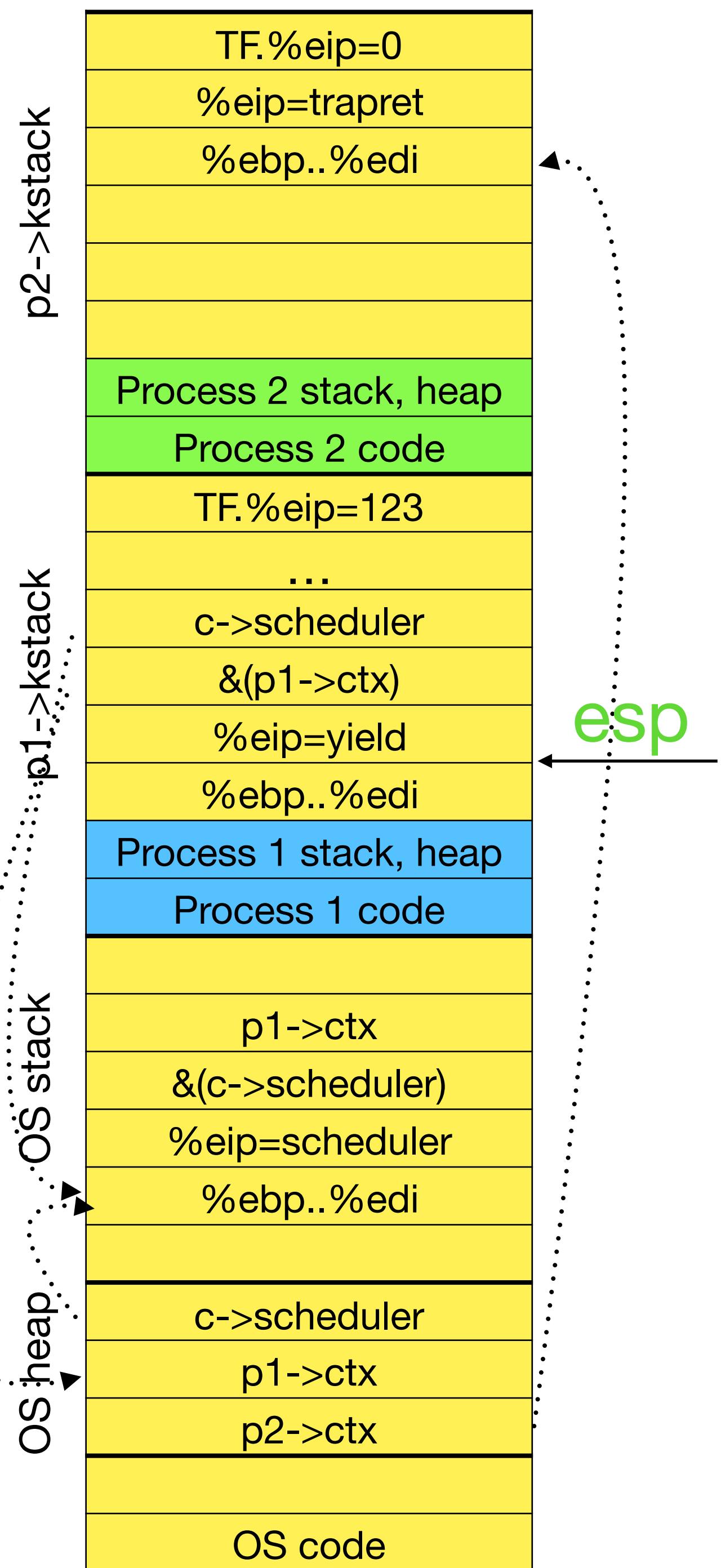
```
.globl swtch          void scheduler(void) {  
swtch:  
    movl 4(%esp), %eax  
    movl 8(%esp), %edx  
    pushl %ebp  
    pushl %ebx  
    pushl %esi  
    pushl %edi  
    movl %esp, (%eax)  
    movl %edx, %esp  
    popl %edi  
    popl %esi  
    popl %ebx  
    popl %ebp  
    void yield(void) {  
        struct proc *p = myproc();  
        p->state = RUNNABLE;  
        ret  
        eip  
        → swtch(&p->context, c->scheduler);  
    }  
    struct proc *p;  struct cpu *c = mycpu();  
    for(;;){  
        for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){  
            if(p->state != RUNNABLE)  
                continue;  
            p->state = RUNNING;  
            switchuvm(p);  
            swtch(&(c->scheduler), p->context);  
        }  
    }  
}  
void trap(struct trapframe *tf) {  
    ..  
    if(tf->trapno == T_IRQ0+IRQ_TIMER)  
        yield();  
}
```



