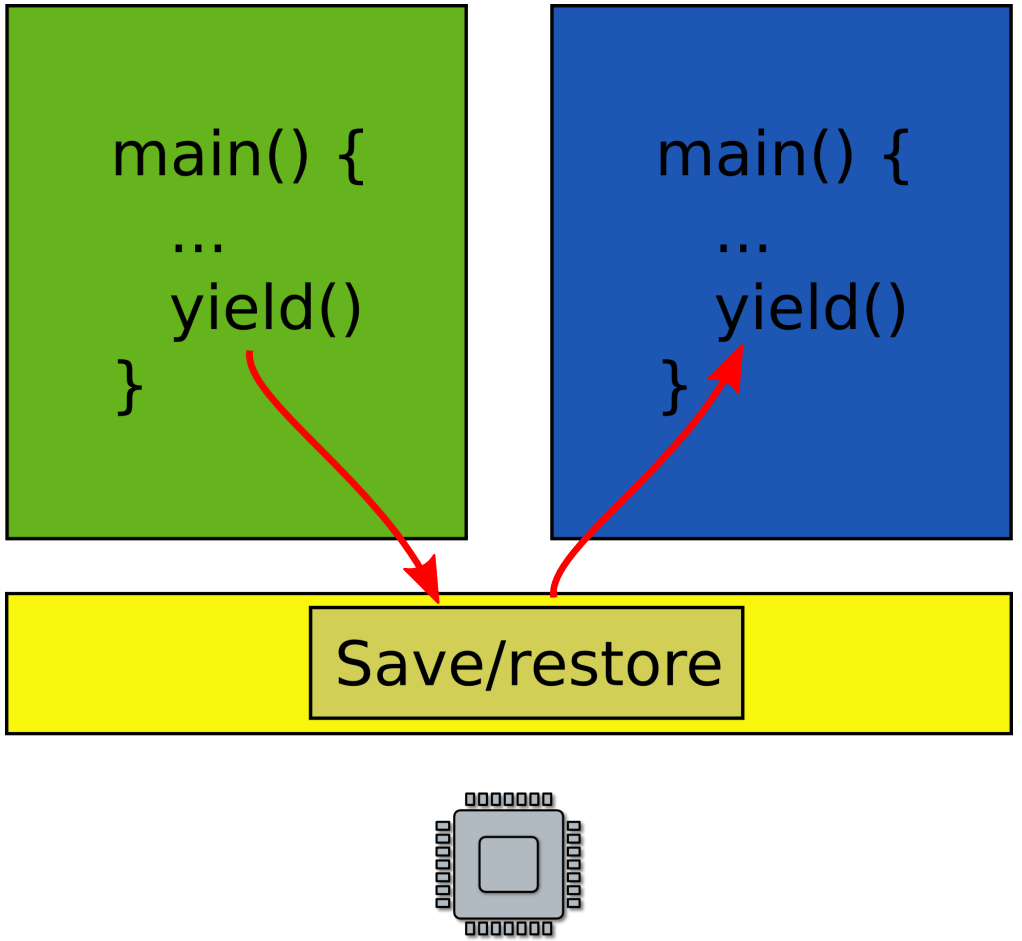
143A: Principles of Operating Systems

Lecture 5: Address translation

Anton Burtsev
October, 2017

Two programs one memory

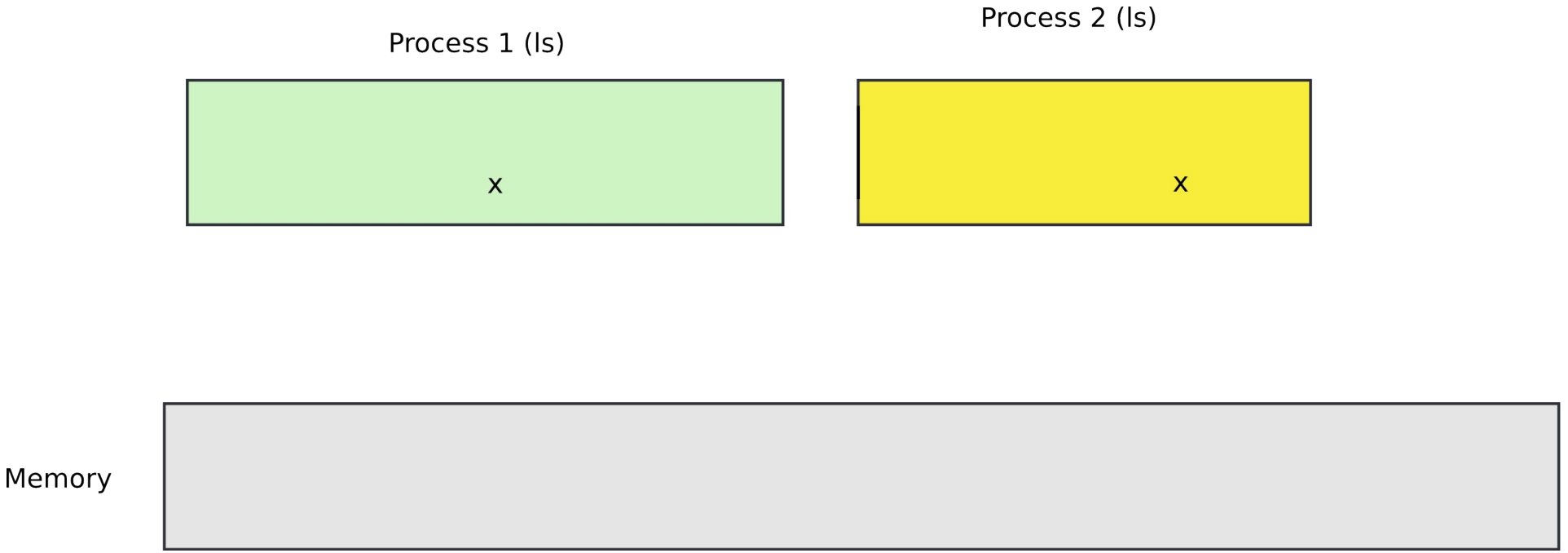


Very much like car sharing

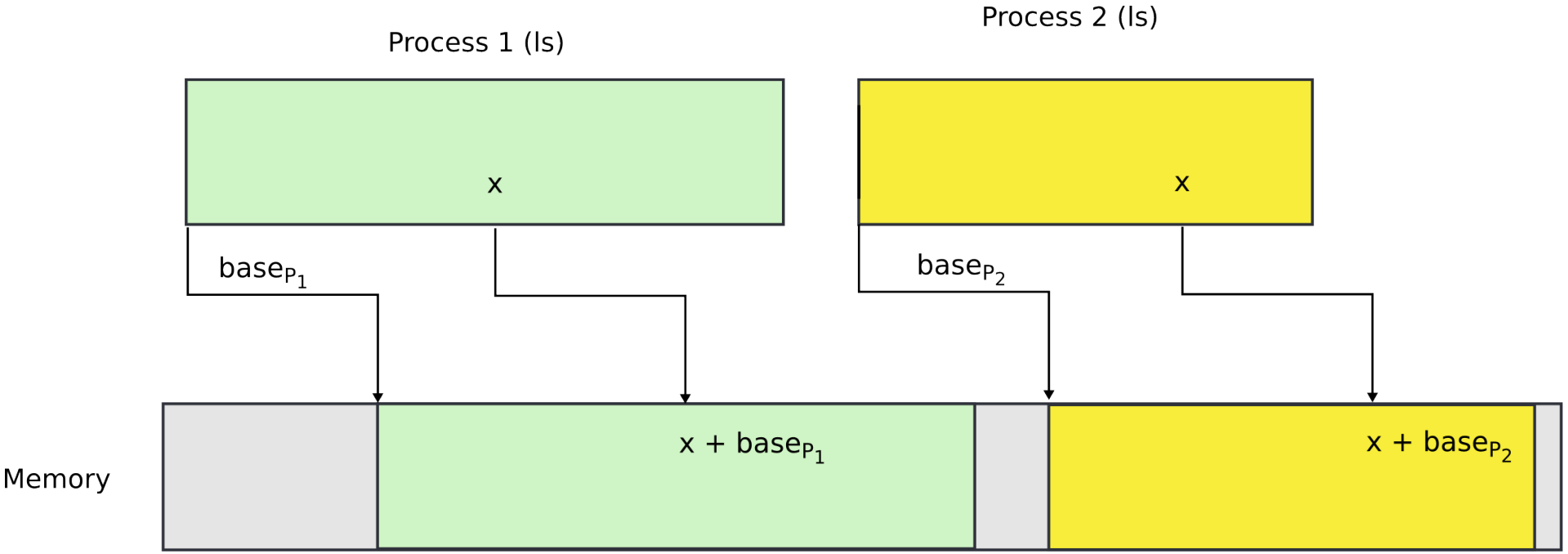
What are we aiming for?

- Illusion of a private address space
  - Identical copy of an address space in multiple programs
    - Remember `fork()`?
  - Simplifies software architecture
    - One program is not restricted by the memory layout of the others

Two processes, one memory?



