CSC 252: Computer Organization
Spring 2019: Lecture 23

Instructor: Yuhao Zhu

Department of Computer Science
University of Rochester

Action Items:

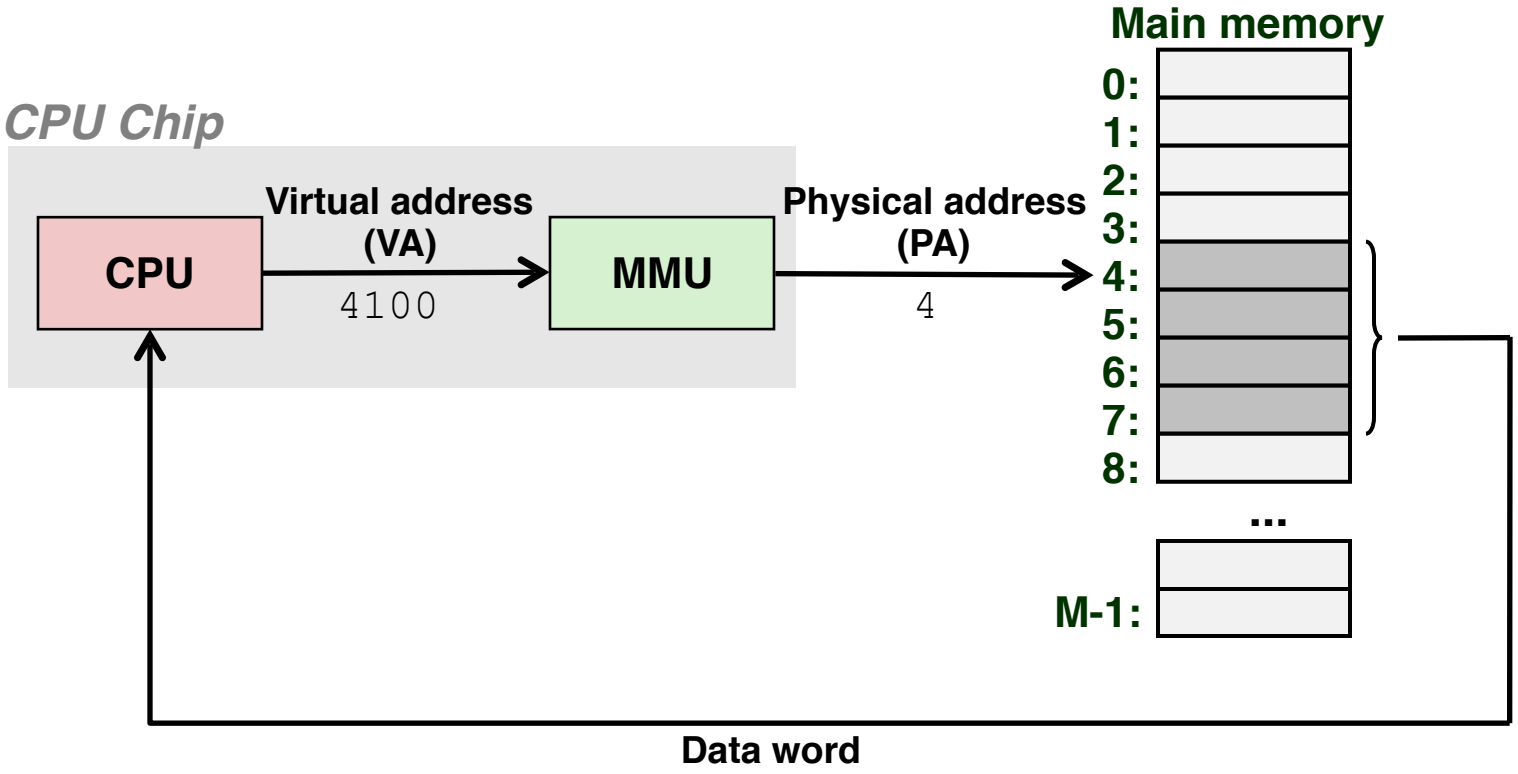
- **Lab 6 is out. Due on May 2.**

Announcement

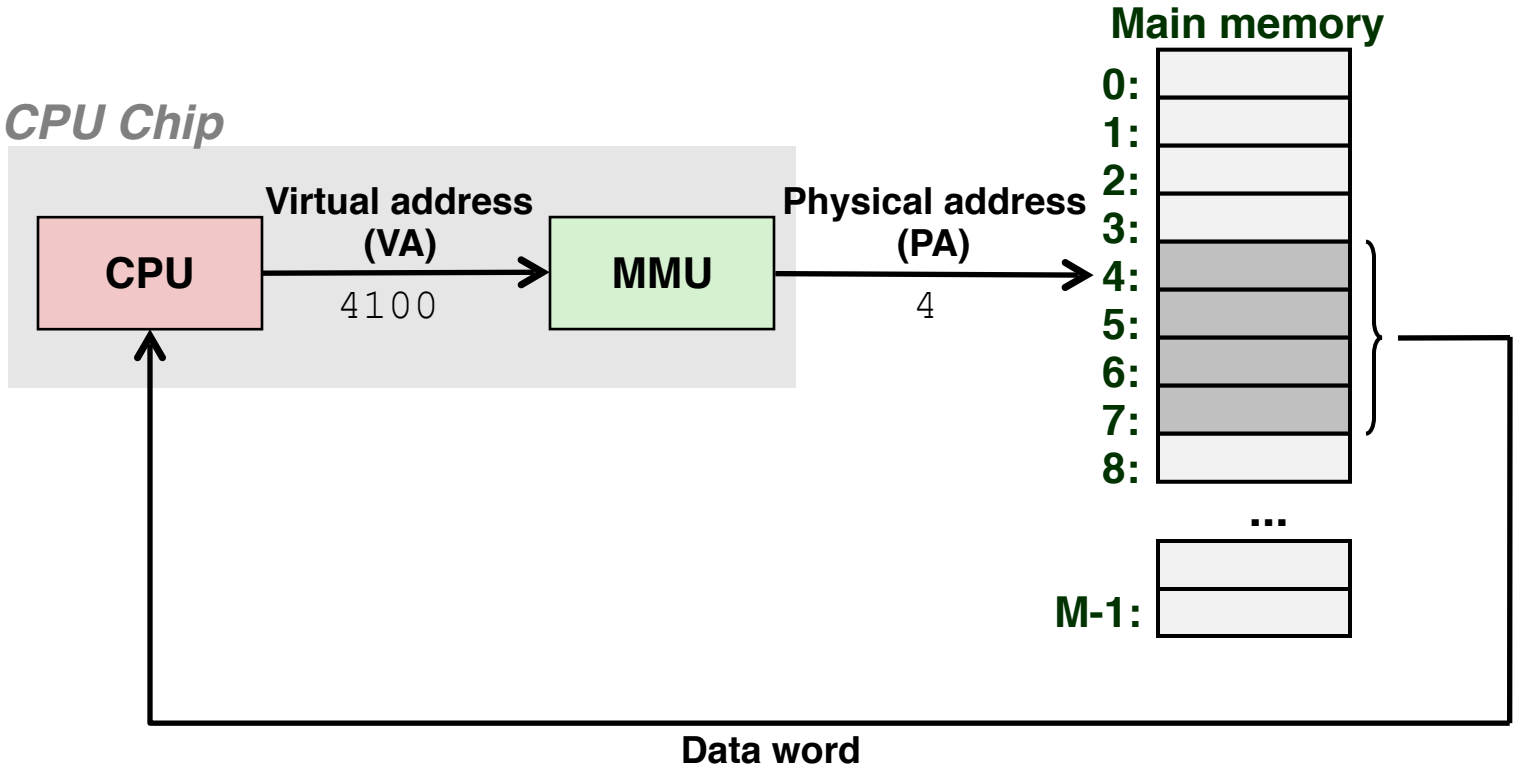
- Programming Assignment 6 is out
  - Main assignment: 11:59pm, **Thursday, May 2.**

14	15	16 **Today**	17	18 **Trivia**	19	20
21	22	23	24	25	26	27
28	29	30	May 1	2 **Due**	3	4

A System Using Virtual Addressing



A System Using Virtual Addressing



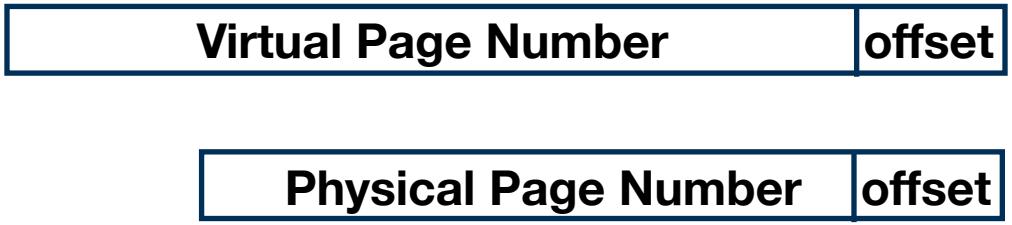
- On a 64-bit machine, virtual memory size = 2^{64}
- Physical memory size is much much smaller:
  - iPhone 8: 2 GB (2^{31})
  - 15-inch Macbook Pro 2017: 16 GB (2^{34})

VM Concepts

- Conceptually, *virtual memory* is an array of N *pages* stored on disk.
- The physical memory is an array of M *pages* stored in DRAM.
- M << N
- So store only the most frequently used pages in the physical memory, and if a page is not on the physical memory, have to first swap it from the disk to the DRAM.

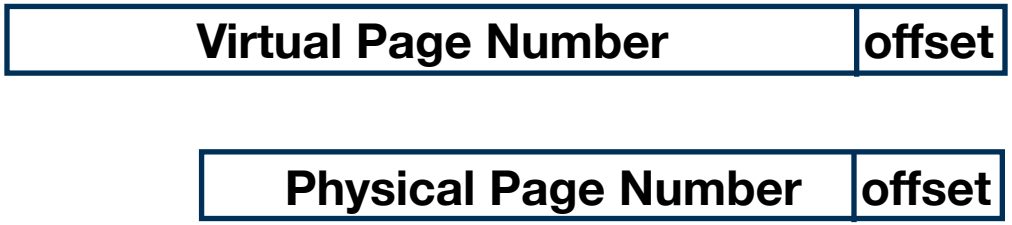
VM Concepts

- Divide both virtual memory (VM) and physical memory (PM) into “pages”



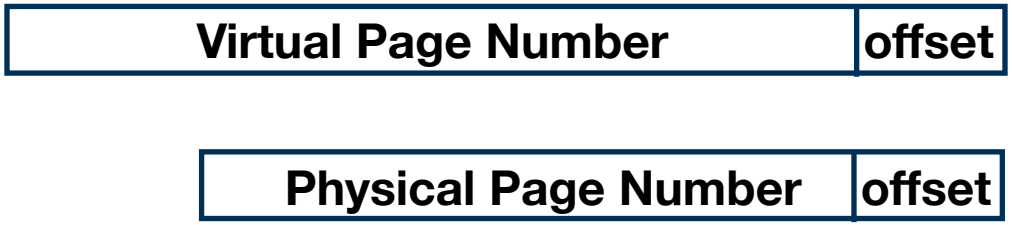
VM Concepts

- Divide both virtual memory (VM) and physical memory (PM) into “pages”
- Page size is the same for VM and PM



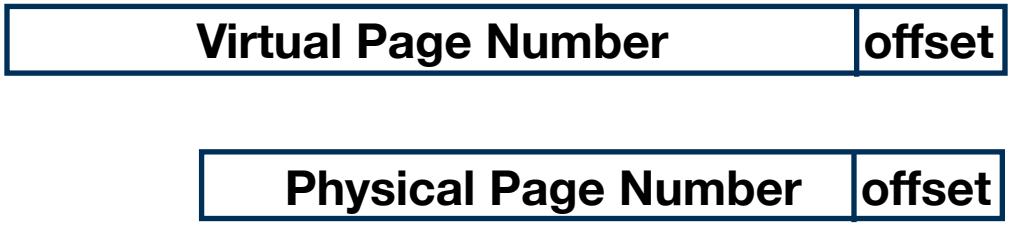
VM Concepts

- Divide both virtual memory (VM) and physical memory (PM) into “pages”
- Page size is the same for VM and PM
- In a 64-bit machine, VA is 64-bit long. Assuming PM is 4 GB. Assuming 4KB page size.



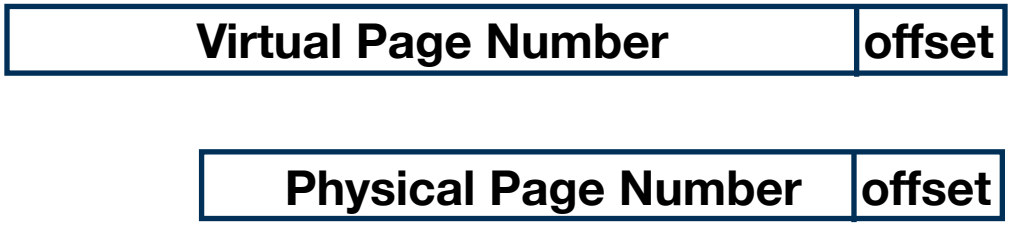
VM Concepts

- Divide both virtual memory (VM) and physical memory (PM) into “pages”
- Page size is the same for VM and PM
- In a 64-bit machine, VA is 64-bit long. Assuming PM is 4 GB. Assuming 4KB page size.
- How many bits for offset?
  - 12. Same for VM and PM



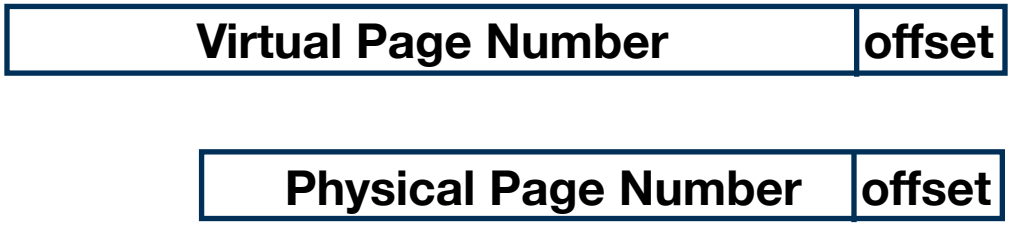
VM Concepts

- Divide both virtual memory (VM) and physical memory (PM) into “pages”
- Page size is the same for VM and PM
- In a 64-bit machine, VA is 64-bit long. Assuming PM is 4 GB. Assuming 4KB page size.
- How many bits for offset?
  - 12. Same for VM and PM
- How many bits for Virtual Page Number?
  - 52



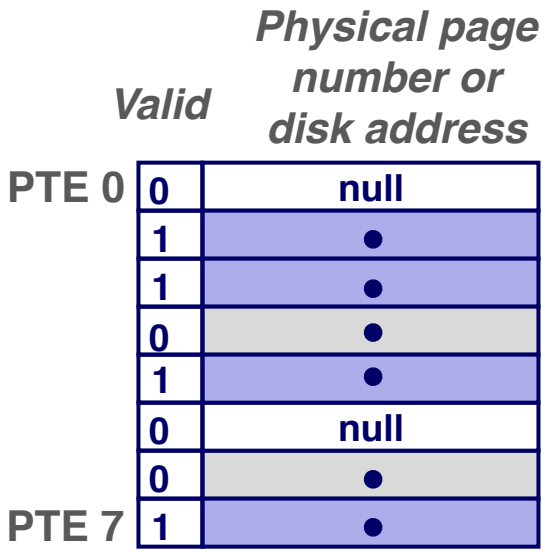
VM Concepts

- Divide both virtual memory (VM) and physical memory (PM) into “pages”
- Page size is the same for VM and PM
- In a 64-bit machine, VA is 64-bit long. Assuming PM is 4 GB. Assuming 4KB page size.
- How many bits for offset?
  - 12. Same for VM and PM
- How many bits for Virtual Page Number?
  - 52
- How many bits for Physical Page Number?
  - 20



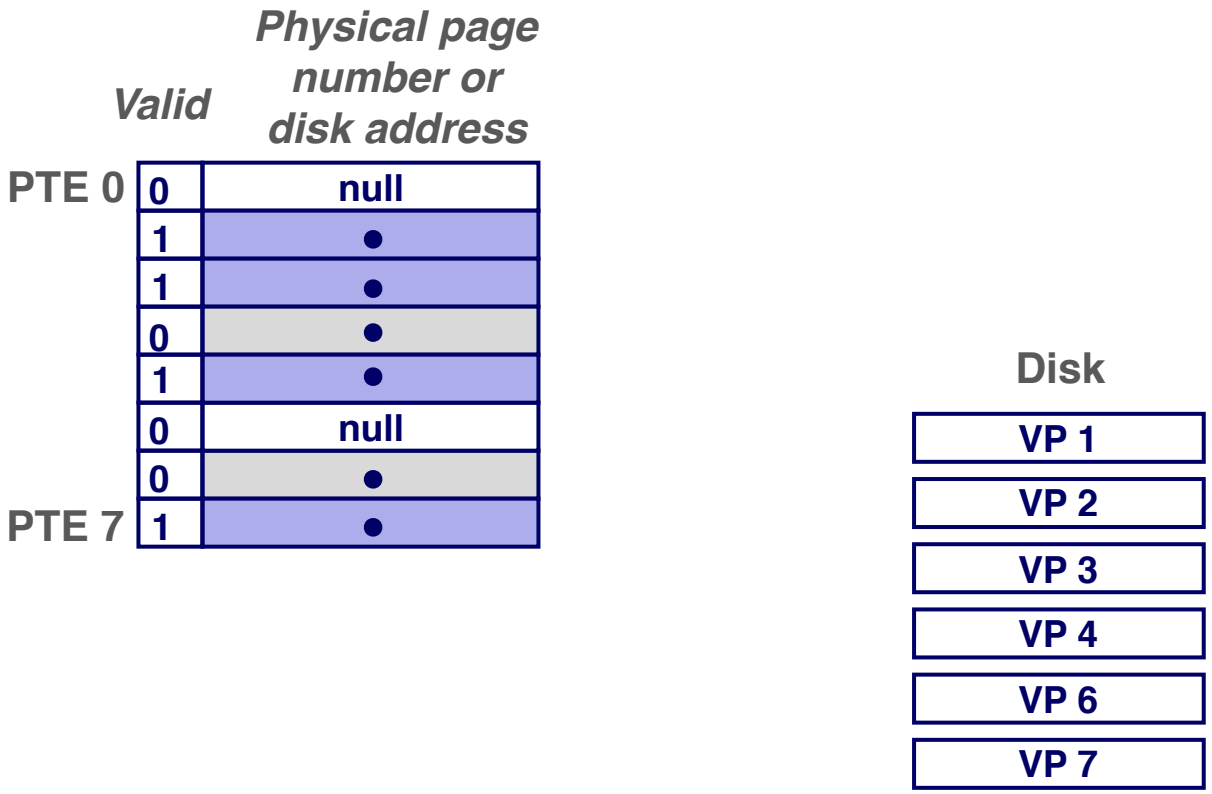
Page Table: Enabling VA to PA Translation

- A page table is an array of page table entries (PTEs) that maps every virtual page to its physical page.



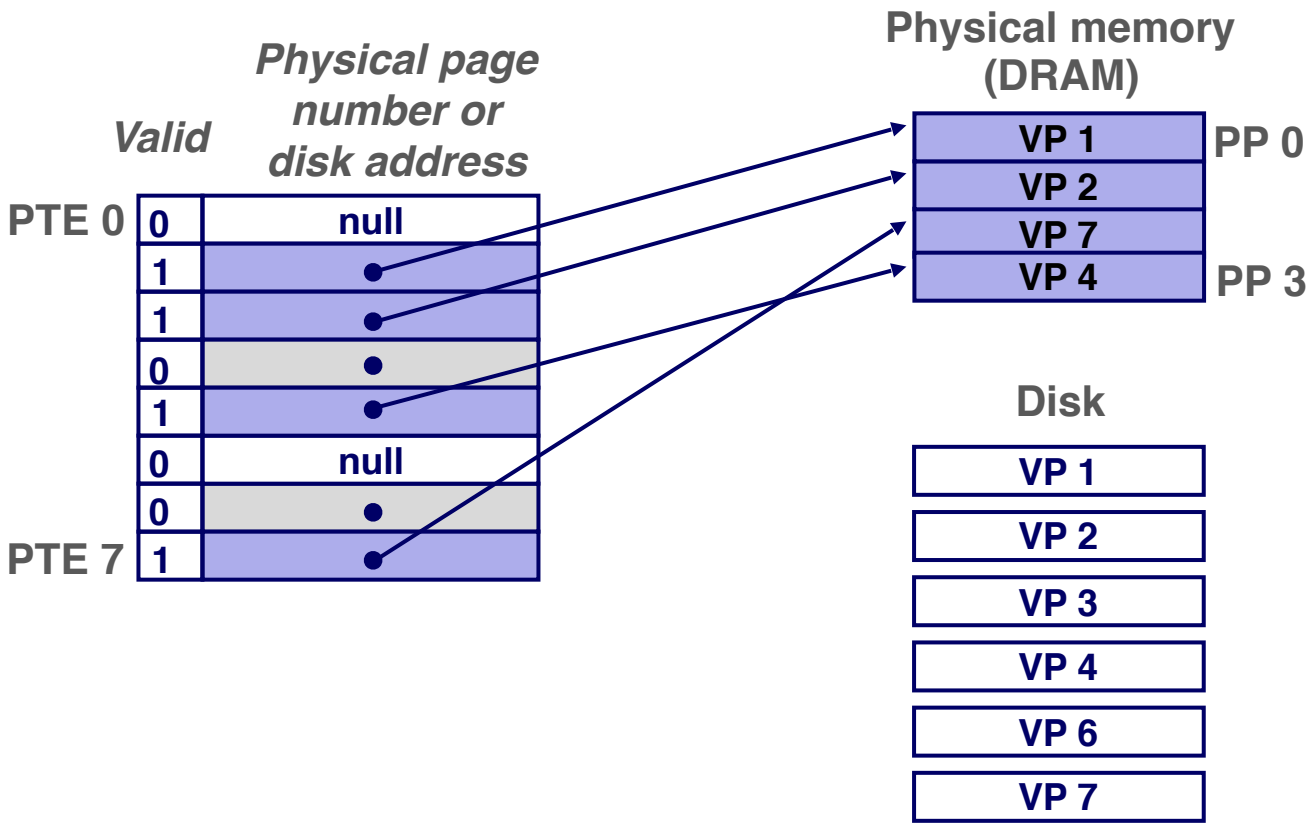
Page Table: Enabling VA to PA Translation

- A page table is an array of page table entries (PTEs) that maps every virtual page to its physical page.



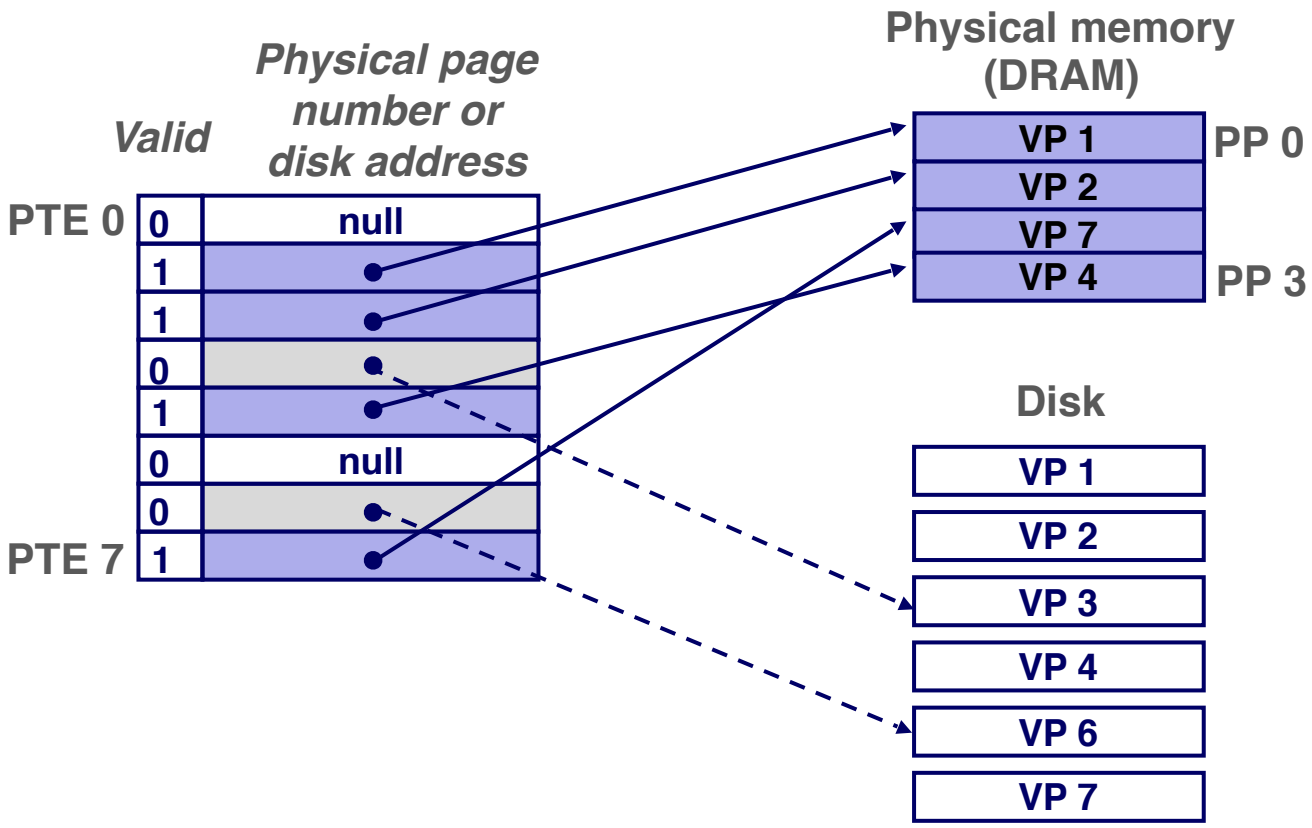
Page Table: Enabling VA to PA Translation

- A page table is an array of page table entries (PTEs) that maps every virtual page to its physical page.



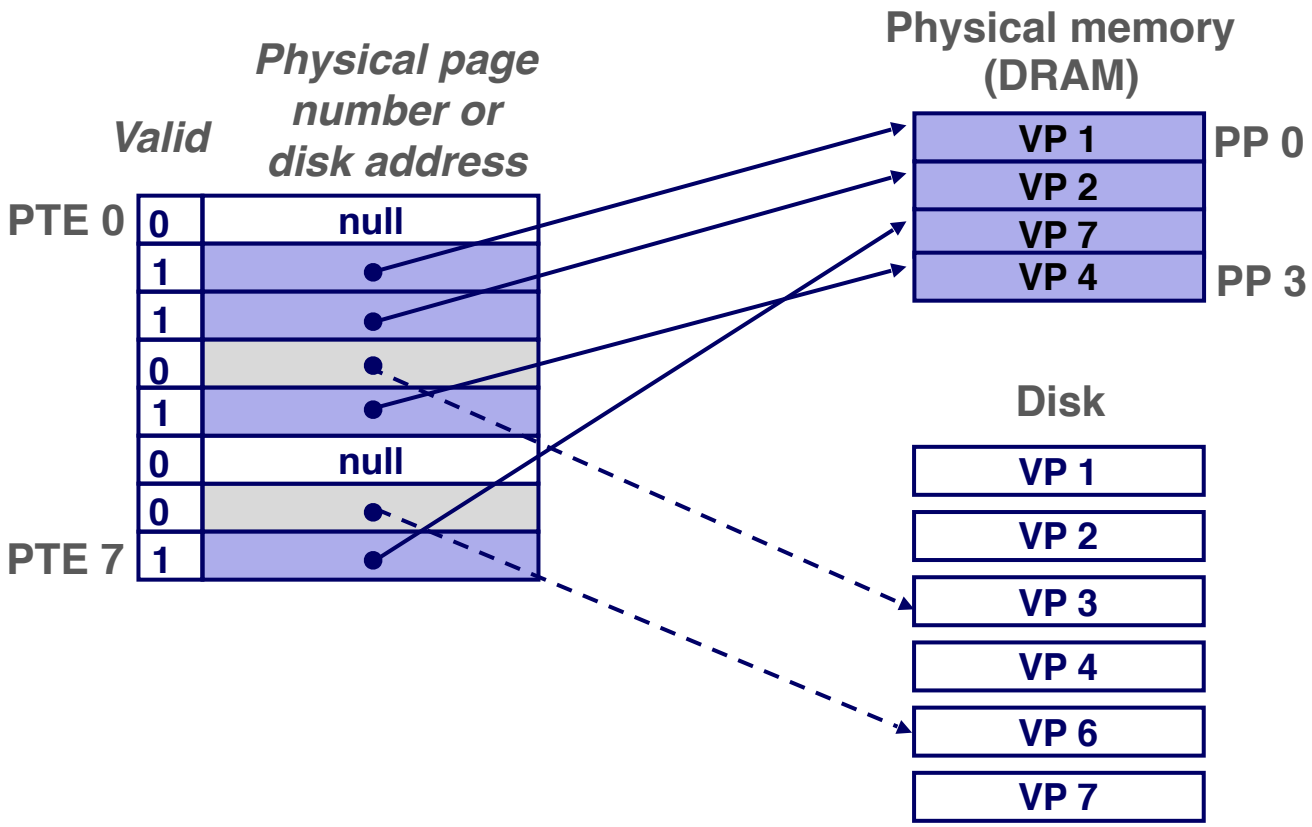
Page Table: Enabling VA to PA Translation

- A page table is an array of page table entries (PTEs) that maps every virtual page to its physical page.

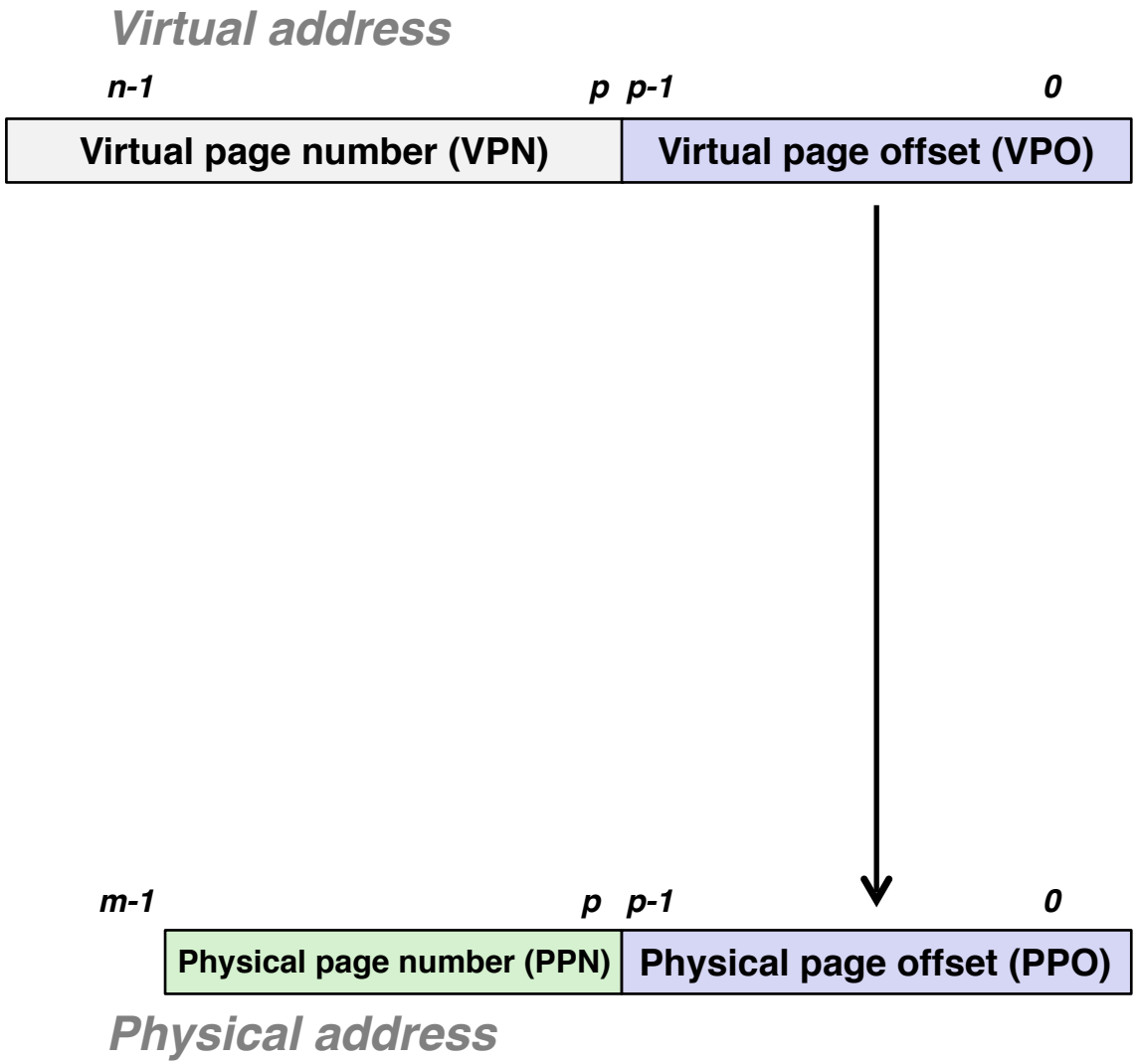


Page Table: Enabling VA to PA Translation

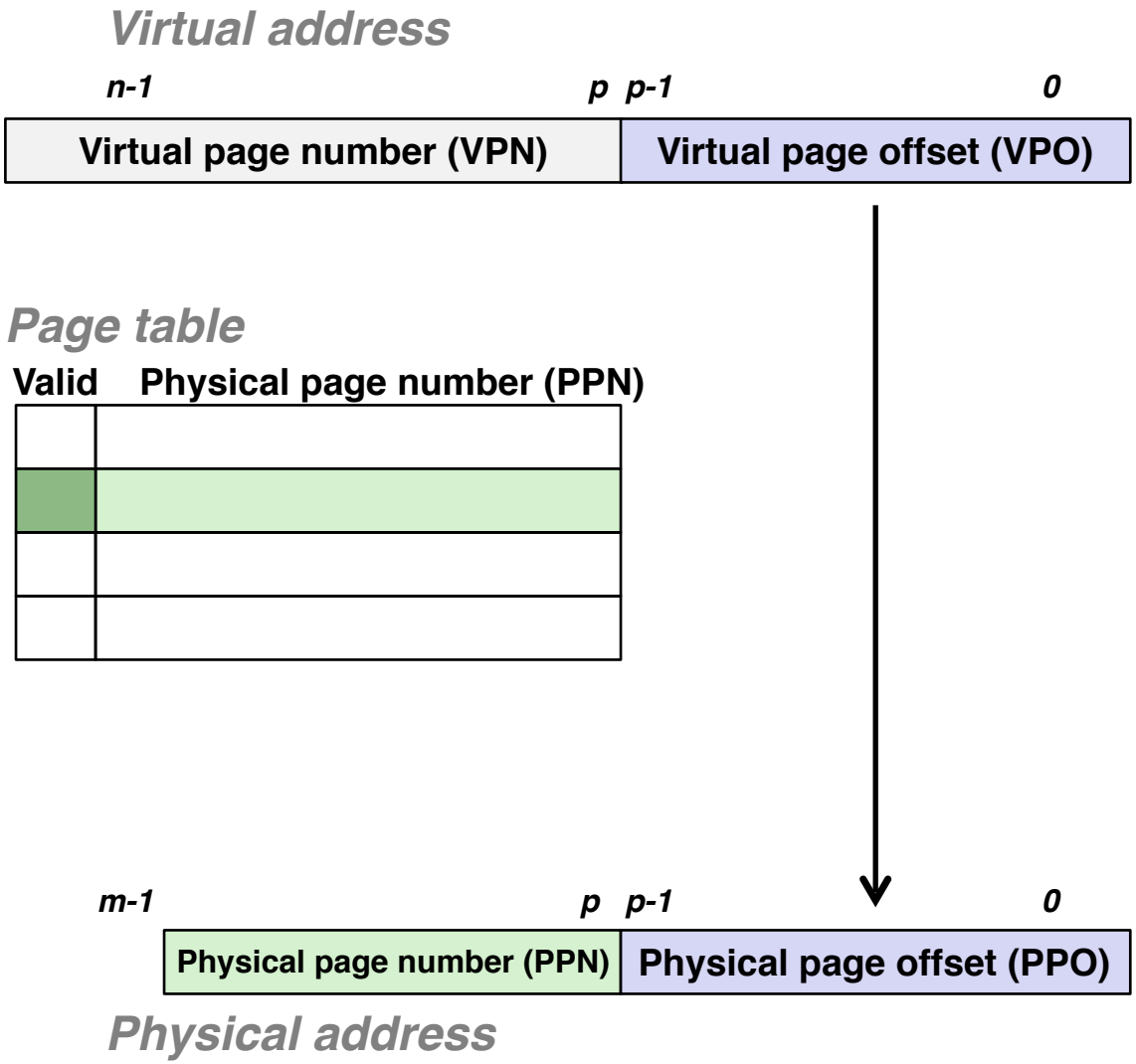
- A page table is an array of page table entries (PTEs) that maps every virtual page to its physical page.
- 64-bit machine, 4KB page size, how many PTEs?
  - Every page has a PTE, so 2^{52} PTEs.