Context switching in action: taking control

p18-sched

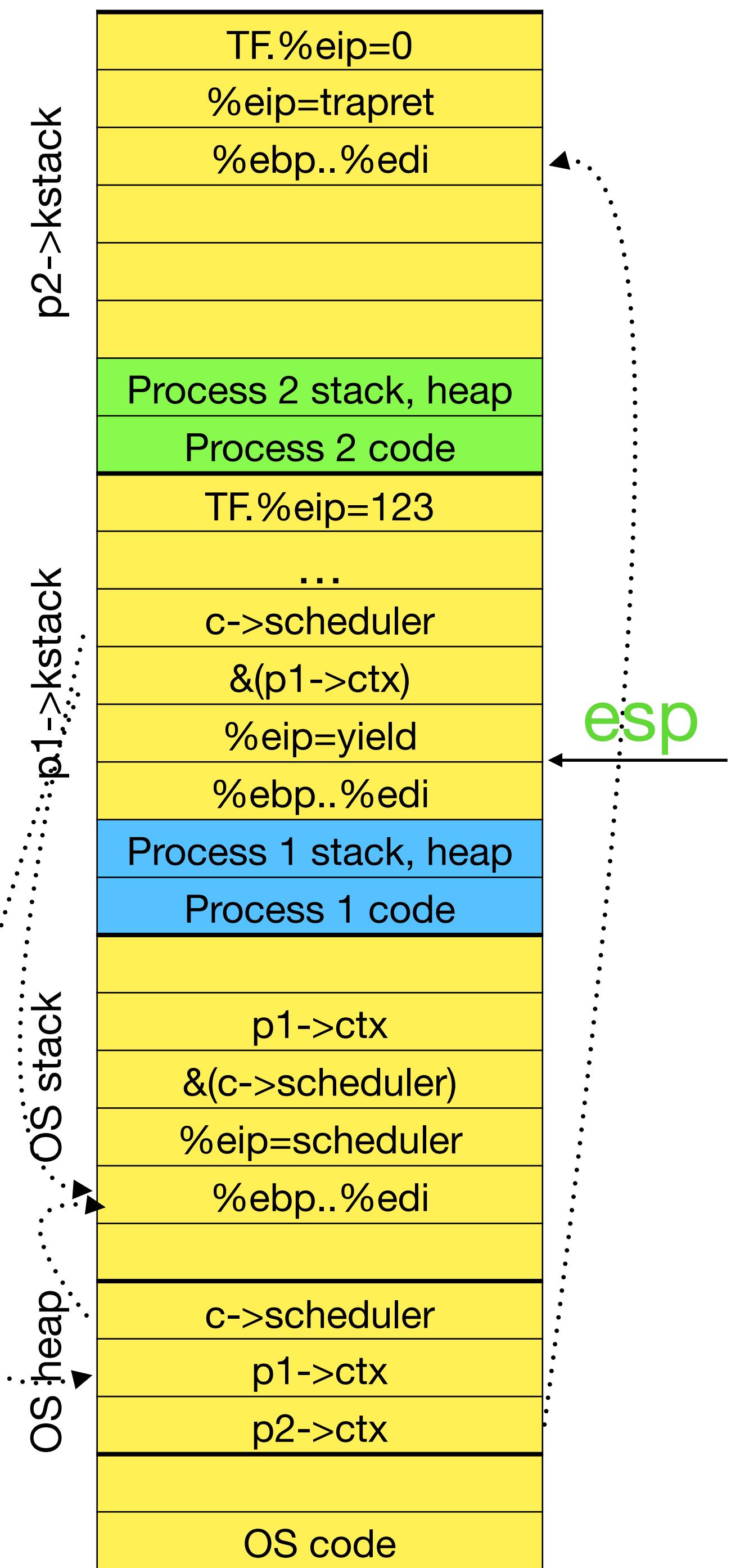
```
.globl swtch          void scheduler(void) {  
swtch:  
    movl 4(%esp), %eax  
    movl 8(%esp), %edx  
    pushl %ebp  
    pushl %ebx  
    pushl %esi  
    pushl %edi  
    movl %esp, (%eax)  
    movl %edx, %esp  
    popl %edi  
    popl %esi  
    popl %ebx  
    popl %ebp  
    void yield(void) {  
        struct proc *p = myproc();  
        p->state = RUNNABLE;  
        ret  
        eip → swtch(&p->context, c->scheduler);  
    }  
    struct proc *p;  struct cpu *c = mycpu();  
    for(;;){  
        for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){  
            if(p->state != RUNNABLE)  
                continue;  
            p->state = RUNNING;  
            switchuvm(p);  
            swtch(&(c->scheduler), p->context);  
        }  
    }  
}  
void trap(struct trapframe *tf) {  
    ..  
    if(tf->trapno == T_IRQ0+IRQ_TIMER)  
        yield();  
}
```



Context switching in action: taking control

p18-sched

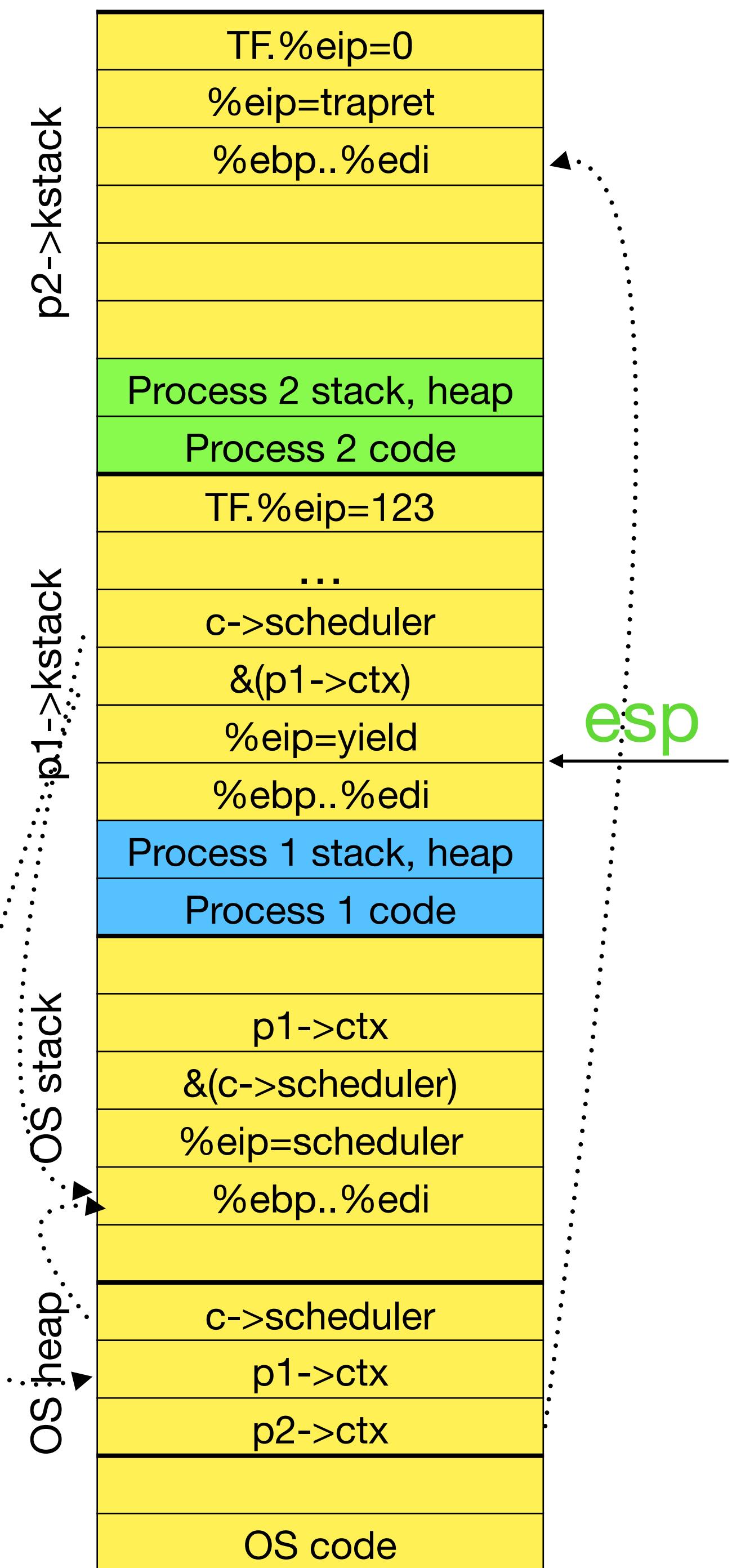
```
.globl swtch
swtch:
eip    movl 4(%esp), %eax
       movl 8(%esp), %edx
       pushl %ebp
       pushl %ebx
       pushl %esi
       pushl %edi
       movl %esp, (%eax)
       movl %edx, %esp
       popl %edi
       popl %esi
       popl %ebx
       popl %ebp
       ret
void scheduler(void) {
    struct proc *p; struct cpu *c = mycpu();
    for(;;){
        for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
            if(p->state != RUNNABLE)
                continue;
            p->state = RUNNING;
            switchuvm(p);
            swtch(&(c->scheduler), p->context);
        }
    }
}
void yield(void) {
    struct proc *p = myproc();
    p->state = RUNNABLE;
    swtch(&p->context, c->scheduler);
}
void trap(struct trapframe *tf) {
    ...
    if(tf->trapno == T_IRQ0+IRQ_TIMER)
        yield();
}
```