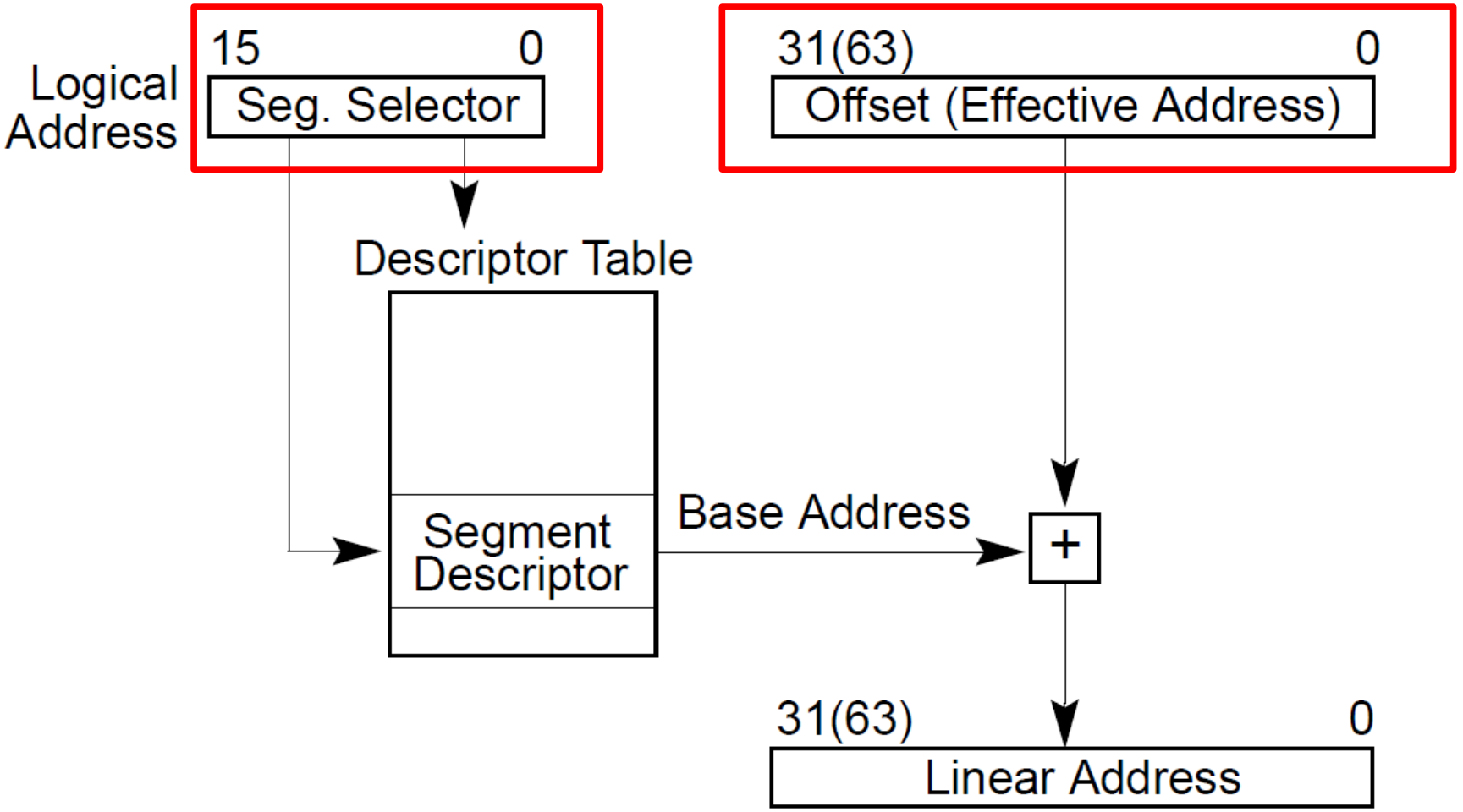
Two processes, one memory?



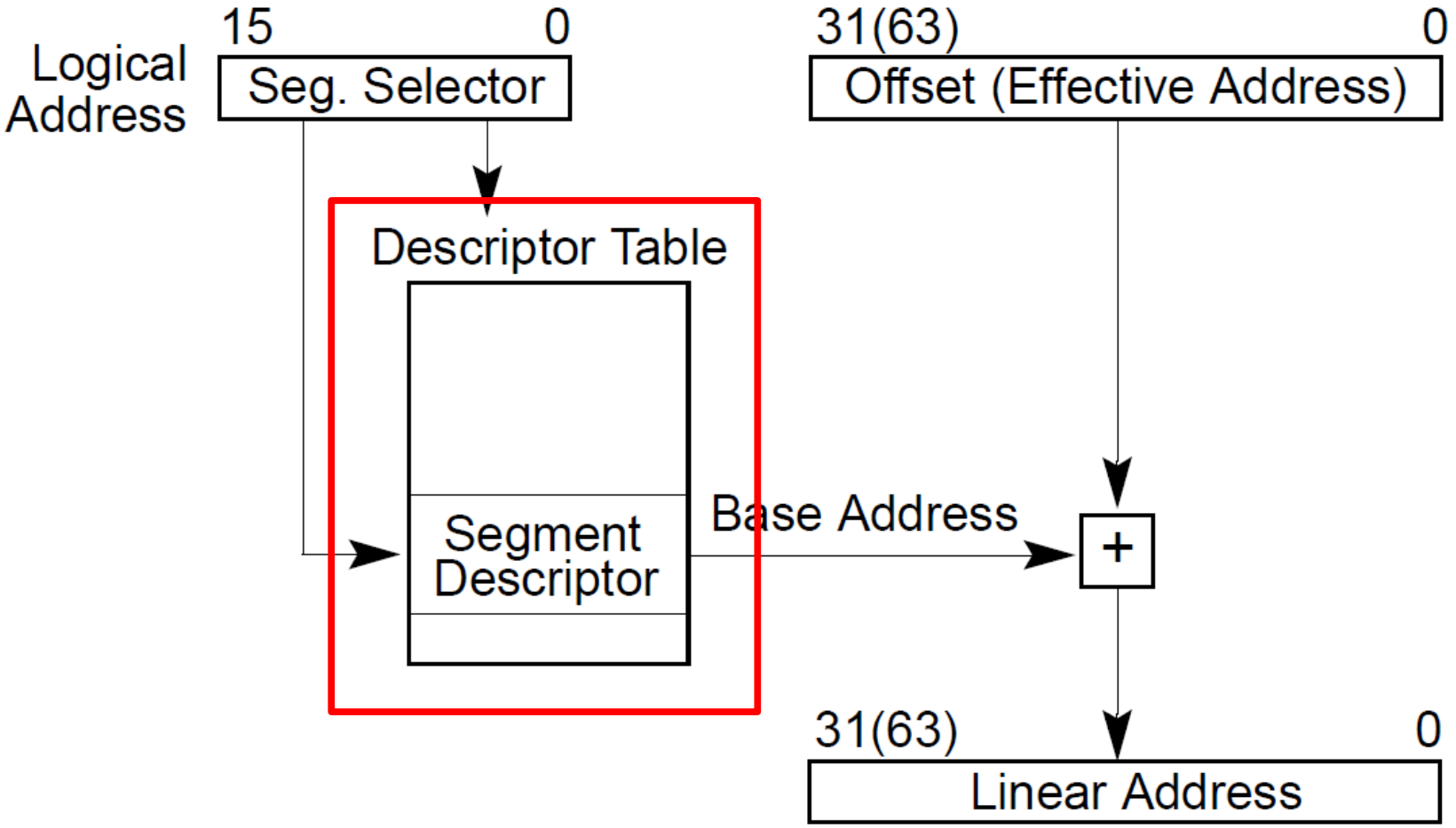
This is called segmentation

All addresses are logical address

- They consist of two parts
  - Segment selector (16 bit) + offset (32 bi

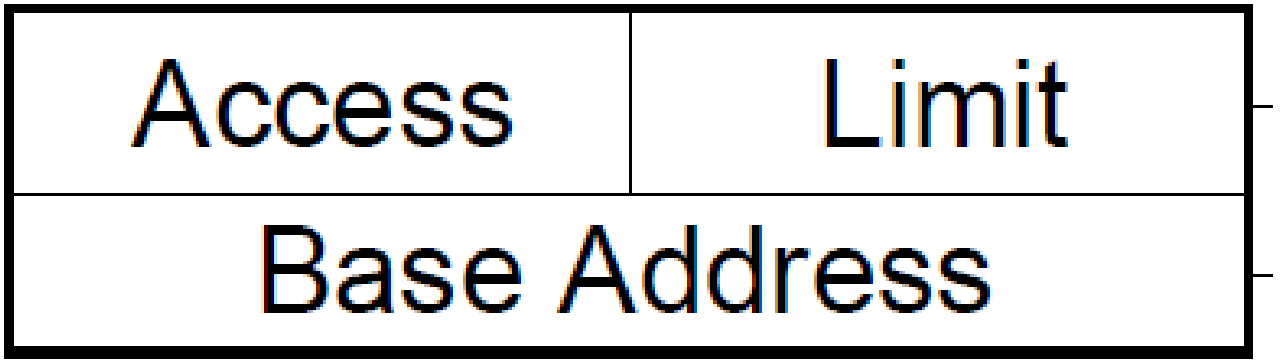


- Segment selector (16 bit)
  - Is simply an index into an array (Descriptor Table)
  - That holds segment descriptors
    - Base and limit (size) for each segment

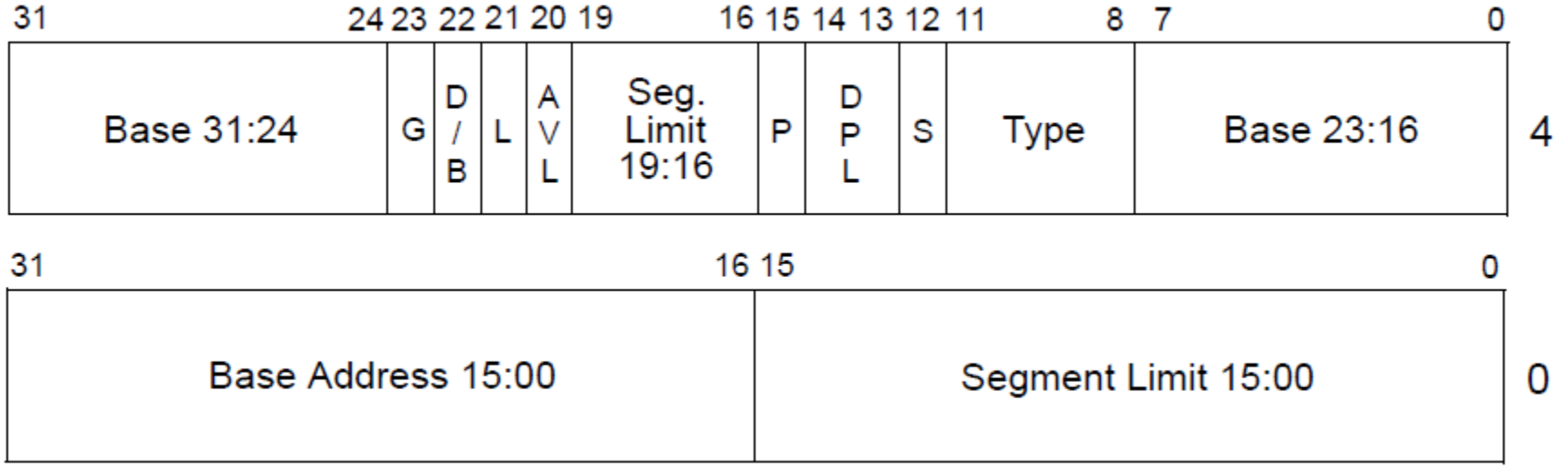


Elements of that array are **segment descriptors**

- Base address
  - 0 – 4 GB
- Limit (size)
  - 0 – 4 GB
- Access rights
  - Executable, readable, writable
  - Privilege level (0 - 3)