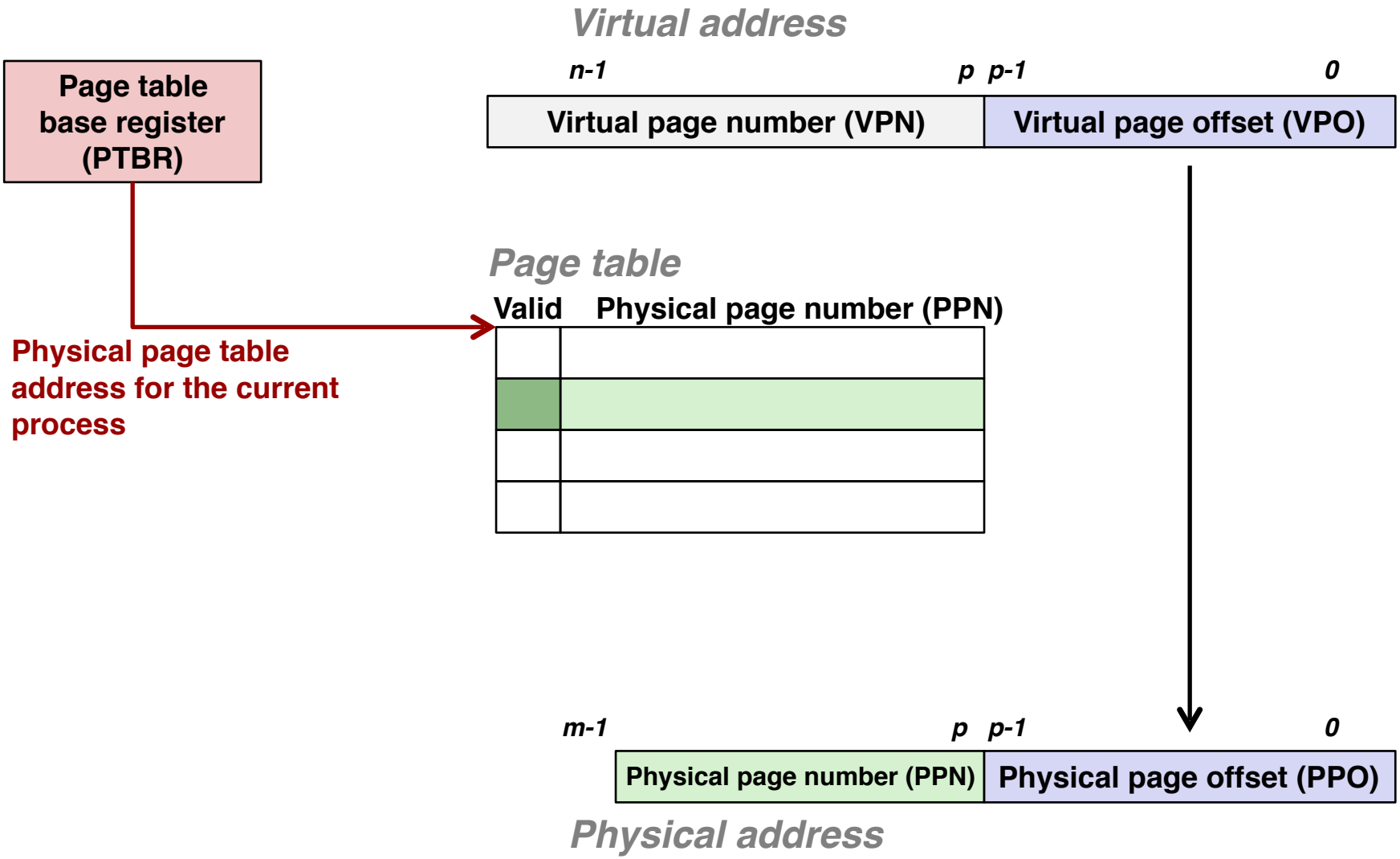
Address Translation With a Page Table



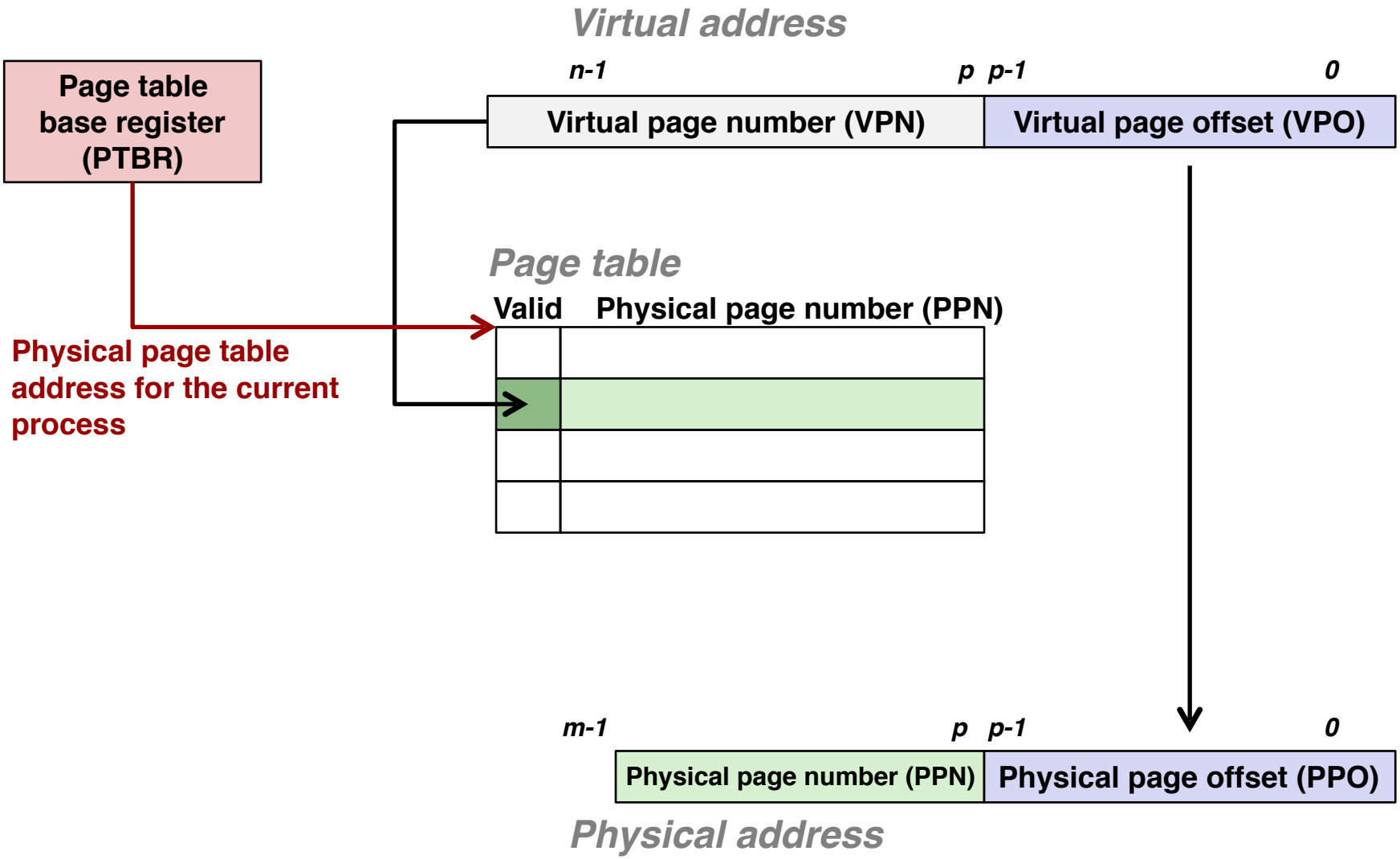
Address Translation With a Page Table



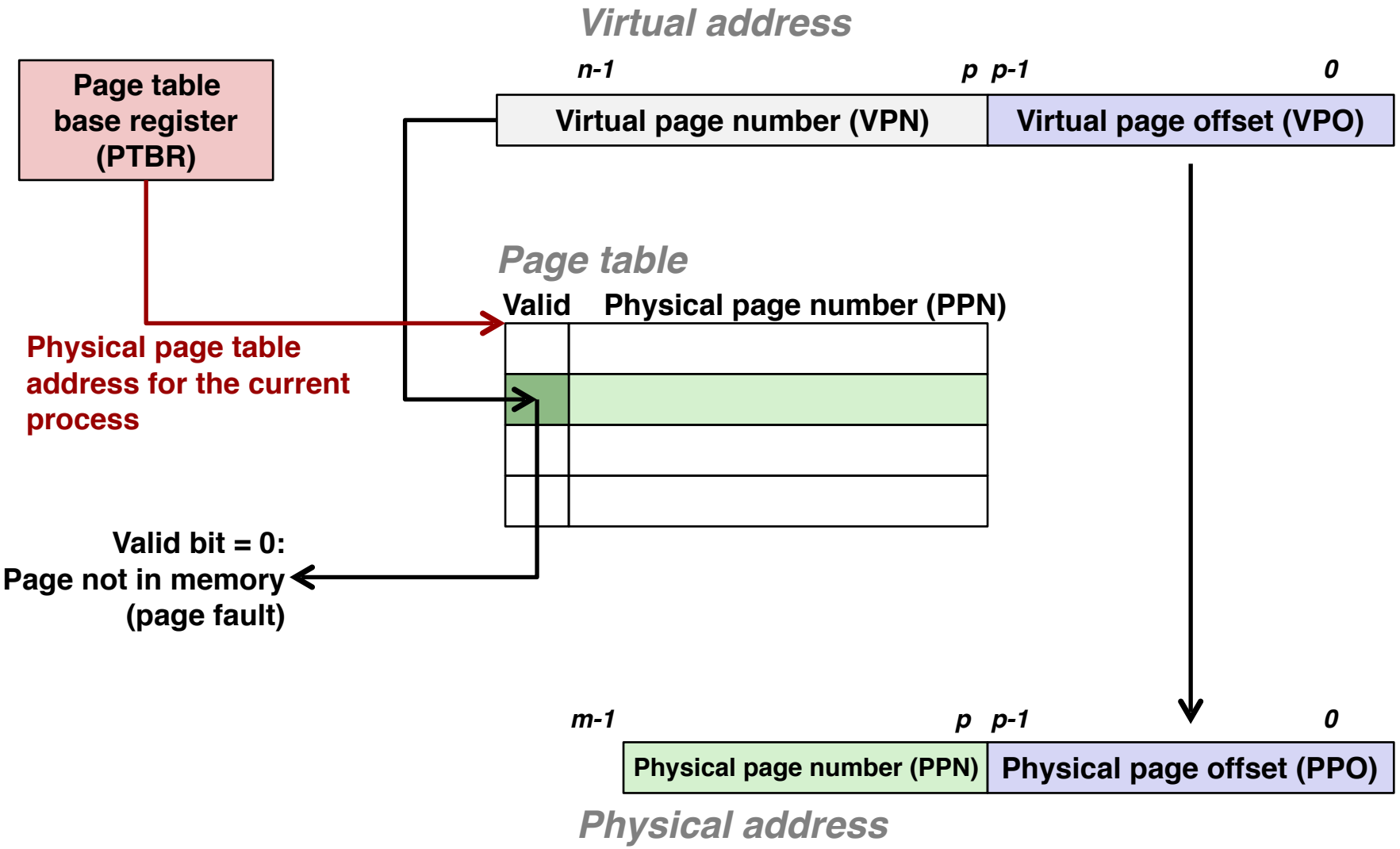
Address Translation With a Page Table



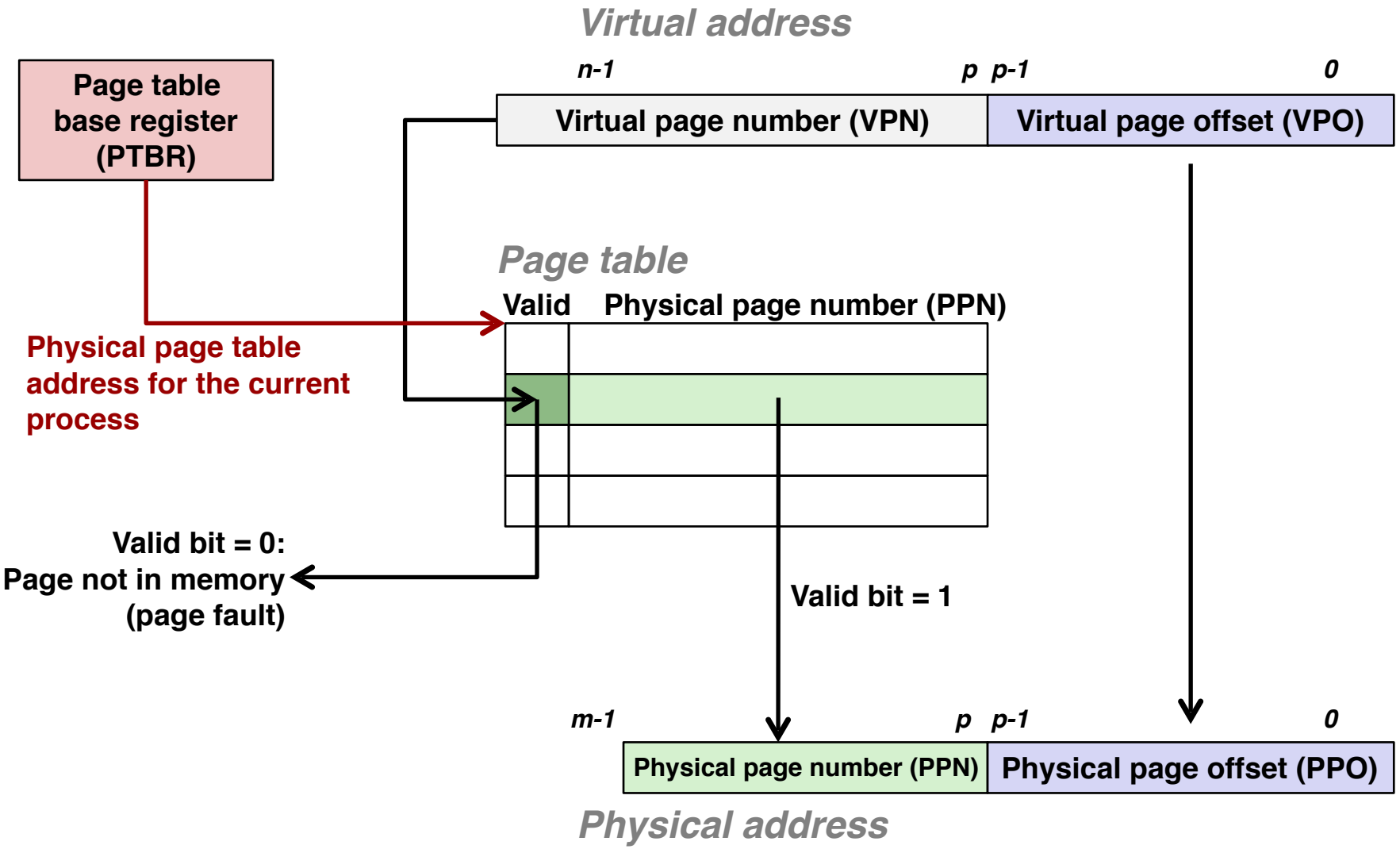
Address Translation With a Page Table



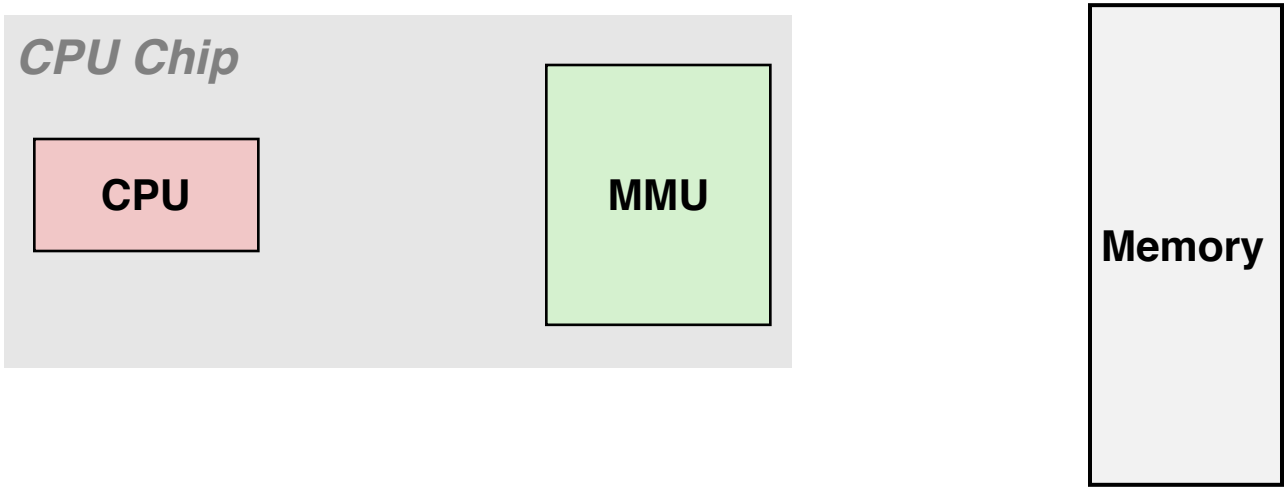
Address Translation With a Page Table



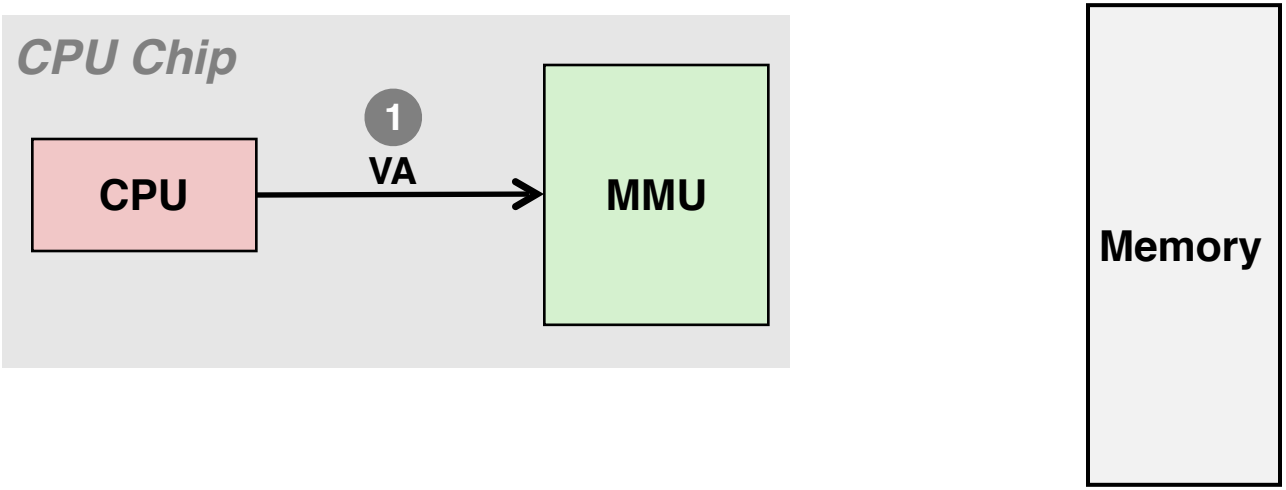
Address Translation With a Page Table



Address Translation: Page Hit

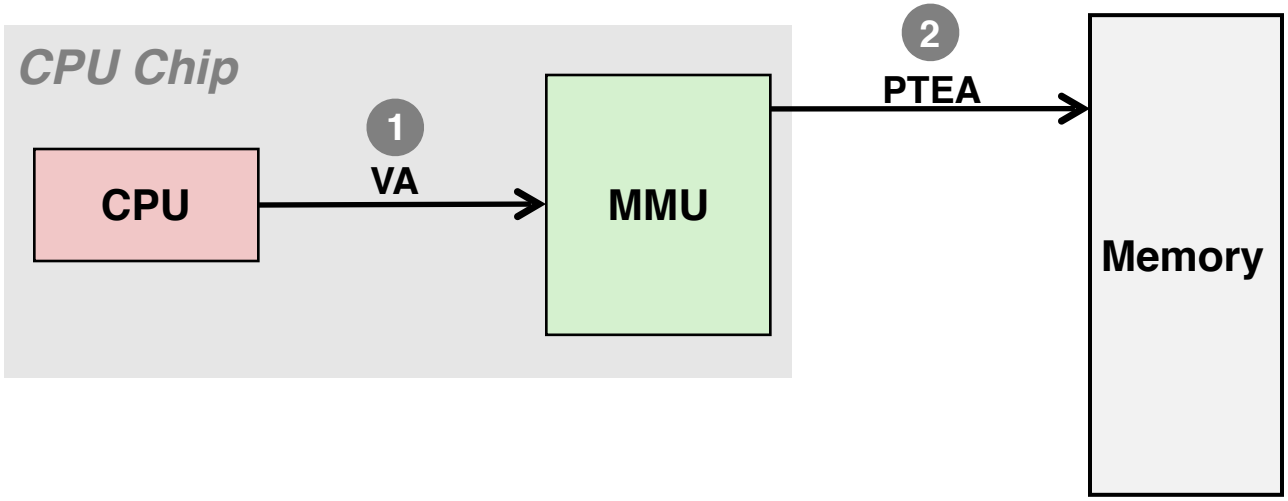


Address Translation: Page Hit



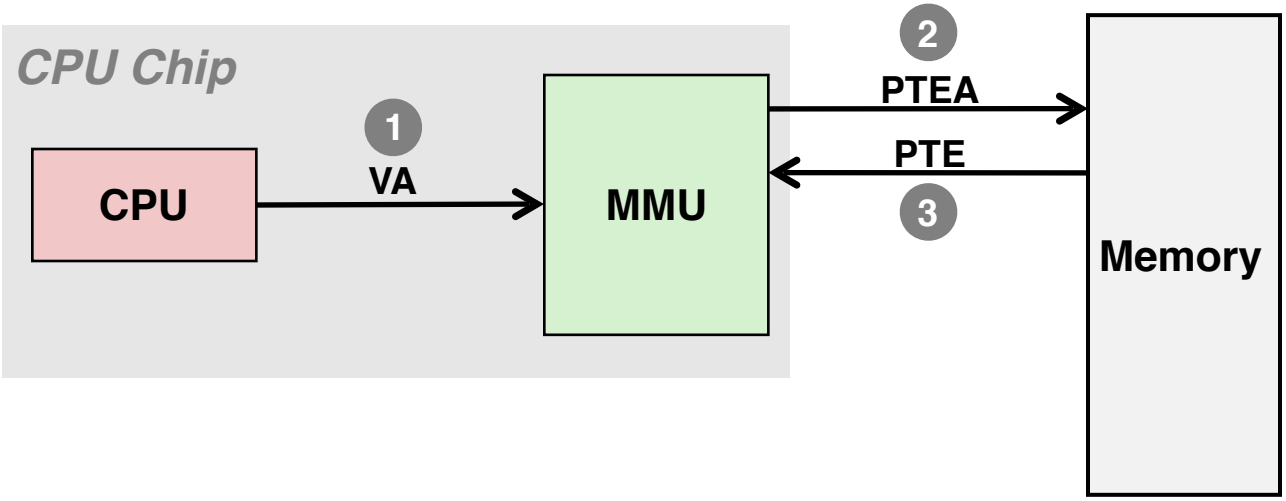
1) Processor sends virtual address to MMU

Address Translation: Page Hit



1) Processor sends virtual address to MMU

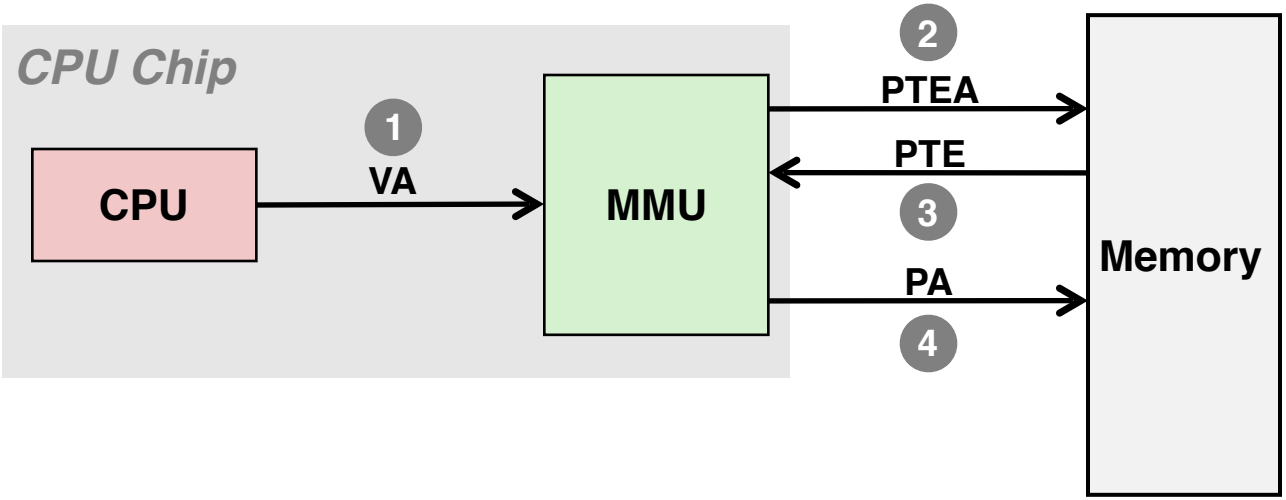
Address Translation: Page Hit



1) Processor sends virtual address to MMU

2-3) MMU fetches PTE from page table in memory

Address Translation: Page Hit

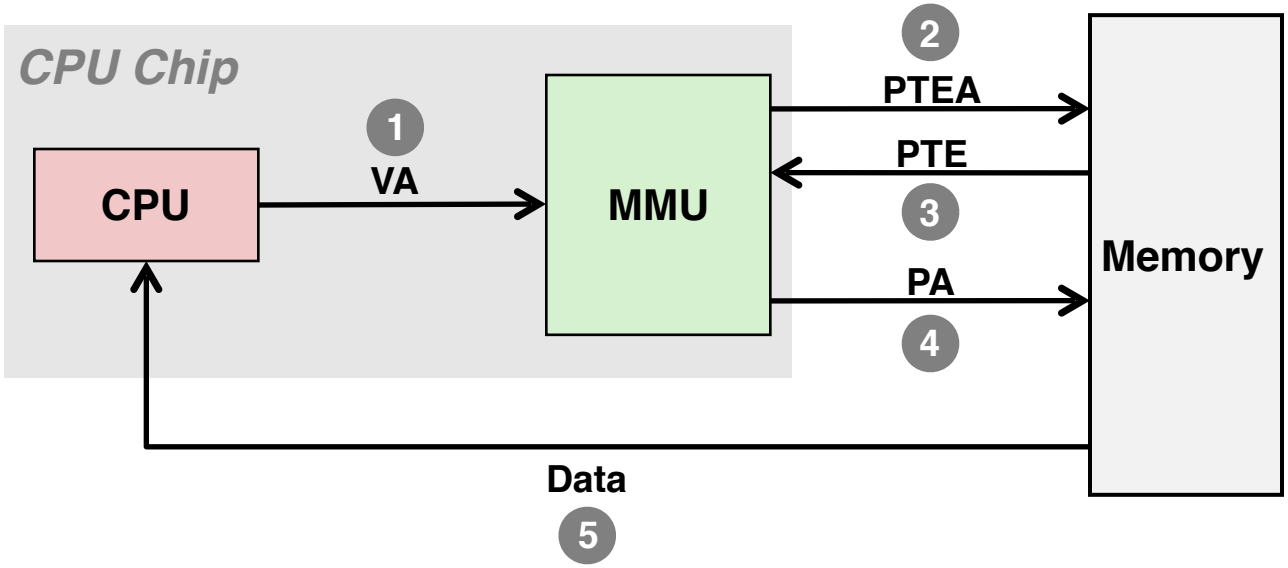


1) Processor sends virtual address to MMU

2-3) MMU fetches PTE from page table in memory

4) MMU sends physical address to cache/memory

Address Translation: Page Hit



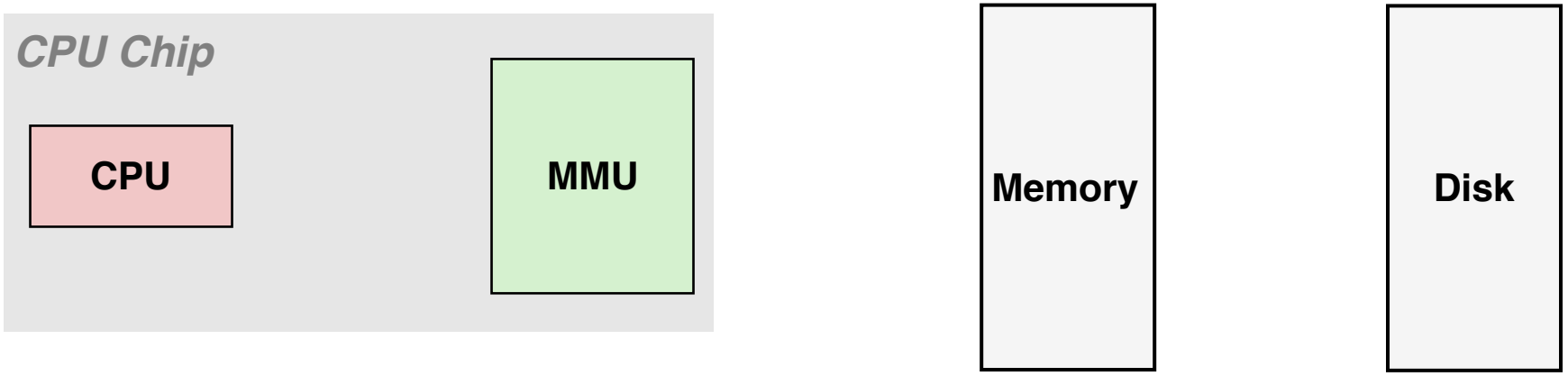
1) Processor sends virtual address to MMU

2-3) MMU fetches PTE from page table in memory

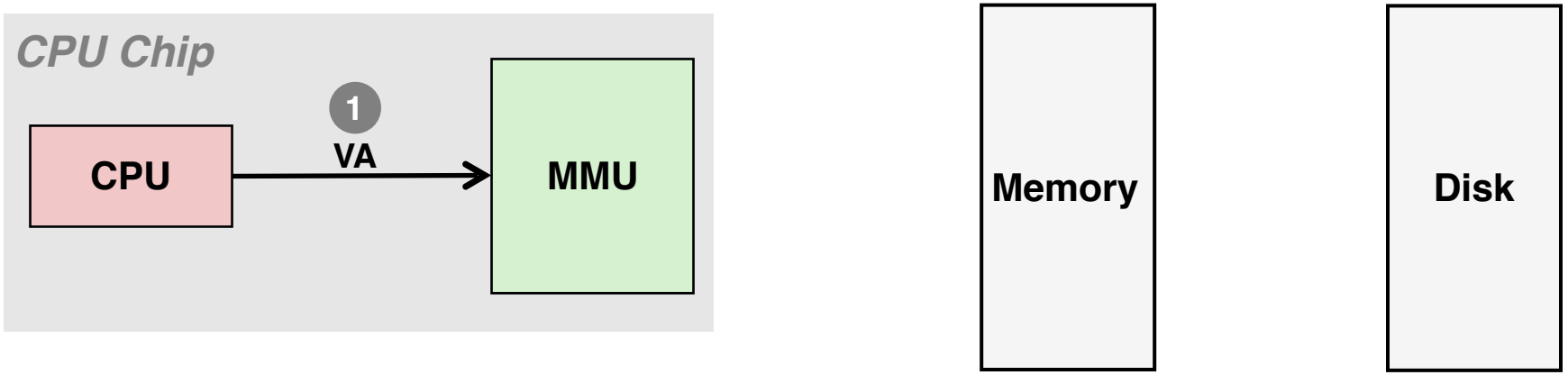
4) MMU sends physical address to cache/memory

5) Cache/memory sends data word to processor

Address Translation: Page Fault

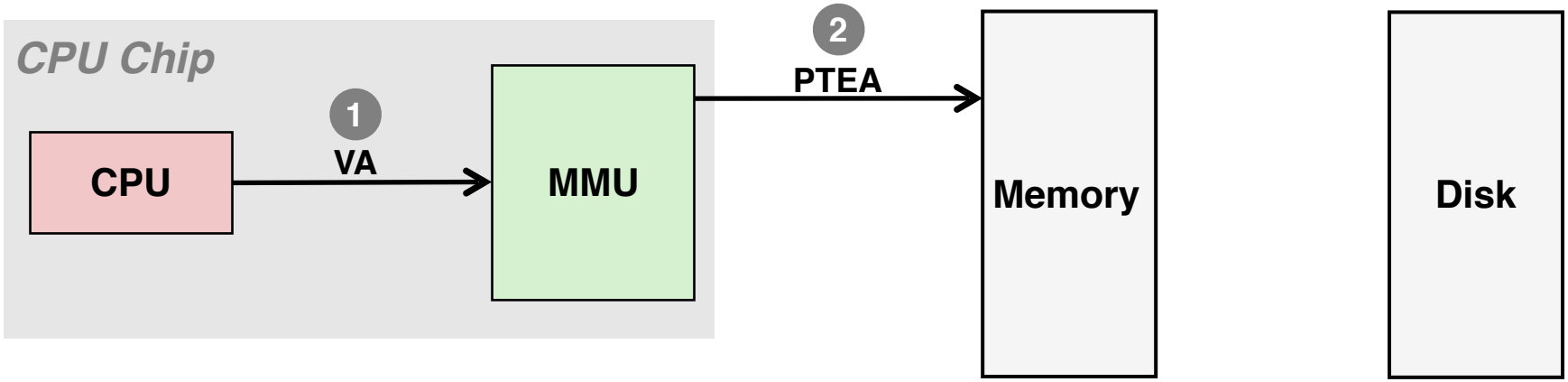


Address Translation: Page Fault



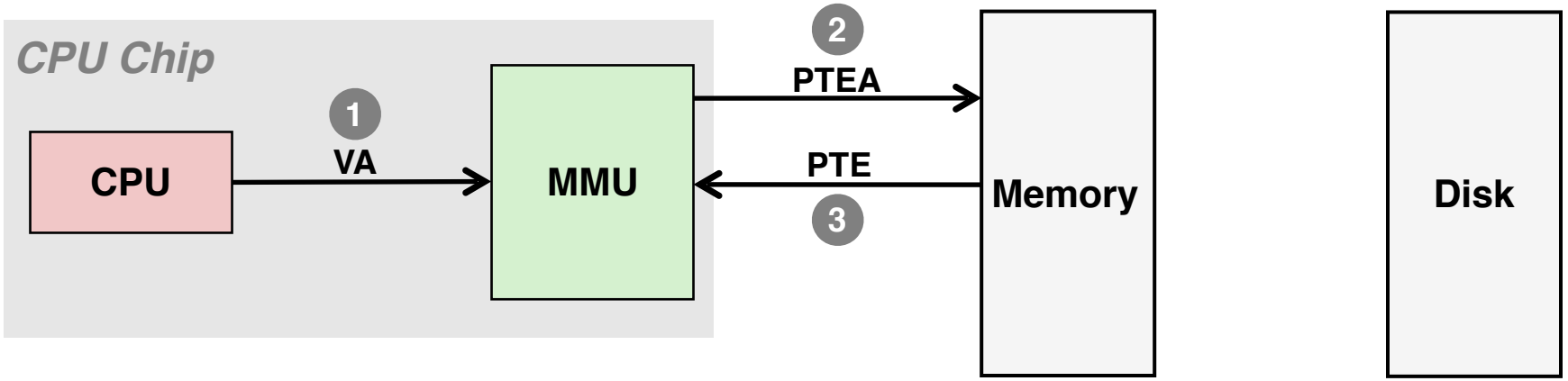
1) Processor sends virtual address to MMU