Context switching in action: taking control

p18-sched

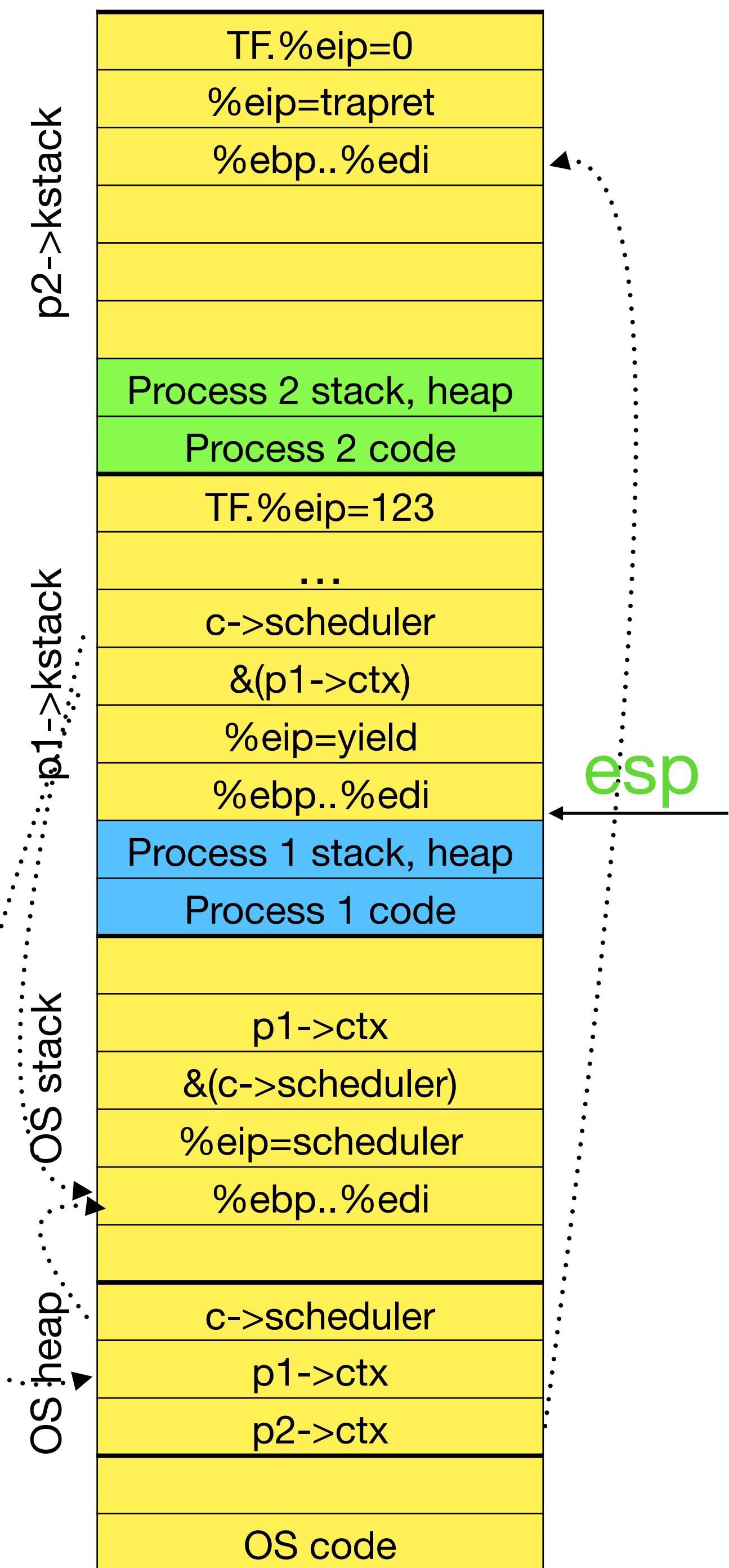
```
.globl swtch          void scheduler(void) {  
swtch:  
    movl 4(%esp), %eax  
    movl 8(%esp), %edx  
    pushl %ebp  
    pushl %ebx  
    pushl %esi  
    pushl %edi  
    movl %esp, (%eax)  
    movl %edx, %esp  
    popl %edi  
    popl %esi  
    popl %ebx  
    popl %ebp  
    ret  
    }  
    void yield(void) {  
        struct proc *p = myproc();  
        p->state = RUNNABLE;  
        swtch(&p->context, c->scheduler);  
    }  
    void trap(struct trapframe *tf) {  
        ..  
        if(tf->trapno == T_IRQ0+IRQ_TIMER)  
            yield();  
    }
```



Context switching in action: taking control

p18-sched

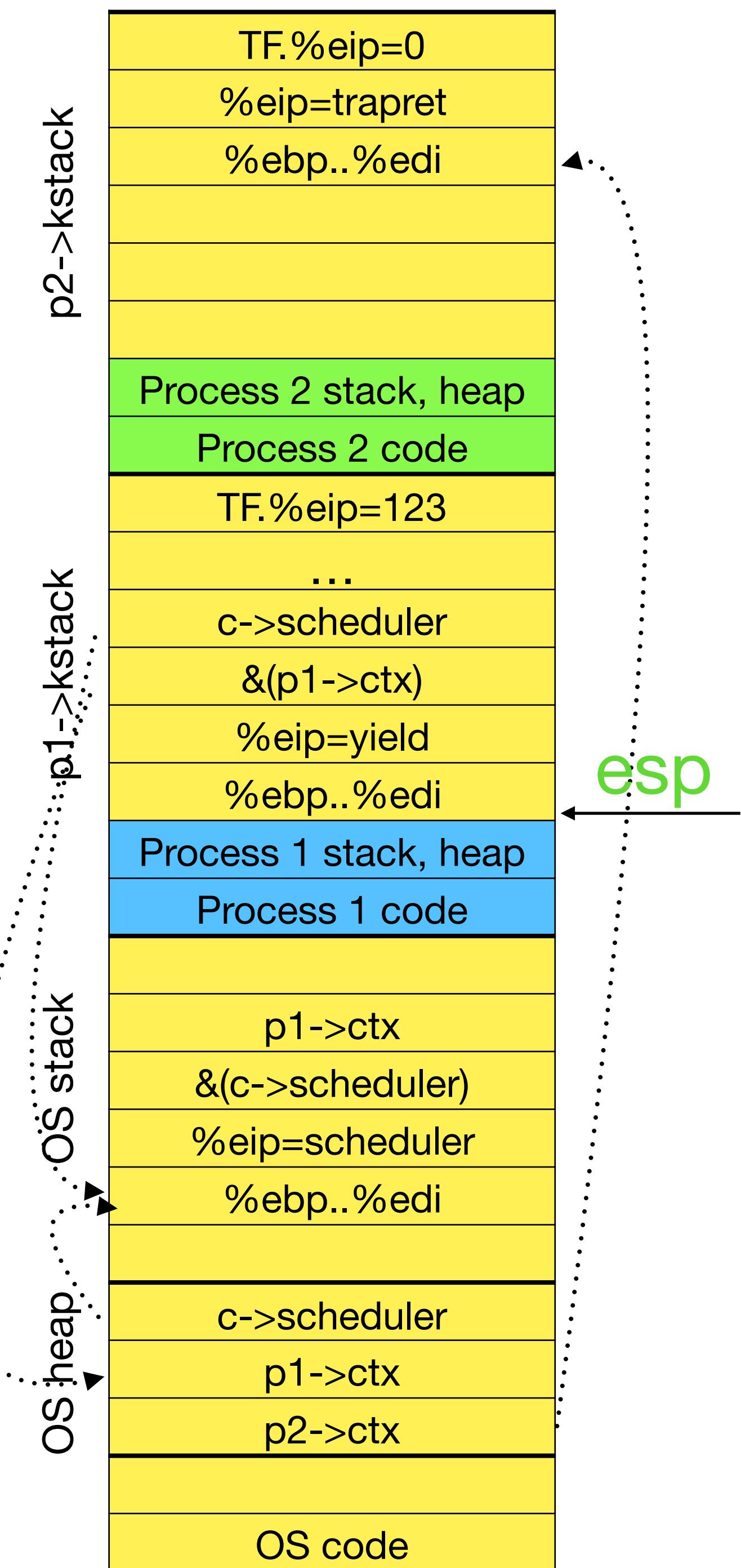
```
.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret
        void scheduler(void) {
            struct proc *p; struct cpu *c = mycpu();
            for(;;){
                for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                    if(p->state != RUNNABLE)
                        continue;
                    p->state = RUNNING;
                    switchuvm(p);
                    swtch(&(c->scheduler), p->context);
                }
            }
        }
        void yield(void) {
            struct proc *p = myproc();
            p->state = RUNNABLE;
            swtch(&p->context, c->scheduler);
        }
        void trap(struct trapframe *tf) {
            ..
            if(tf->trapno == T_IRQ0+IRQ_TIMER)
                yield();
        }
    }
```