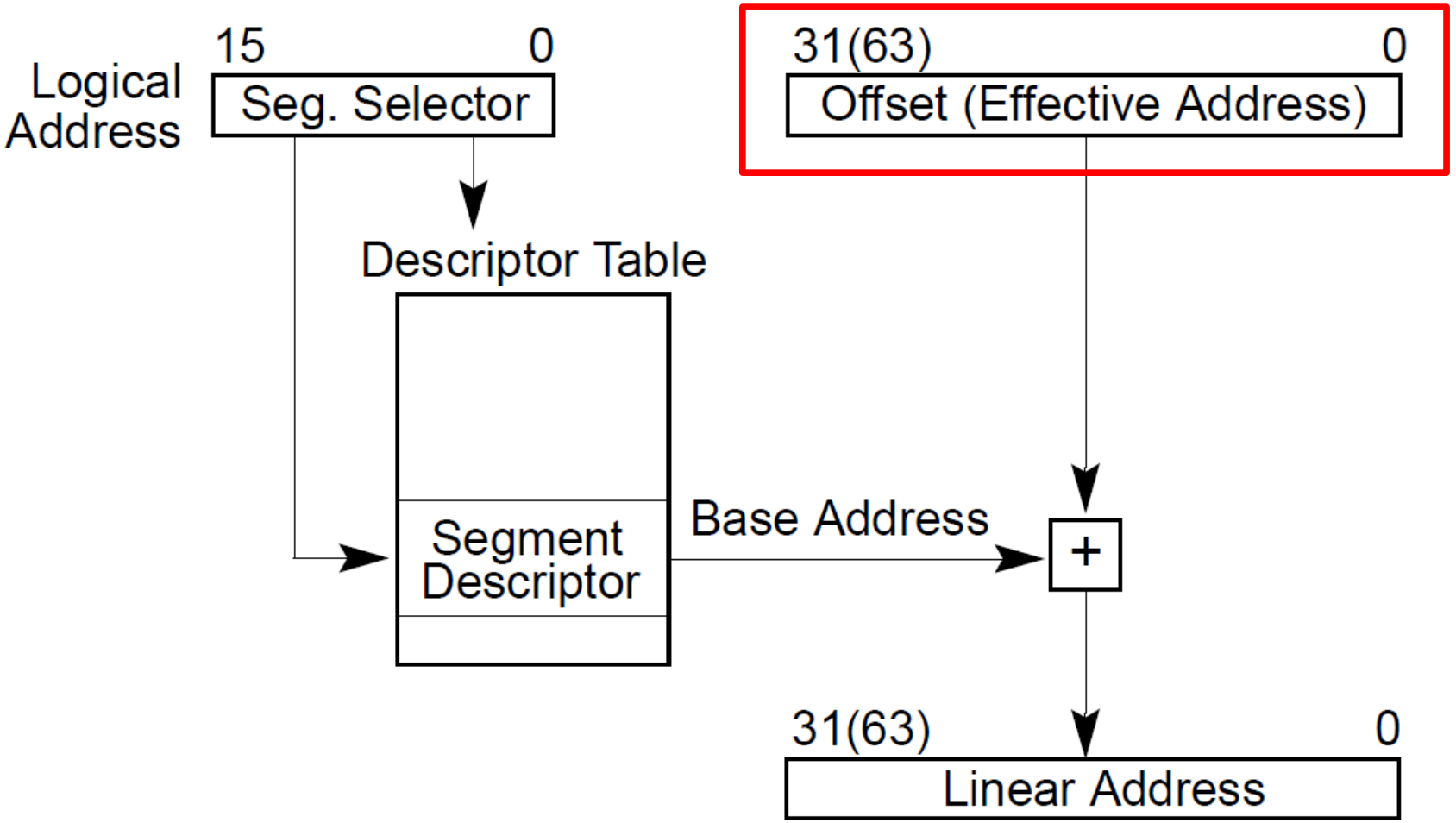


Segment descriptors

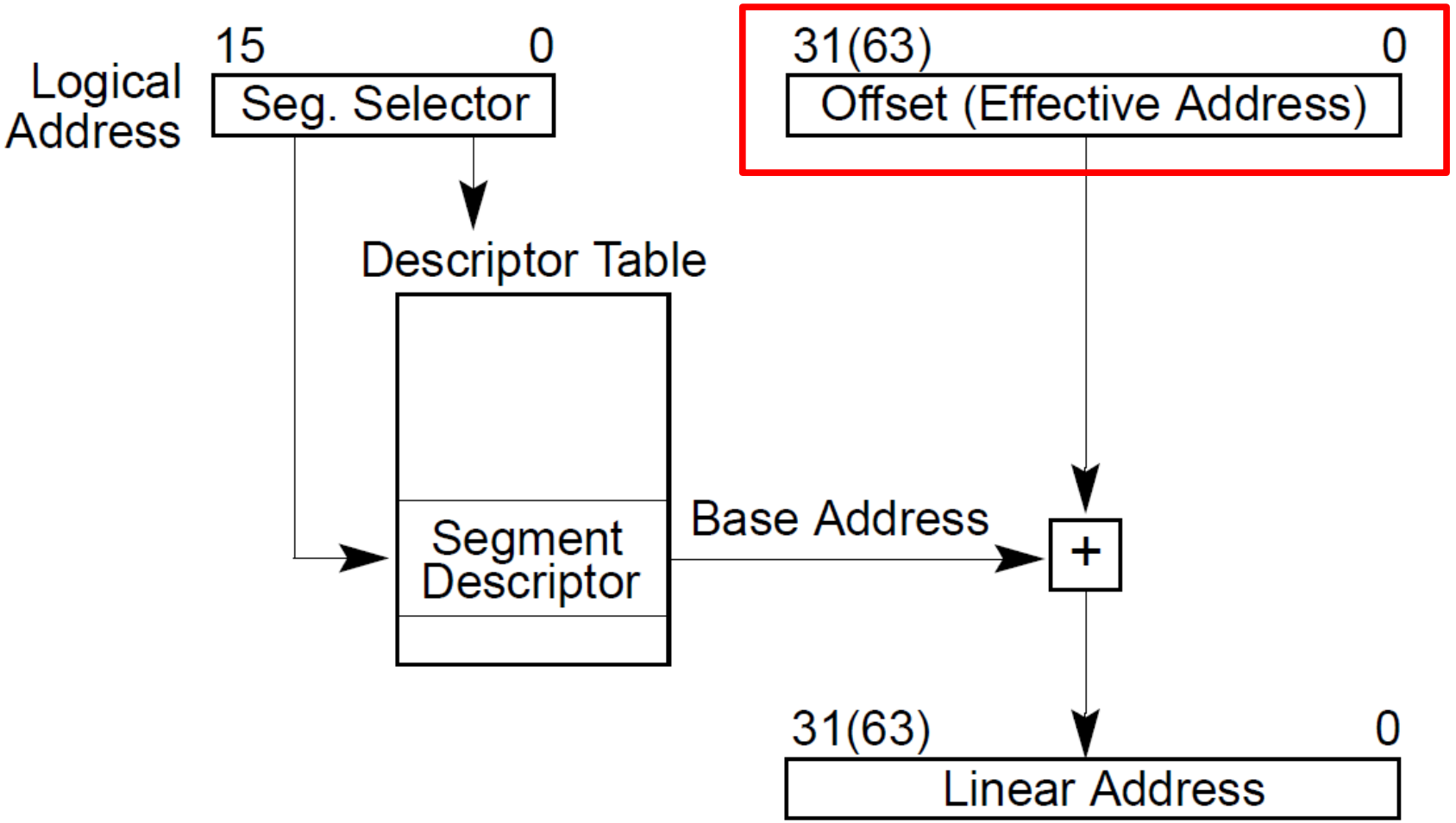


L — 64-bit code segment (IA-32e mode only)
AVL — Available for use by system software
BASE — Segment base address
D/B — Default operation size (0 = 16-bit segment; 1 = 32-bit segment)
DPL — Descriptor privilege level
G — Granularity
LIMIT — Segment Limit
P — Segment present
S — Descriptor type (0 = system;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

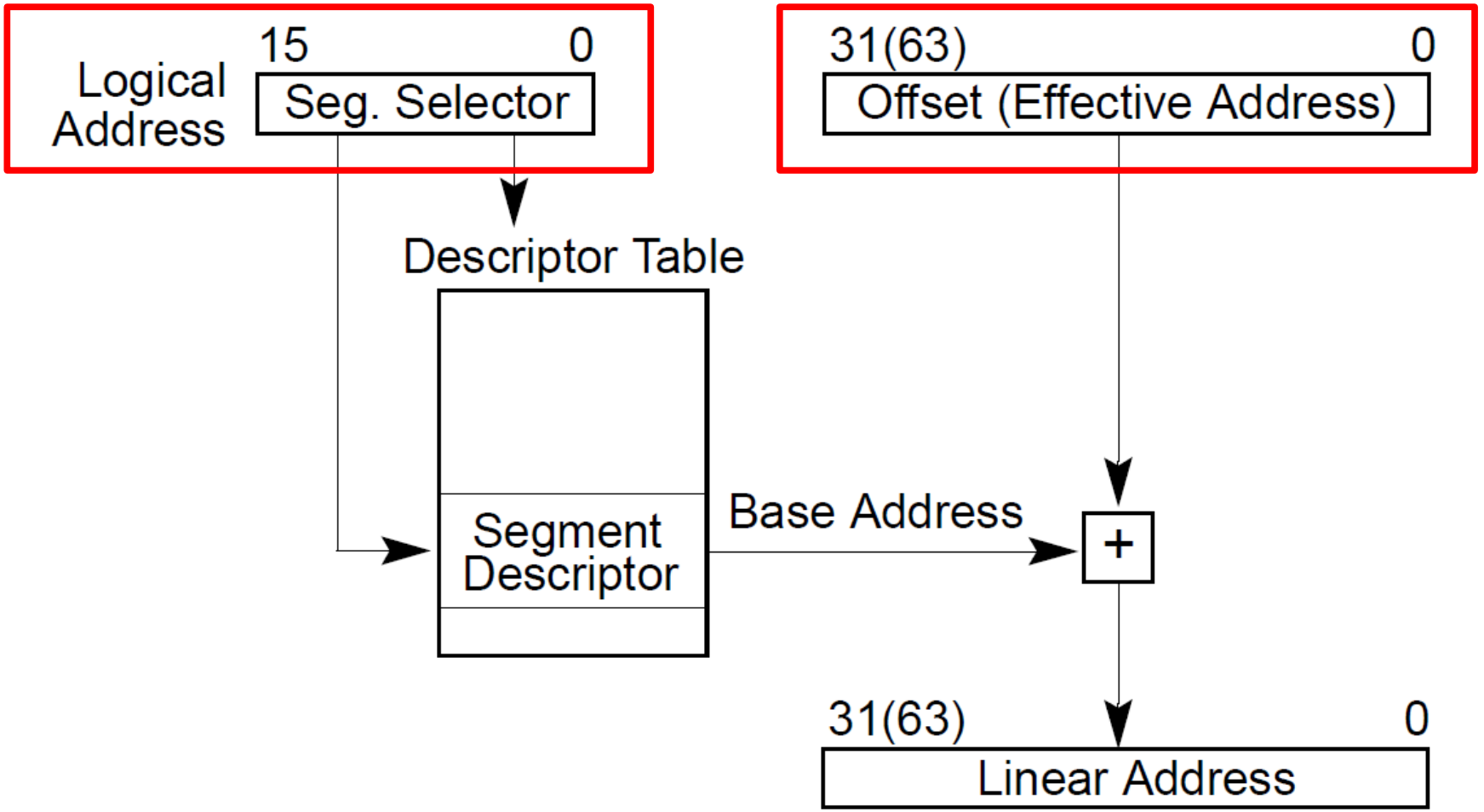
- Offsets into segments (x in our example) or “Effective addresses” are in registers