Address Translation: Page Fault



1) Processor sends virtual address to MMU

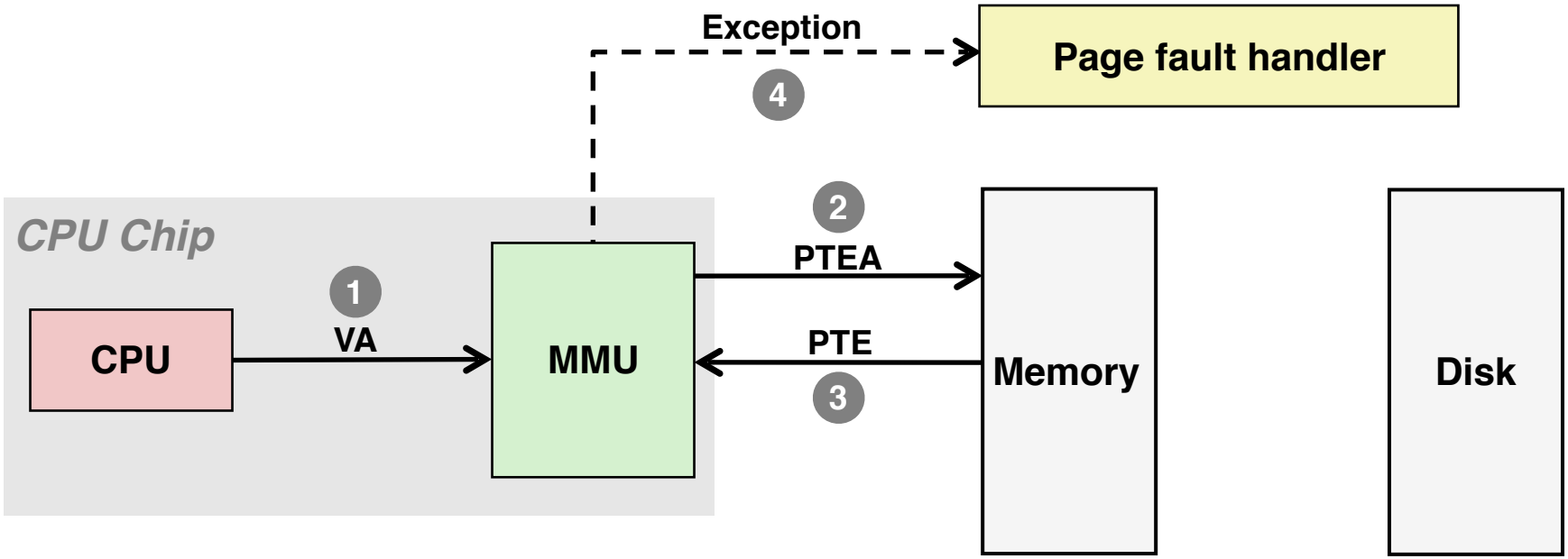
Address Translation: Page Fault



1) Processor sends virtual address to MMU

2-3) MMU fetches PTE from page table in memory

Address Translation: Page Fault

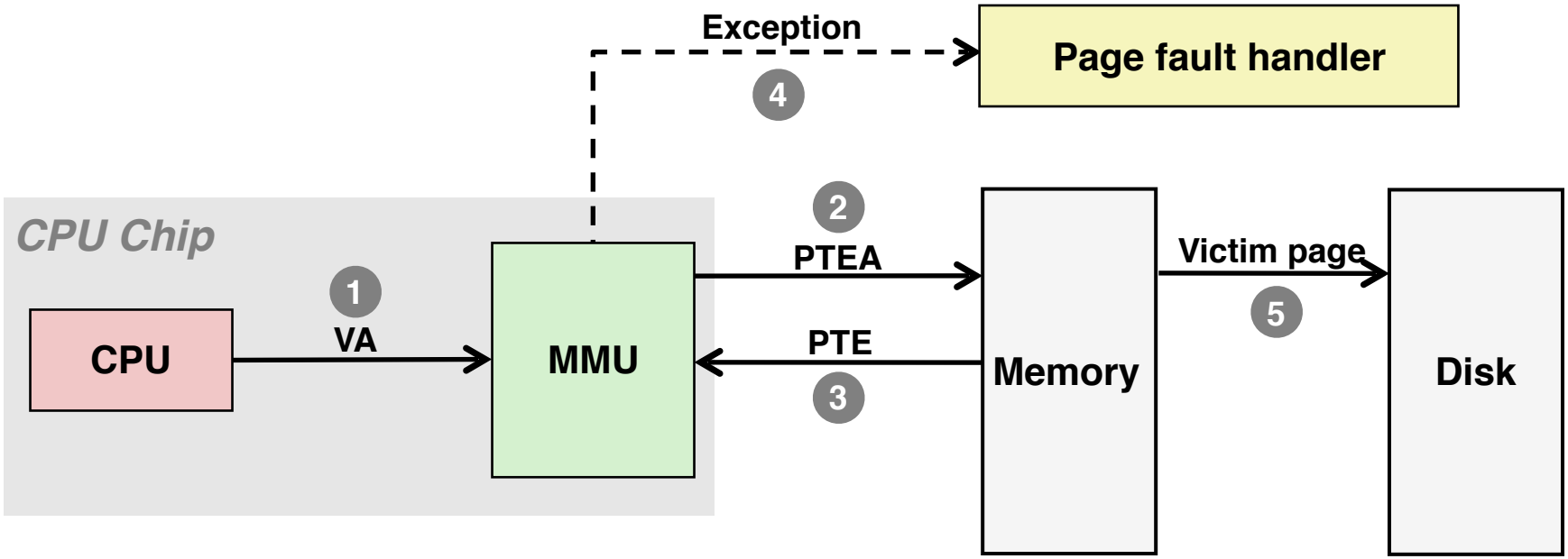


1) Processor sends virtual address to MMU

2-3) MMU fetches PTE from page table in memory

4) Valid bit is zero, so MMU triggers page fault exception

Address Translation: Page Fault



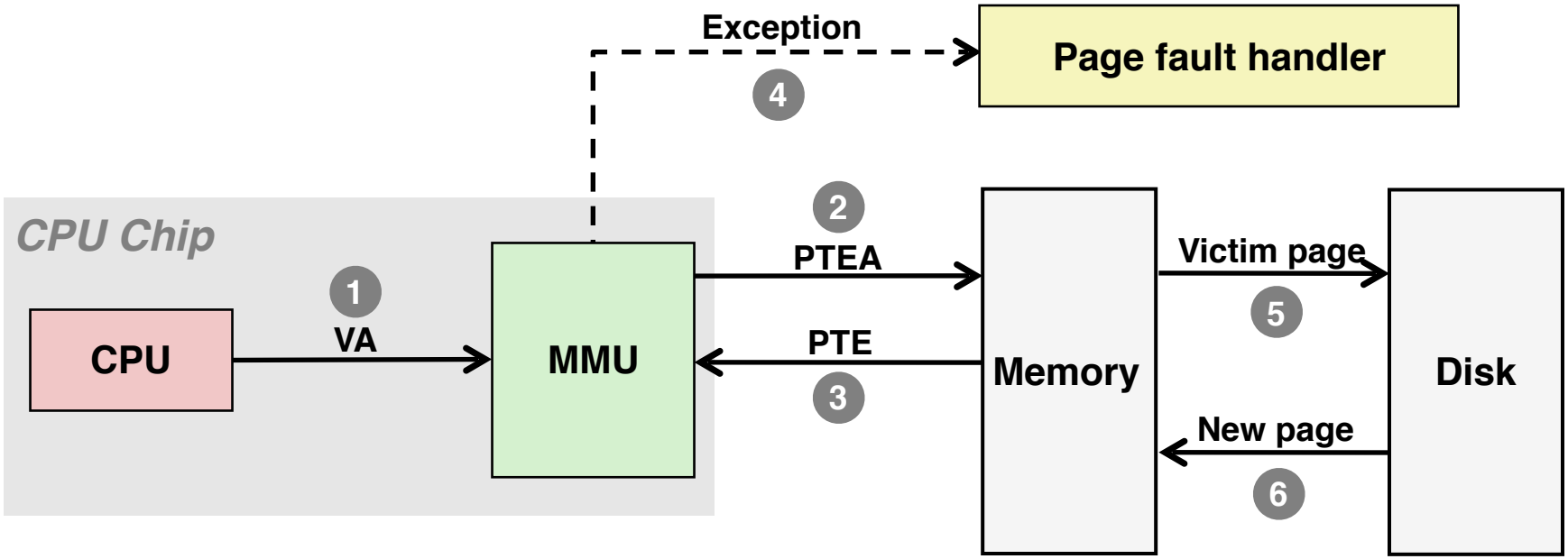
1) Processor sends virtual address to MMU

2-3) MMU fetches PTE from page table in memory

4) Valid bit is zero, so MMU triggers page fault exception

5) Handler identifies victim (and, if dirty, pages it out to disk)

Address Translation: Page Fault



1) Processor sends virtual address to MMU

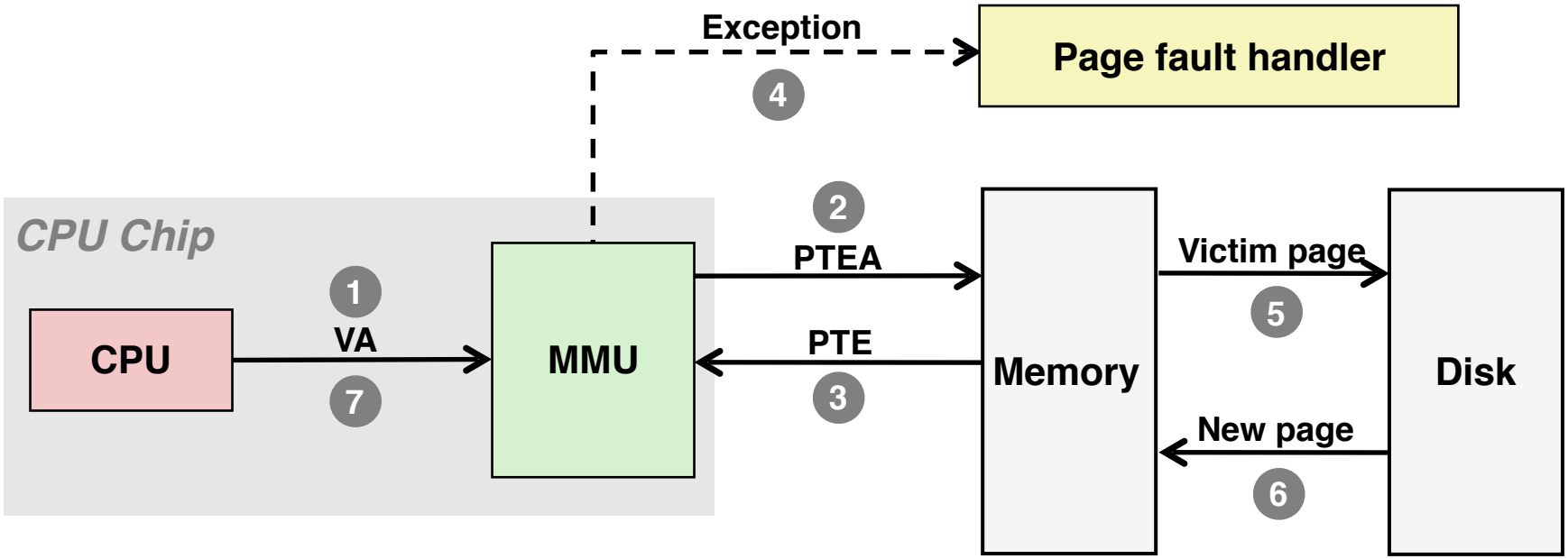
2-3) MMU fetches PTE from page table in memory

4) Valid bit is zero, so MMU triggers page fault exception

5) Handler identifies victim (and, if dirty, pages it out to disk)

6) Handler pages in new page and updates PTE in memory

Address Translation: Page Fault



1) Processor sends virtual address to MMU

2-3) MMU fetches PTE from page table in memory

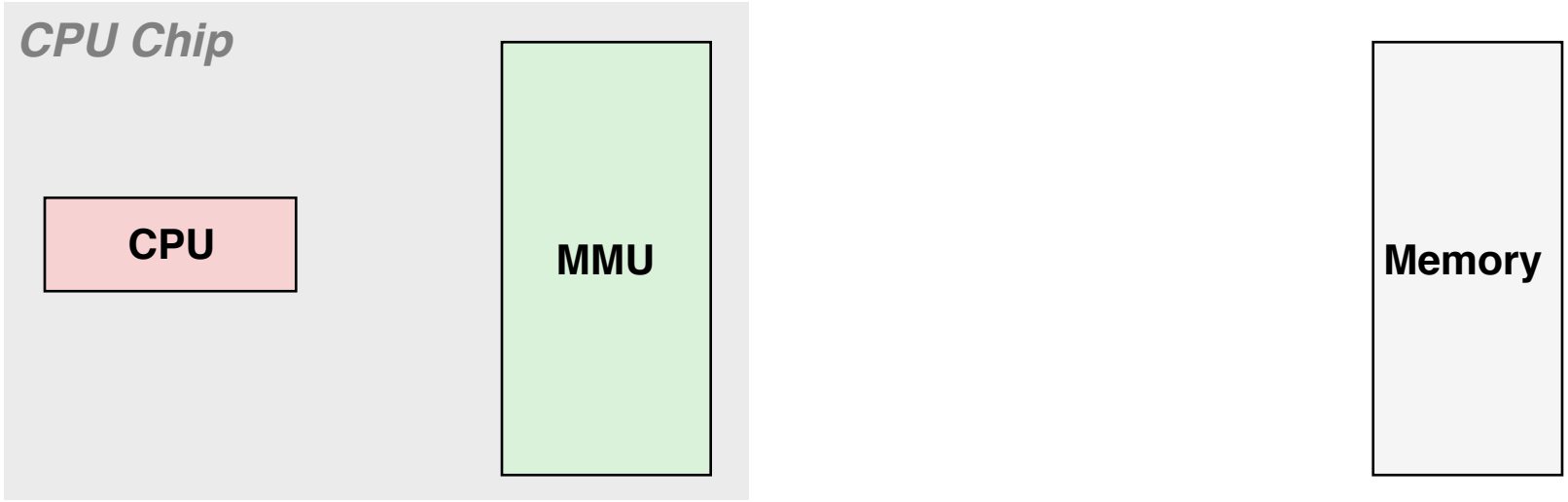
4) Valid bit is zero, so MMU triggers page fault exception

5) Handler identifies victim (and, if dirty, pages it out to disk)

6) Handler pages in new page and updates PTE in memory

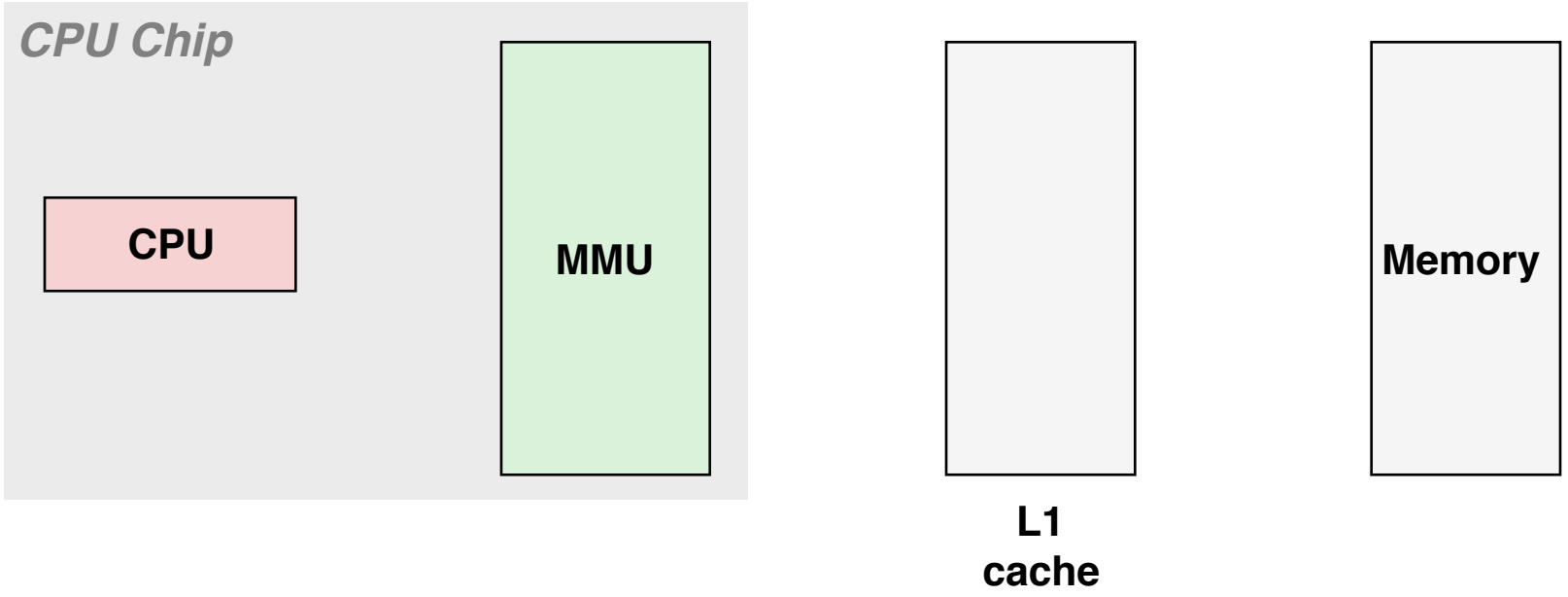
7) Handler returns to original process, restarting faulting instruction

Integrating VM and Cache



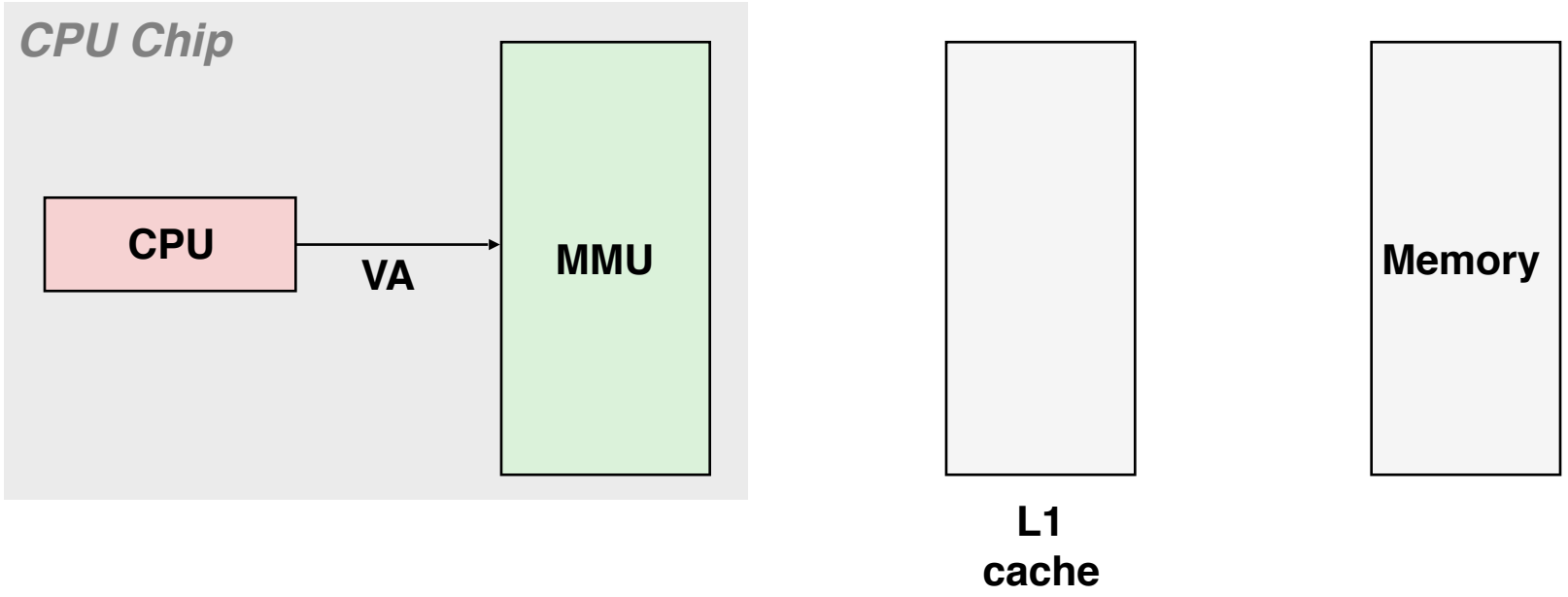
VA: virtual address, PA: physical address, PTE: page table entry, PTEA = PTE address

Integrating VM and Cache



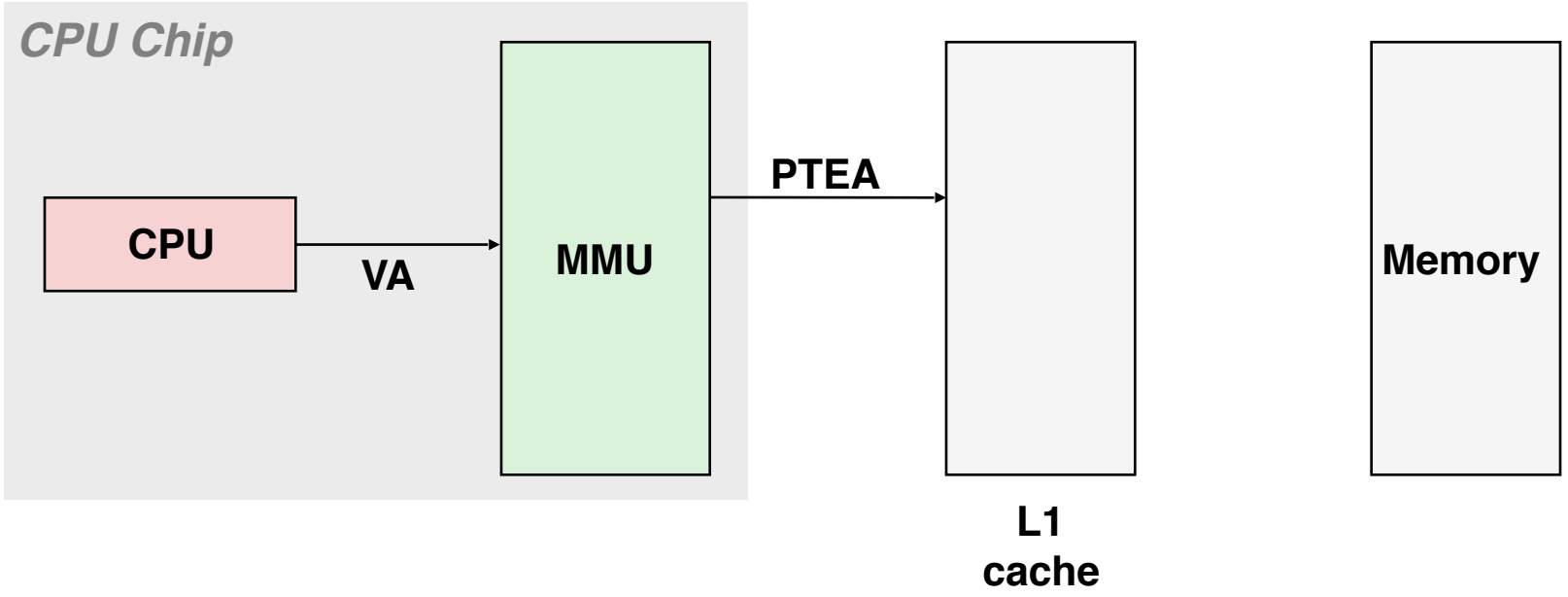
VA: virtual address, PA: physical address, PTE: page table entry, PTEA = PTE address

Integrating VM and Cache



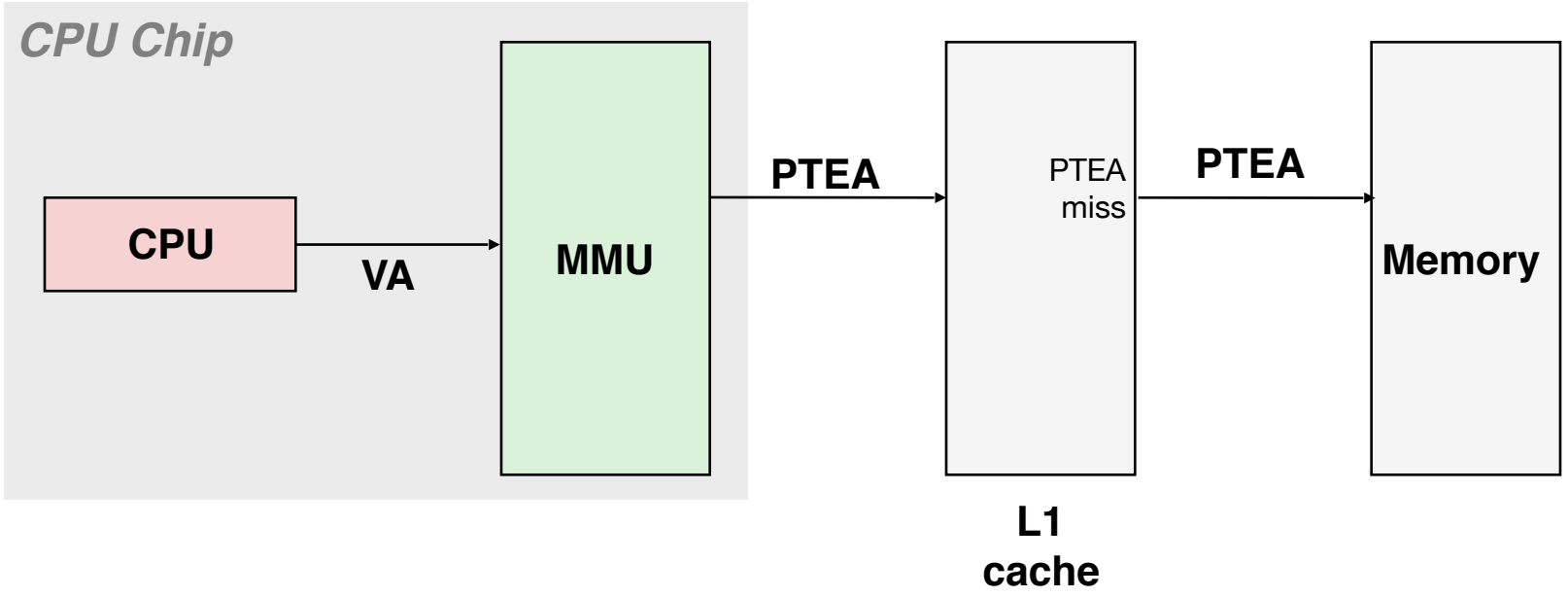
VA: virtual address, PA: physical address, PTE: page table entry, PTEA = PTE address

Integrating VM and Cache



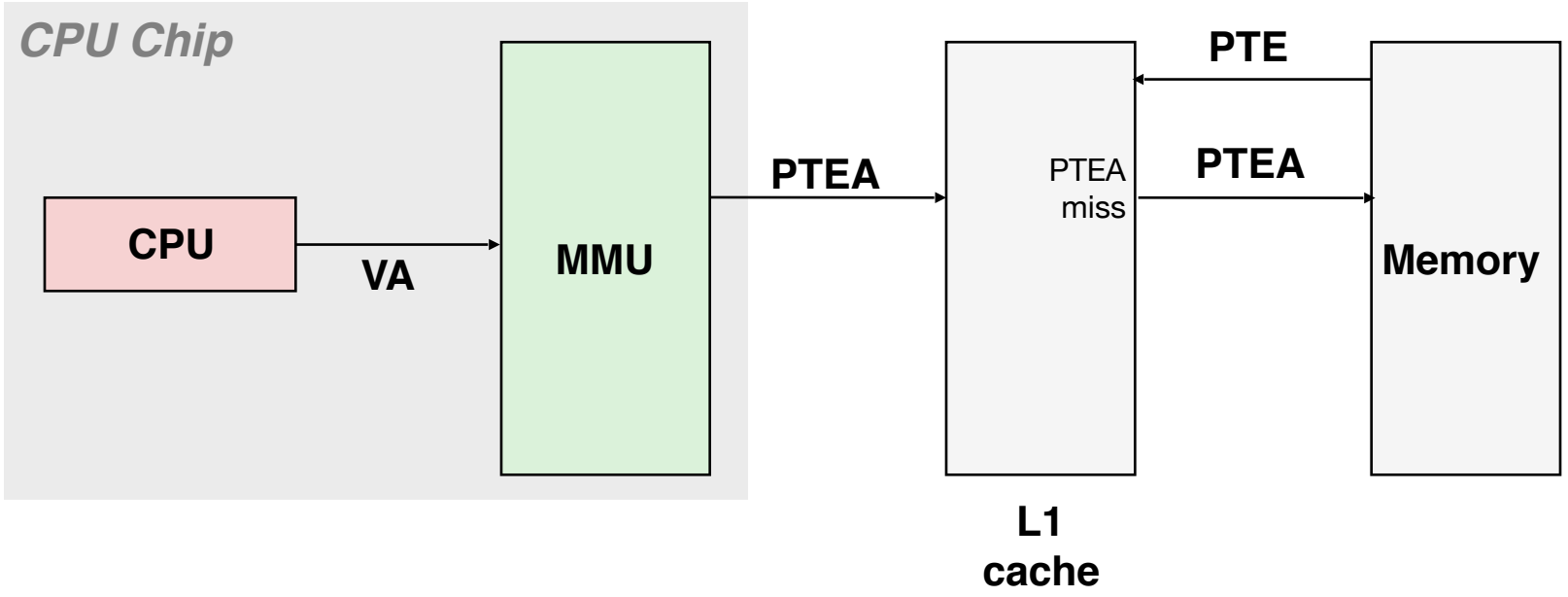
VA: virtual address, PA: physical address, PTE: page table entry, PTEA = PTE address

Integrating VM and Cache



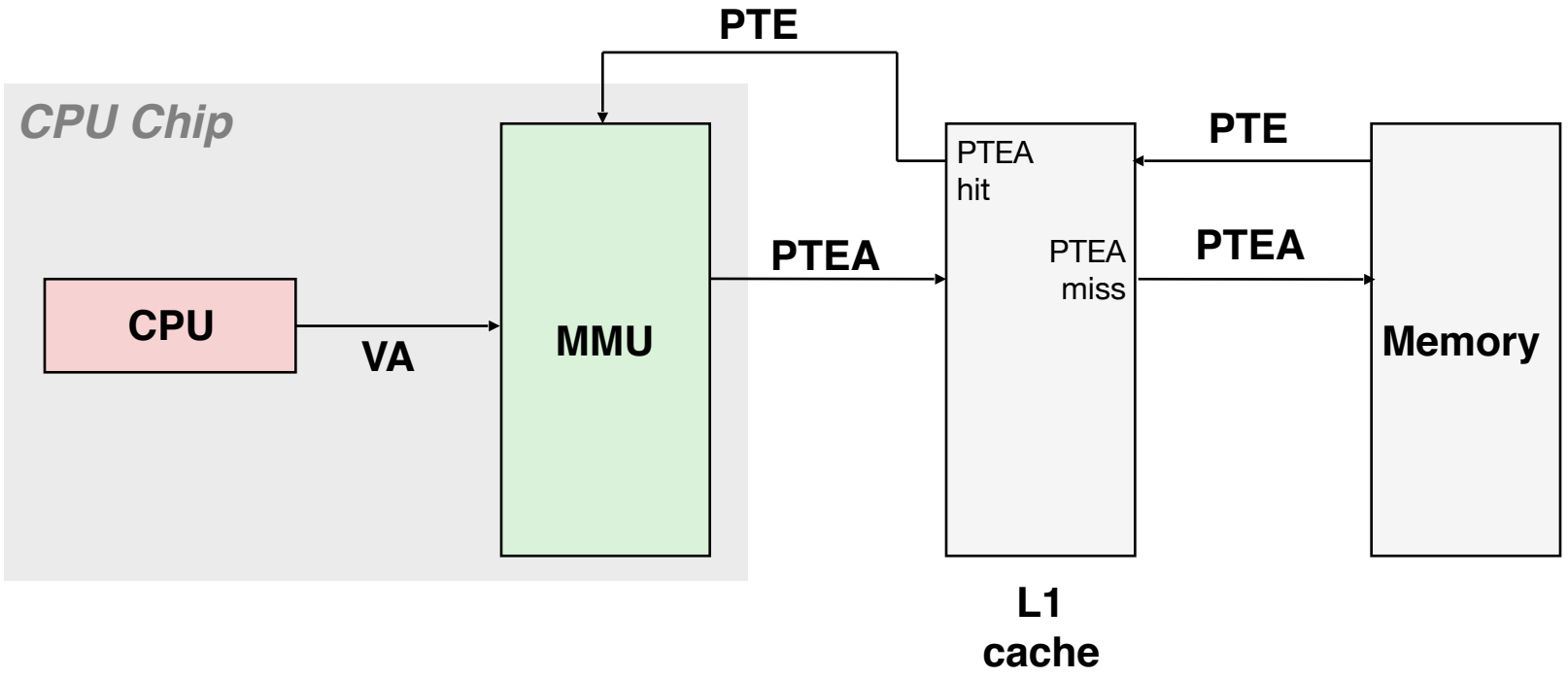
VA: virtual address, PA: physical address, PTE: page table entry, PTEA = PTE address