Context switching in action: taking control

p18-sched

```
.globl swtch          void scheduler(void) {  
swtch:  
    movl 4(%esp), %eax  
    movl 8(%esp), %edx  
    pushl %ebp  
    pushl %ebx  
    pushl %esi  
    pushl %edi  
    movl %esp, (%eax)  
    movl %edx, %esp  
    popl %edi  
    popl %esi  
    popl %ebx  
    popl %ebp  
    ret  
    }  
  
    void yield(void) {  
        struct proc *p = myproc();  
        p->state = RUNNABLE;  
        swtch(&p->context, c->scheduler);  
    }  
  
    void trap(struct trapframe *tf) {  
        ..  
        if(tf->trapno == T_IRQ0+IRQ_TIMER)  
            yield();  
    }  
}
```



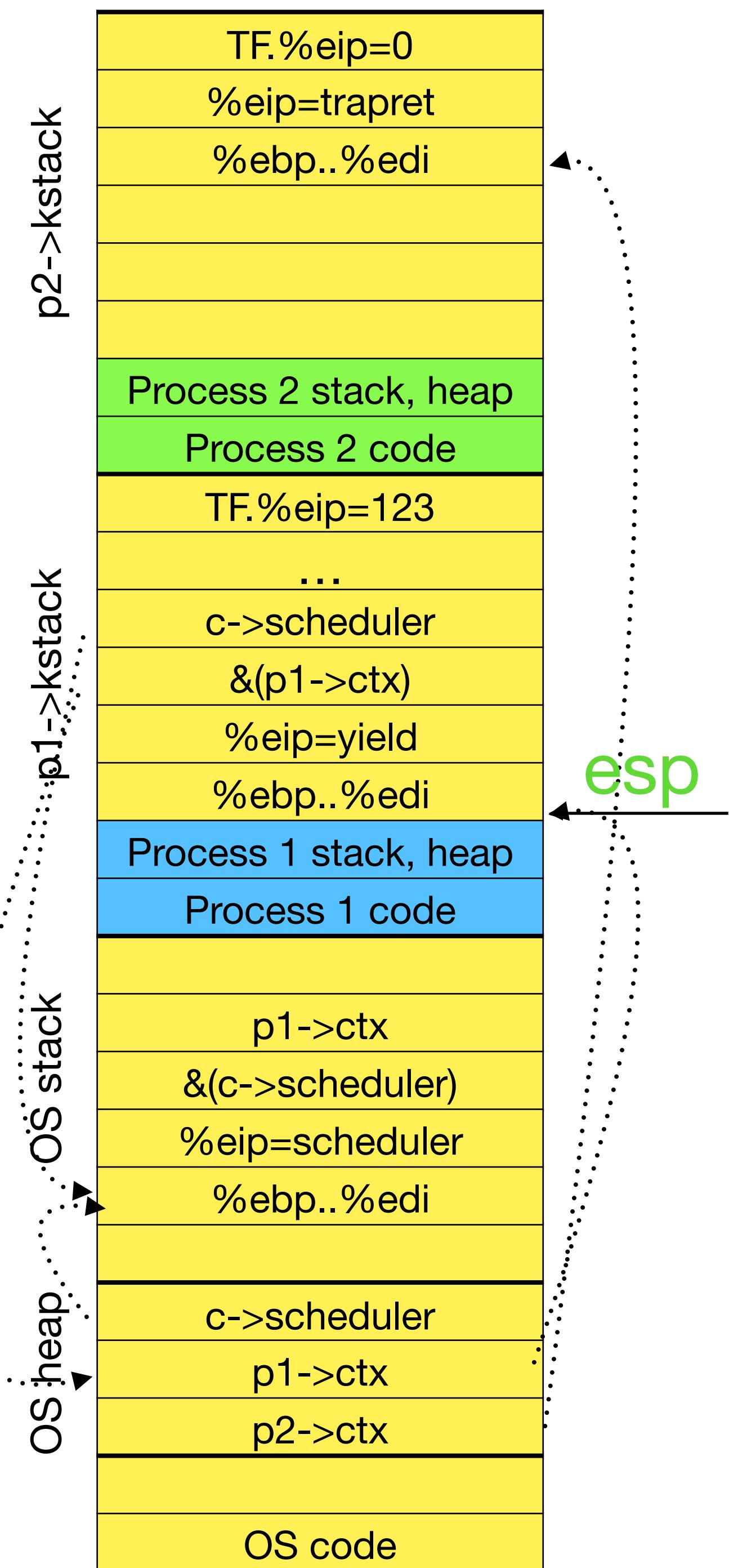
Context switching in action: taking control