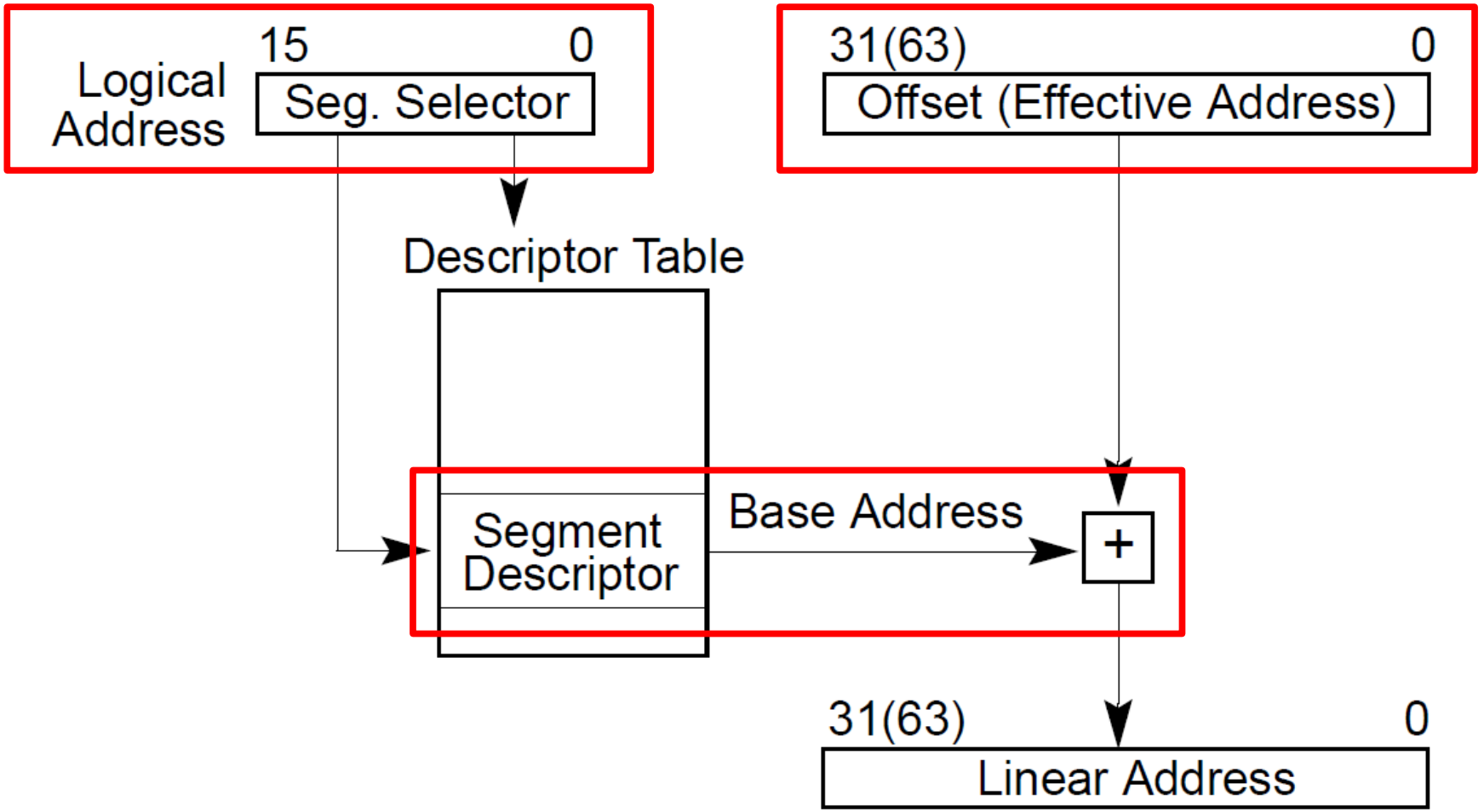
- Logical addresses are translated into physical
  - *Effective address + DescriptorTable[selector].Base*



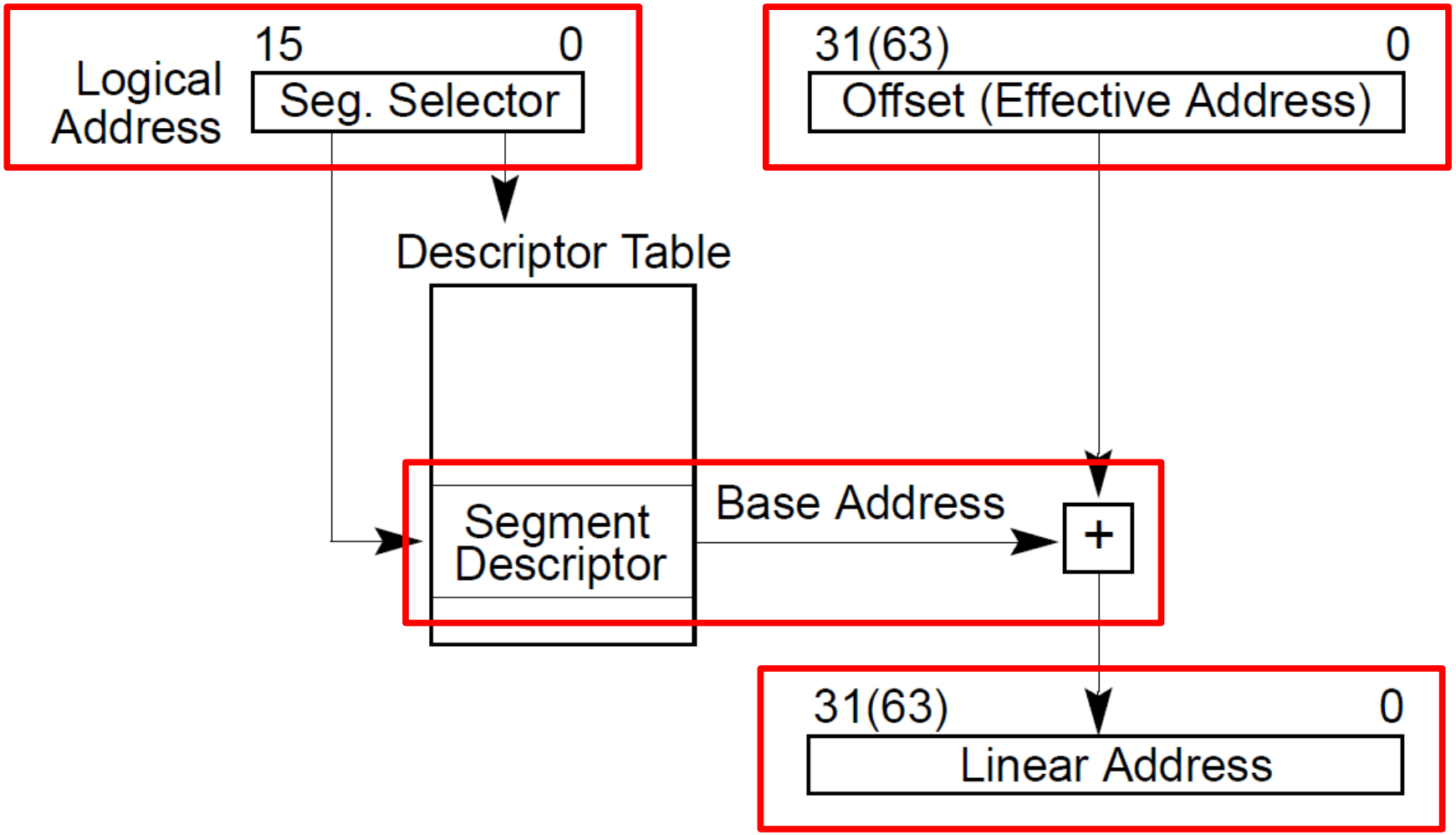
- Logical addresses are translated into physical
  - Effective address + DescriptorTable[selector].Base



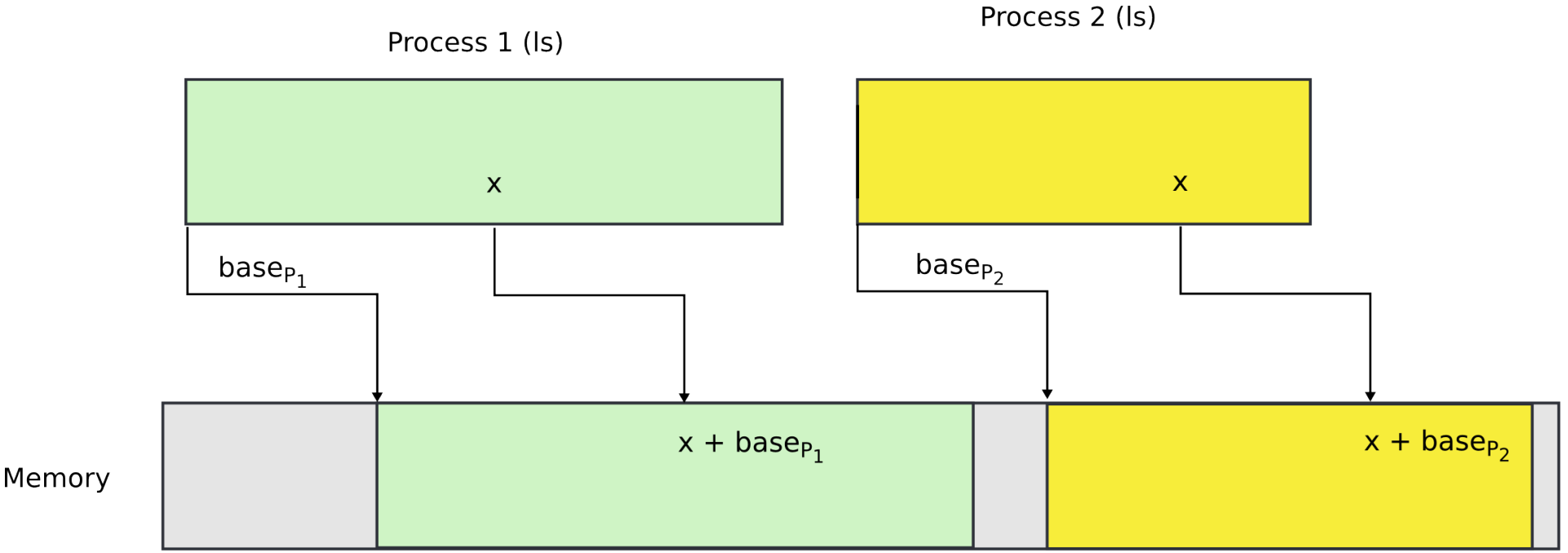
- Logical addresses are translated into physical
  - Effective address + DescriptorTable[selector].Base