Integrating VM and Cache



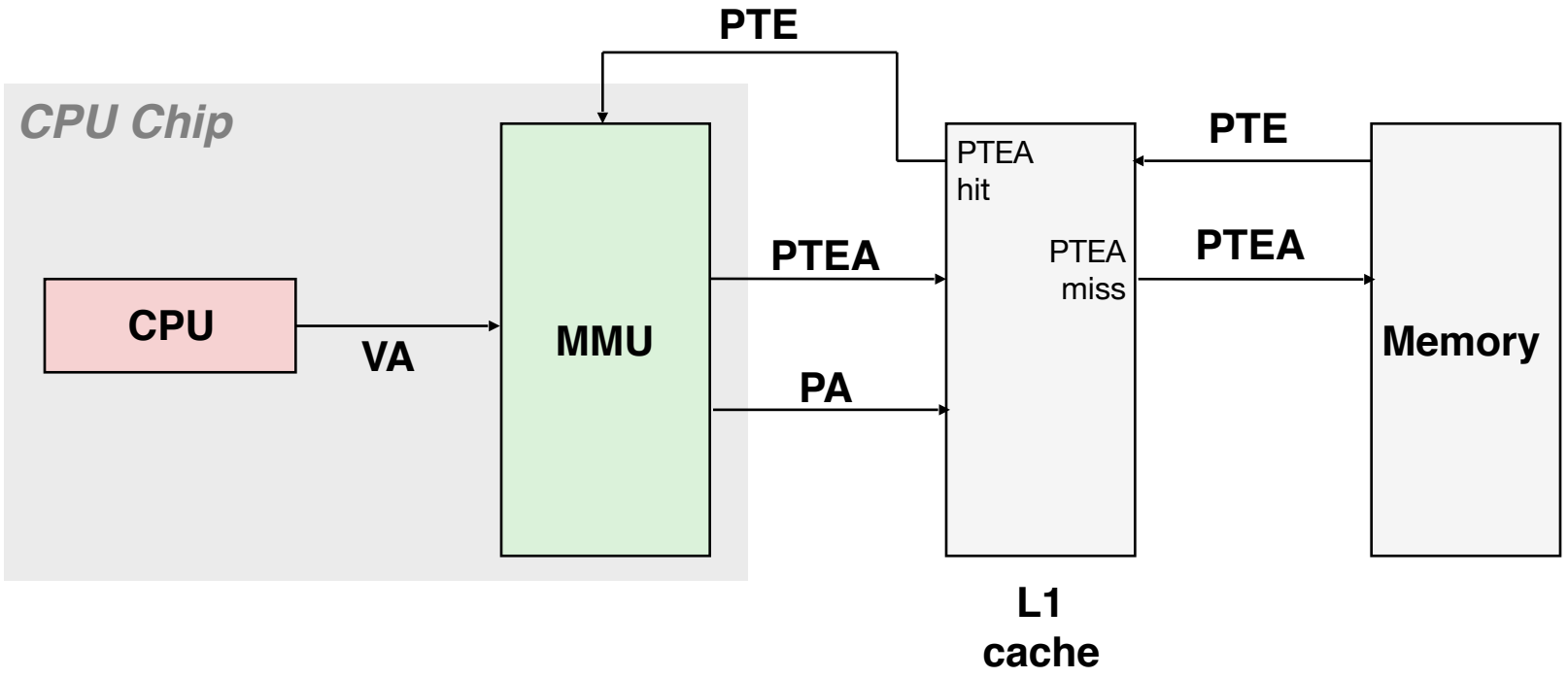
VA: virtual address, PA: physical address, PTE: page table entry, PTEA = PTE address

Integrating VM and Cache



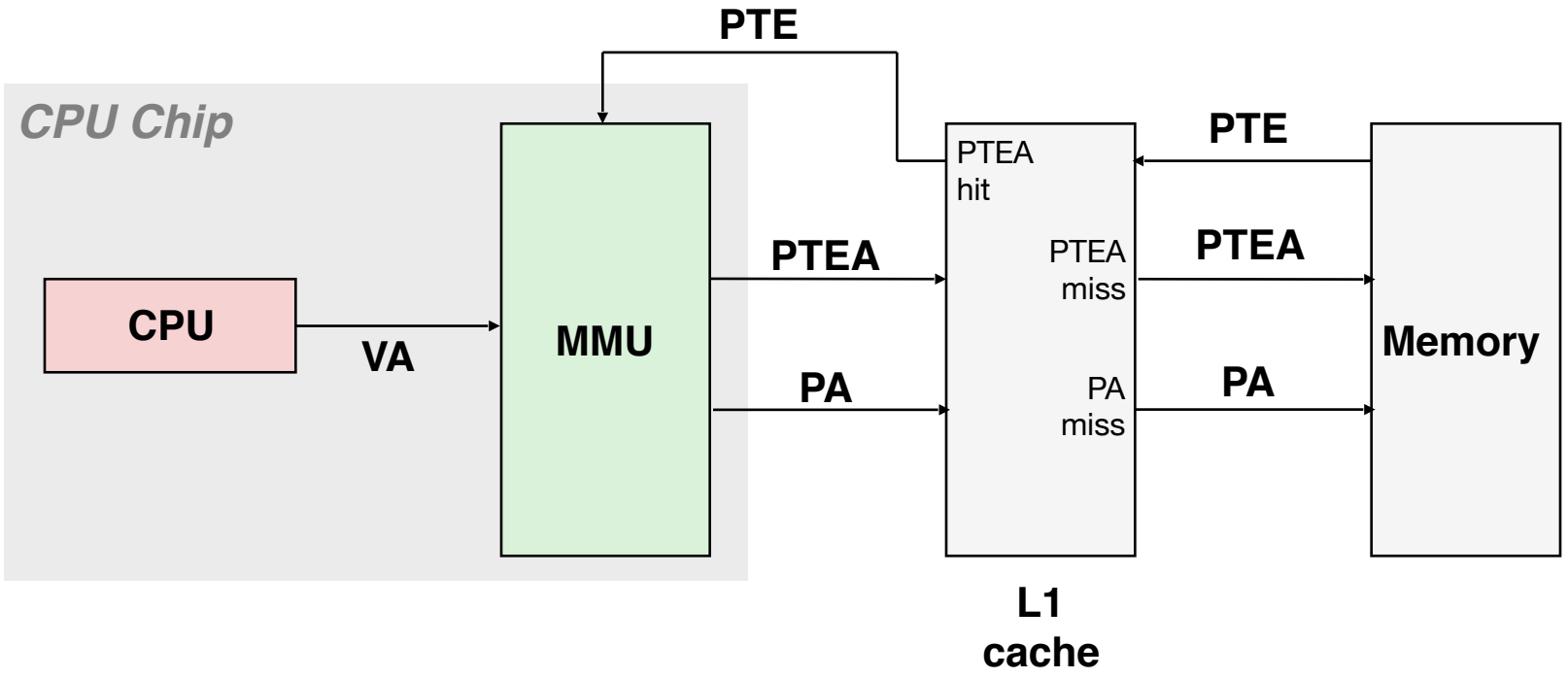
VA: virtual address, PA: physical address, PTE: page table entry, PTEA = PTE address

Integrating VM and Cache



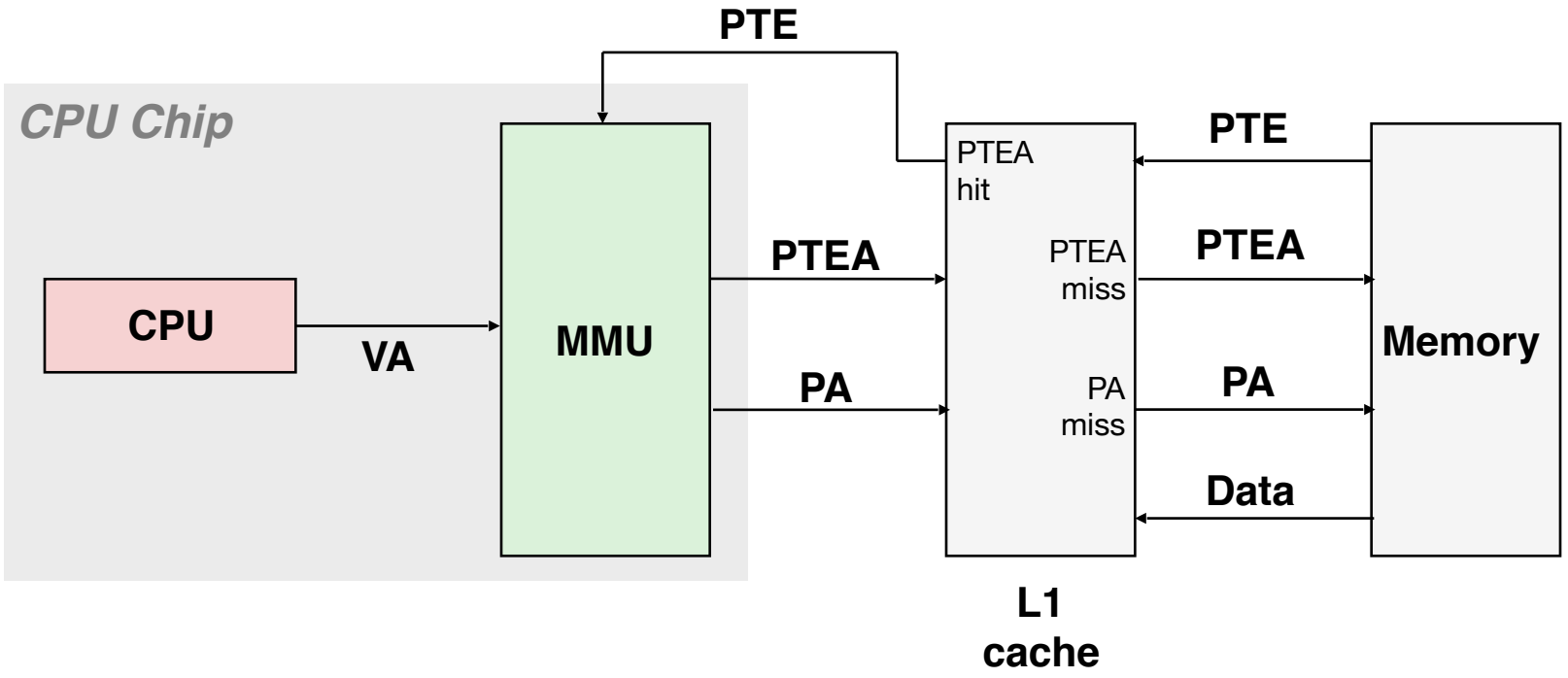
VA: virtual address, PA: physical address, PTE: page table entry, PTEA = PTE address

Integrating VM and Cache



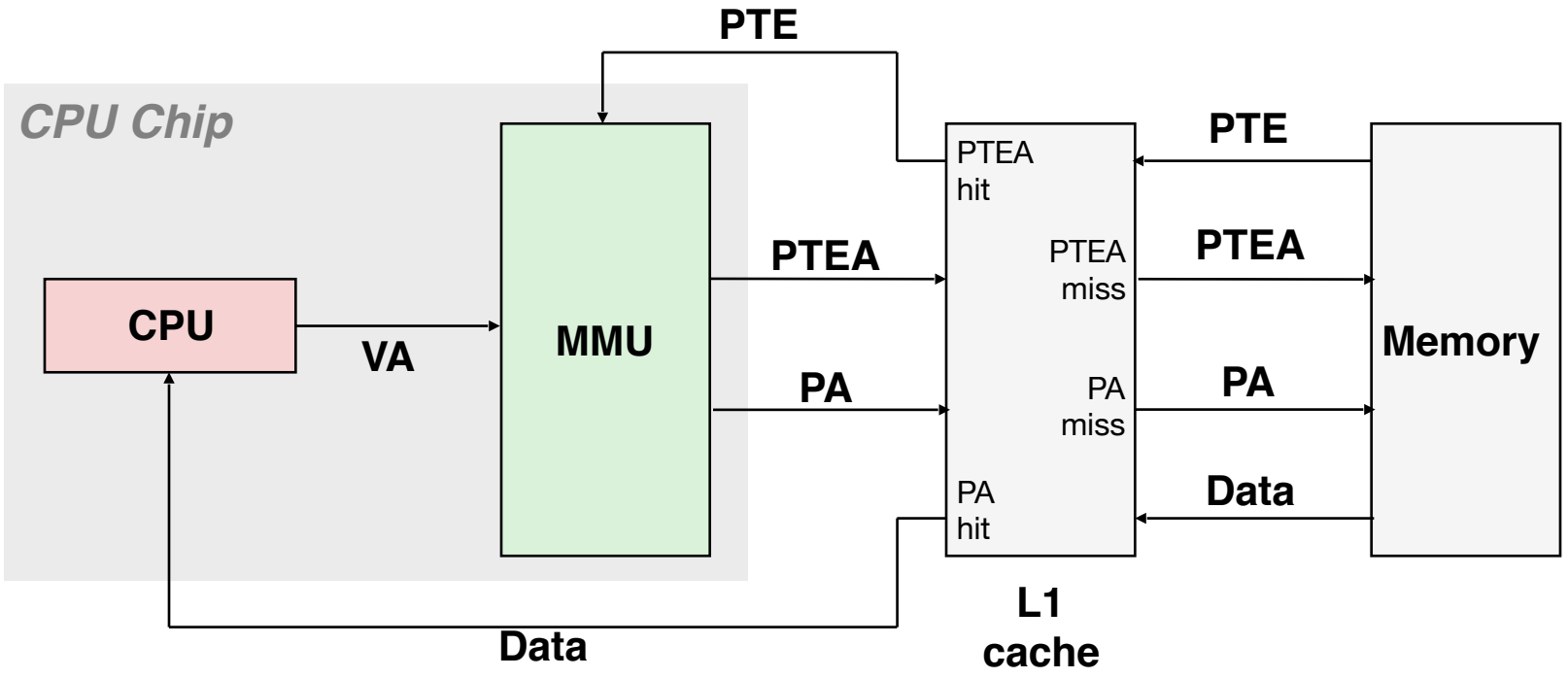
VA: virtual address, PA: physical address, PTE: page table entry, PTEA = PTE address

Integrating VM and Cache



VA: virtual address, PA: physical address, PTE: page table entry, PTEA = PTE address

Integrating VM and Cache



VA: virtual address, PA: physical address, PTE: page table entry, PTEA = PTE address

Today

- Three Virtual Memory Optimizations
  - TLB
  - Page the page table (a.k.a., multi-level page table)
  - Virtually-indexed, physically-tagged cache
- Case-study: Intel Core i7/Linux example

Speeding up Address Translation

Speeding up Address Translation

- Problem: Every memory load/store requires two memory accesses: one for PTE, another for real
  - The PTE access is kind of an overhead
  - Can we speed it up?

Speeding up Address Translation

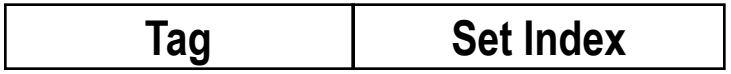
- Problem: Every memory load/store requires two memory accesses: one for PTE, another for real
  - The PTE access is kind of an overhead
  - Can we speed it up?
- Page table entries (PTEs) are already cached in L1 data cache like any other memory data. But:
  - PTEs may be evicted by other data references
  - PTE hit still requires a small L1 delay

Speeding up Translation with a TLB

- Solution: *Translation Lookaside Buffer* (TLB)
  - Think of it as a dedicated cache for page table
  - Small set-associative hardware cache in MMU
  - Contains complete page table entries for a small number of pages

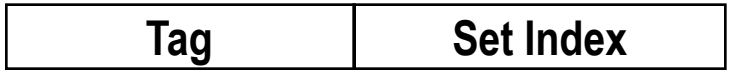
Speeding up Translation with a TLB

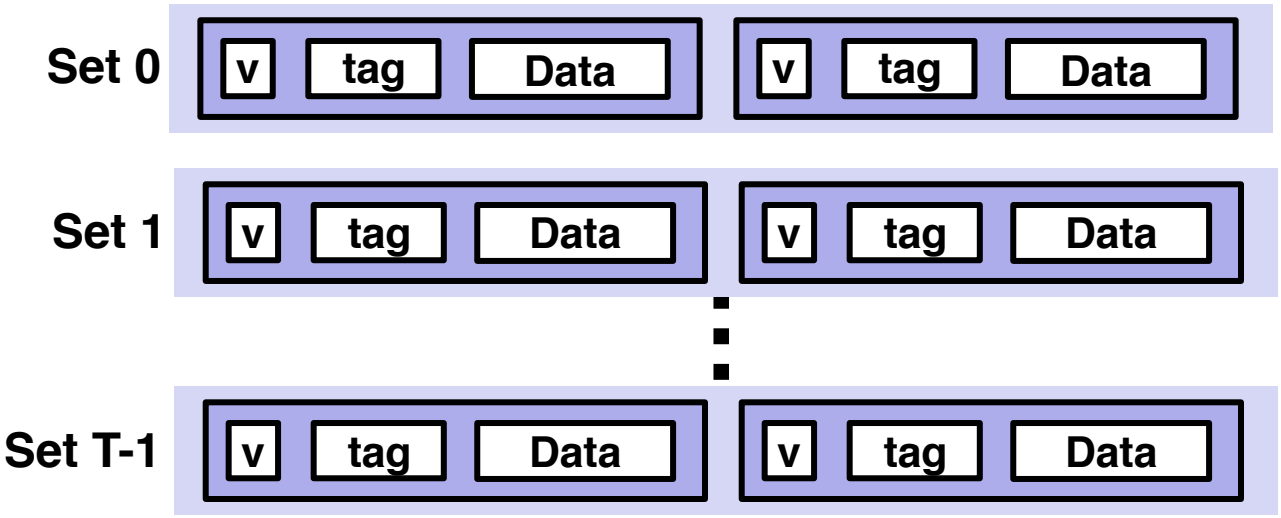
- Solution: *Translation Lookaside Buffer* (TLB)
  - Think of it as a dedicated cache for page table
  - Small set-associative hardware cache in MMU
  - Contains complete page table entries for a small number of pages



Speeding up Translation with a TLB

- Solution: *Translation Lookaside Buffer* (TLB)
  - Think of it as a dedicated cache for page table
  - Small set-associative hardware cache in MMU
  - Contains complete page table entries for a small number of pages

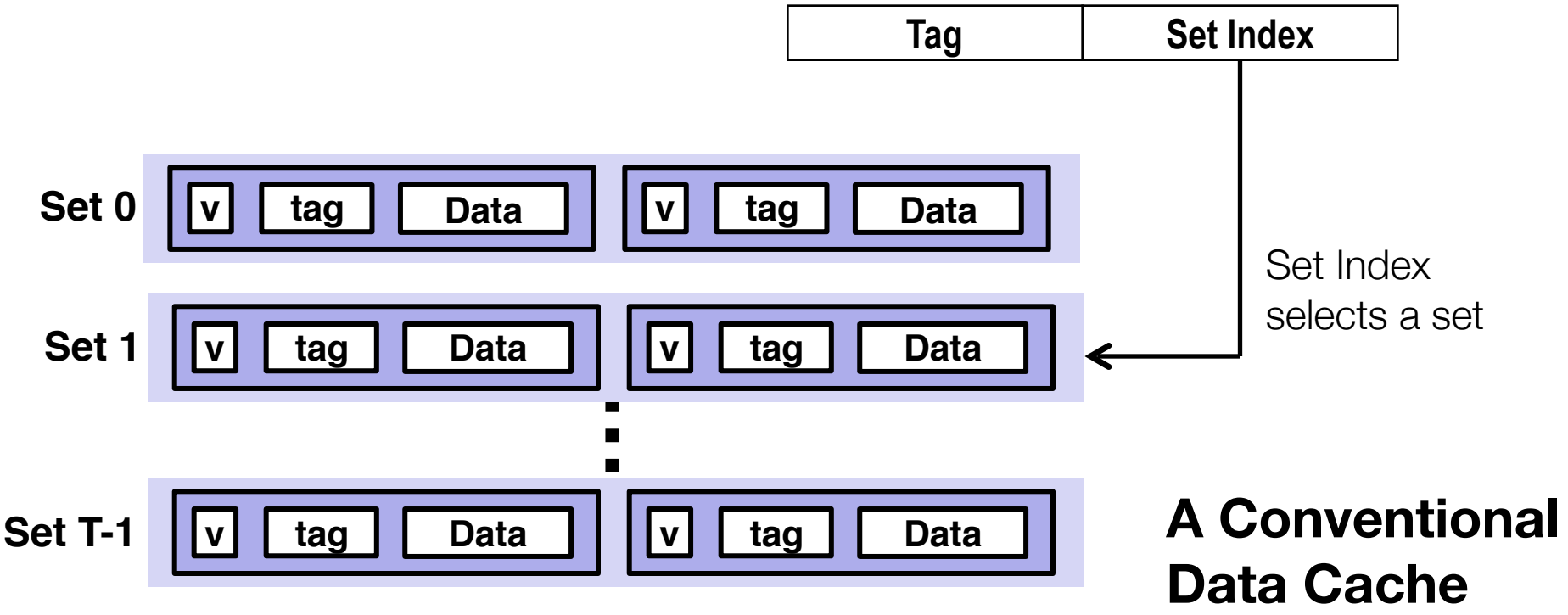




A Conventional Data Cache

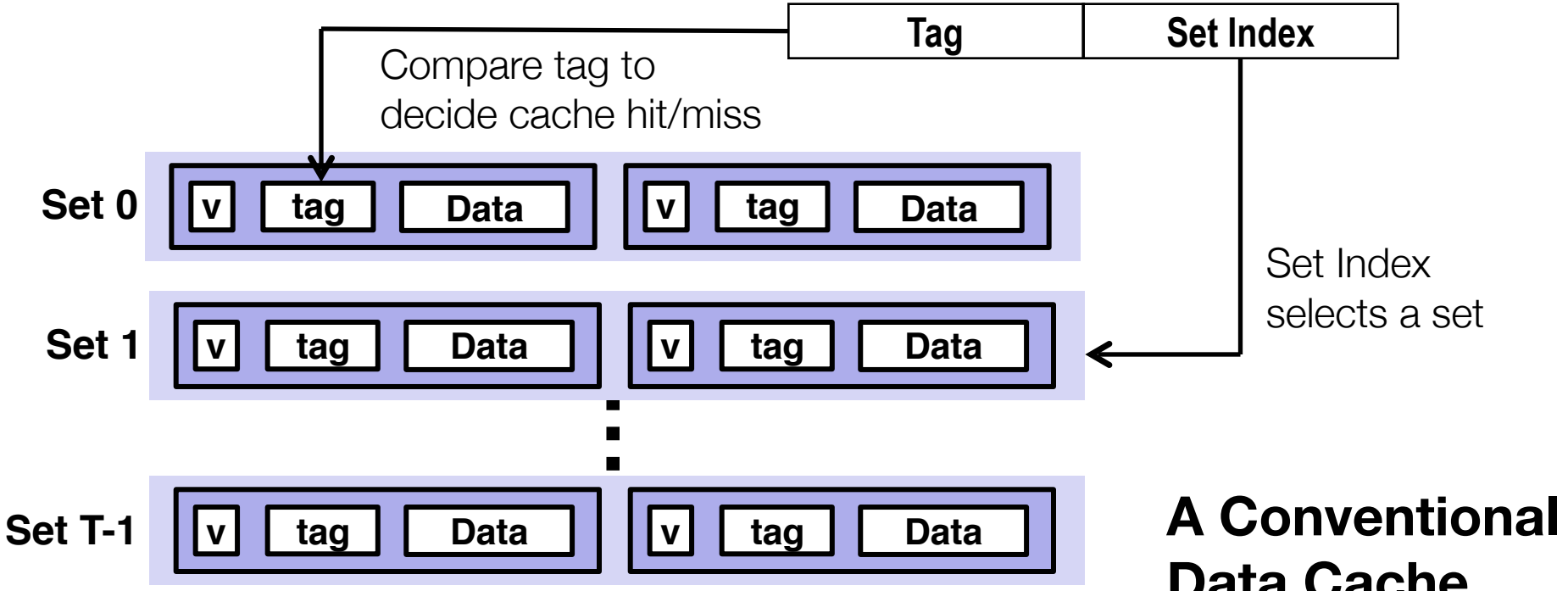
Speeding up Translation with a TLB

- Solution: *Translation Lookaside Buffer* (TLB)
  - Think of it as a dedicated cache for page table
  - Small set-associative hardware cache in MMU
  - Contains complete page table entries for a small number of pages



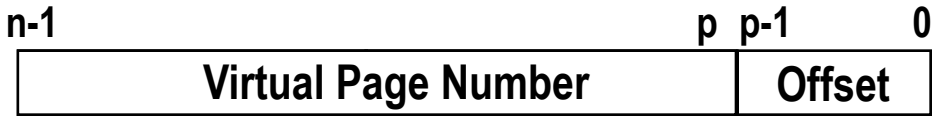
Speeding up Translation with a TLB

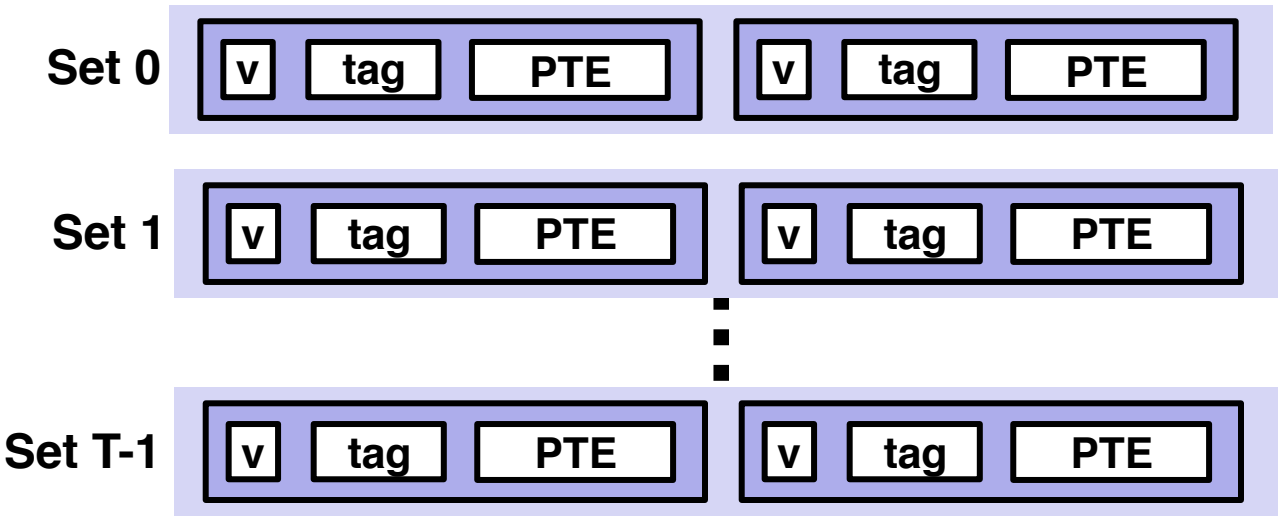
- Solution: *Translation Lookaside Buffer* (TLB)
  - Think of it as a dedicated cache for page table
  - Small set-associative hardware cache in MMU
  - Contains complete page table entries for a small number of pages



Accessing the TLB

- MMU uses the Virtual Page Number portion of the virtual address to access the TLB:

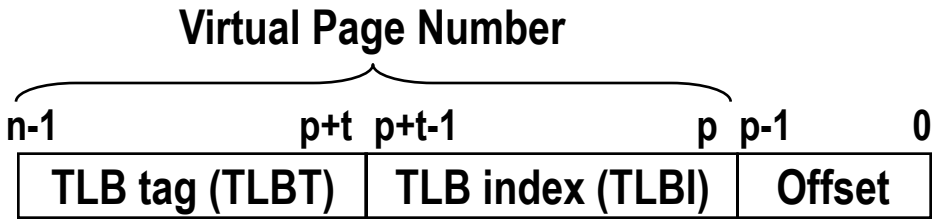


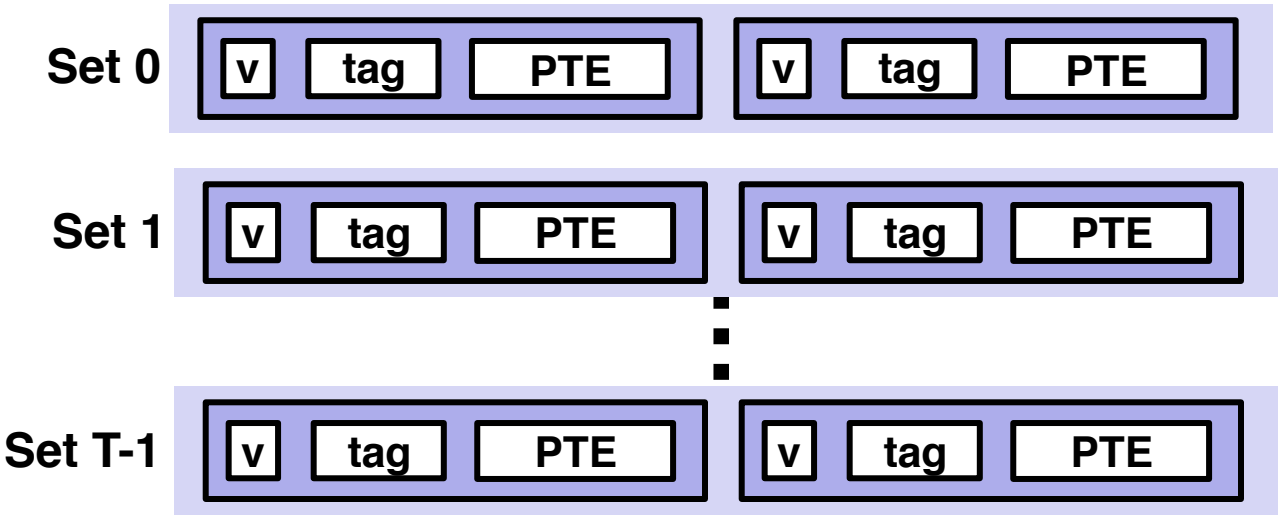


A Page Table Cache

Accessing the TLB

- MMU uses the Virtual Page Number portion of the virtual address to access the TLB:

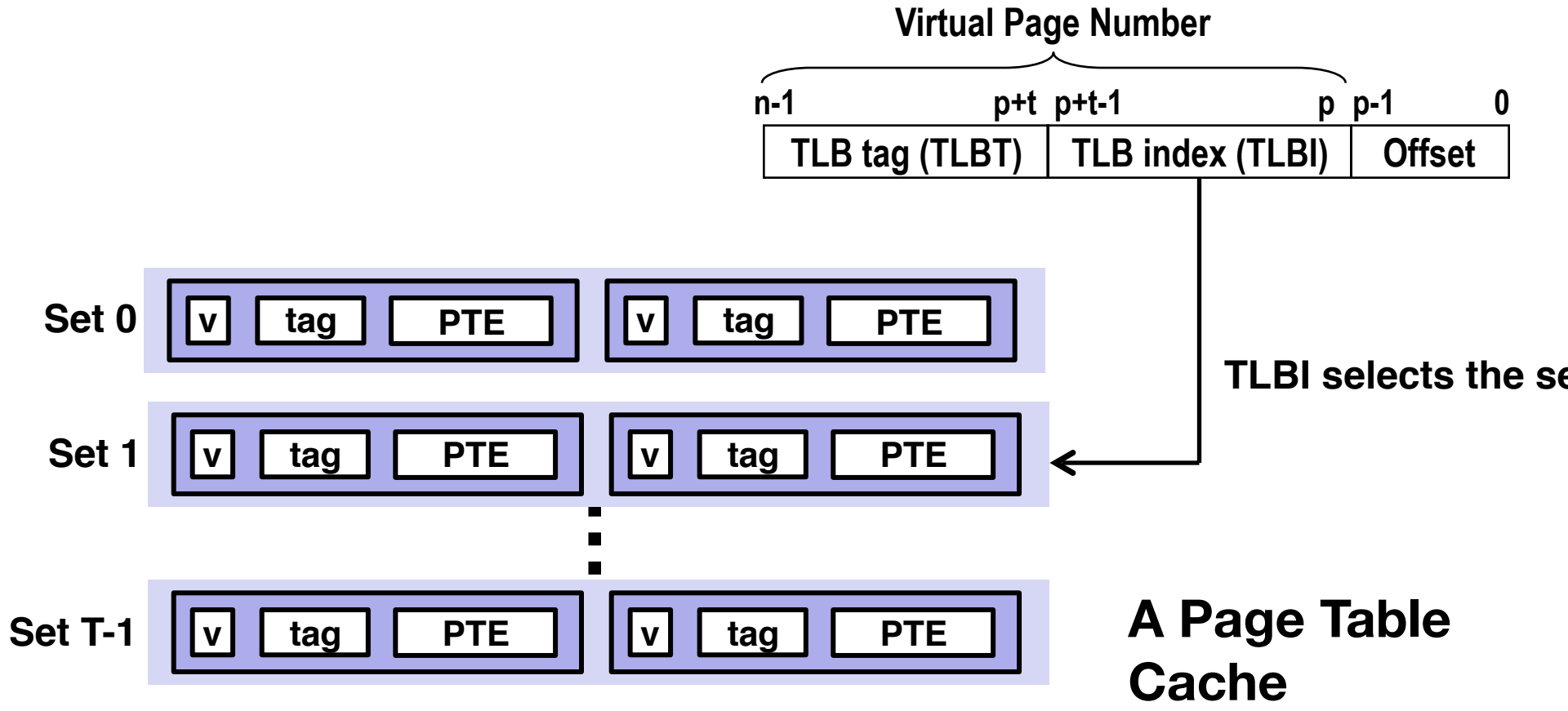




A Page Table Cache

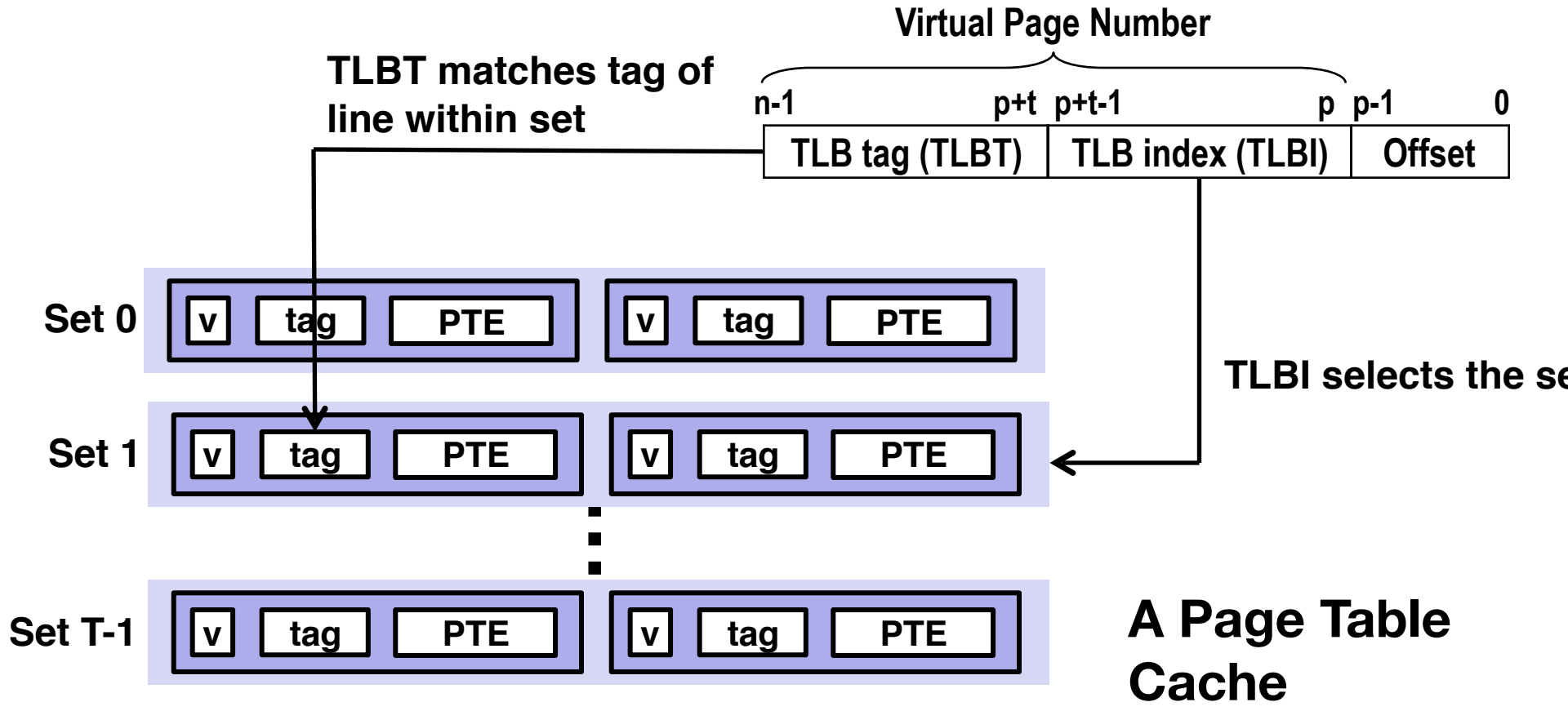
Accessing the TLB

- MMU uses the Virtual Page Number portion of the virtual address to access the TLB:

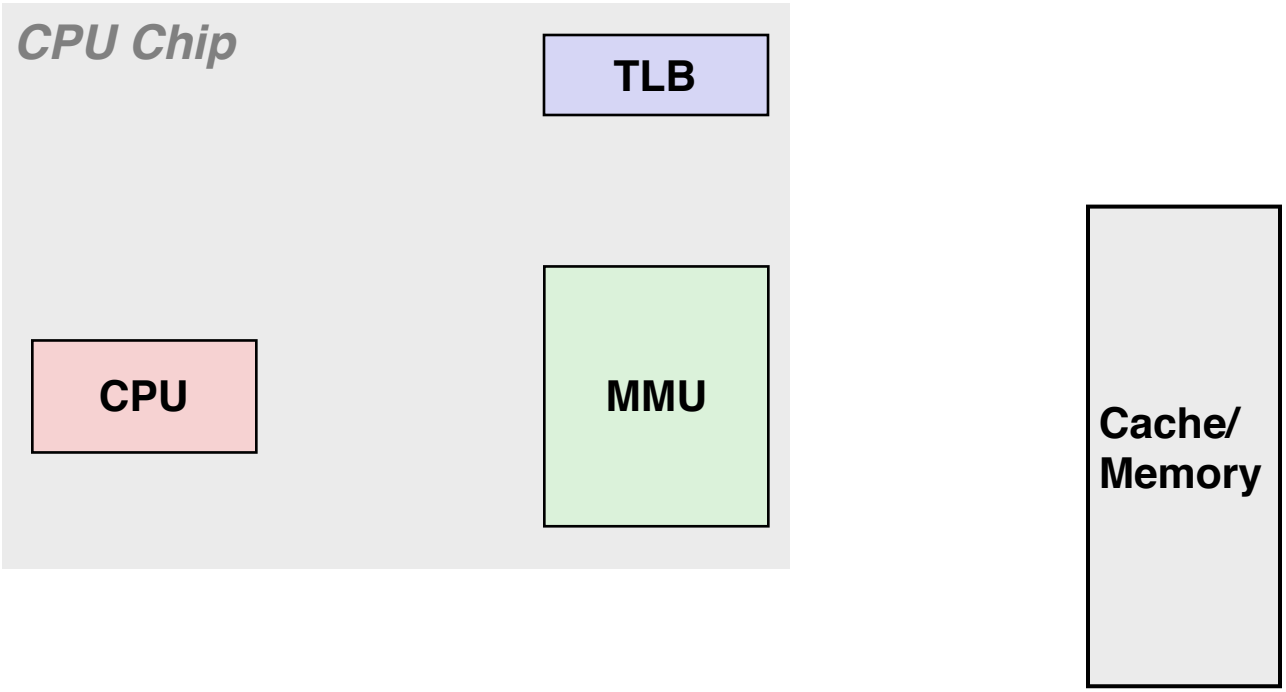


Accessing the TLB

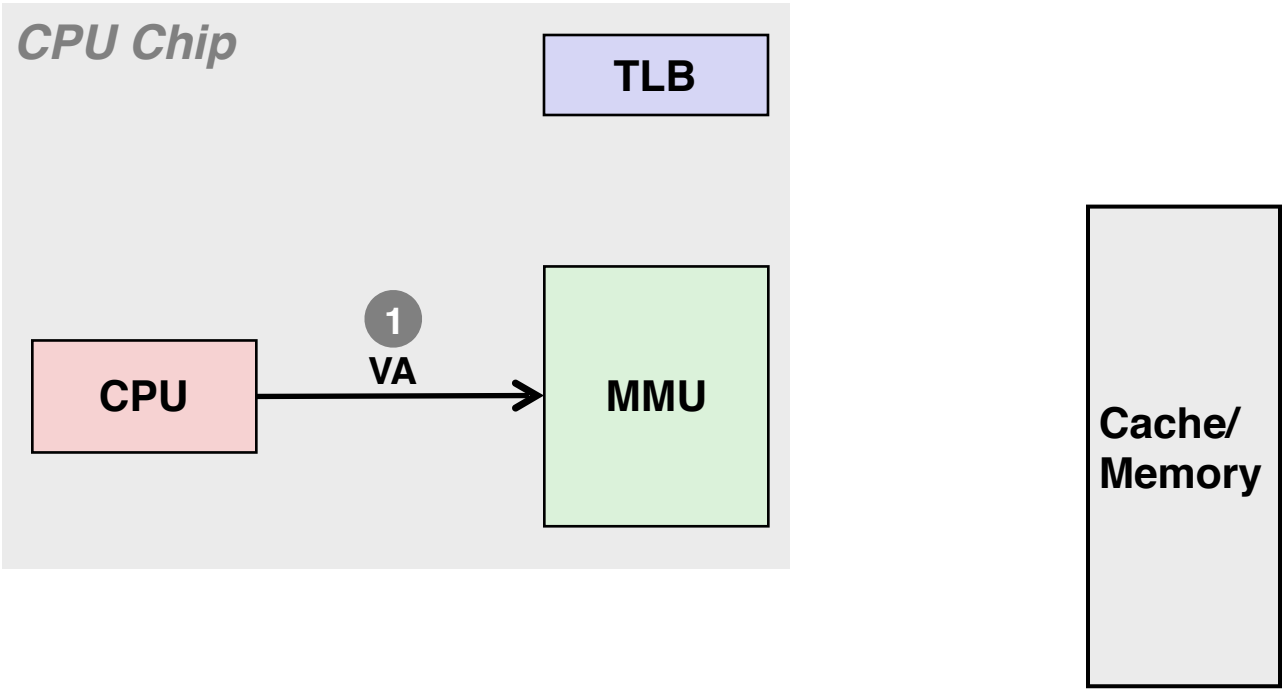
- MMU uses the Virtual Page Number portion of the virtual address to access the TLB:



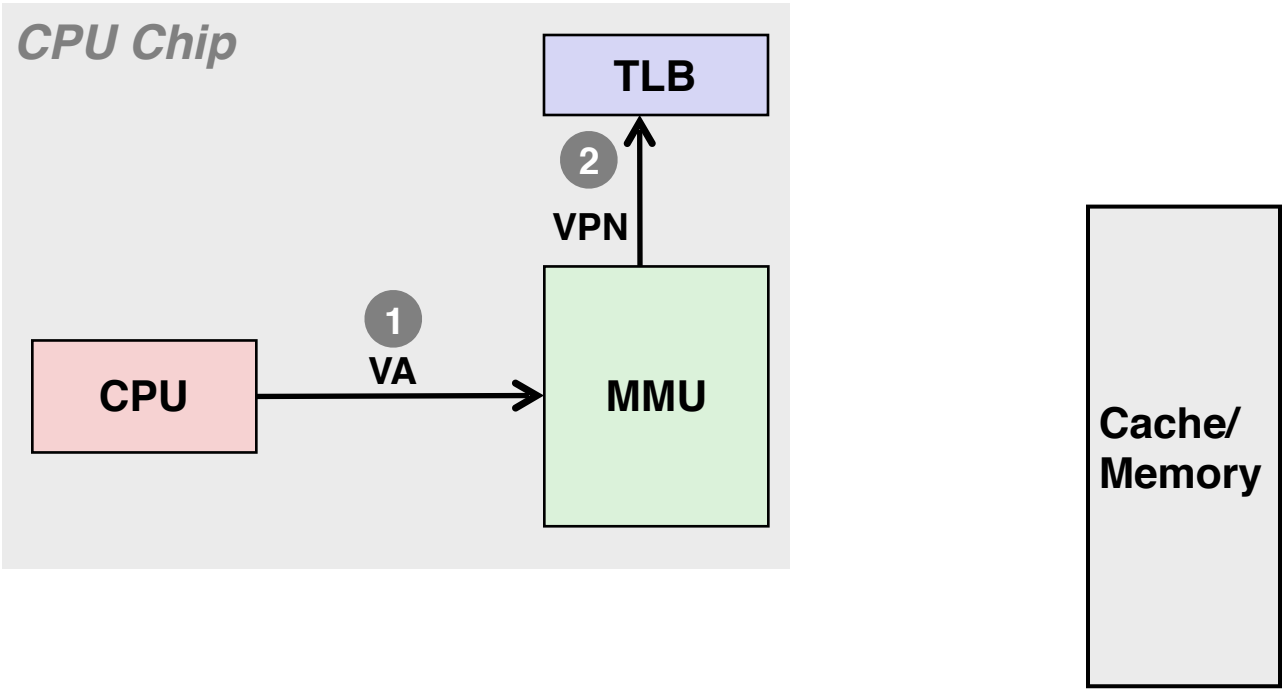
TLB Hit



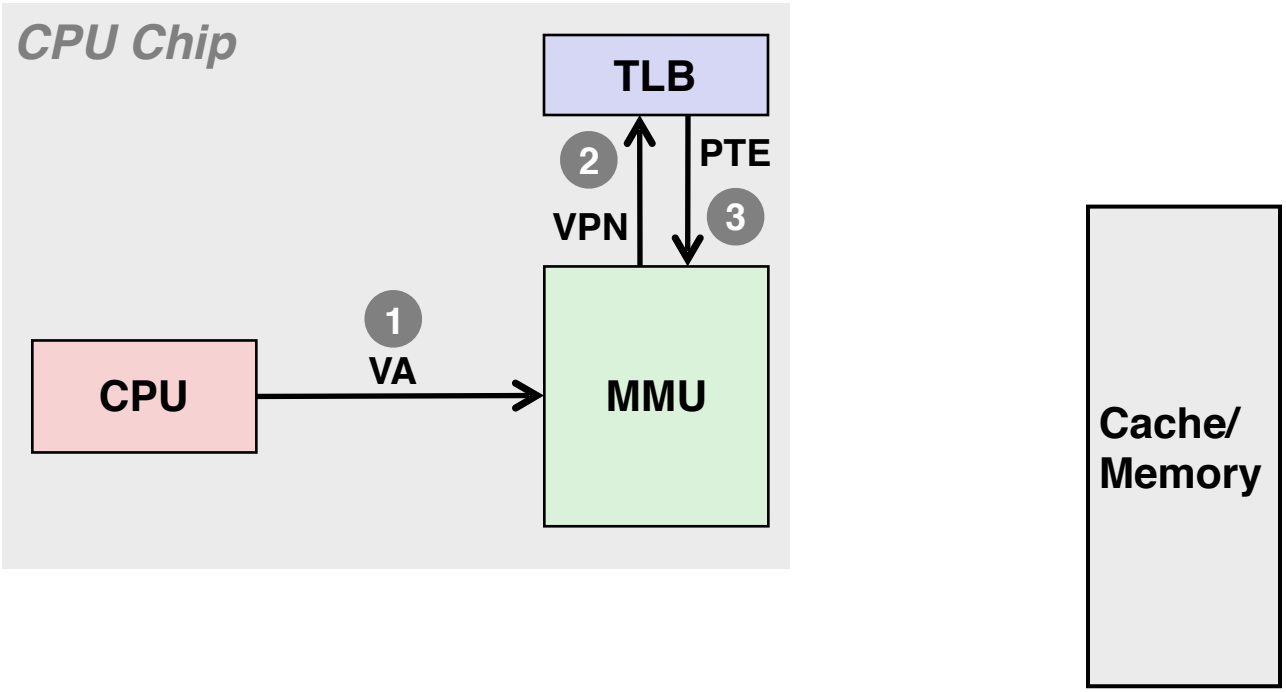
TLB Hit



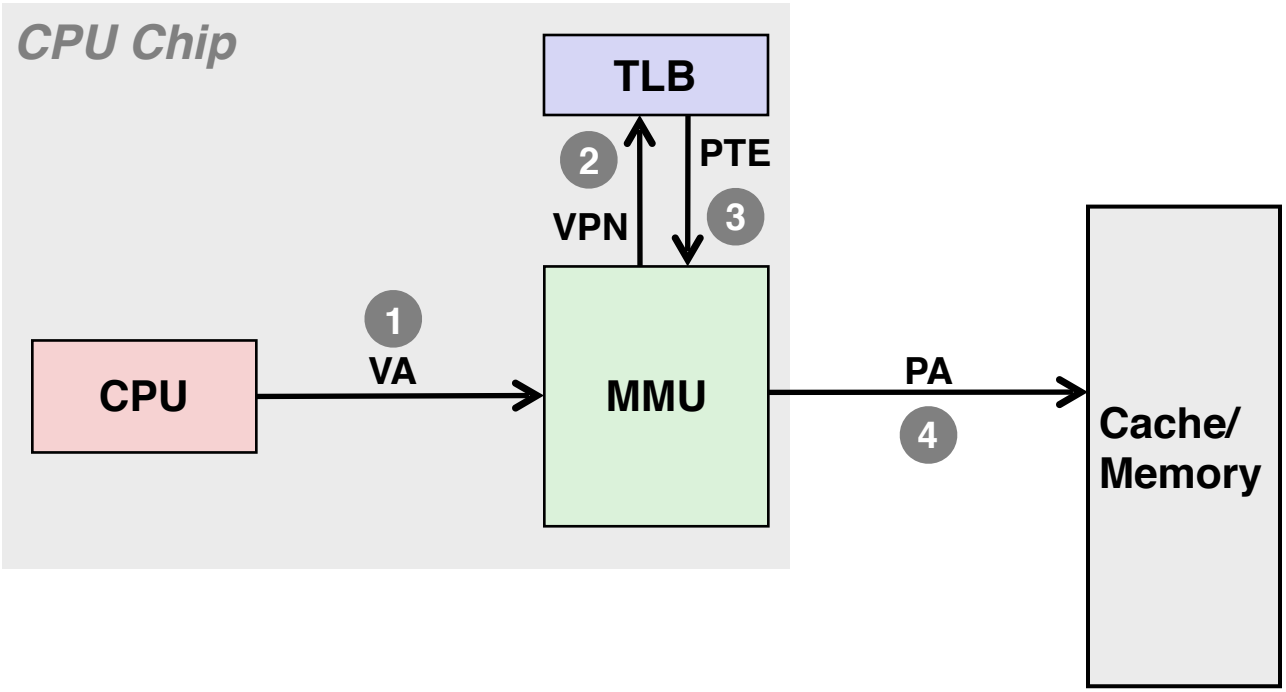
TLB Hit



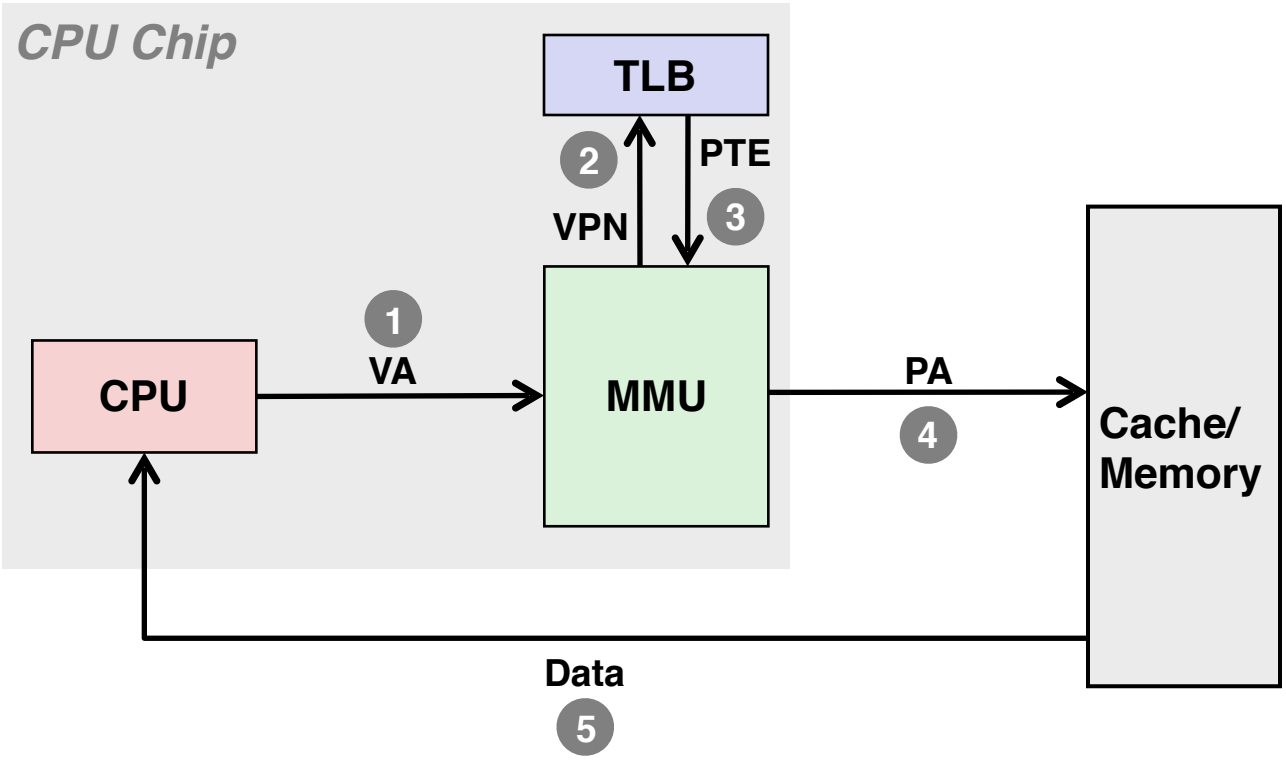
TLB Hit



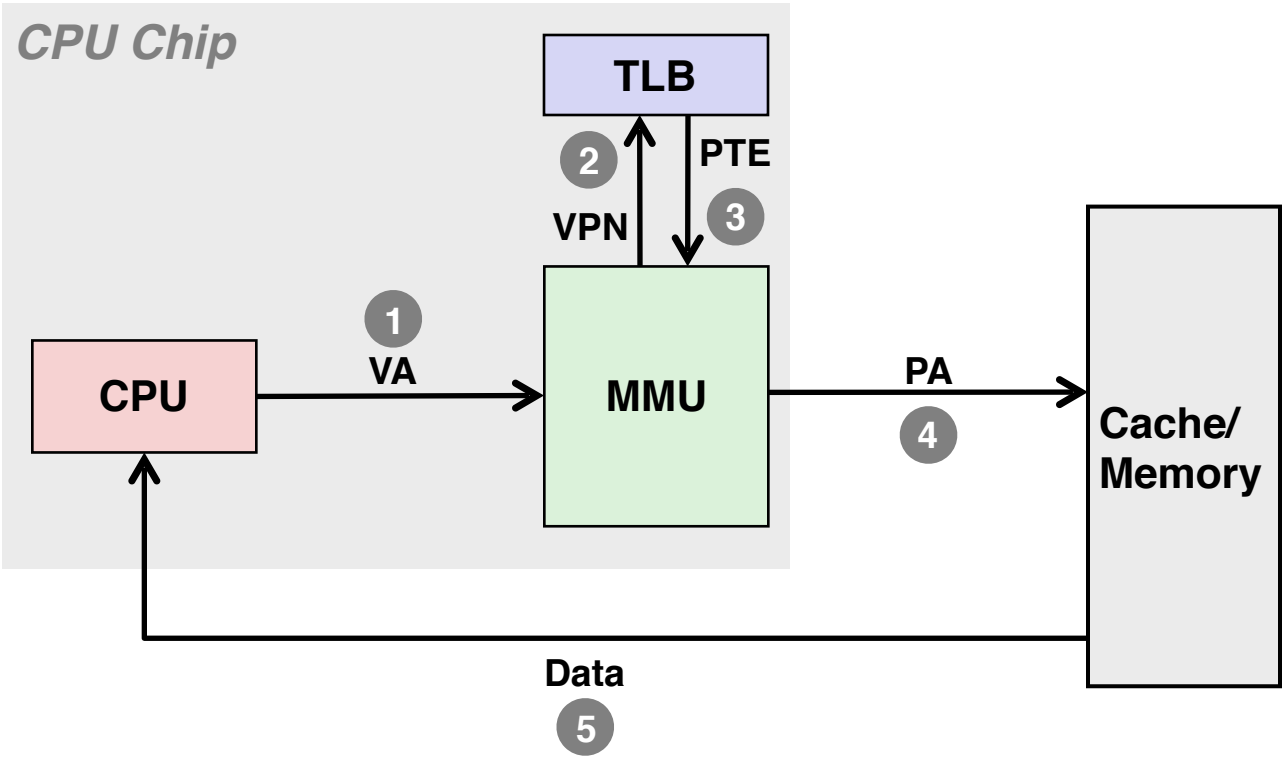
TLB Hit



TLB Hit

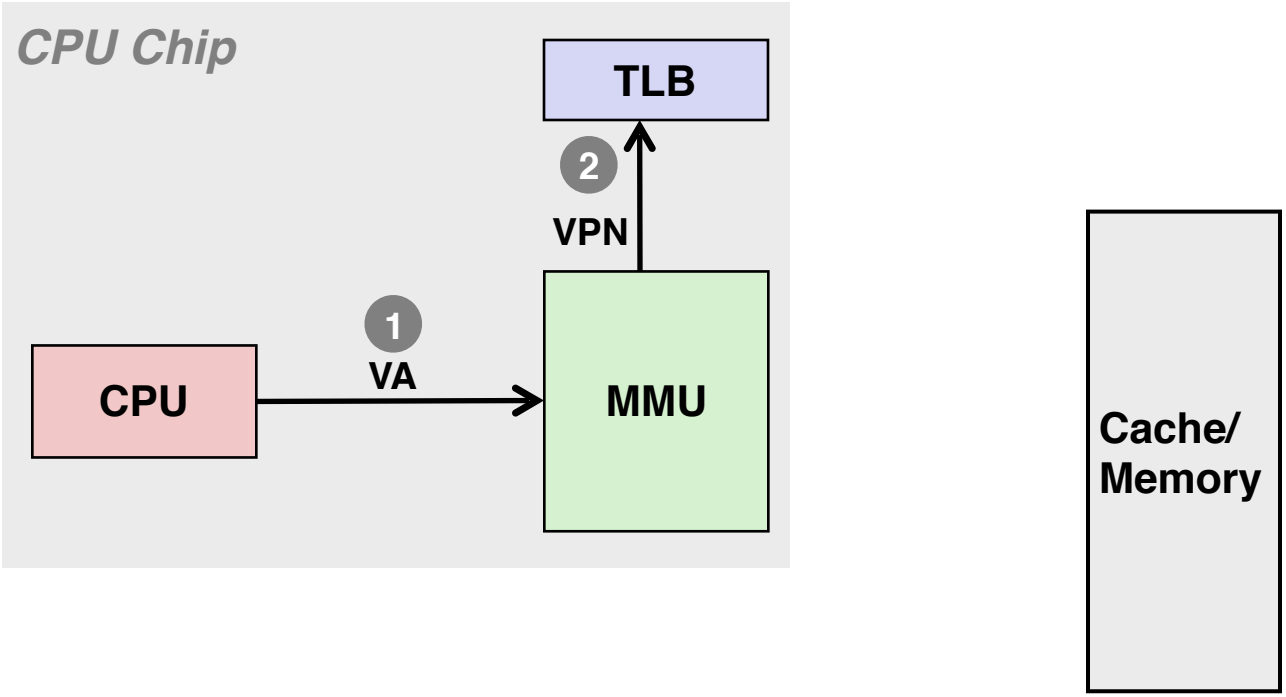


TLB Hit

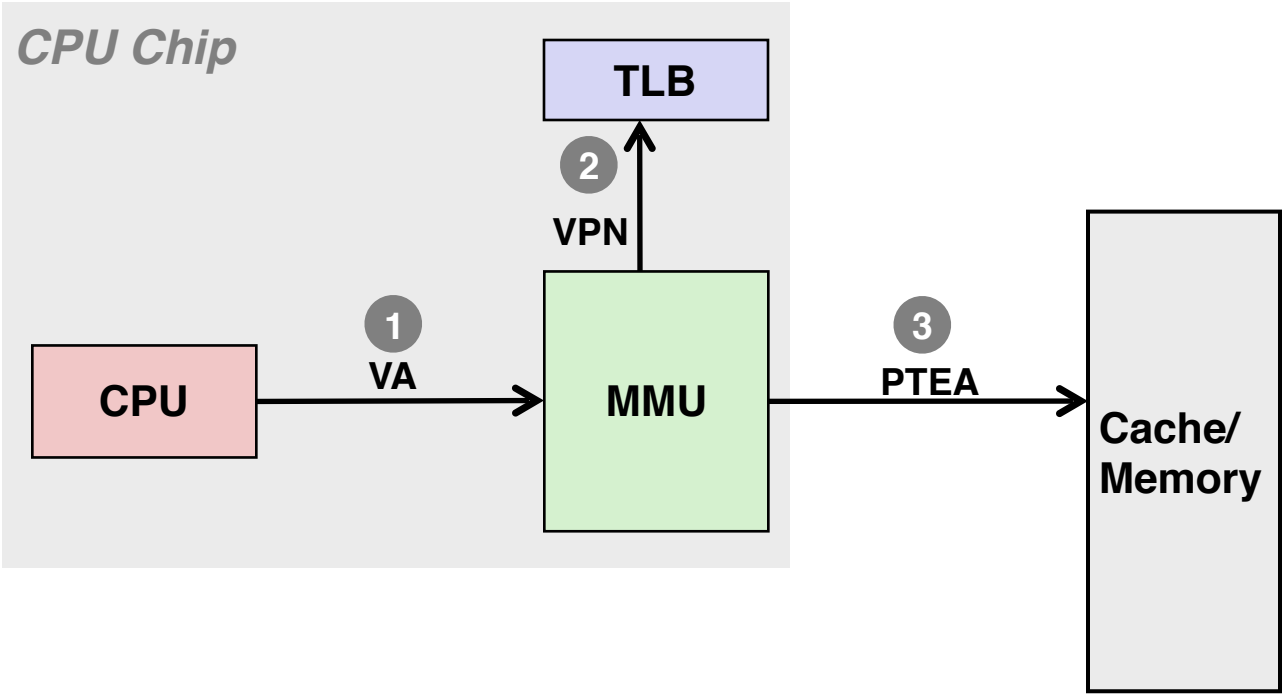


A TLB hit eliminates a memory access

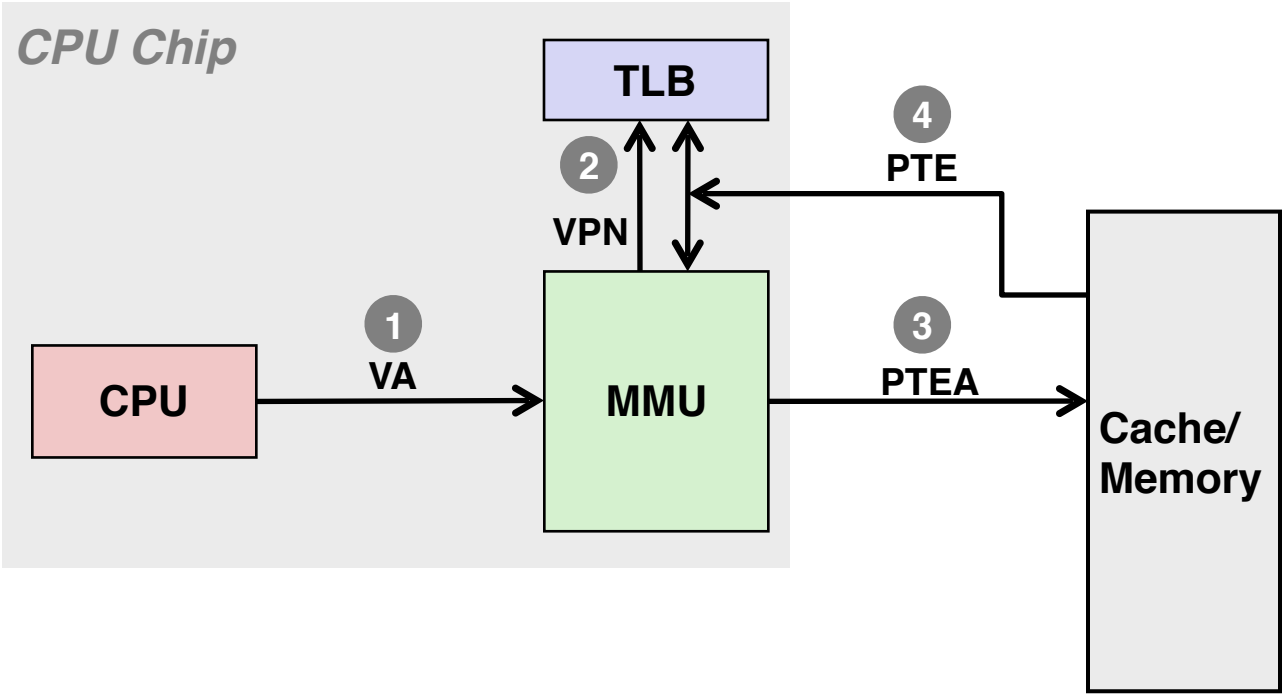
TLB Miss



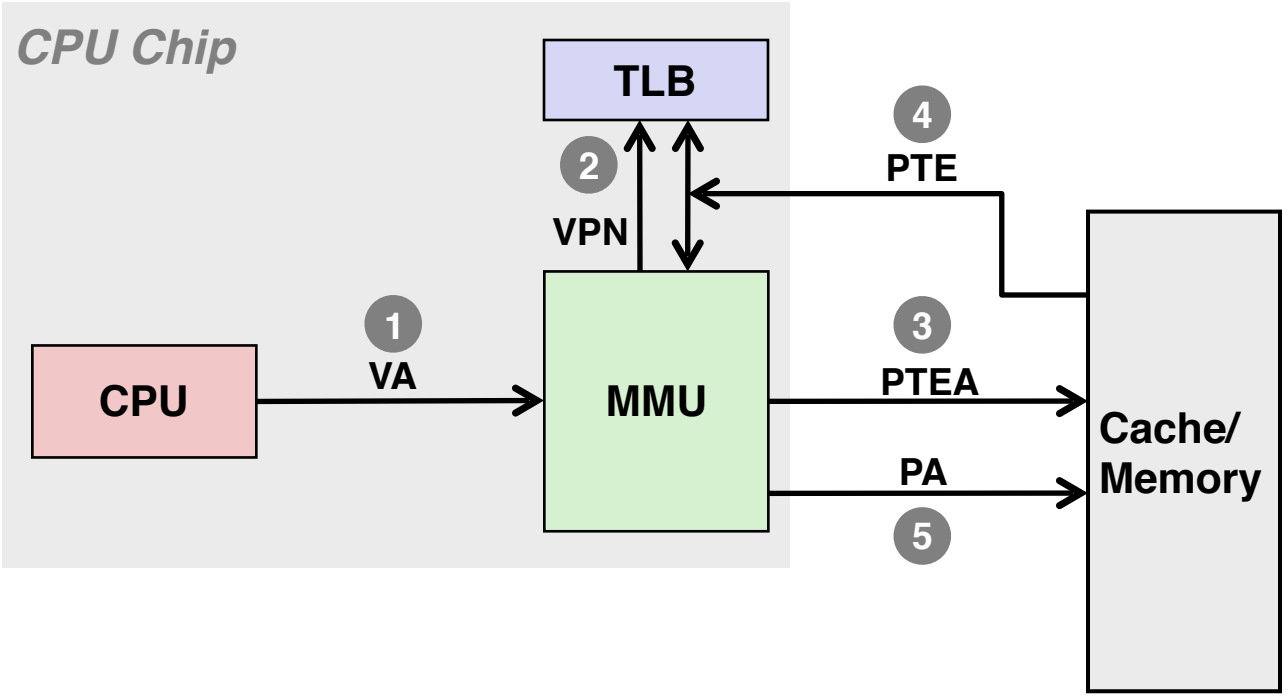
TLB Miss



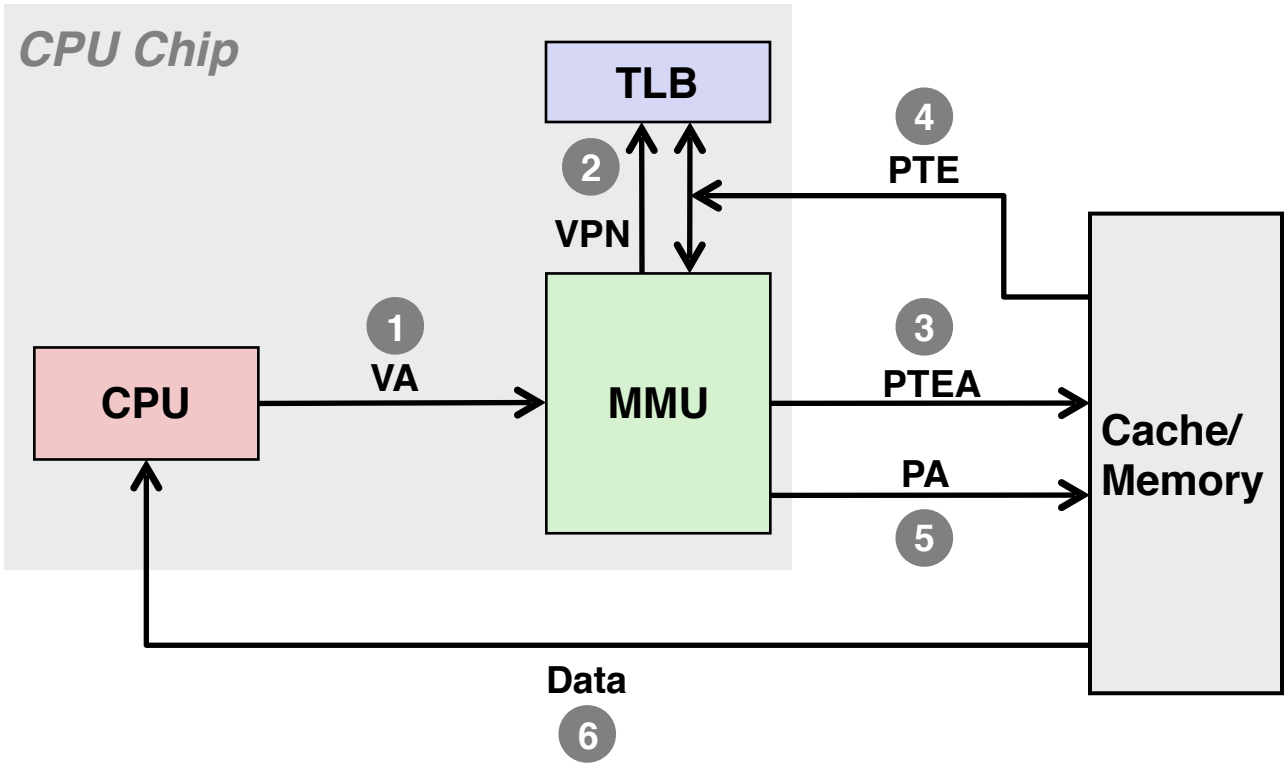
TLB Miss



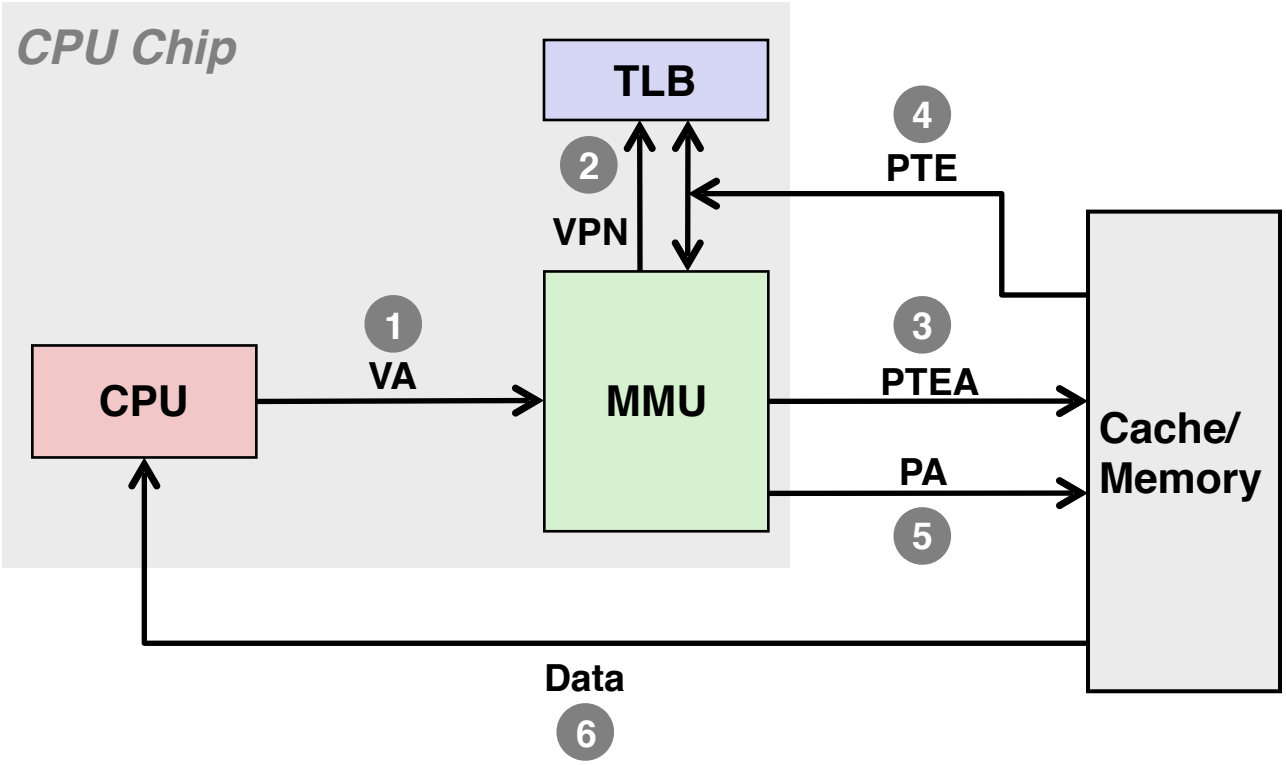
TLB Miss



TLB Miss


CPU Chip
TLB
4
PTE
2
VPN
1
VA
3
PTEA
CPU
MMU
Cache/
Memory
PA
5
Data
6

TLB Miss



A TLB miss incurs an additional memory access (the PTE)
Fortunately, TLB misses are rare.