p18-sched

```

.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret
    void scheduler(void) {
        struct proc *p; struct cpu *c = mycpu();
        for(;;){
            for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                if(p->state != RUNNABLE)
                    continue;
                p->state = RUNNING;
                switchuvm(p);
                swtch(&(c->scheduler), p->context);
            }
        }
    }
    void yield(void) {
        struct proc *p = myproc();
        p->state = RUNNABLE;
        swtch(&p->context, c->scheduler);
    }
    void trap(struct trapframe *tf) {
        ..
        if(tf->trapno == T_IRQ0+IRQ_TIMER)
            yield();
    }

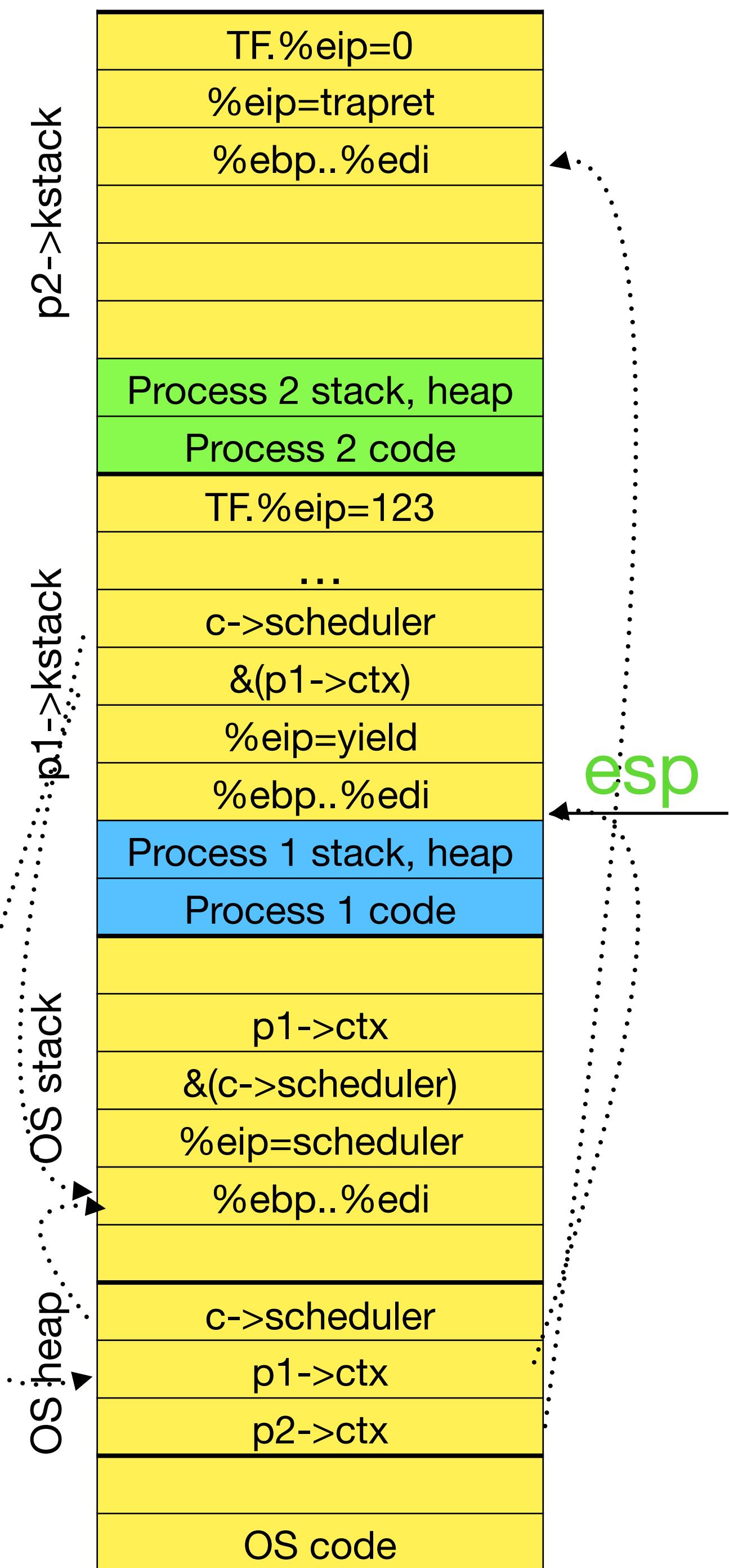
```



Context switching in action: taking control

p18-sched

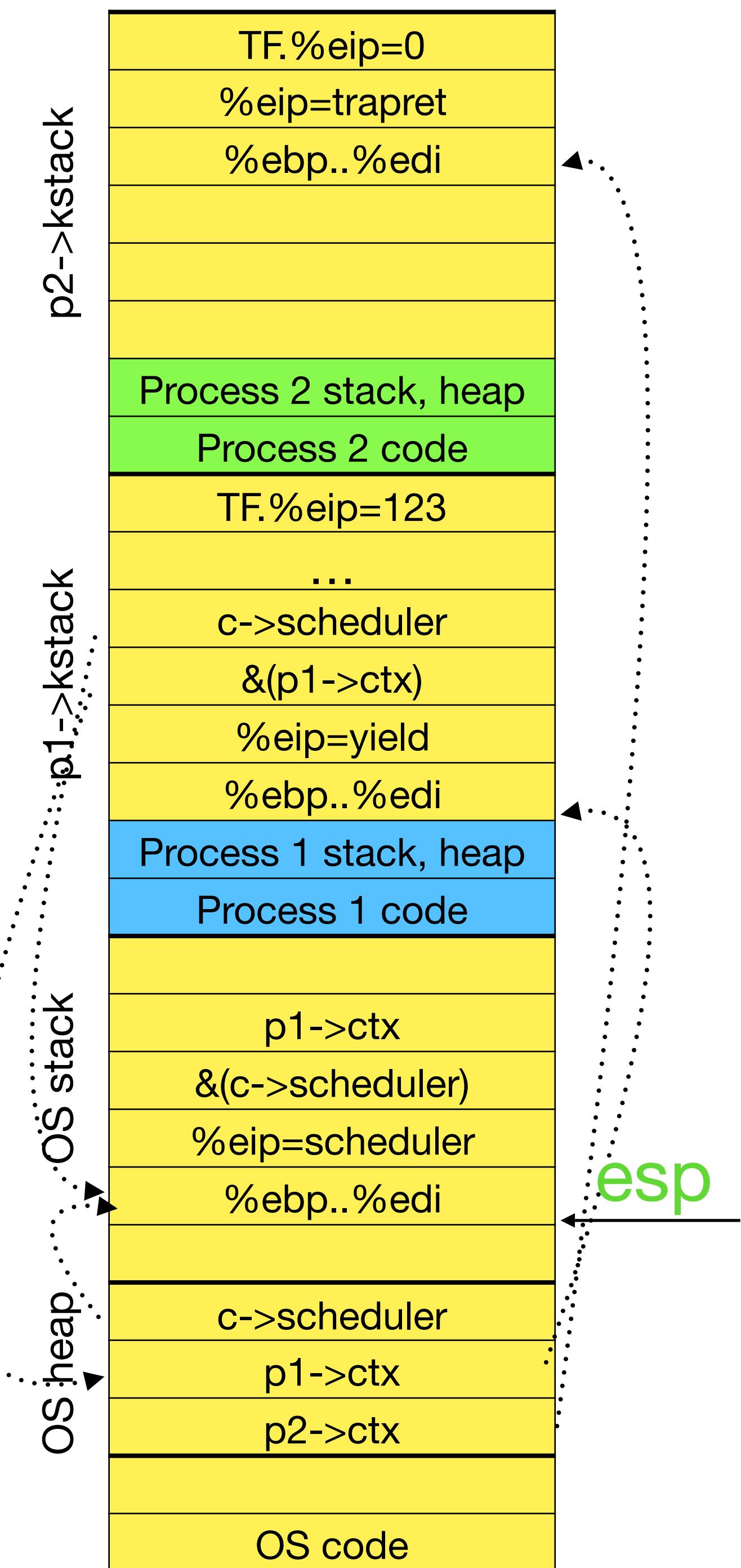
```
.globl swtch          void scheduler(void) {  
swtch:  
    movl 4(%esp), %eax  
    movl 8(%esp), %edx  
    pushl %ebp  
    pushl %ebx  
    pushl %esi  
    pushl %edi  
    movl %esp, (%eax)  
    movl %edx, %esp  
    popl %edi  
    popl %esi  
    popl %ebx  
    popl %ebp  
    ret  
    }  
    void yield(void) {  
        struct proc *p = myproc();  
        p->state = RUNNABLE;  
        swtch(&p->context, c->scheduler);  
    }  
    void trap(struct trapframe *tf) {  
        ..  
        if(tf->trapno == T_IRQ0+IRQ_TIMER)  
            yield();  
    }  
}
```