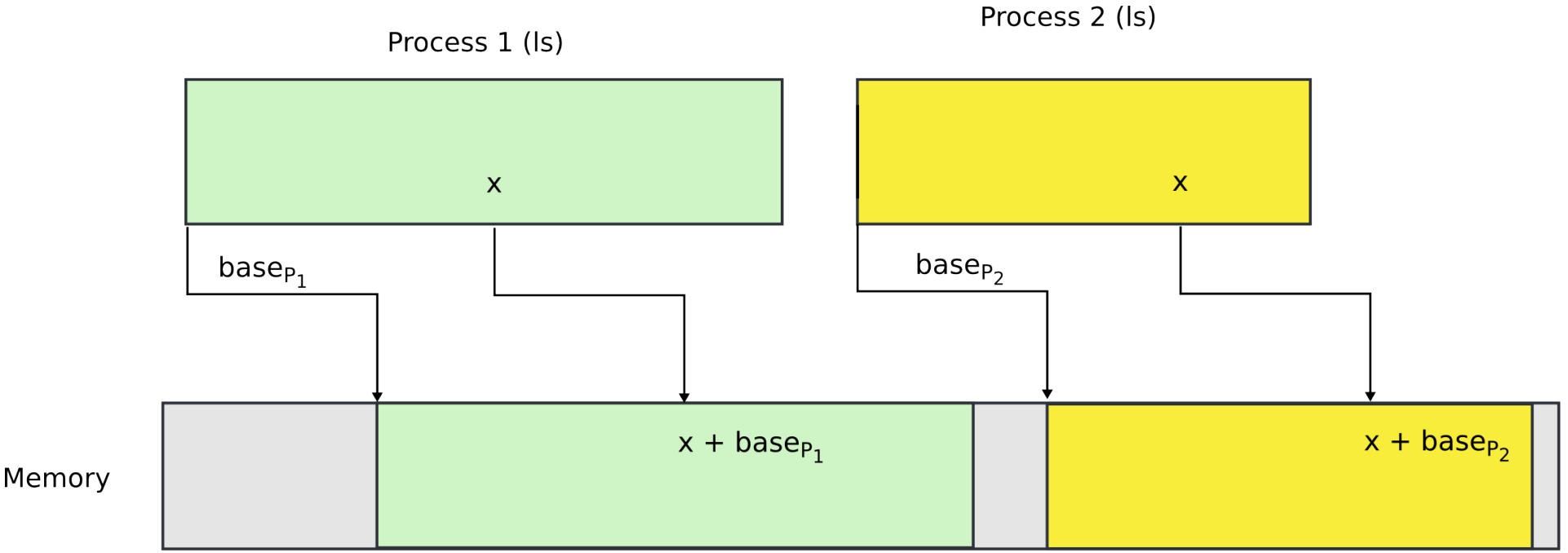
- Logical addresses are translated into physical
  - Effective address + DescriptorTable[selector].Base



Same picture



- Offsets (effective addresses) are in registers
  - *Effective address + DescriptorTable[selector].Base*
  - **But where is the selector?**



Right! Segment registers

- Hold 16 bit segment selectors
  - Pointers into a special table
  - Global or local descriptor table
- Segments are associated with one of three types of storage
  - Code
  - Data
  - Stack

Programming model

- Segments for: code, data, stack, “extra”
  - A program can have up to 6 total segments
  - Segments identified by registers: cs, ds, ss, es, fs, gs

- Prefix all memory accesses with desired segment:
  - `mov eax, ds:0x80` (load offset 0x80 from data into eax)
  - `jmp cs:0xab8` (jump execution to code offset 0xab8)
  - `mov ss:0x40, ecx` (move ecx to stack offset 0x40)

Segmented programming (not real)

```
static int x = 1;
int y; // stack
if (x) {
    y = 1;
    printf (“Boo”);
} else
    y = 0;
```

```
ds:x = 1; // data
ss:y;     // stack
if (ds:x) {
   ss:y = 1;
   cs:printf(ds:“Boo”);
} else
   ss:y = 0;
```

Programming model, cont.

- This is cumbersome, so infer code, data and stack segments by instruction type:
  - Control-flow instructions use code segment (jump, call)
  - Stack management (push/pop) uses stack
  - Most loads/stores use data segment
- Extra segments (es, fs, gs) must be used explicitly

Code segment

- Code
  - CS register
  - EIP is an offset inside the segment stored in CS
- Can only be changed with
  - procedure calls,
  - interrupt handling, or
  - task switching

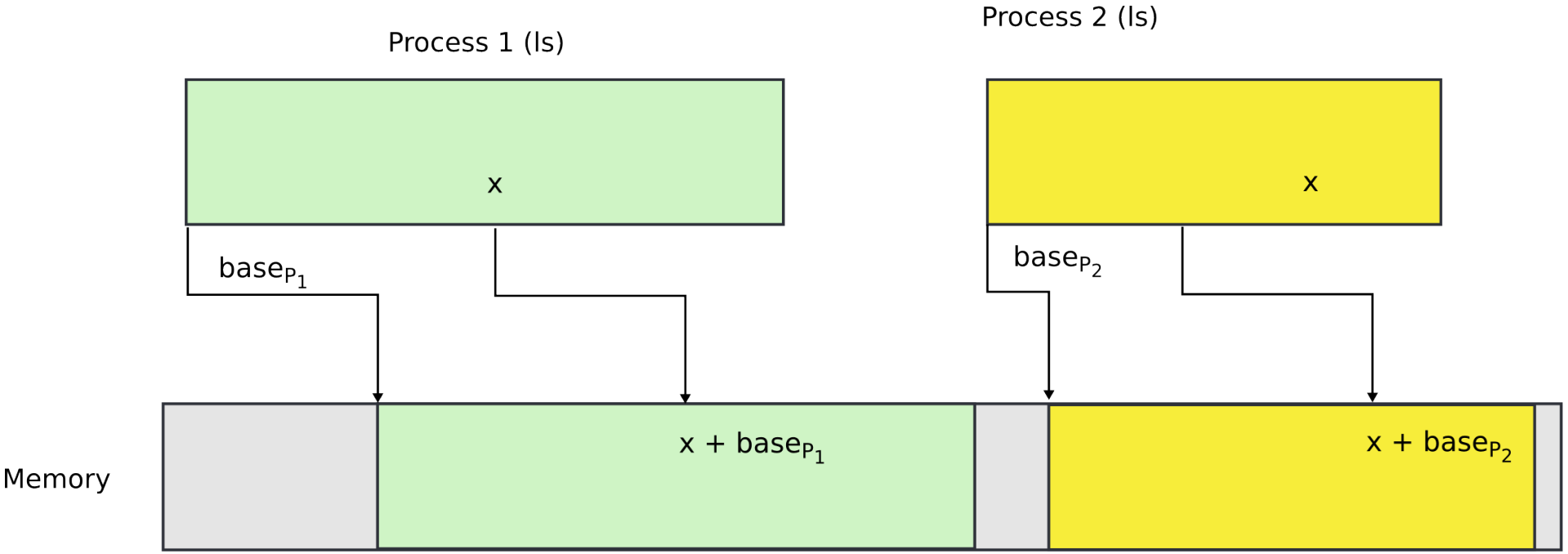
Data segment

- Data
  - DS, ES, FS, GS
  - 4 possible data segments can be used at the same time

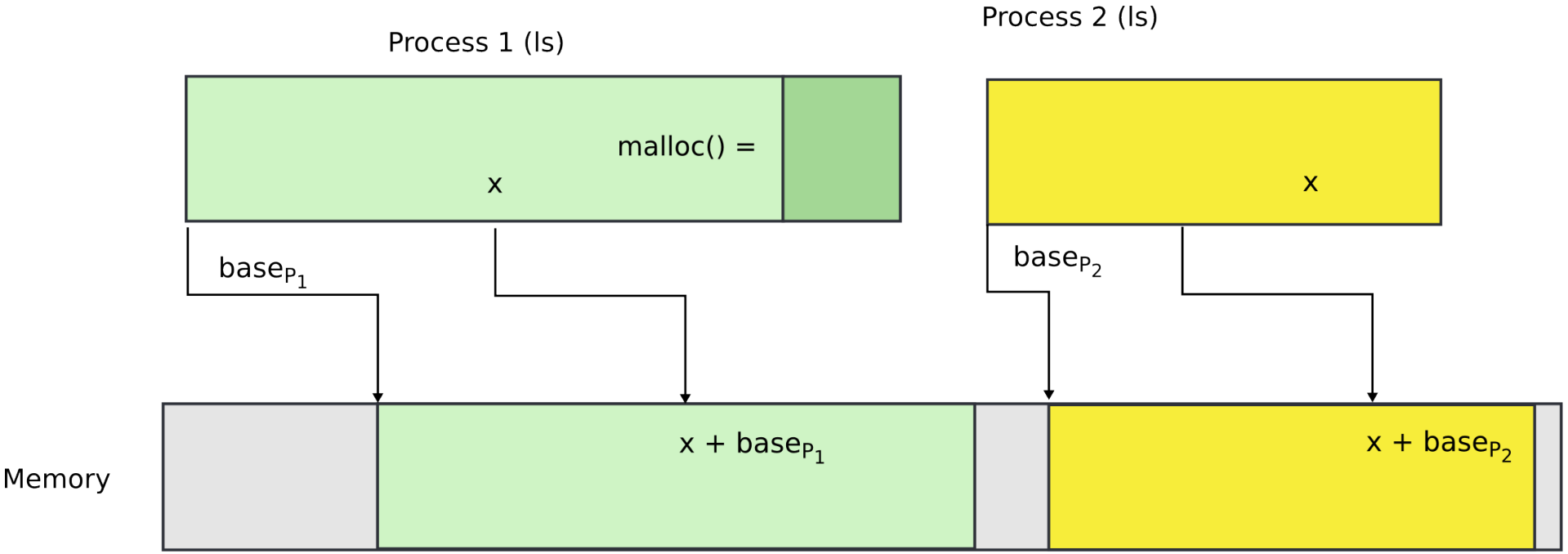
Stack segment

- Stack
  - SS
- Can be loaded explicitly
  - OS can set up multiple stacks
  - Of course, only one is accessible at a time