Today

- Three Virtual Memory Optimizations
  - TLB
  - Page the page table (a.k.a., multi-level page table)
  - Virtually-indexed, physically-tagged cache
- Case-study: Intel Core i7/Linux example

Where Does Page Table Live?

Where Does Page Table Live?

- It needs to be at a specific location where we can find it
  - In main memory, with its start address stored in a special register (PTBR)

Where Does Page Table Live?

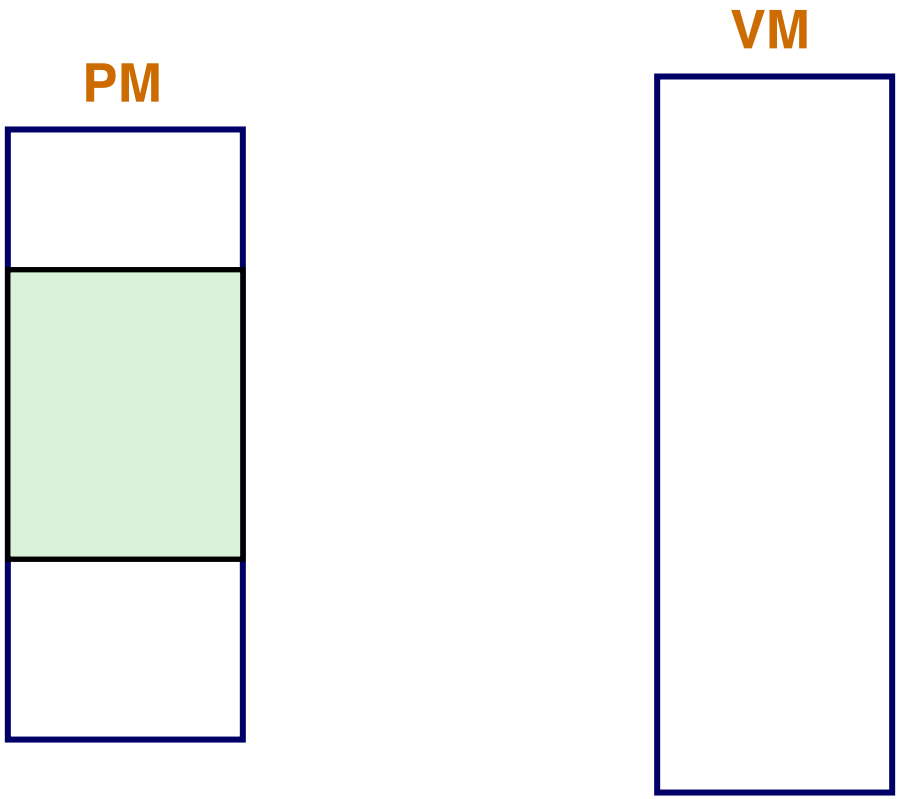
- It needs to be at a specific location where we can find it
  - In main memory, with its start address stored in a special register (PTBR)
- Assume 4KB page, 48-bit virtual memory, each PTE is 8 Bytes
  - 2^{36} PTEs in a page table
  - 512 GB total size per page table??!!

Where Does Page Table Live?

- It needs to be at a specific location where we can find it
  - In main memory, with its start address stored in a special register (PTBR)
- Assume 4KB page, 48-bit virtual memory, each PTE is 8 Bytes
  - 2^{36} PTEs in a page table
  - 512 GB total size per page table??!!
- Problem: Page tables are huge
  - One table per process!
  - Storing them all in main memory wastes space

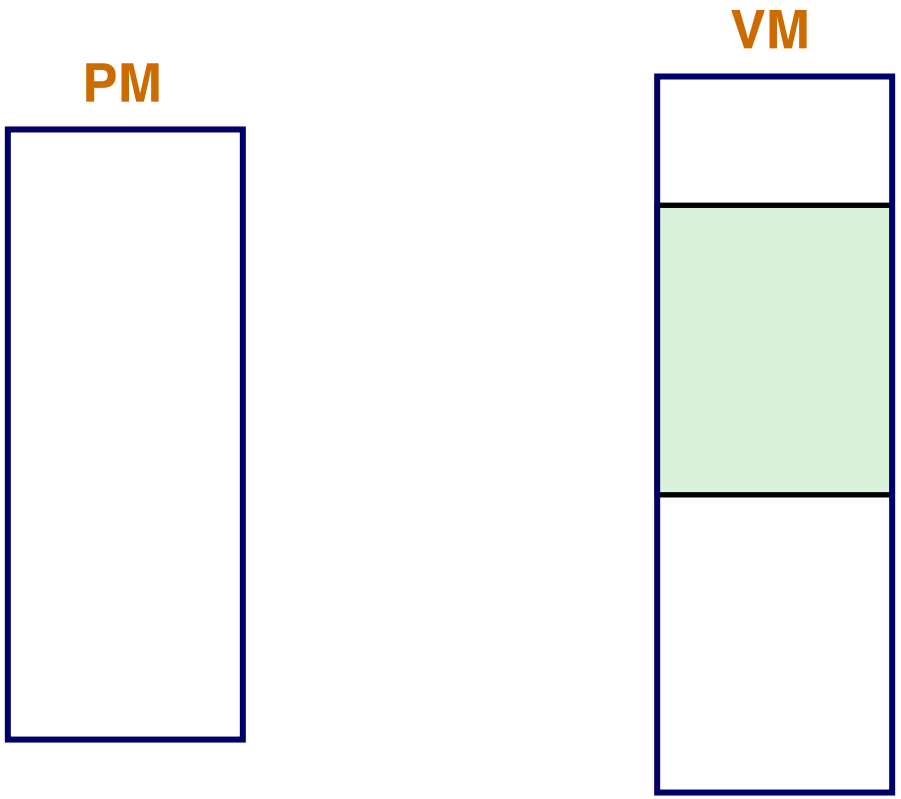
Solution: Page the Page Table

- Observation: Only a small number of pages (working set) are accessed during a certain period of time, due to locality
- Put only the relevant page table entires in main memory
- Idea: Put Page Table in Virtual Memory and swap it just like data



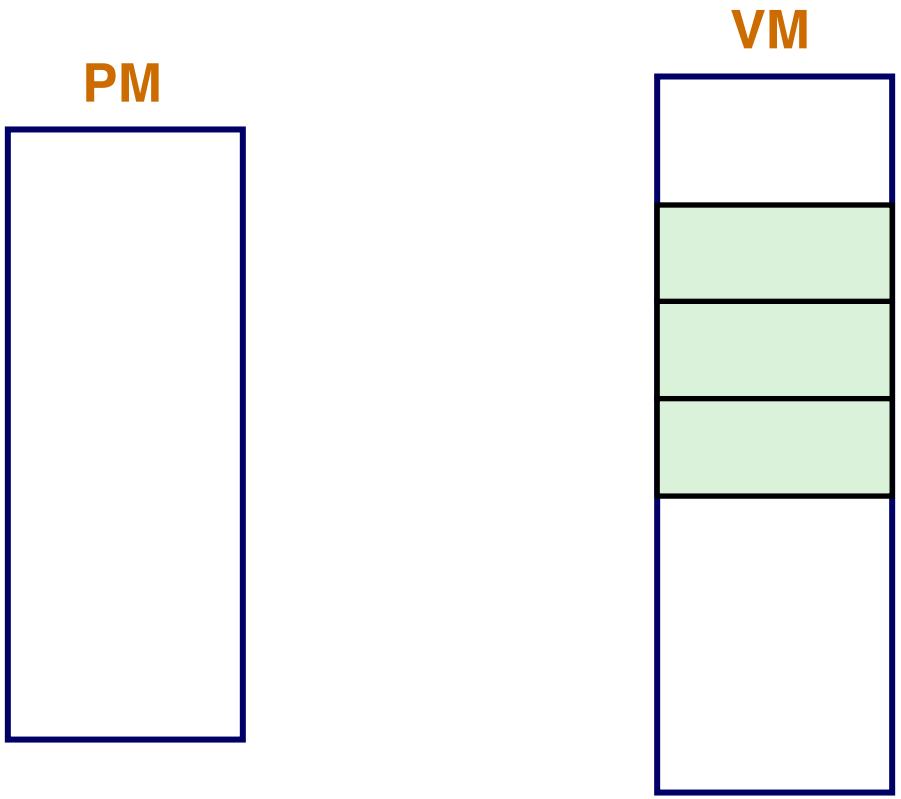
Solution: Page the Page Table

- Observation: Only a small number of pages (working set) are accessed during a certain period of time, due to locality
- Put only the relevant page table entires in main memory
- Idea: Put Page Table in Virtual Memory and swap it just like data



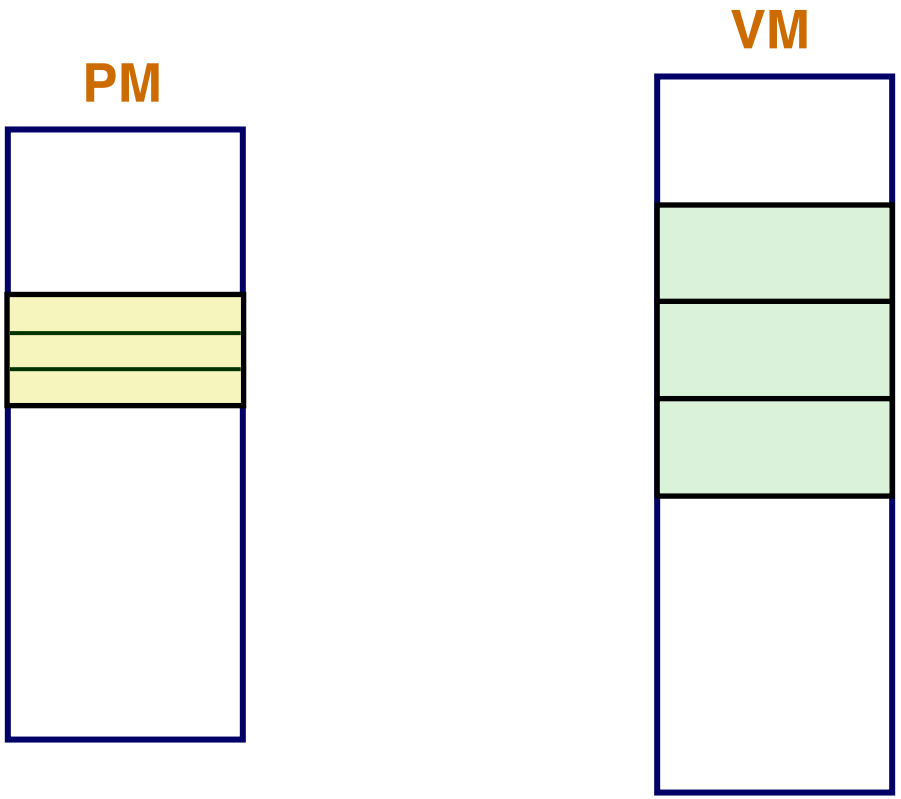
Solution: Page the Page Table

- Observation: Only a small number of pages (working set) are accessed during a certain period of time, due to locality
- Put only the relevant page table entires in main memory
- Idea: Put Page Table in Virtual Memory and swap it just like data



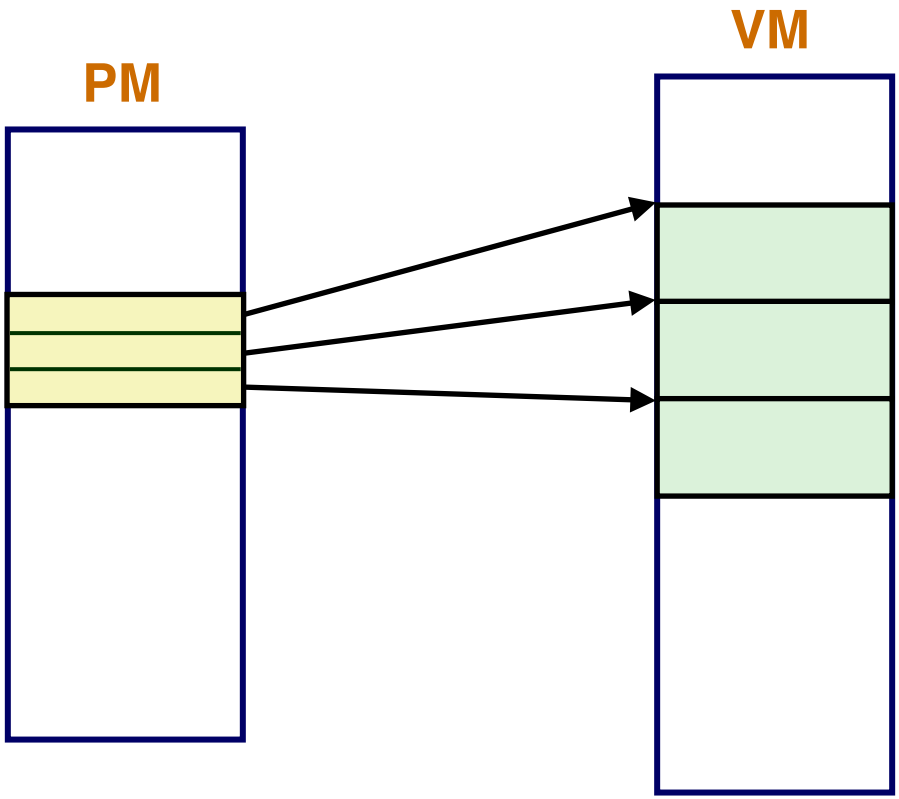
Solution: Page the Page Table

- Observation: Only a small number of pages (working set) are accessed during a certain period of time, due to locality
- Put only the relevant page table entires in main memory
- Idea: Put Page Table in Virtual Memory and swap it just like data



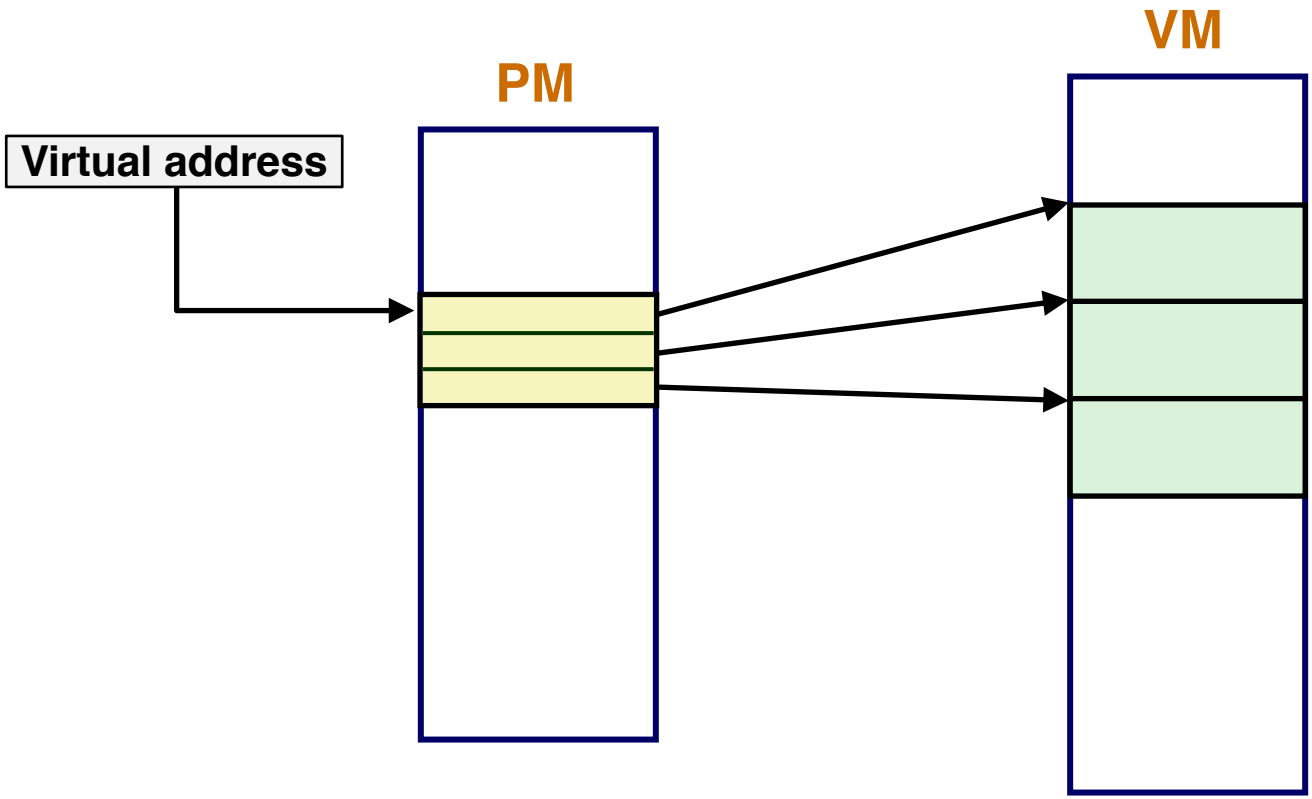
Solution: Page the Page Table

- Observation: Only a small number of pages (working set) are accessed during a certain period of time, due to locality
- Put only the relevant page table entires in main memory
- Idea: Put Page Table in Virtual Memory and swap it just like data



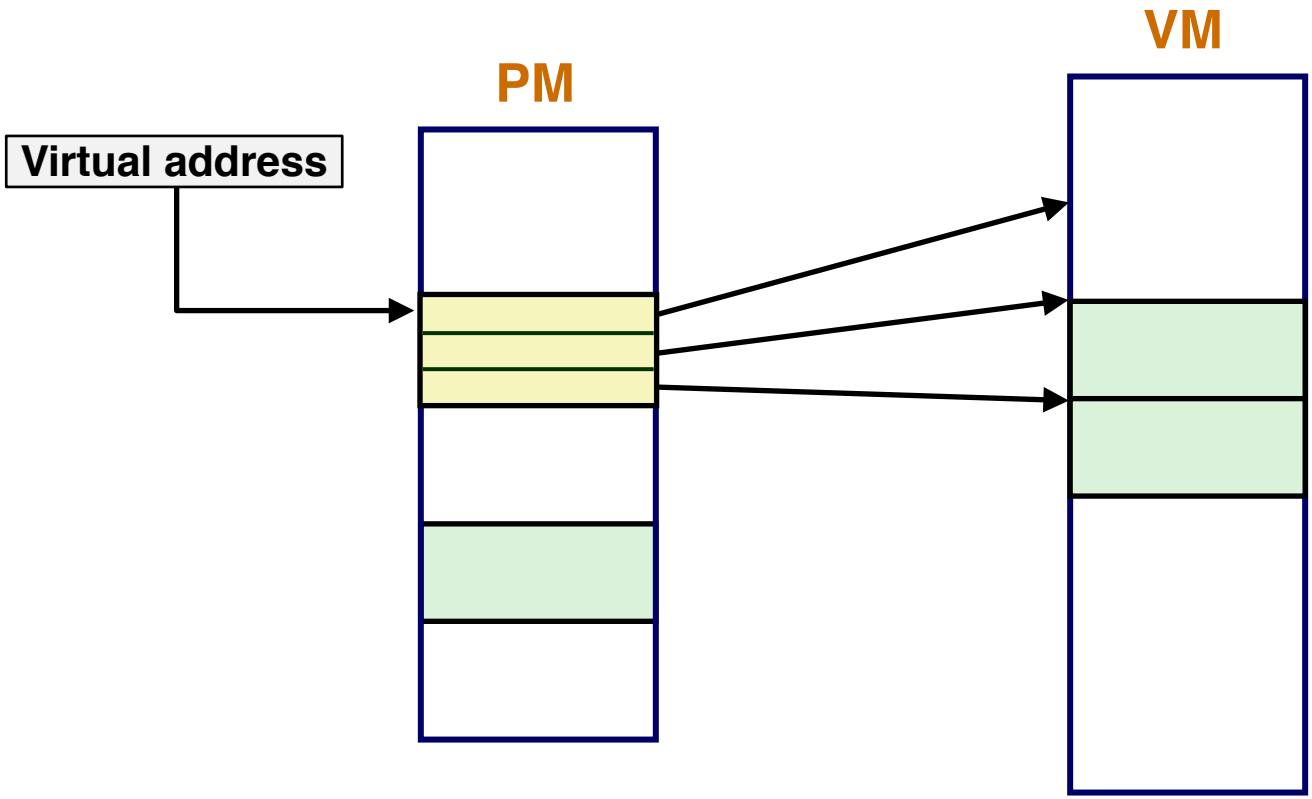
Solution: Page the Page Table

- Observation: Only a small number of pages (working set) are accessed during a certain period of time, due to locality
- Put only the relevant page table entires in main memory
- Idea: Put Page Table in Virtual Memory and swap it just like data



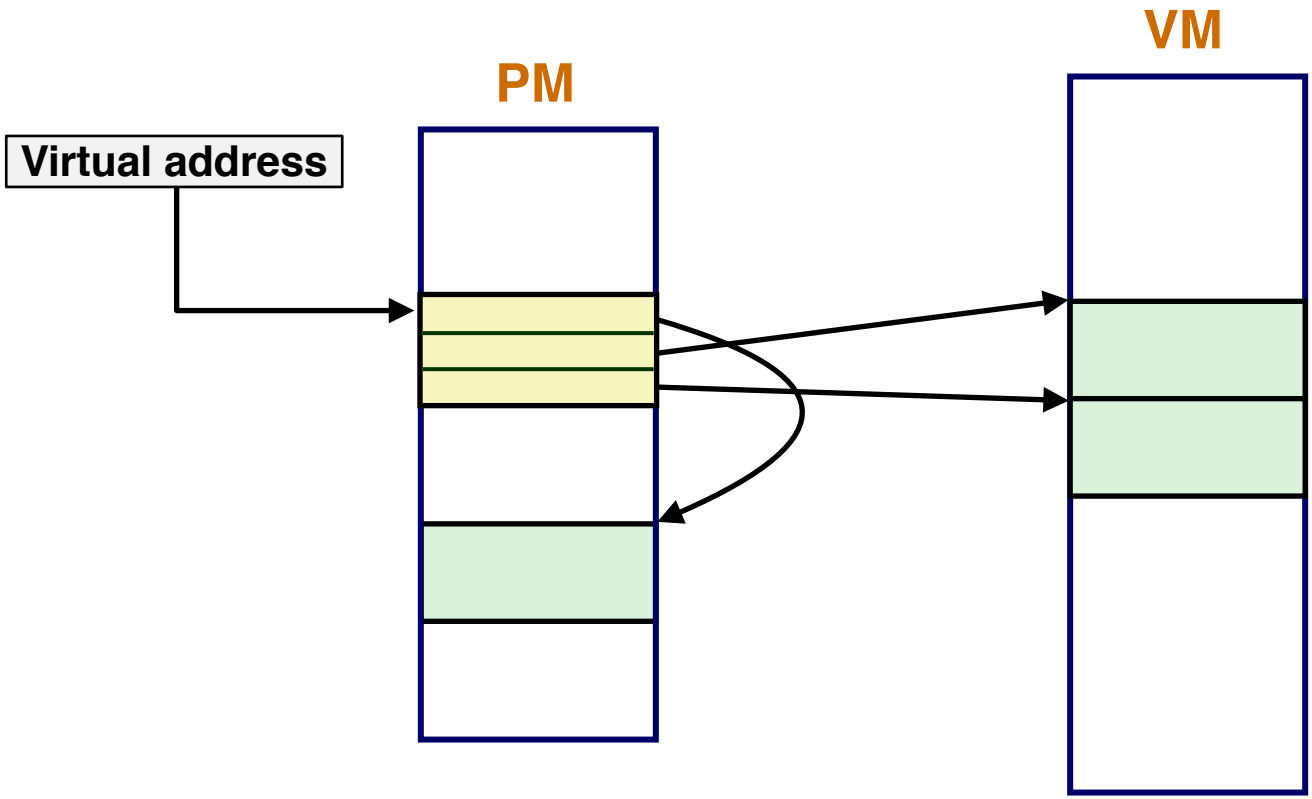
Solution: Page the Page Table

- Observation: Only a small number of pages (working set) are accessed during a certain period of time, due to locality
- Put only the relevant page table entires in main memory
- Idea: Put Page Table in Virtual Memory and swap it just like data



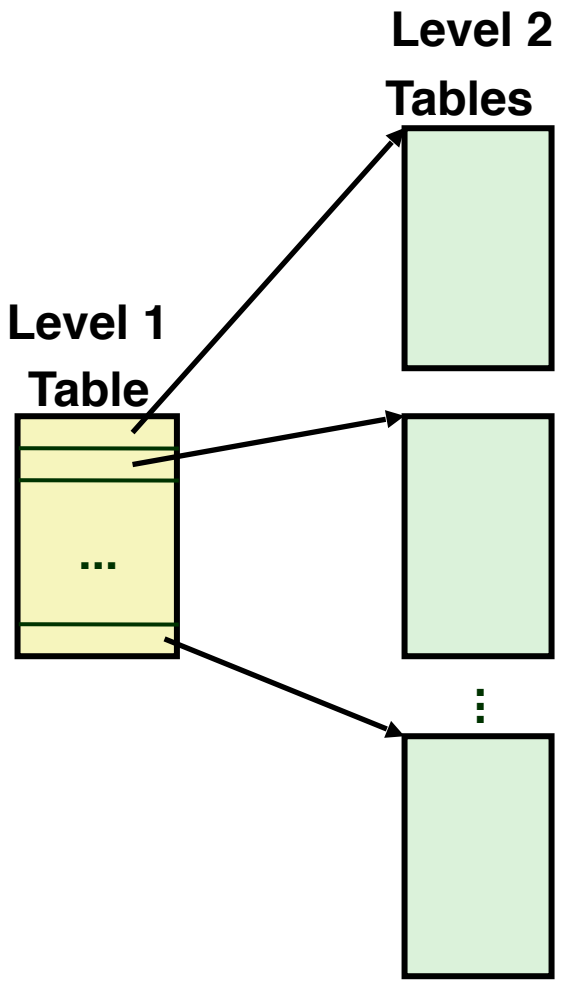
Solution: Page the Page Table

- Observation: Only a small number of pages (working set) are accessed during a certain period of time, due to locality
- Put only the relevant page table entires in main memory
- Idea: Put Page Table in Virtual Memory and swap it just like data

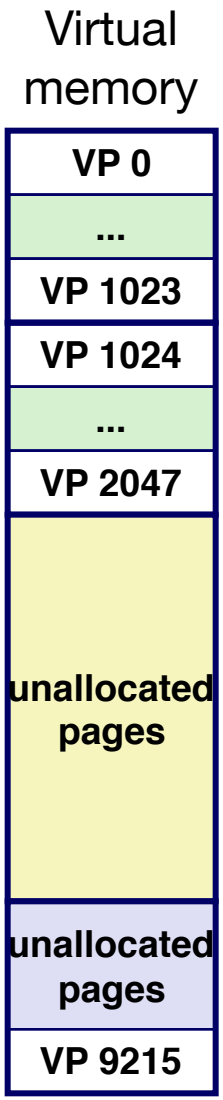


Effectively: A 2-Level Page Table

- Level 1 table:
  - Always in memory at a known location.
  - Each L1 PTE points to the start address of a L2 page table.
  - Bring that table to memory on-demand.
- Level 2 table:
  - Each PTE points to an actual data page



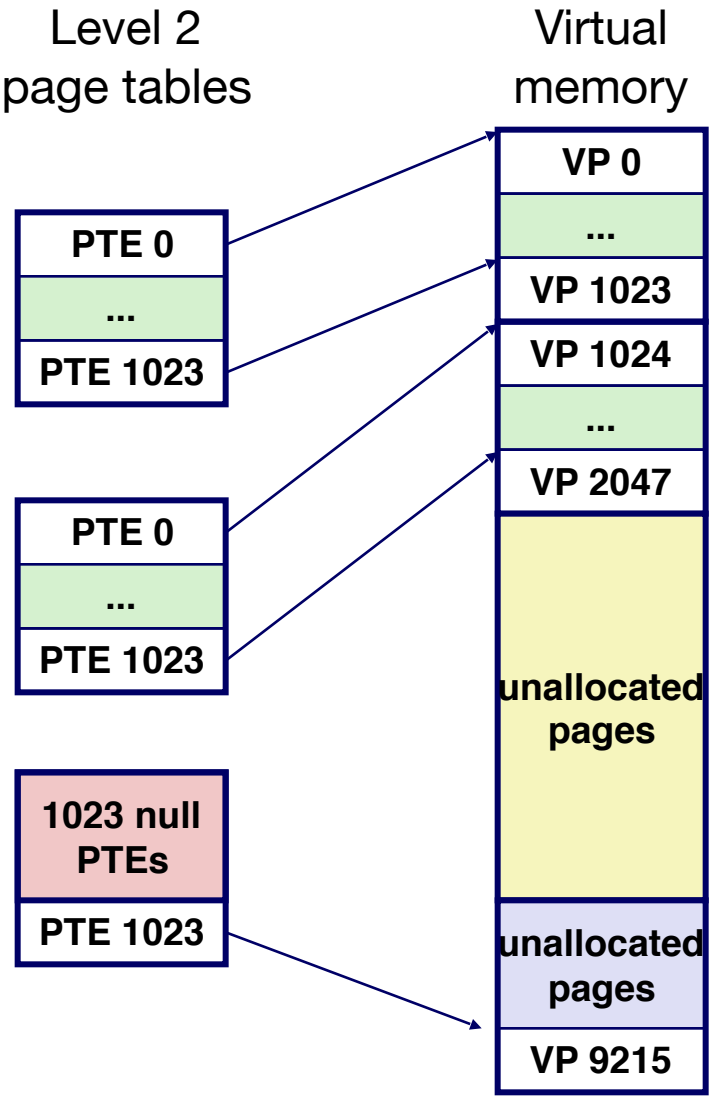
A Two-Level Page Table Hierarchy



32 bit addresses, 4KB pages, 4-byte PTEs

...

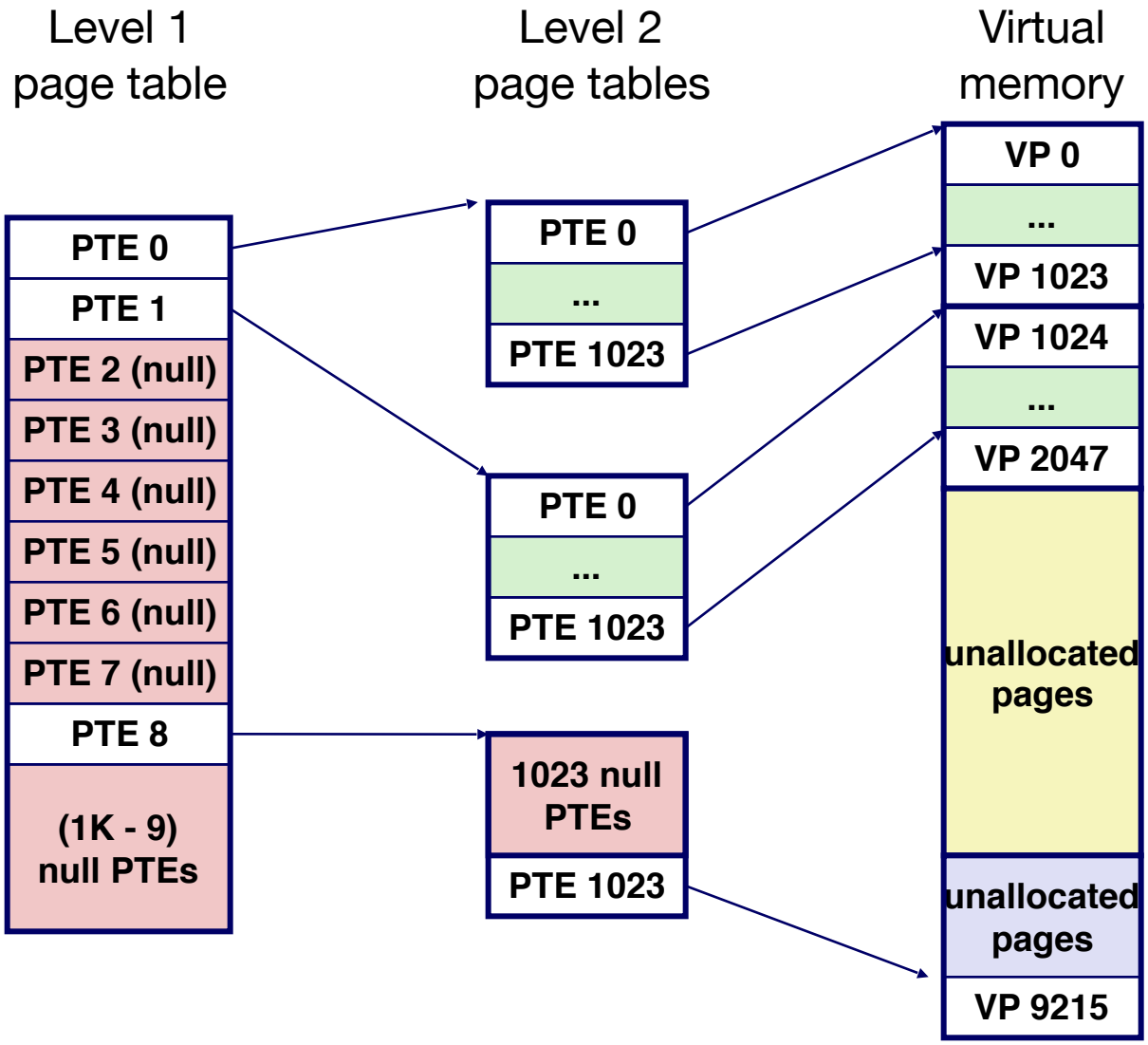
A Two-Level Page Table Hierarchy



32 bit addresses, 4KB pages, 4-byte PTEs

...