Context switching in action: taking control

p18-sched

```
.globl swtch          void scheduler(void) {  
swtch:  
    movl 4(%esp), %eax  
    movl 8(%esp), %edx  
    pushl %ebp  
    pushl %ebx  
    pushl %esi  
    pushl %edi  
    movl %esp, (%eax)  
    movl %edx, %esp  
    popl %edi  
    popl %esi  
    popl %ebx  
    popl %ebp  
    void yield(void) {  
        struct proc *p = myproc();  
        p->state = RUNNABLE;  
        swtch(&p->context, c->scheduler);  
    }  
    ..  
    if(tf->trapno == T_IRQ0+IRQ_TIMER)  
        yield();  
}
```



Context switching in action: taking control

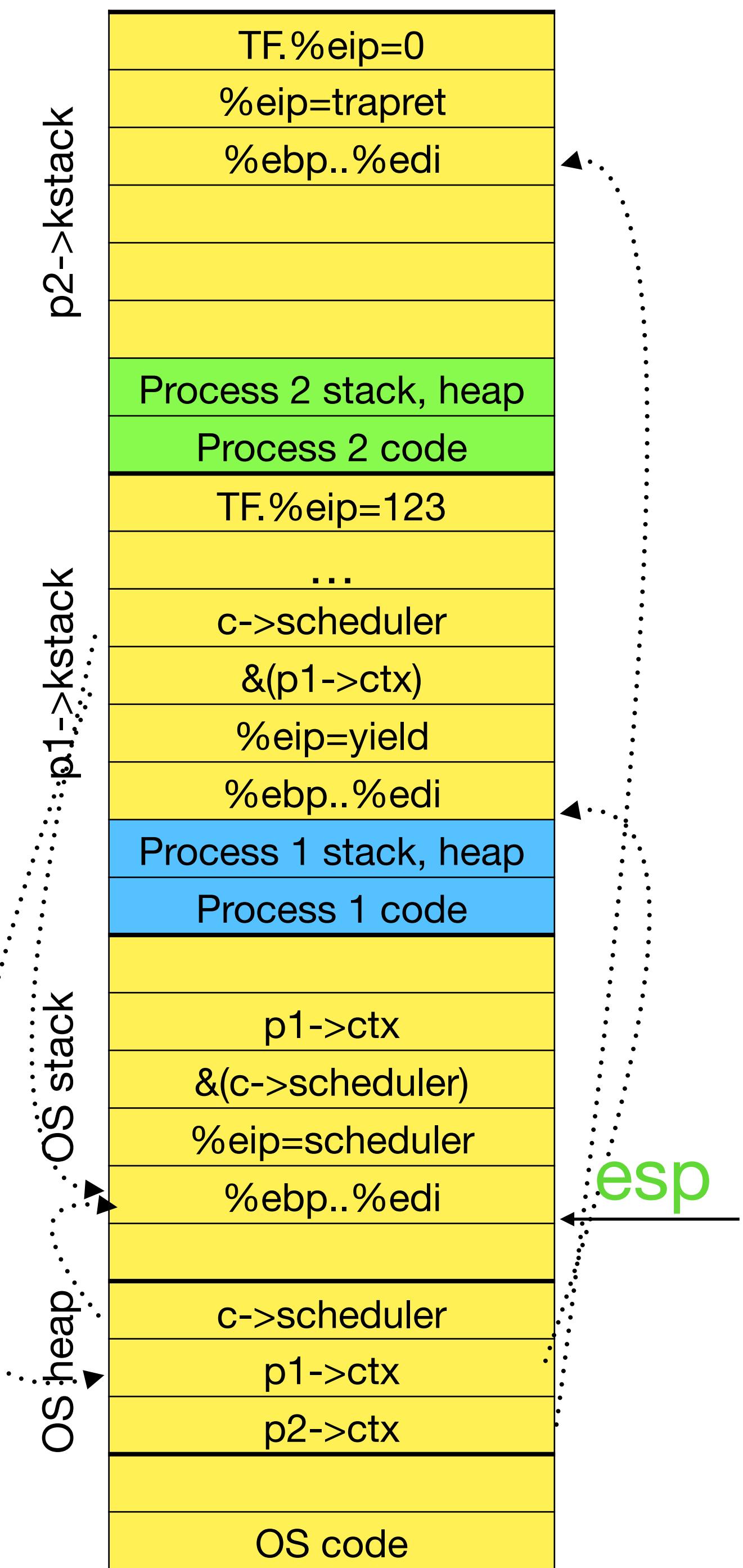
p18-sched

```

.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    ret
        void scheduler(void) {
            struct proc *p; struct cpu *c = mycpu();
            for(;;){
                for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                    if(p->state != RUNNABLE)
                        continue;
                    p->state = RUNNING;
                    switchuvm(p);
                    swtch(&(c->scheduler), p->context);
                }
            }
        }
        void yield(void) {
            struct proc *p = myproc();
            p->state = RUNNABLE;
            swtch(&p->context, c->scheduler);
        }
        void trap(struct trapframe *tf) {
            ..
            if(tf->trapno == T_IRQ0+IRQ_TIMER)
                yield();
        }
    }
}