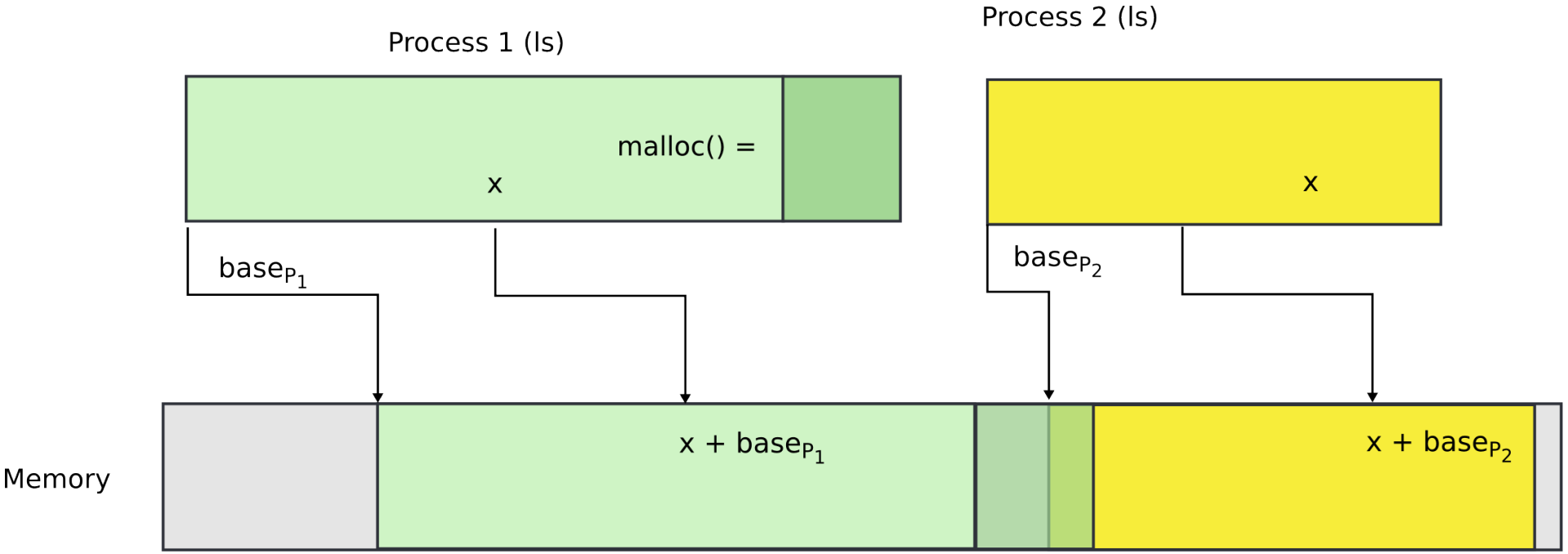
Segmentation is ok... but



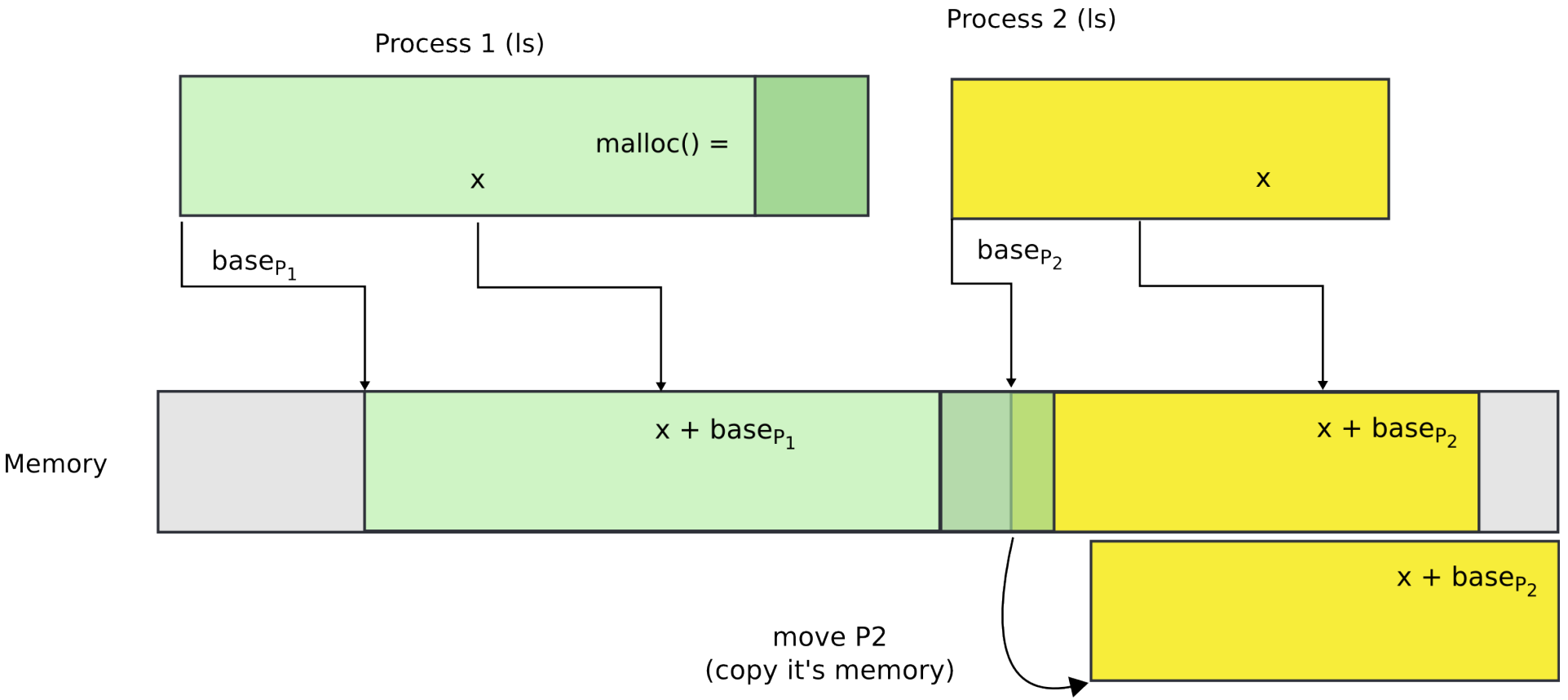
What if process needs more memory?



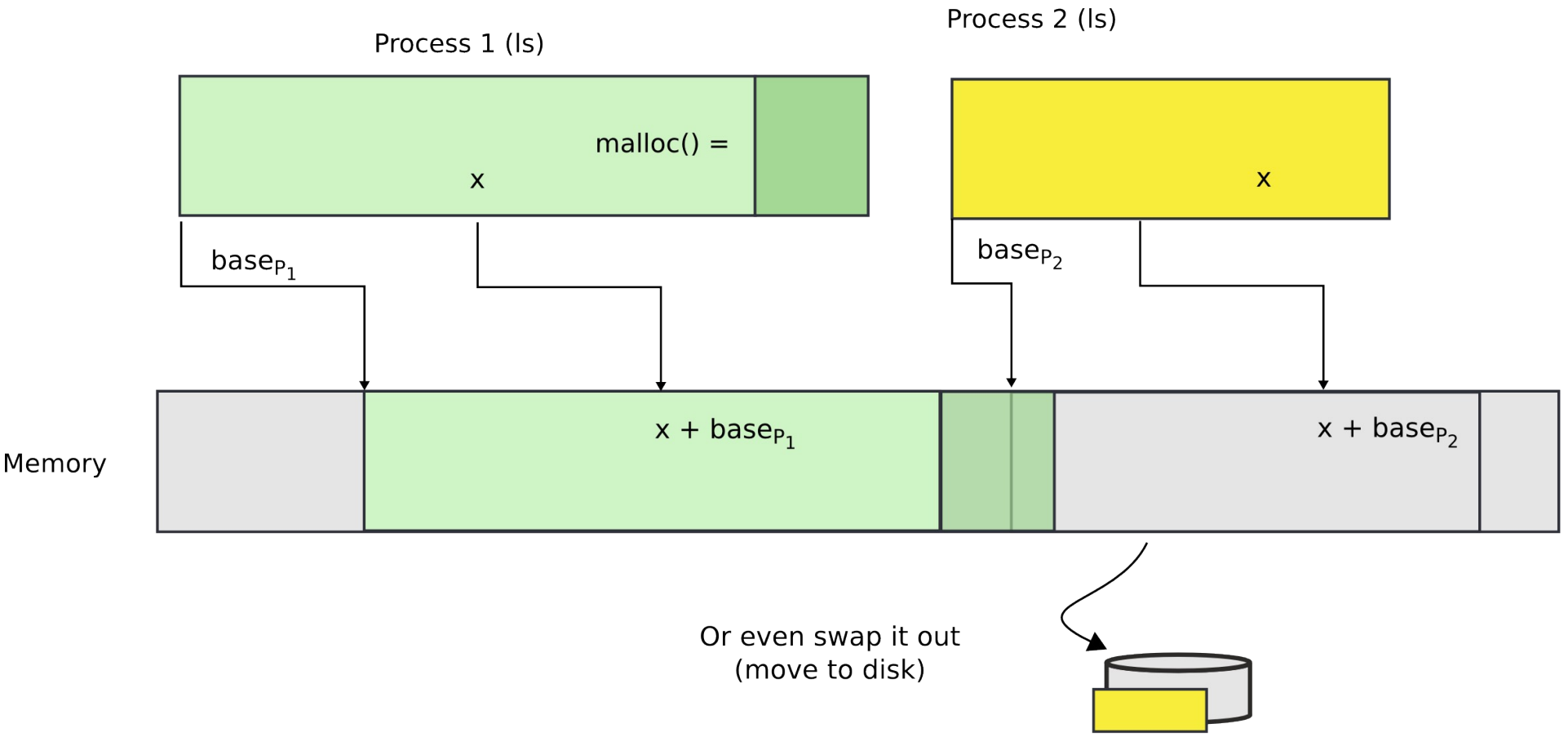
What if process needs more memory?