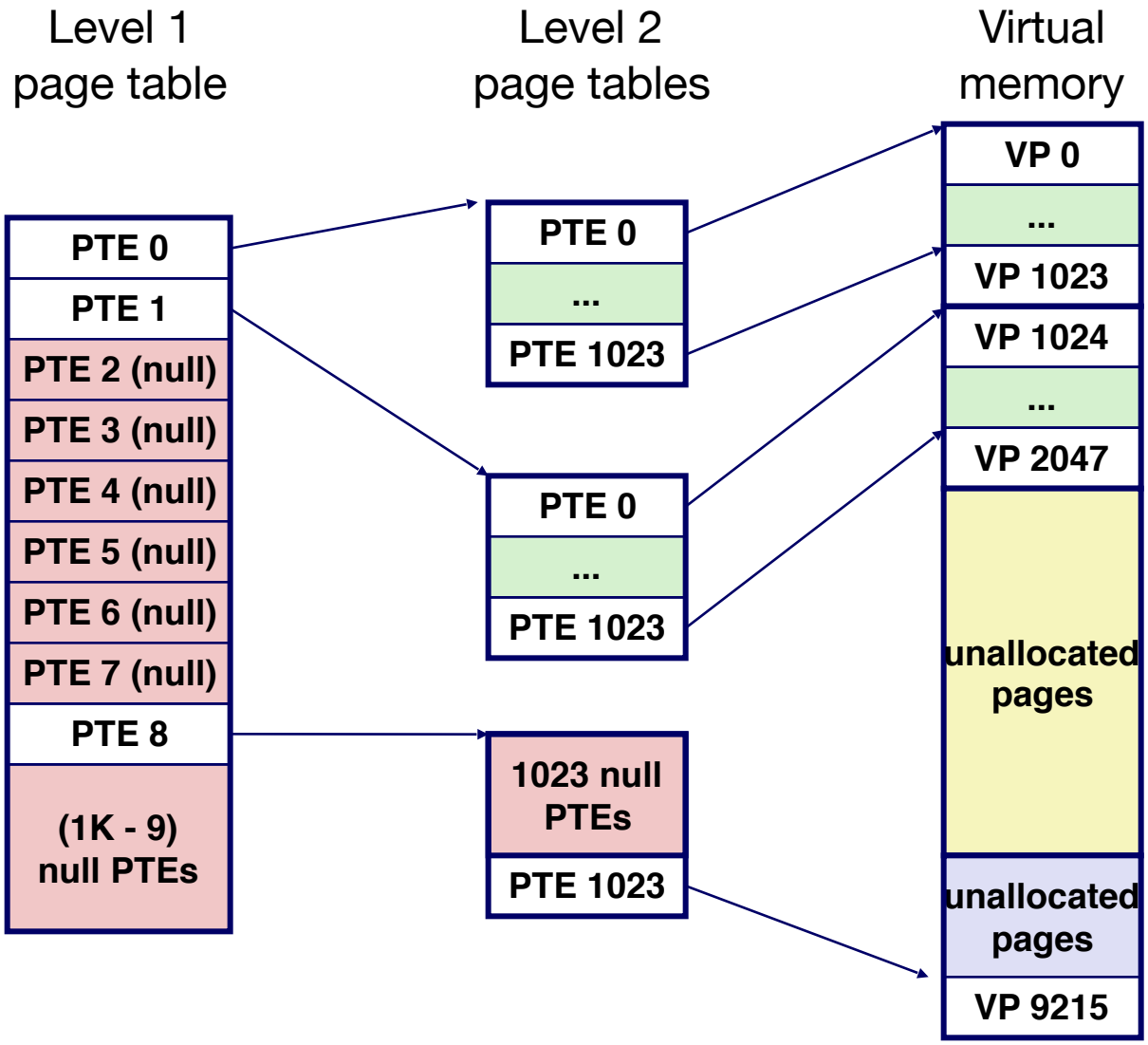
A Two-Level Page Table Hierarchy



32 bit addresses, 4KB pages, 4-byte PTEs

...

A Two-Level Page Table Hierarchy

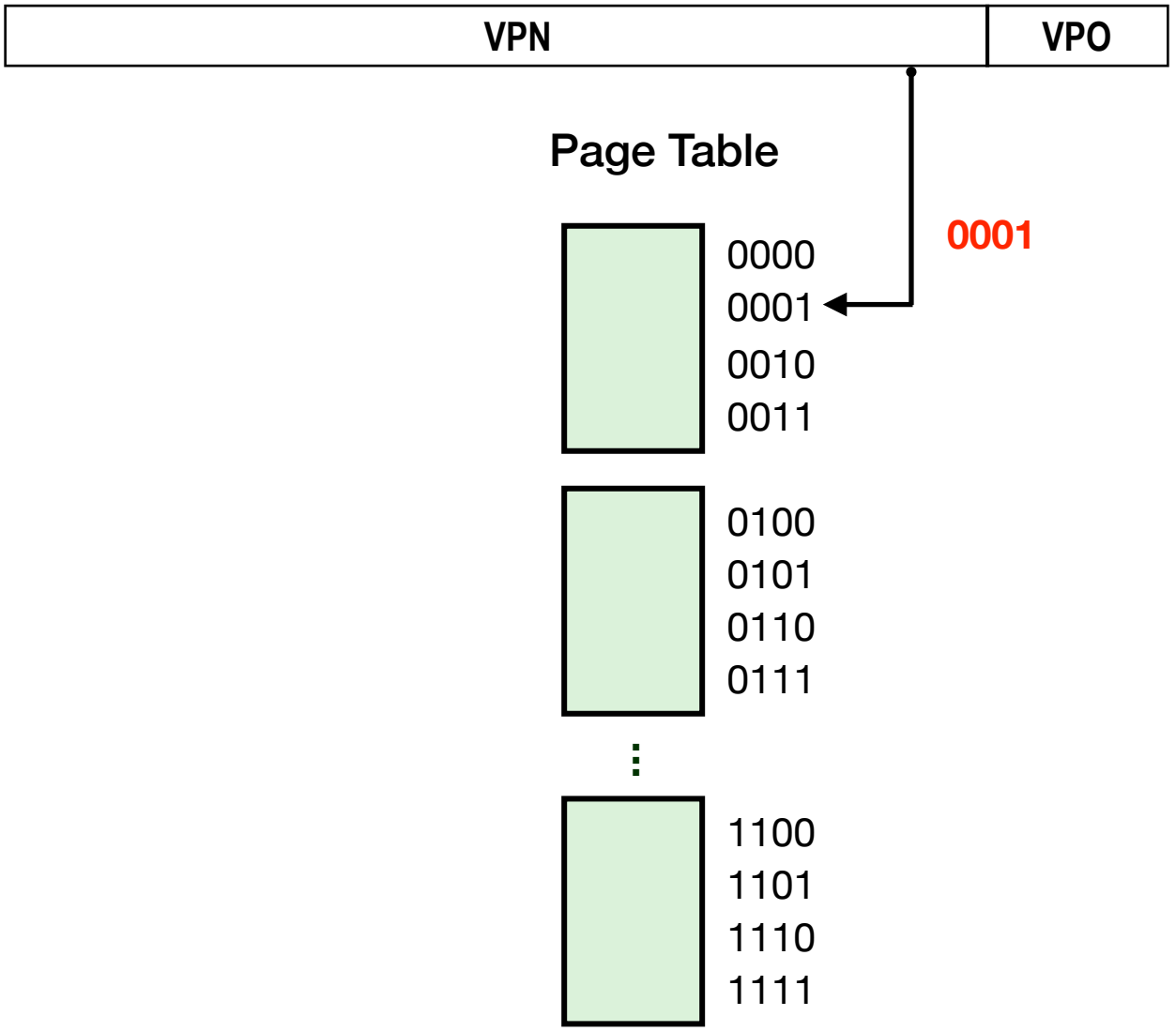


- Level 2 page table size:
  - $2^{32} / 2^{12} * 4 = 4$ MB
- Level 1 page table size:
  - $(2^{32} / 2^{12} * 4) / 2^{12} * 4 = 4$ KB

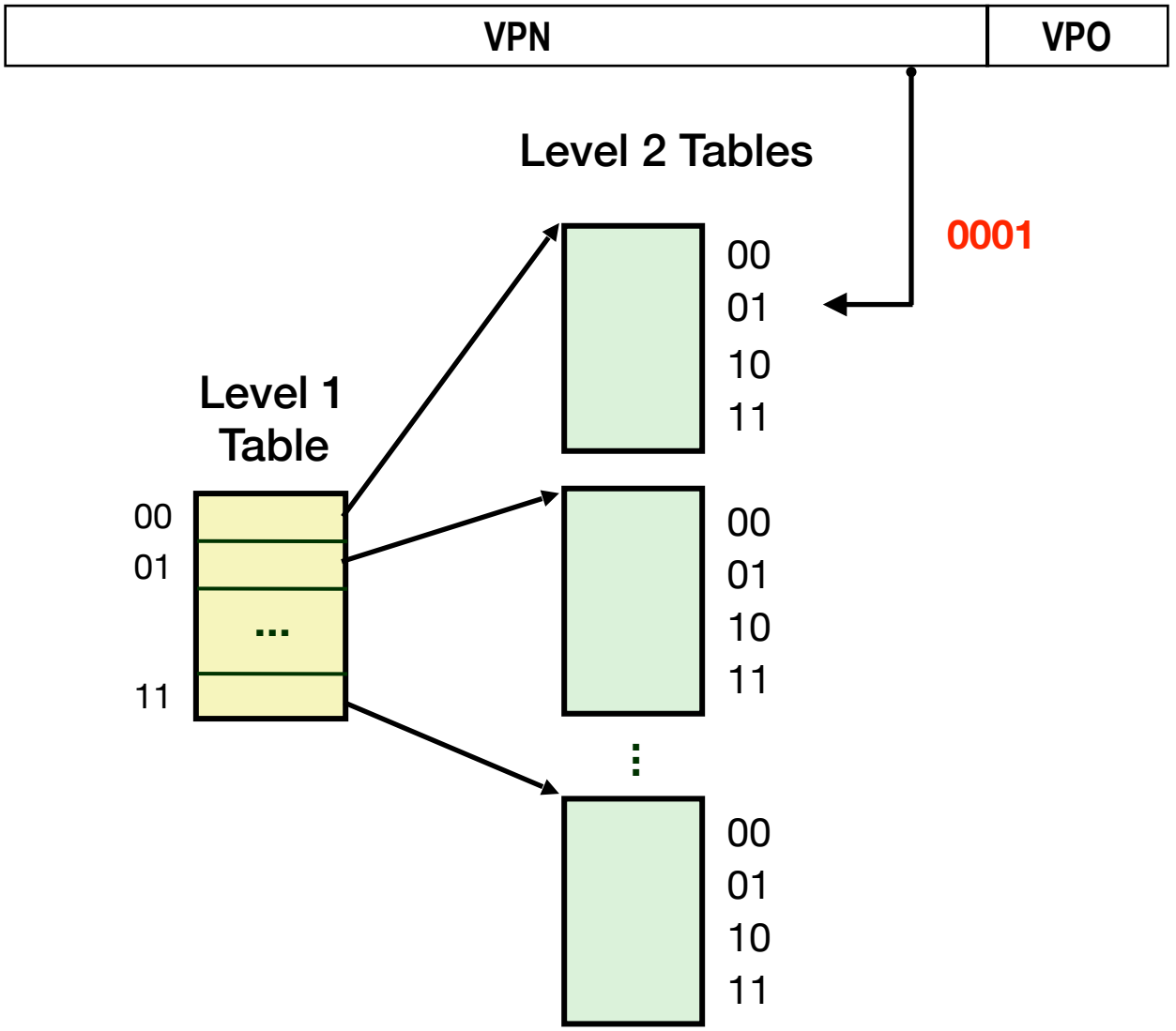
32 bit addresses, 4KB pages, 4-byte PTEs

...

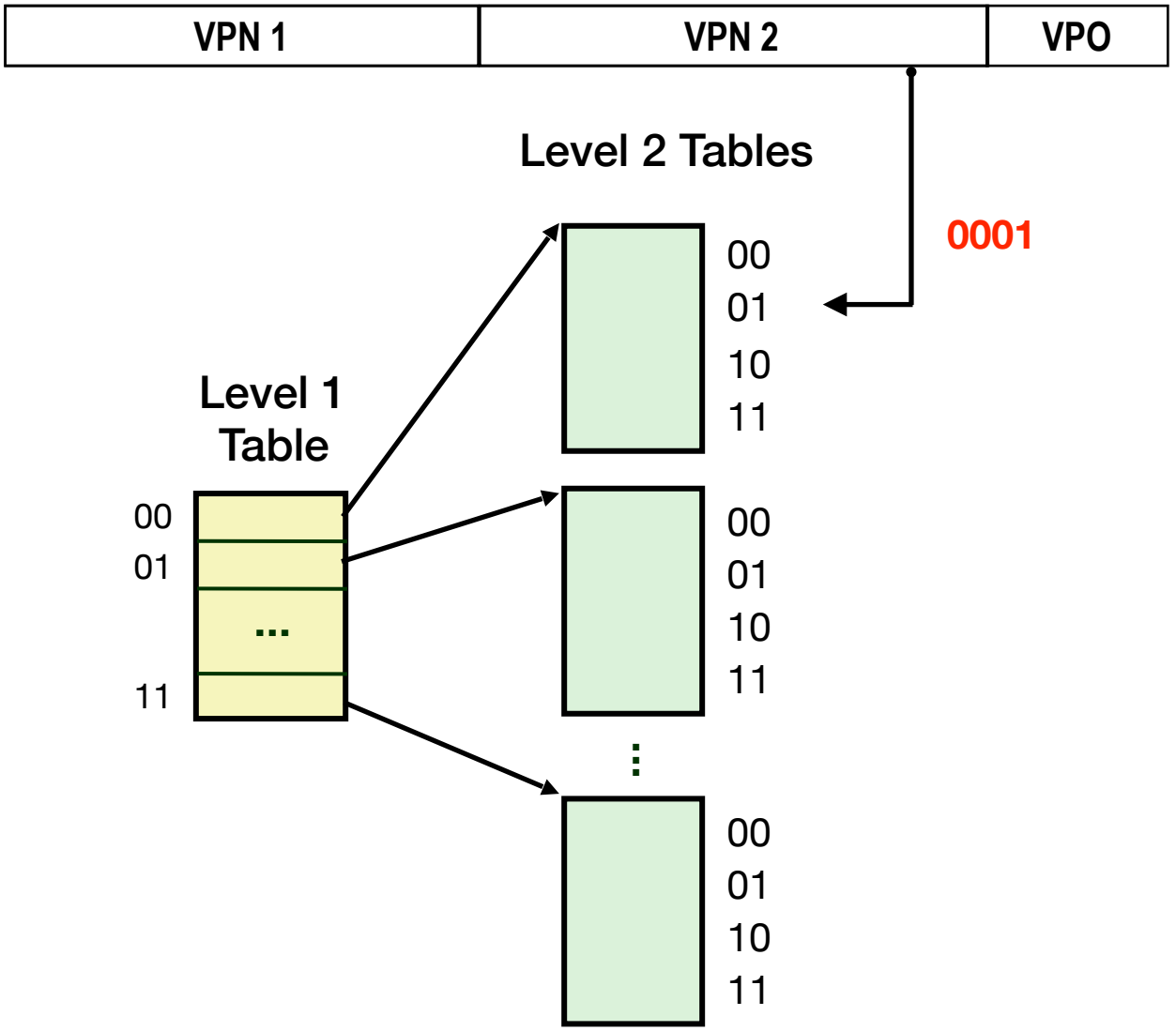
How to Access a 2-Level Page Table?



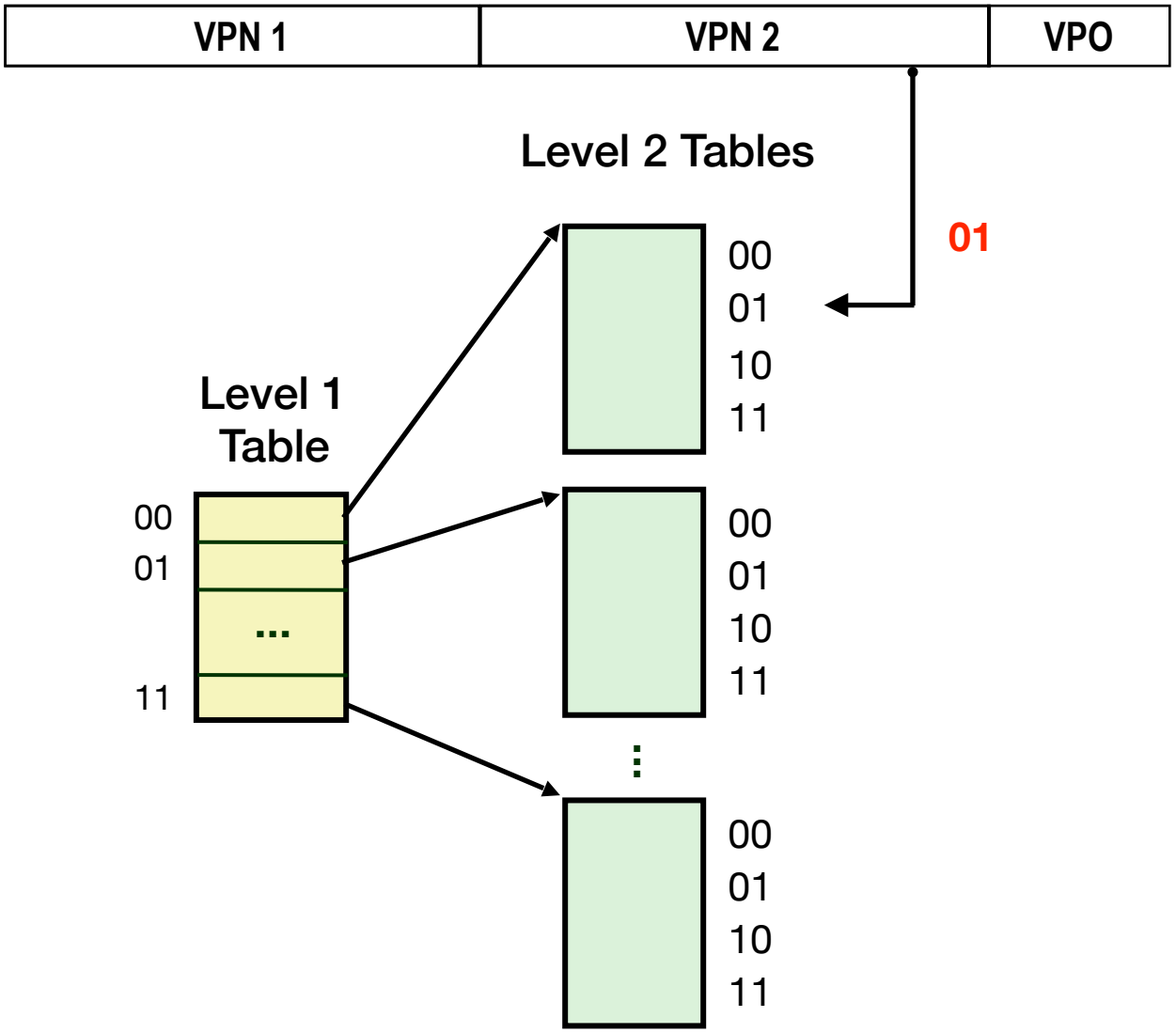
How to Access a 2-Level Page Table?



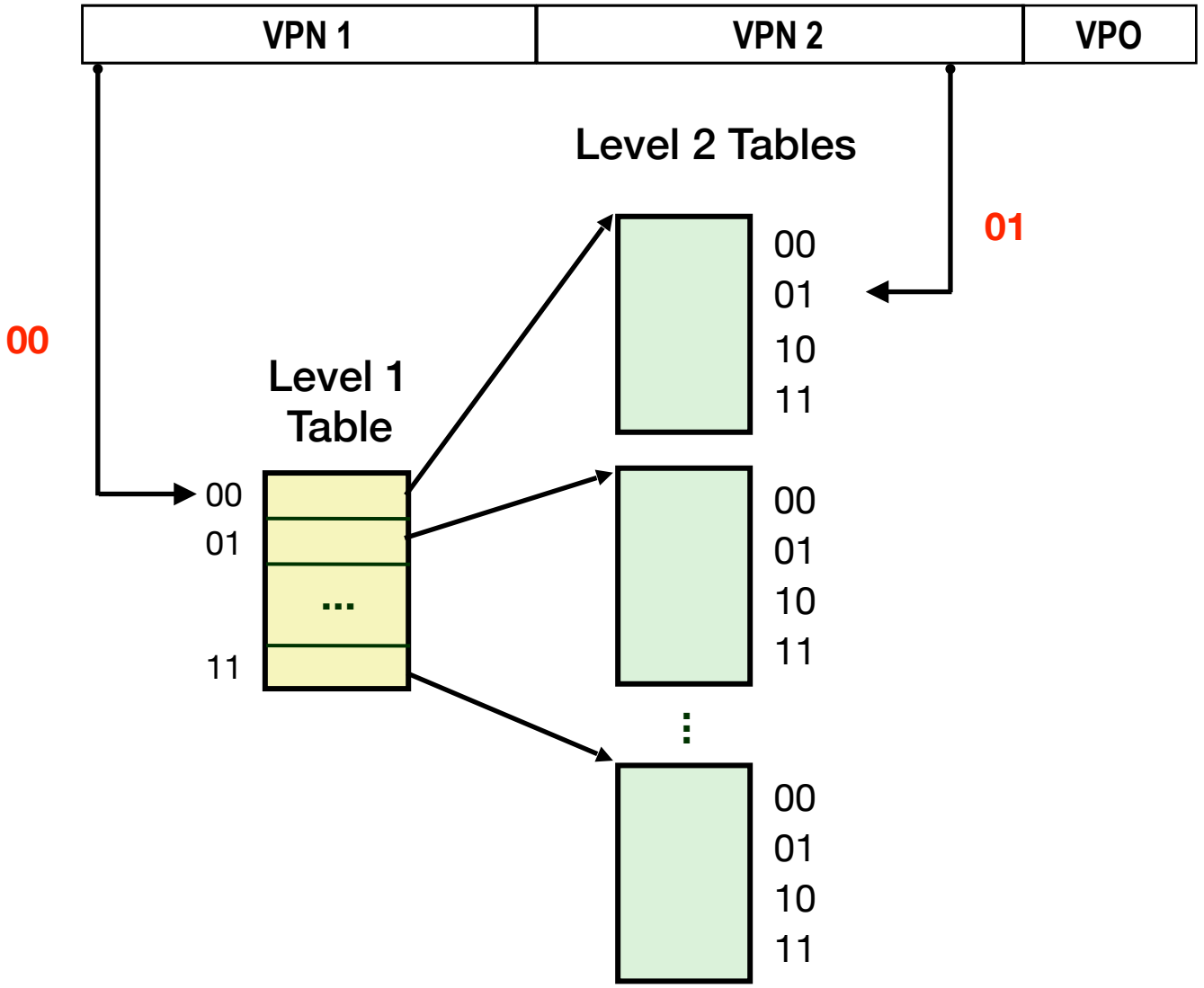
How to Access a 2-Level Page Table?



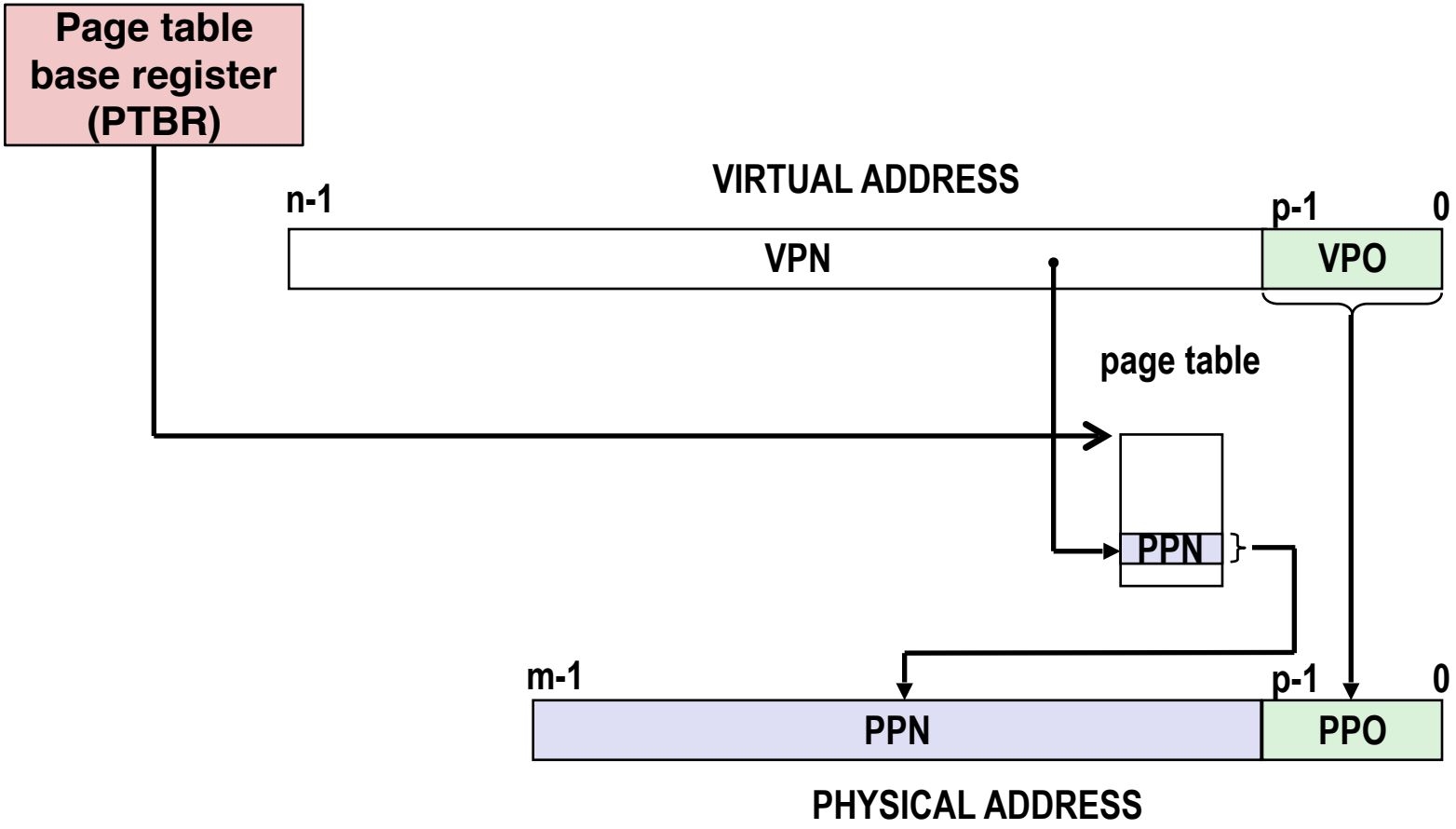
How to Access a 2-Level Page Table?



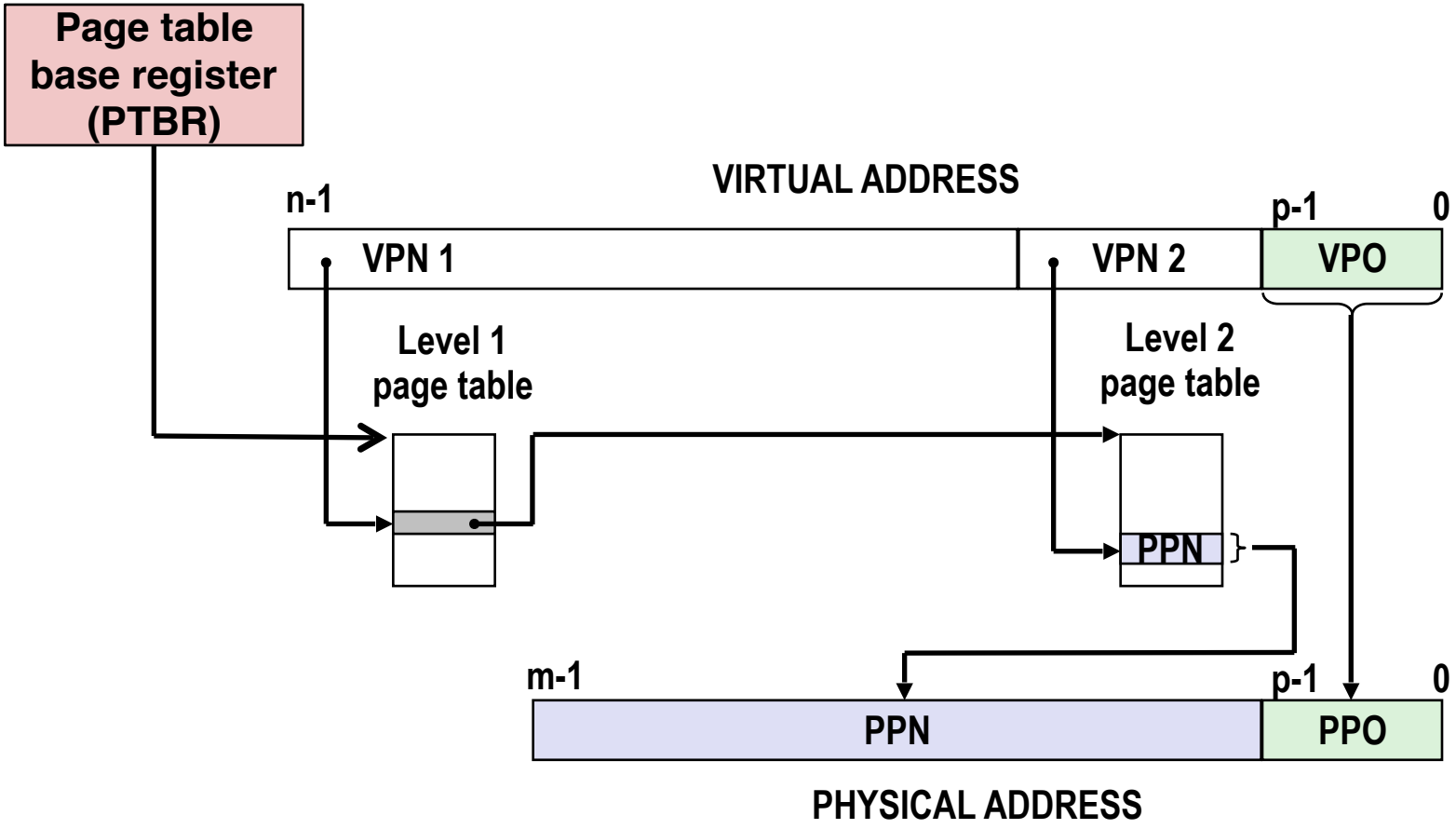
How to Access a 2-Level Page Table?



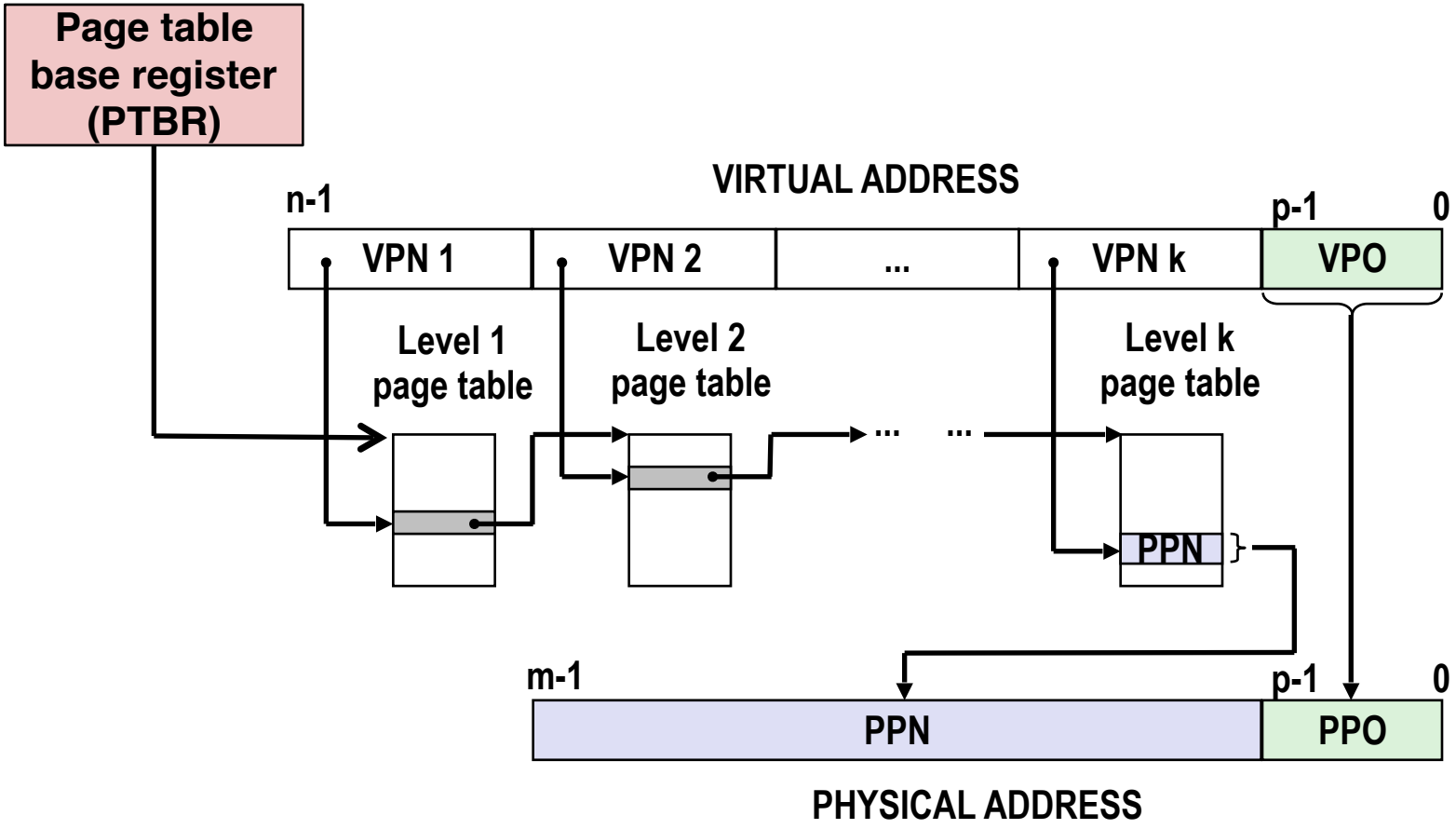
How to Access a 2-Level Page Table?



How to Access a 2-Level Page Table?