```

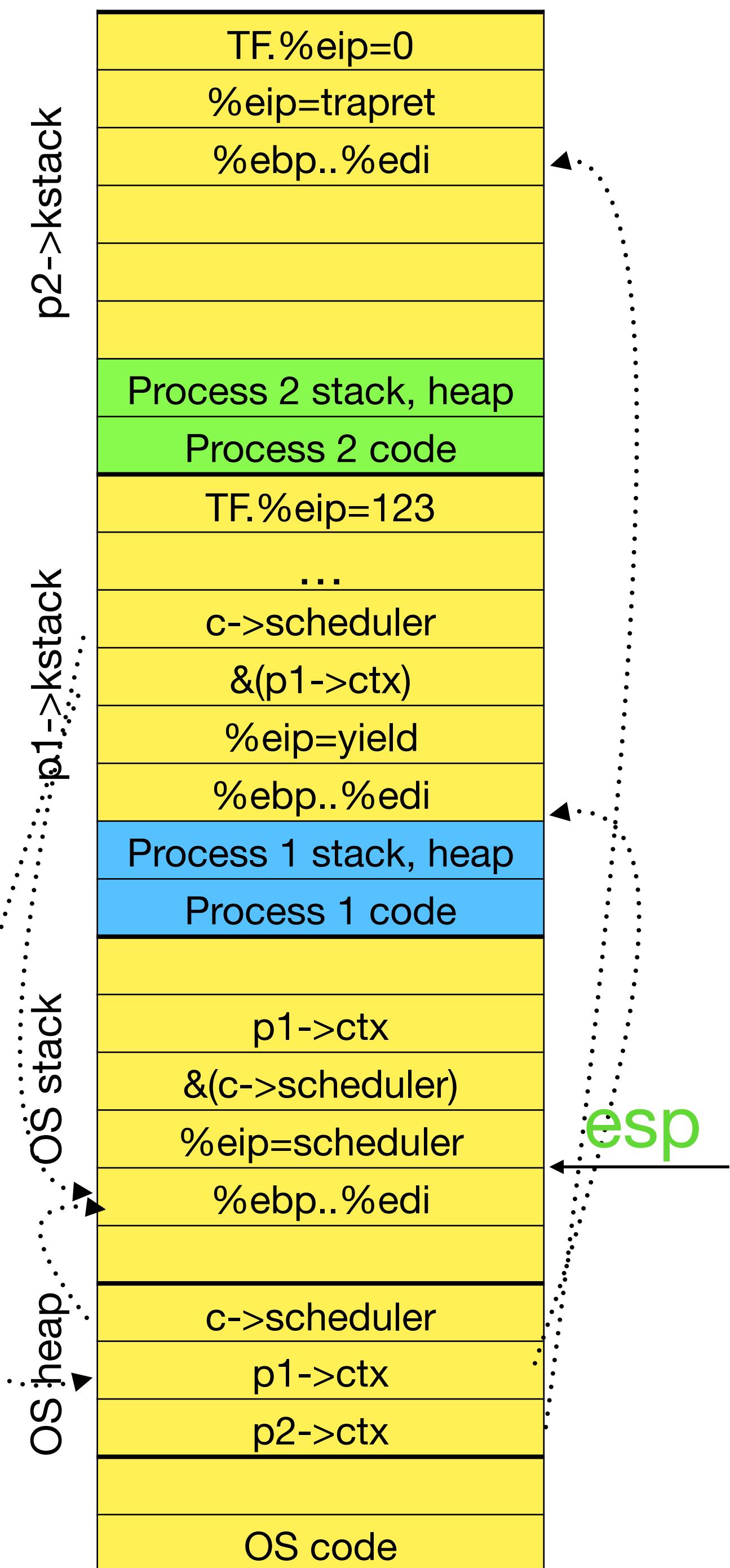
eip



Context switching in action: taking control

p18-sched

```
.globl swtch          void scheduler(void) {  
swtch:  
    movl 4(%esp), %eax  
    movl 8(%esp), %edx  
    pushl %ebp  
    pushl %ebx  
    pushl %esi  
    pushl %edi  
    movl %esp, (%eax)  
    movl %edx, %esp  
    }  
    popl %edi  
    popl %esi  
    popl %ebx  
    popl %ebp  
    void yield(void) {  
        struct proc *p = myproc();  
        p->state = RUNNABLE;  
        swtch(&p->context, c->scheduler);  
    }  
}  
  
void trap(struct trapframe *tf) {  
    ..  
    if(tf->trapno == T_IRQ0+IRQ_TIMER)  
        yield();  
}
```



Context switching in action: taking control

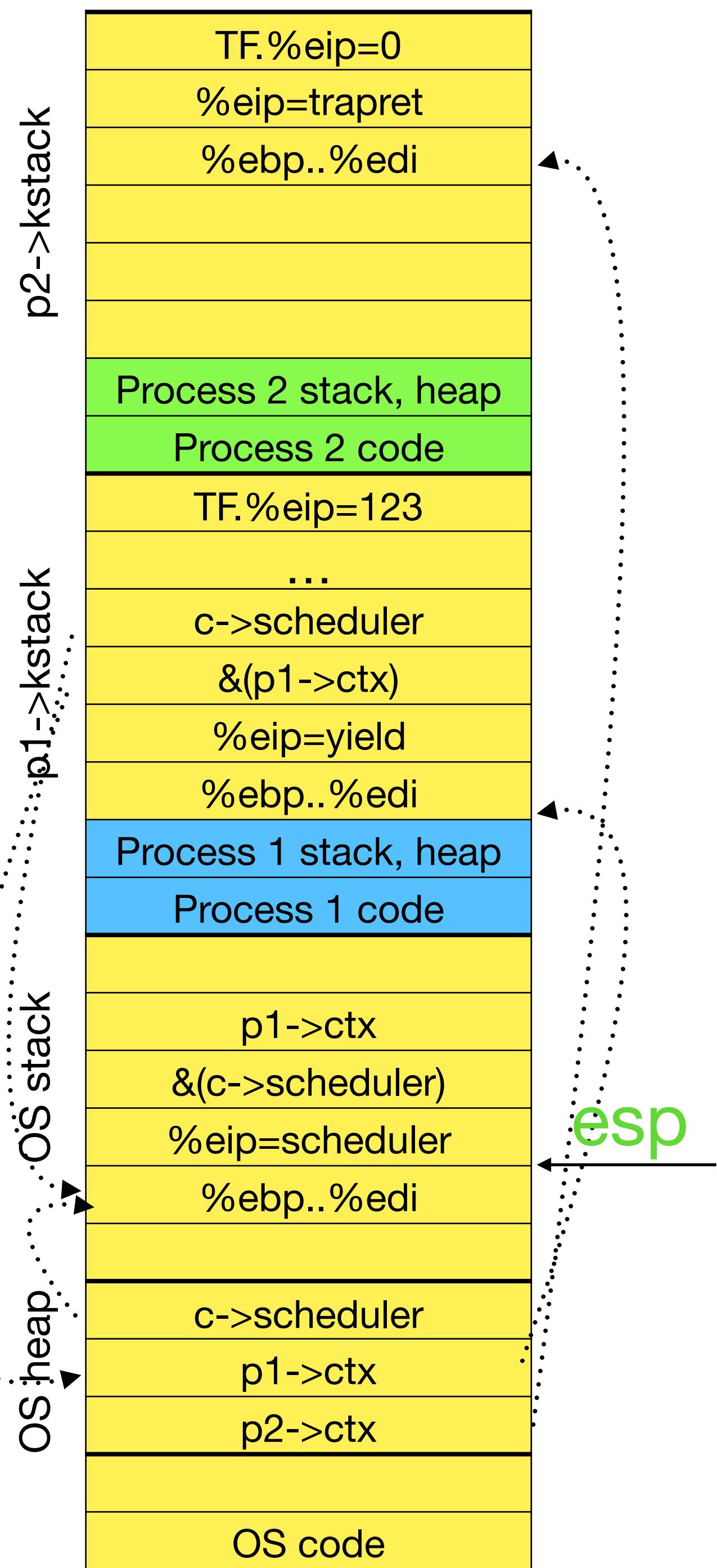
p18-sched

```

.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    void yield(void) {
        struct proc *p = myproc();
        p->state = RUNNABLE;
        swtch(&p->context, c->scheduler);
    }
    void scheduler(void) {
        struct proc *p; struct cpu *c = mycpu();
        for(;;){
            for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                if(p->state != RUNNABLE)
                    continue;
                p->state = RUNNING;
                switchuvm(p);
                swtch(&(c->scheduler), p->context);
            }
        }
    }
    void trap(struct trapframe *tf) {
        ..
        if(tf->trapno == T_IRQ0+IRQ_TIMER)
            yield();
    }
}