You can relocate P2



Or even swap it out to disk

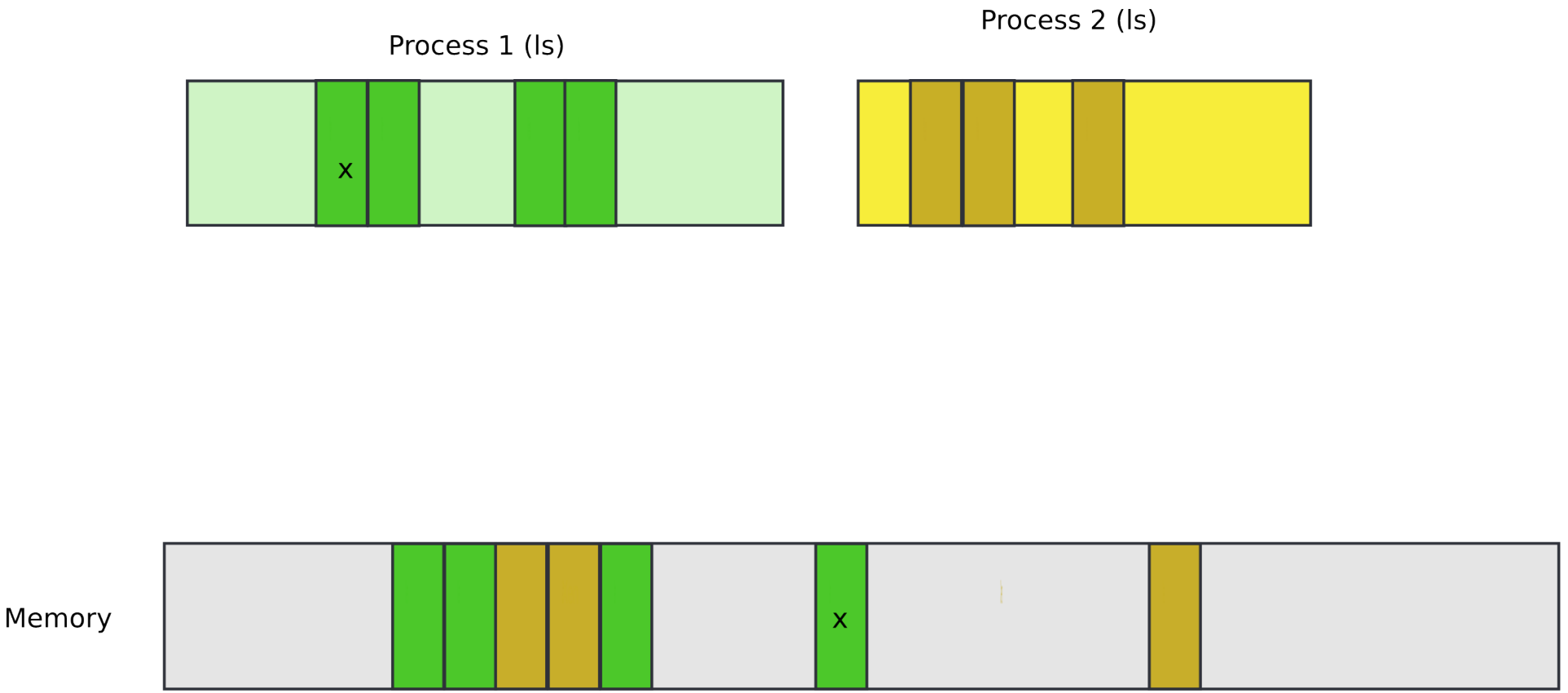


Problems with segments

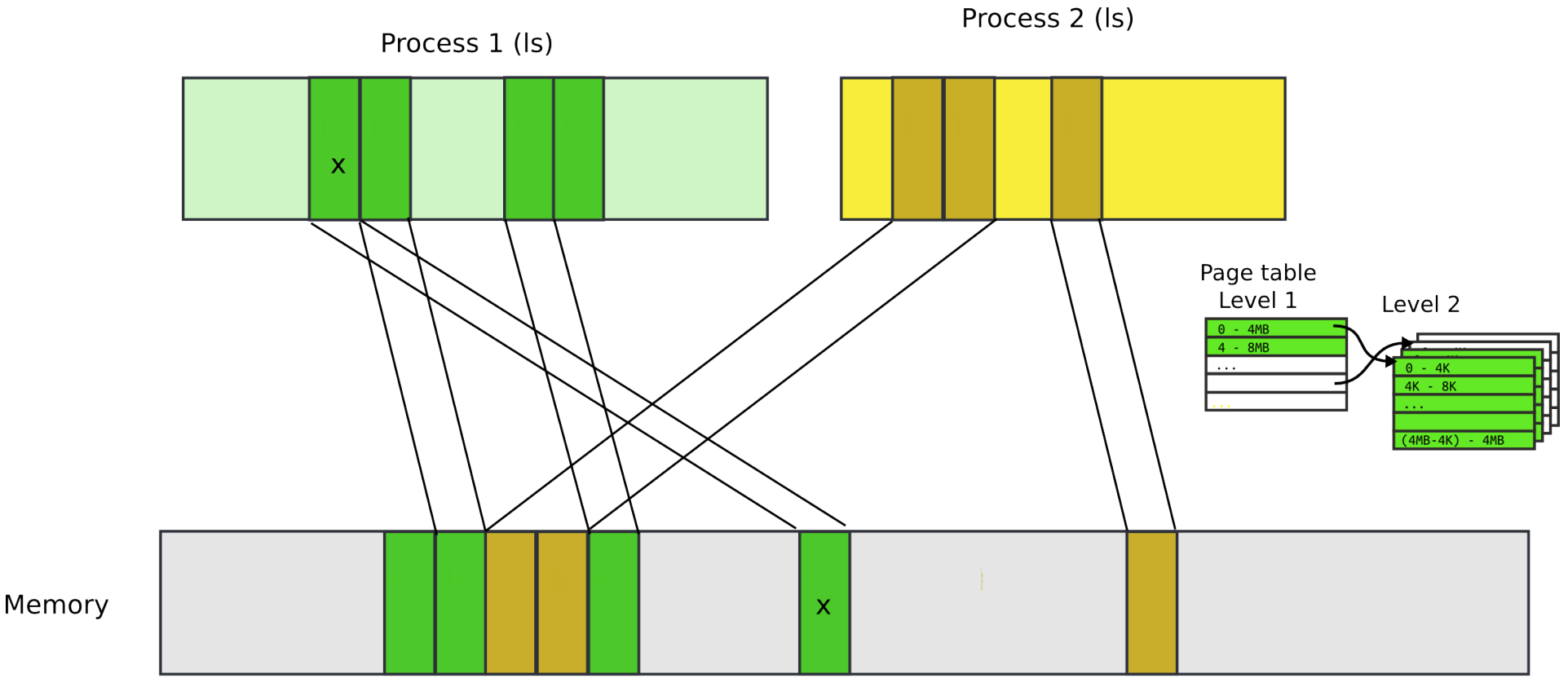
- But it's inefficient
  - Relocating or swapping the entire process takes time
- Memory gets fragmented
  - There might be no space (gap) for the swapped out process to come in
  - Will have to swap out other processes

Paging

Pages



Pages

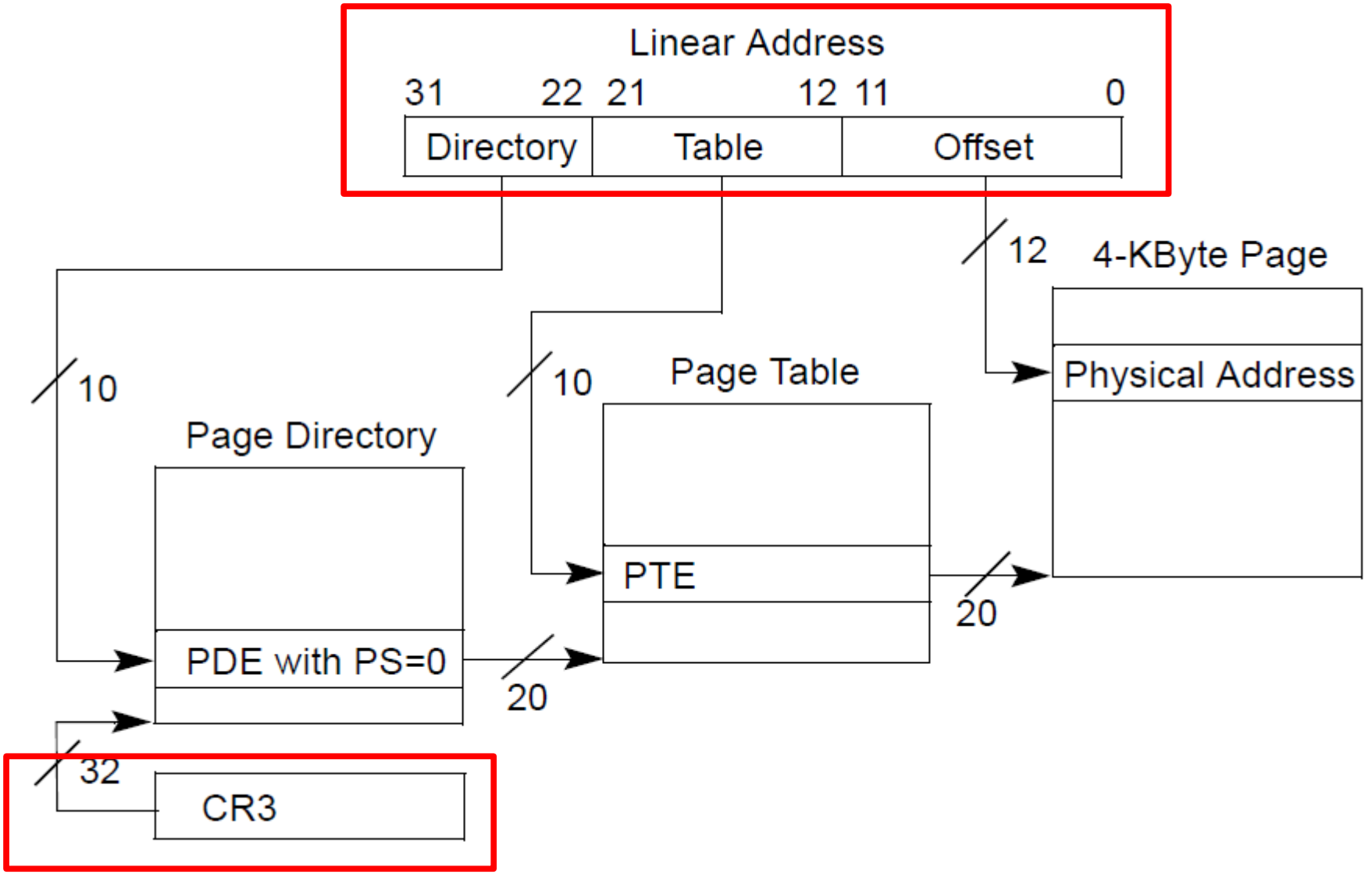

Process 1 (ls)
Process 2 (ls)
x
Page table
Level 1
Level 2
0 - 4MB
4 - 8MB
...
0 - 4K
4K - 8K
...
(4MB-4K) - 4MB
Memory
x