```

eip → ret



Context switching in action: taking control

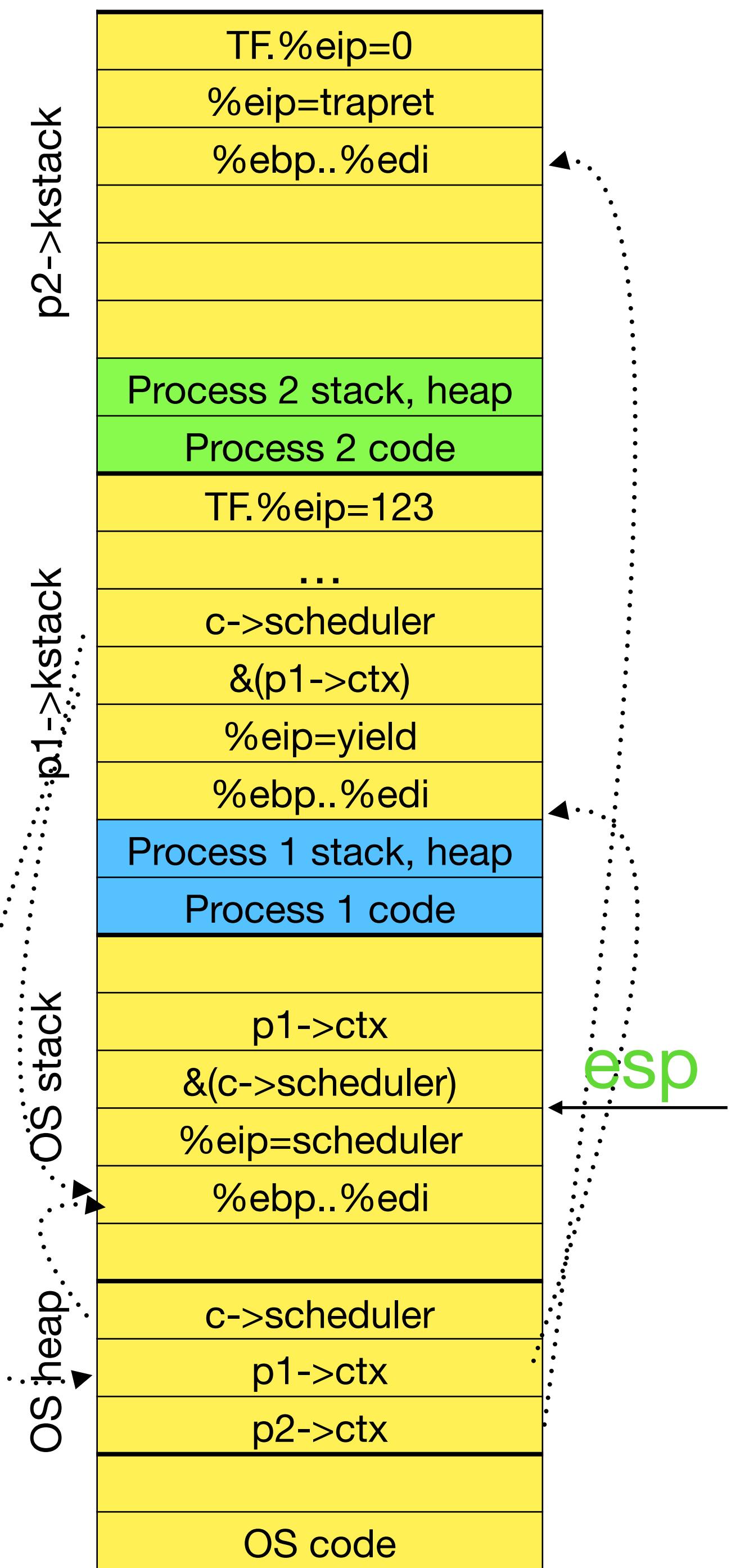
p18-sched

```

.globl swtch
swtch:
    movl 4(%esp), %eax
    movl 8(%esp), %edx
    pushl %ebp
    pushl %ebx
    pushl %esi
    pushl %edi
    movl %esp, (%eax)
    movl %edx, %esp
    popl %edi
    popl %esi
    popl %ebx
    popl %ebp
    void yield(void) {
        struct proc *p = myproc();
        p->state = RUNNABLE;
        swtch(&p->context, c->scheduler);
    }
    void scheduler(void) {
        struct proc *p; struct cpu *c = mycpu();
        for(;;){
            for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){
                if(p->state != RUNNABLE)
                    continue;
                p->state = RUNNING;
                switchuvm(p);
                swtch(&(c->scheduler), p->context);
            }
        }
    }
    void trap(struct trapframe *tf) {
        ..
        if(tf->trapno == T_IRQ0+IRQ_TIMER)
            yield();
    }

```

eip → ret



Context switching in action: taking control

p18-sched

```
.globl swtch          void scheduler(void) {  
swtch:  
    movl 4(%esp), %eax  
    movl 8(%esp), %edx  
    pushl %ebp  
    pushl %ebx  
    pushl %esi  
    pushl %edi  
    movl %esp, (%eax)  
    movl %edx, %esp  
    popl %edi  
    popl %esi  
    popl %ebx  
    popl %ebp  
    ret  
    }  
  
    struct proc *p;  struct cpu *c = mycpu();  
    for(;;){  
        for(p = ptable.proc; p < &ptable.proc[NPROC]; p++){  
            if(p->state != RUNNABLE)  
                continue;  
            p->state = RUNNING;  
            eip  
            switchuvm(p);  
            swtch(&(c->scheduler), p->context);  
        }  
    }  
  
void yield(void) {  
    struct proc *p = myproc();  
    p->state = RUNNABLE;  
    swtch(&p->context, c->scheduler);  
}  
  
void trap(struct trapframe *tf) {  
    ..  
    if(tf->trapno == T_IRQ0+IRQ_TIMER)  
        yield();  
}
```