Paging idea

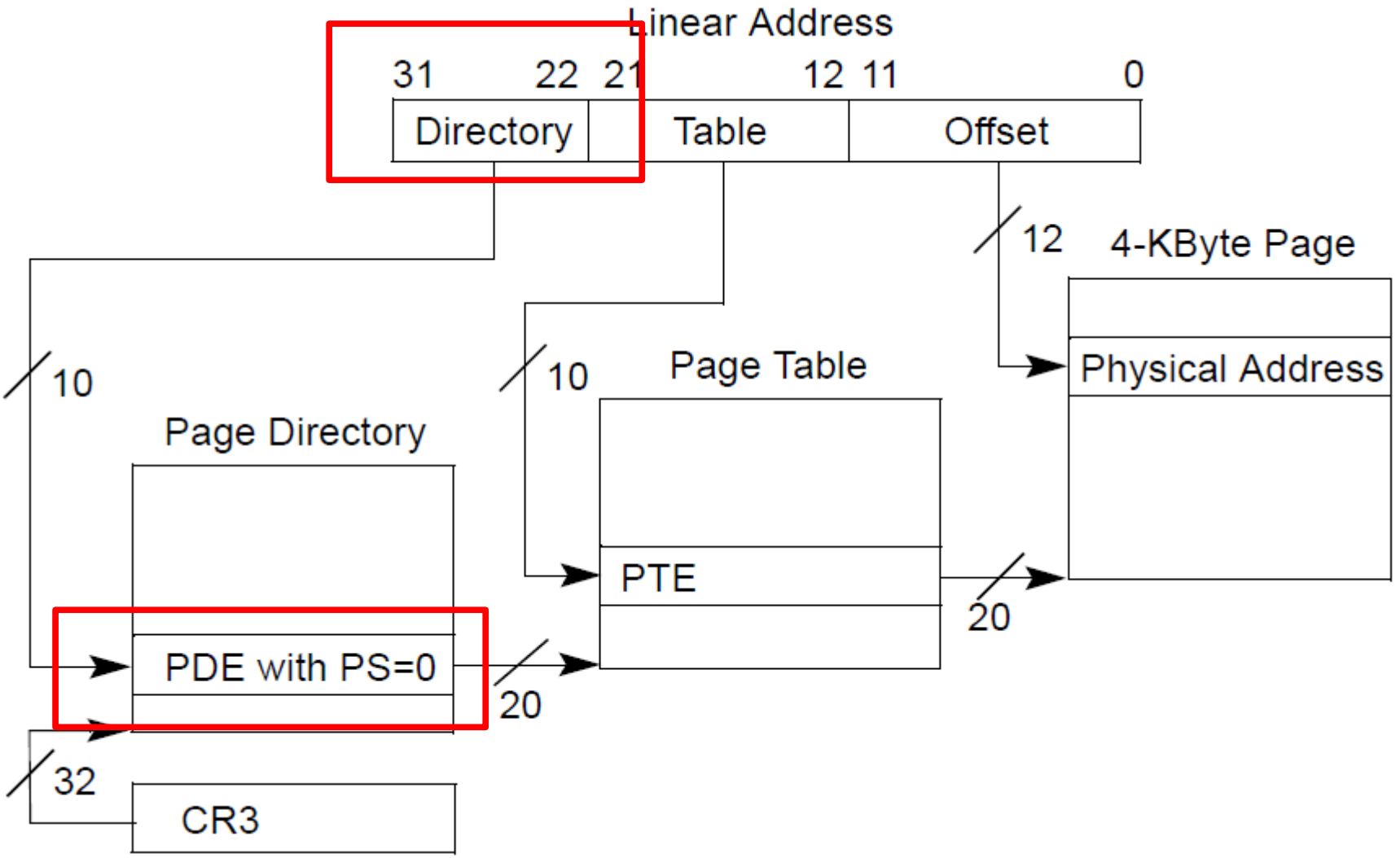
- Break up memory into 4096-byte chunks called pages
  - Modern hardware supports 2MB, 4MB, and 1GB pages
- Independently control mapping for each page of linear address space

- Compare with segmentation (single base + limit)
  - many more degrees of freedom

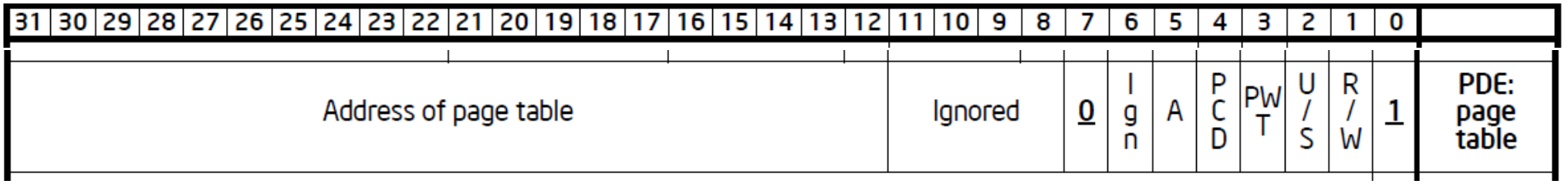
Page translation



Page translation

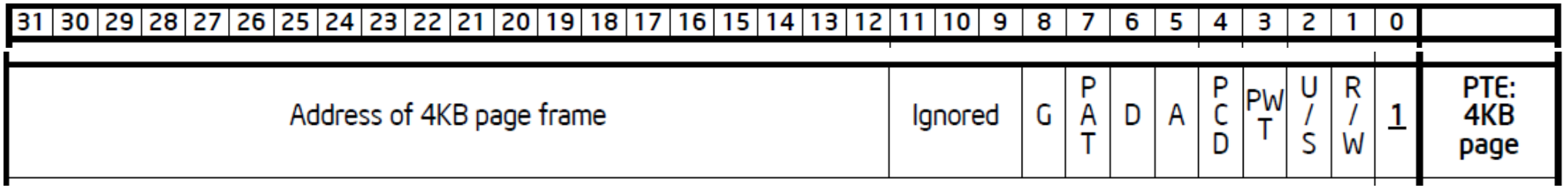


Page directory entry (PDE)



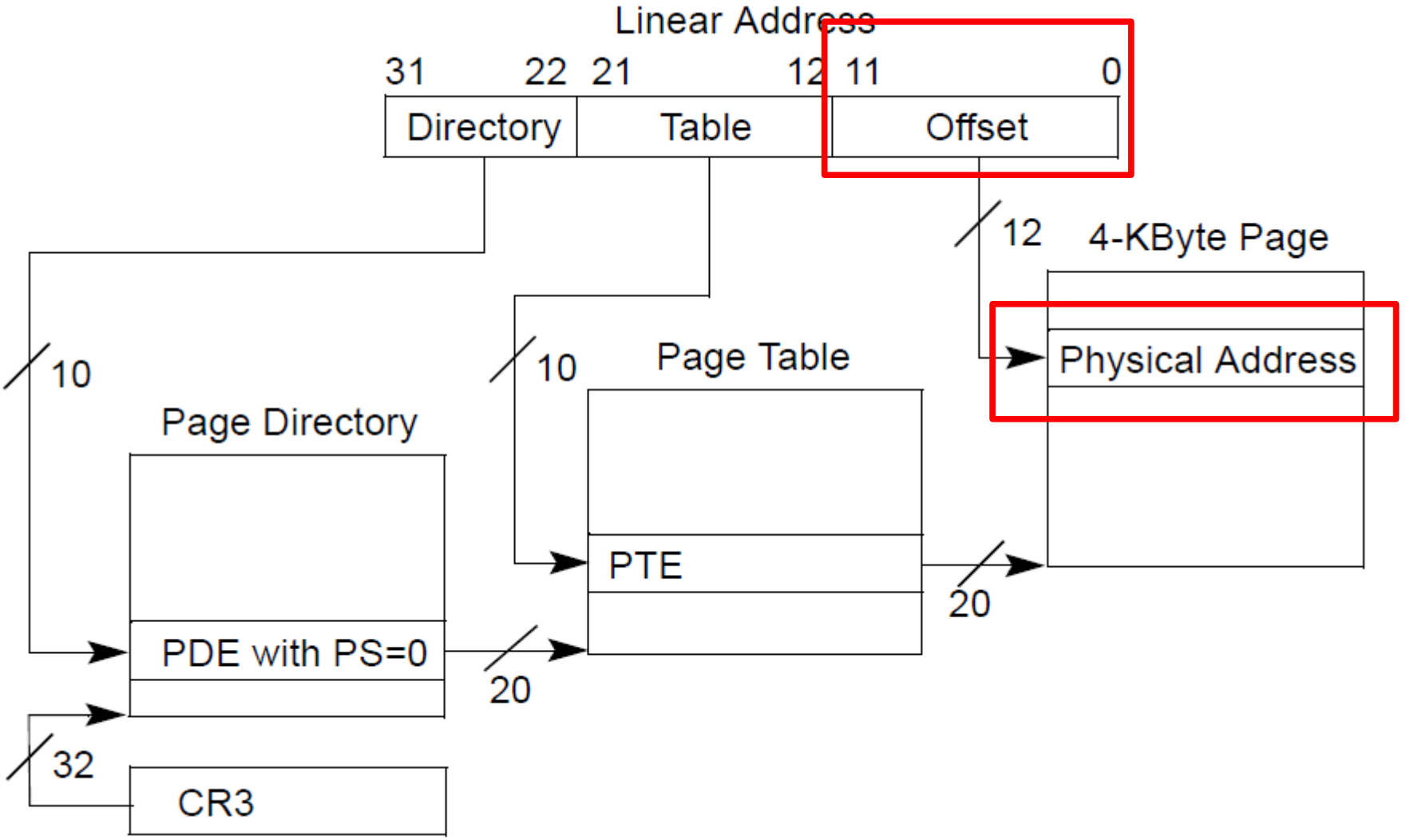
- 20 bit address of the page table
  - Pages 4KB each, we need 1M to cover 4GB
- R/W – writes allowed?
  - To a 4MB region controlled by this entry
- U/S – user/supervisor
  - If 0 – user-mode access is not allowed
- A – accessed

Page table entry (PTE)



- 20 bit address of the 4KB page
  - Pages 4KB each, we need 1M to cover 4GB
- R/W – writes allowed?
  - To a 4KB page
- U/S – user/supervisor
  - If 0 user-mode access is not allowed
- A – accessed
- D – dirty – software has written to this page

Page translation



Back of the envelope

- If a page is 4K and an entry is 4 bytes, how many entries per page?
  - 1k
- How large of an address space can 1 page represent?
  - 1k entries * 1page/entry * 4K/page = 4MB
- How large can we get with a second level of translation?
  - 1k tables/dir * 1k entries/table * 4k/page = 4 GB
  - Nice that it works out that way!