Translating with a k-level Page Table

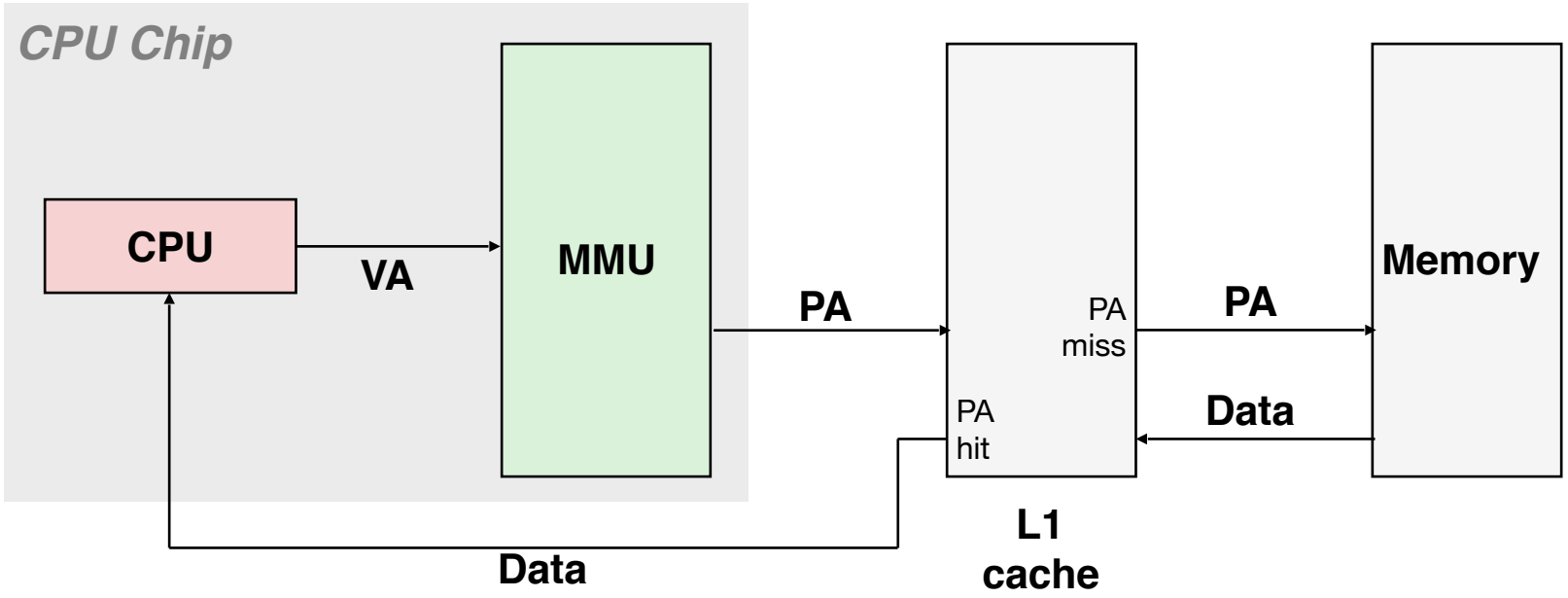


Today

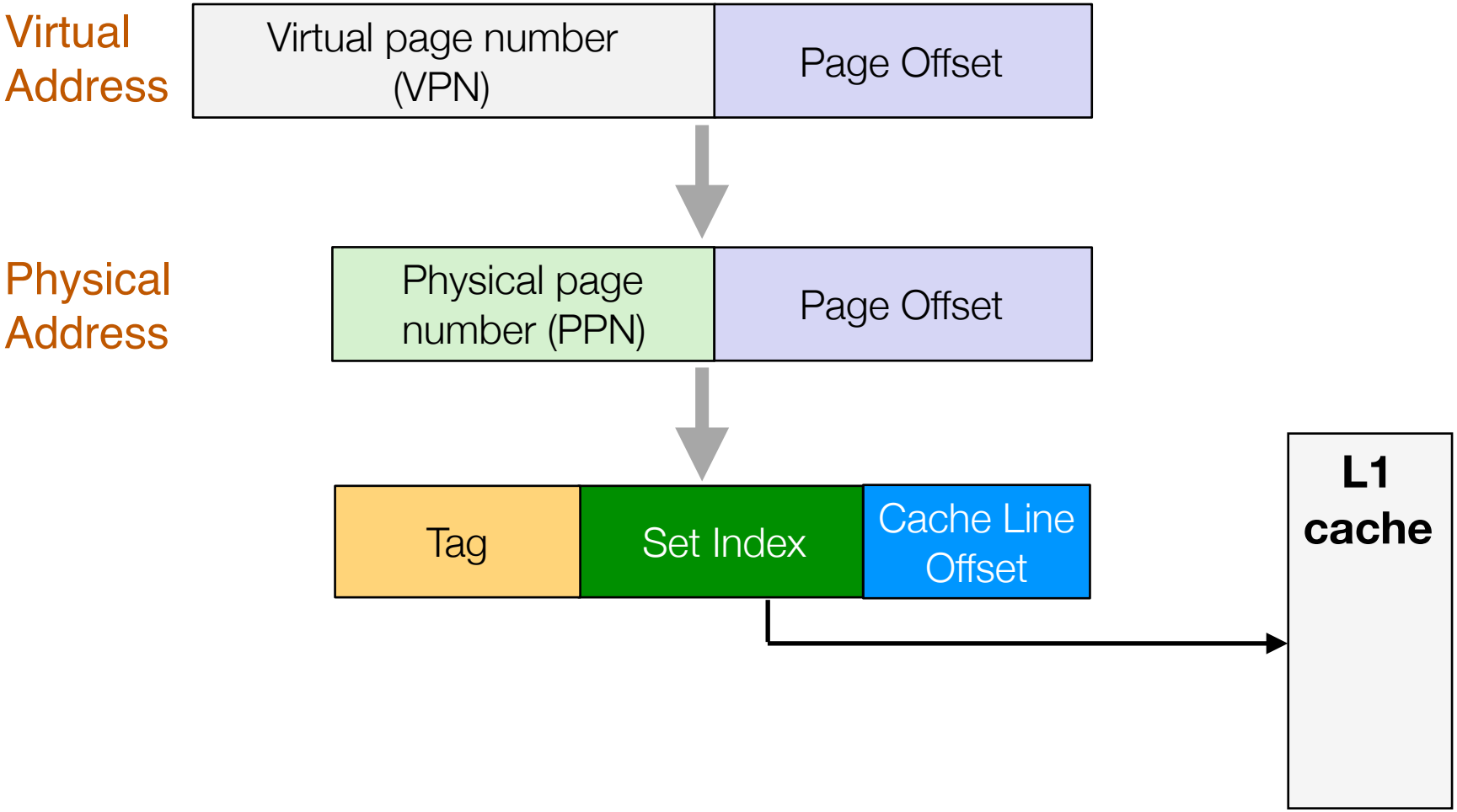
- Three Virtual Memory Optimizations
  - TLB
  - Page the page table (a.k.a., multi-level page table)
  - Virtually-indexed, physically-tagged cache
- Case-study: Intel Core i7/Linux example

Performance Issue in VM

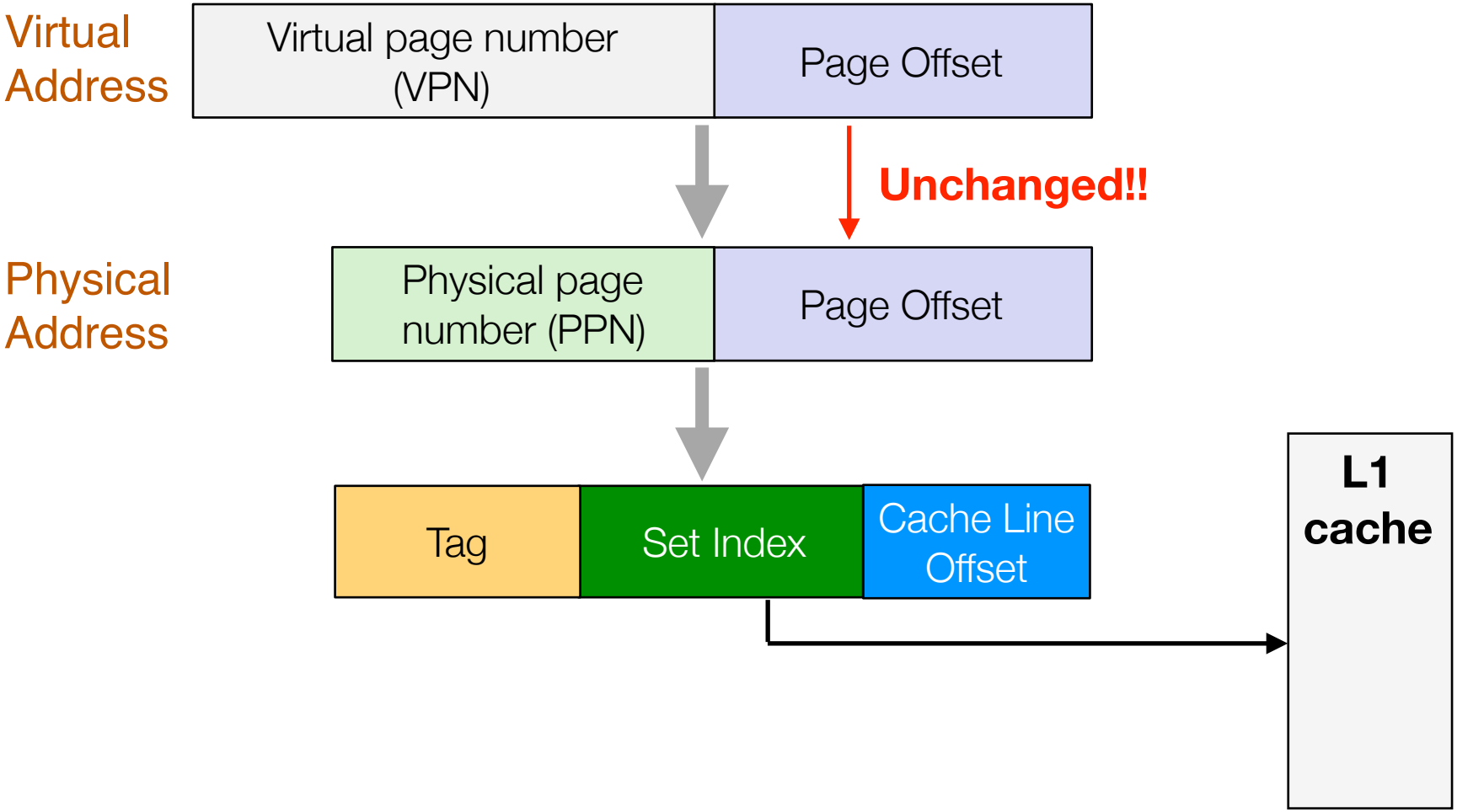
- Address translation and cache accesses are serialized
  - First translate from VA to PA
  - Then use PA to access cache
  - Slow! Can we speed it up?



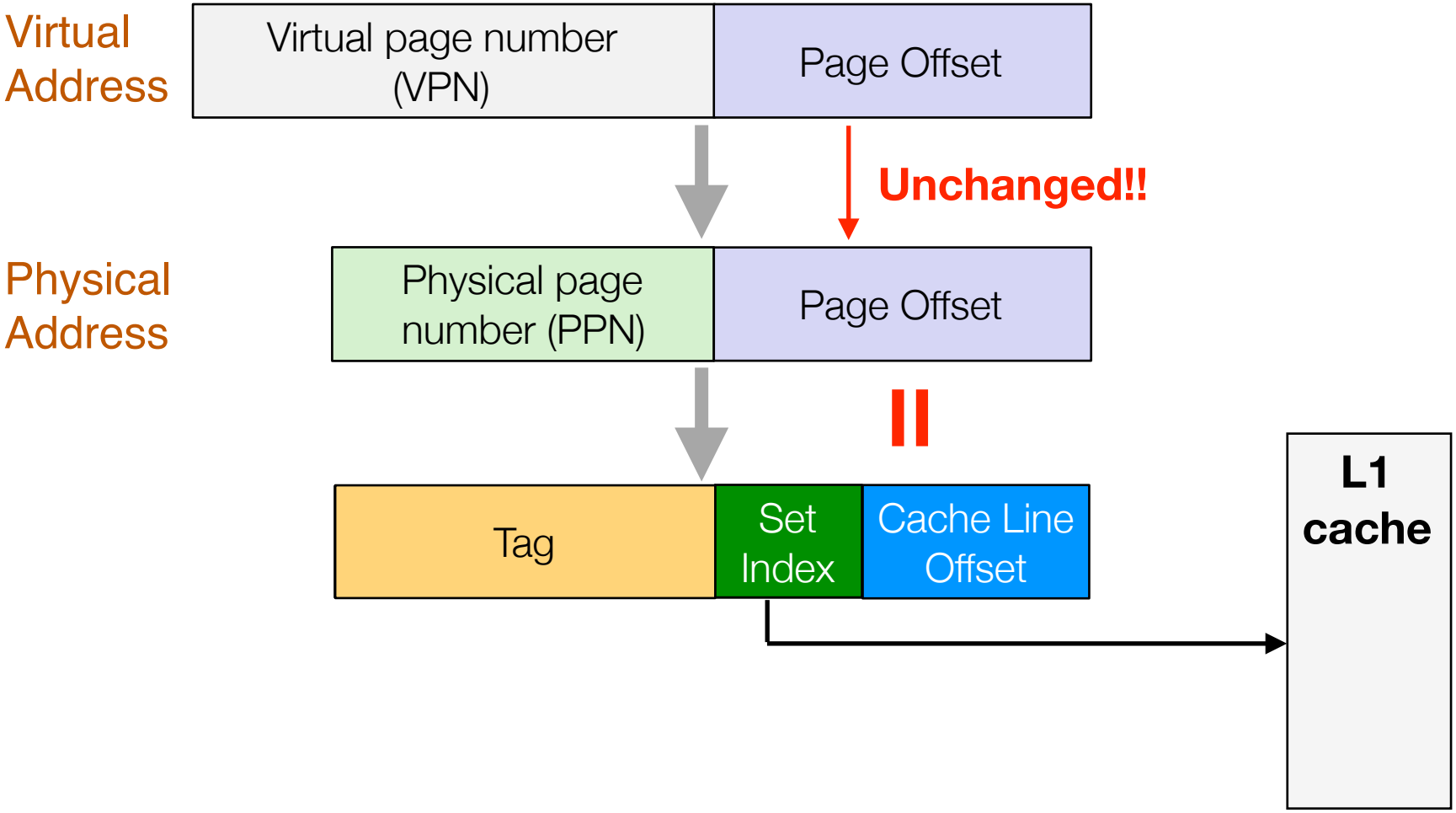
Performance Issue in VM



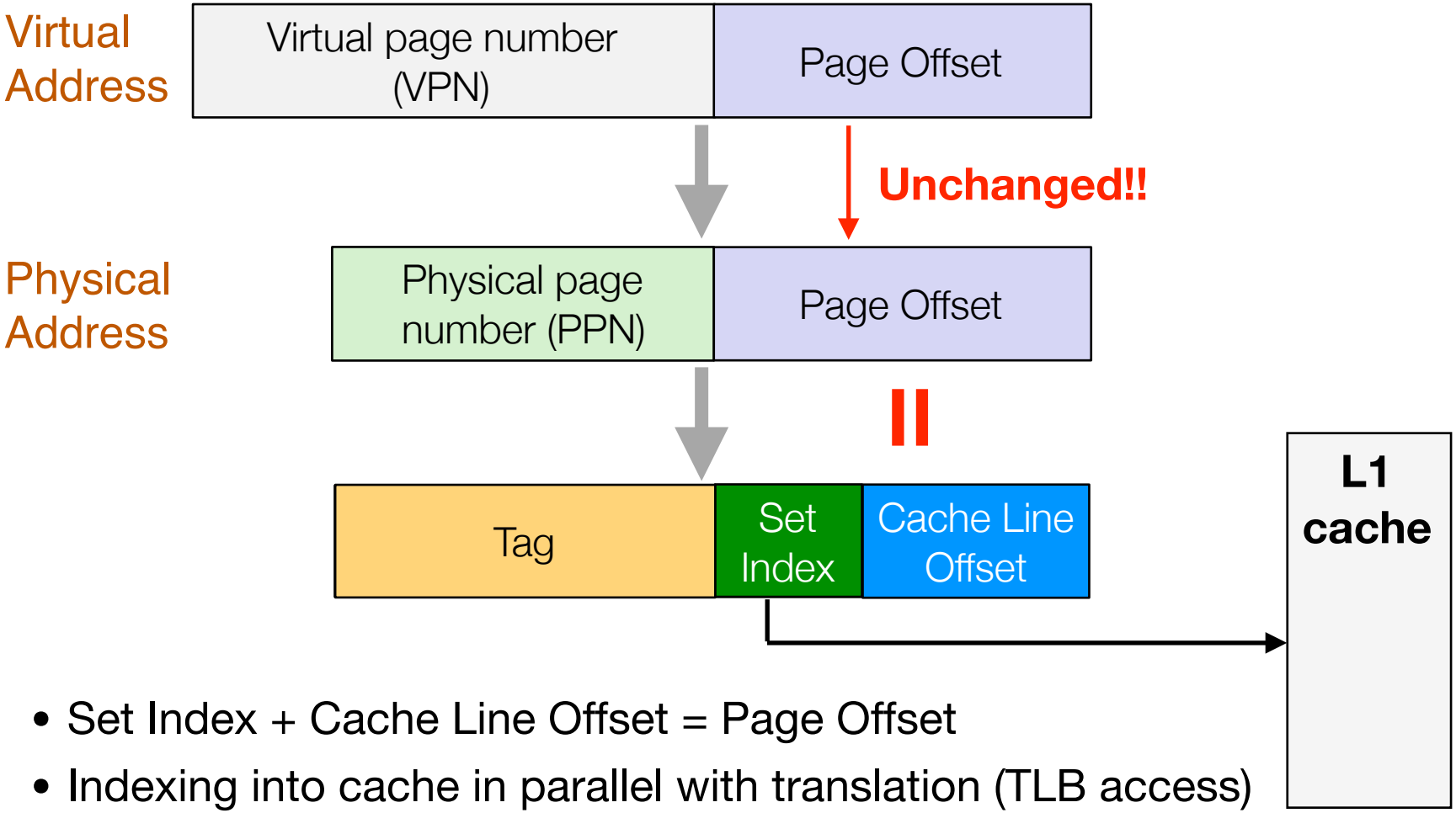
Performance Issue in VM



Performance Issue in VM

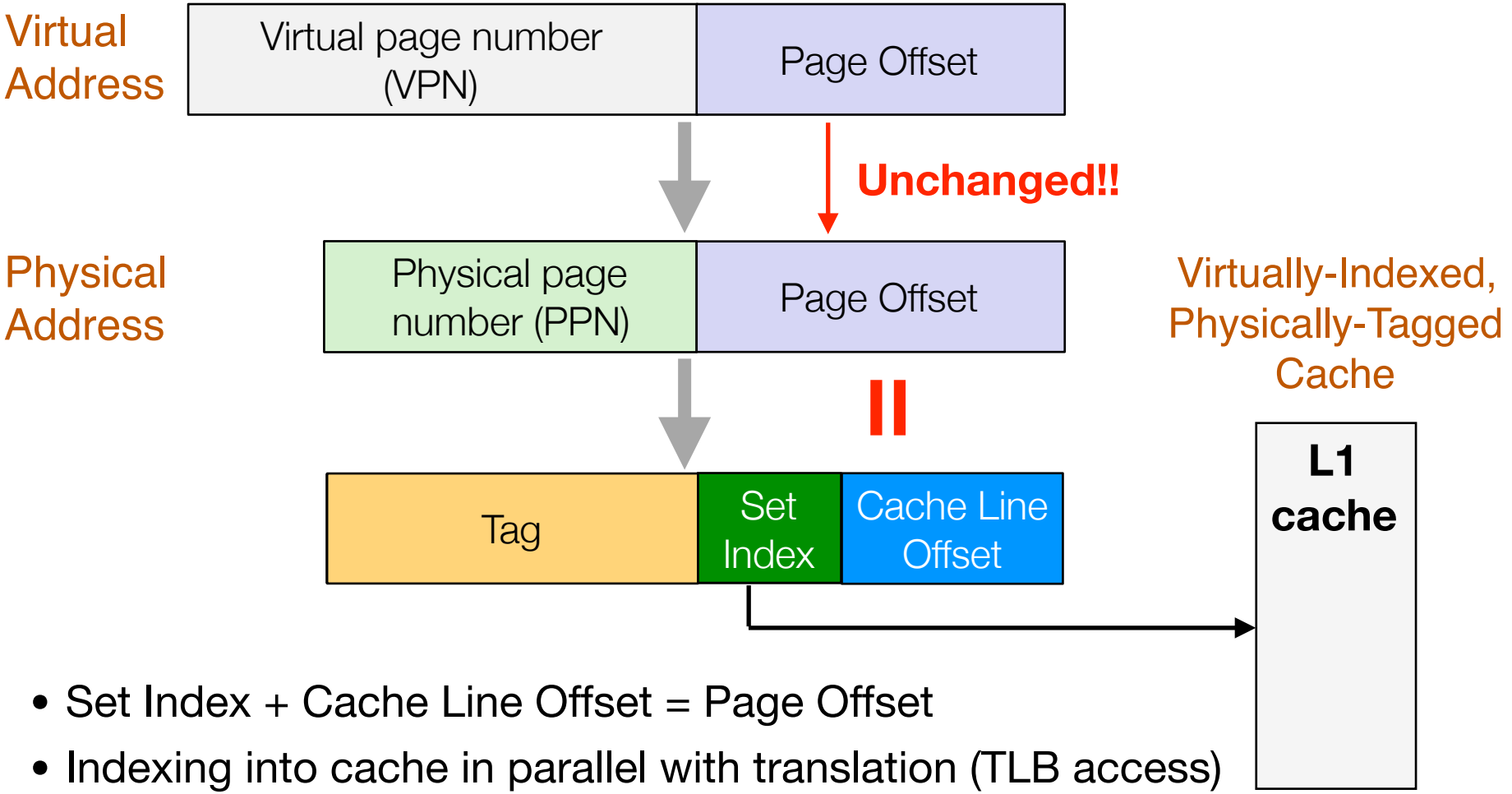


Performance Issue in VM



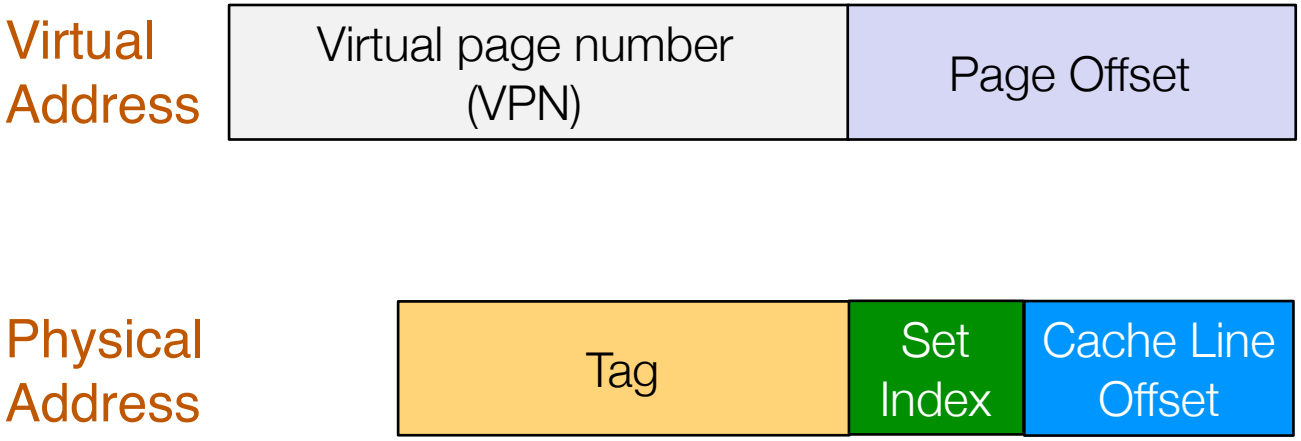
- Set Index + Cache Line Offset = Page Offset
- Indexing into cache in parallel with translation (TLB access)
- If TLB hits, can get the data back in one cycle

Performance Issue in VM

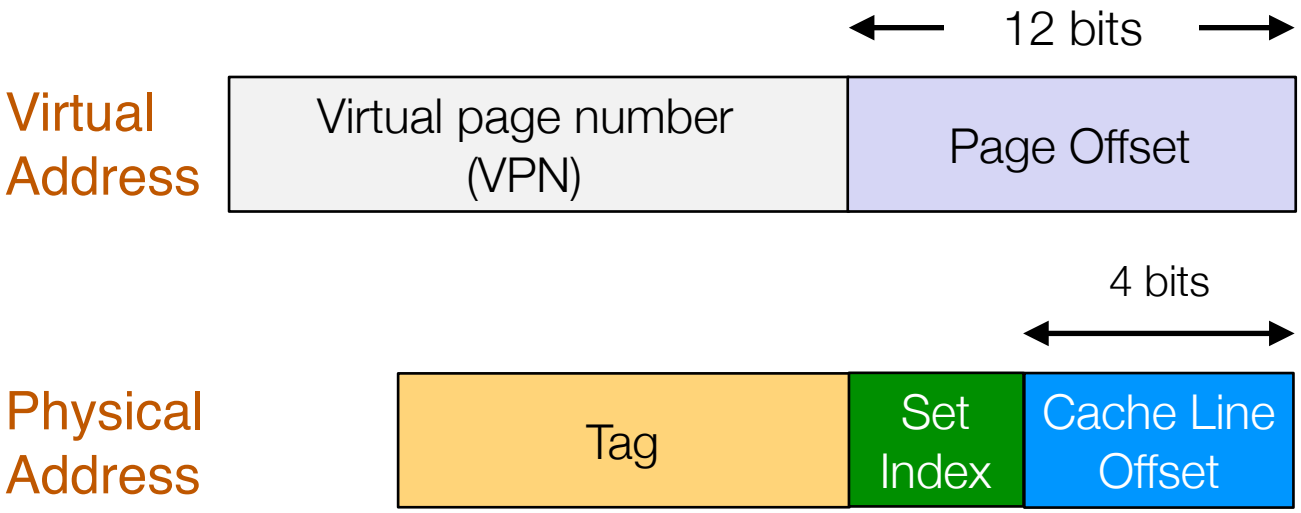


- Set Index + Cache Line Offset = Page Offset
- Indexing into cache in parallel with translation (TLB access)
- If TLB hits, can get the data back in one cycle

Any Implications?

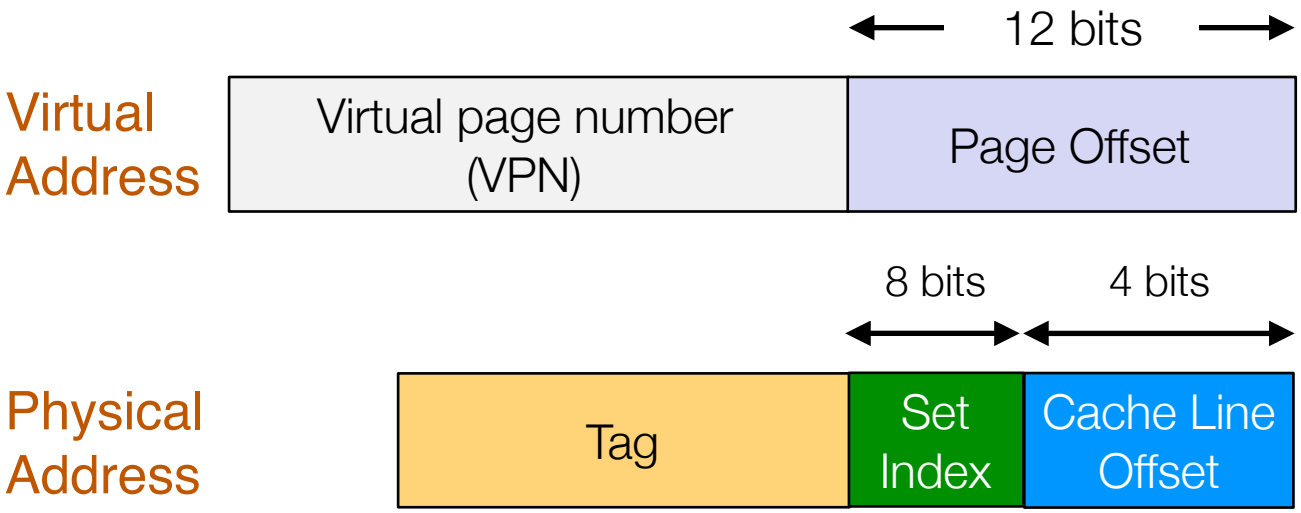


Any Implications?



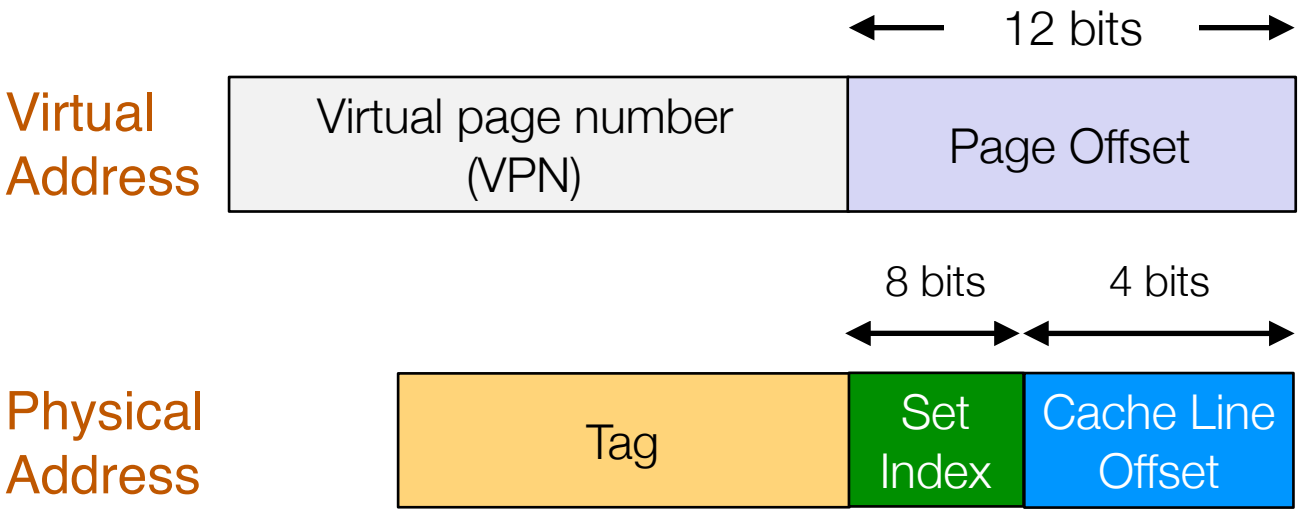
- Assuming 4K page size, cache line size is 16 bytes.

Any Implications?



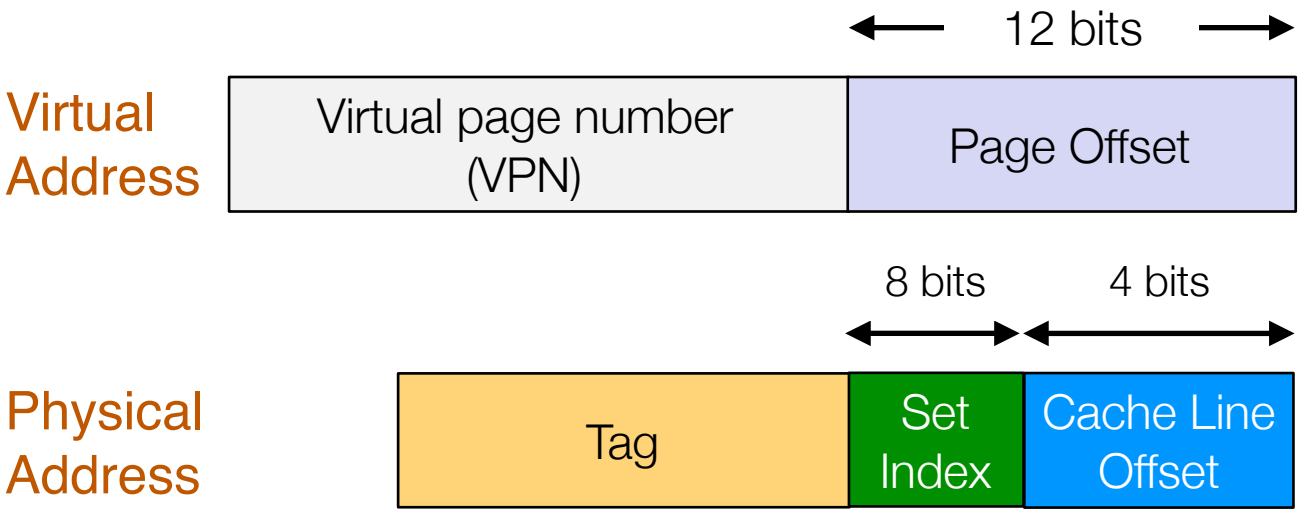
- Assuming 4K page size, cache line size is 16 bytes.
- Set Index = 8 bits. Can only have 256 Sets => Limit cache size

Any Implications?



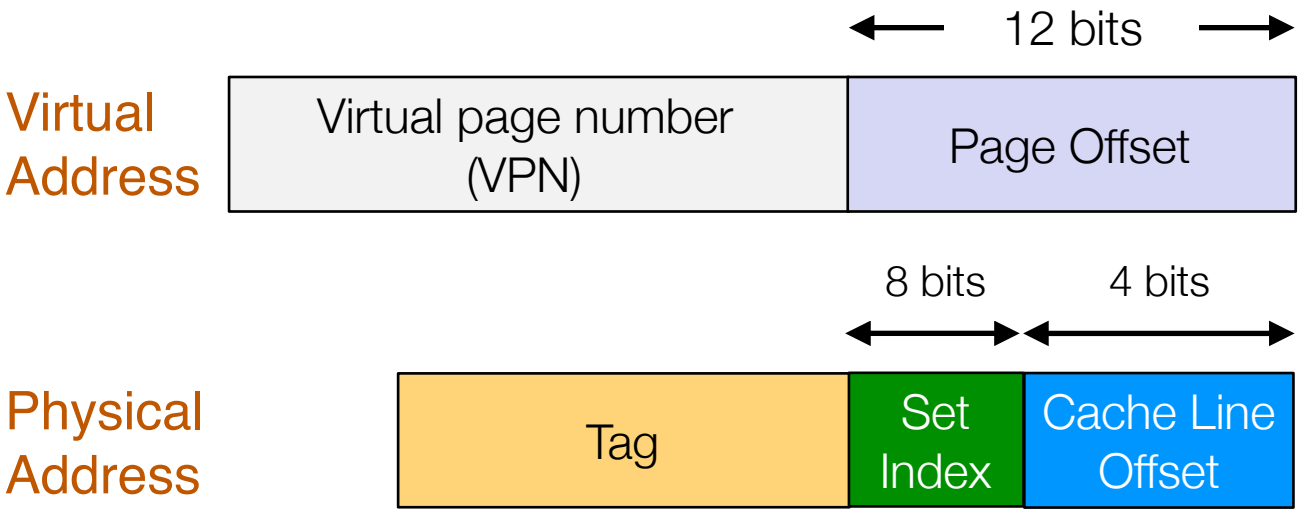
- Assuming 4K page size, cache line size is 16 bytes.
- Set Index = 8 bits. Can only have 256 Sets => Limit cache size
- Increasing cache size then requires increasing associativity

Any Implications?



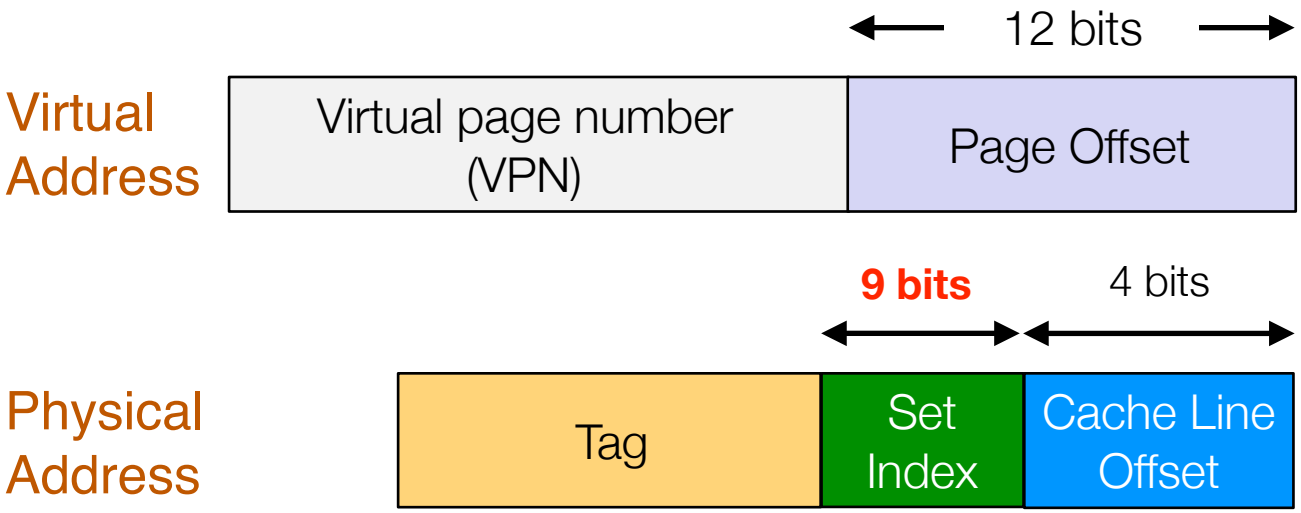
- Assuming 4K page size, cache line size is 16 bytes.
- Set Index = 8 bits. Can only have 256 Sets => Limit cache size
- Increasing cache size then requires increasing associativity
  - Not ideal because that requires comparing more tags

Any Implications?



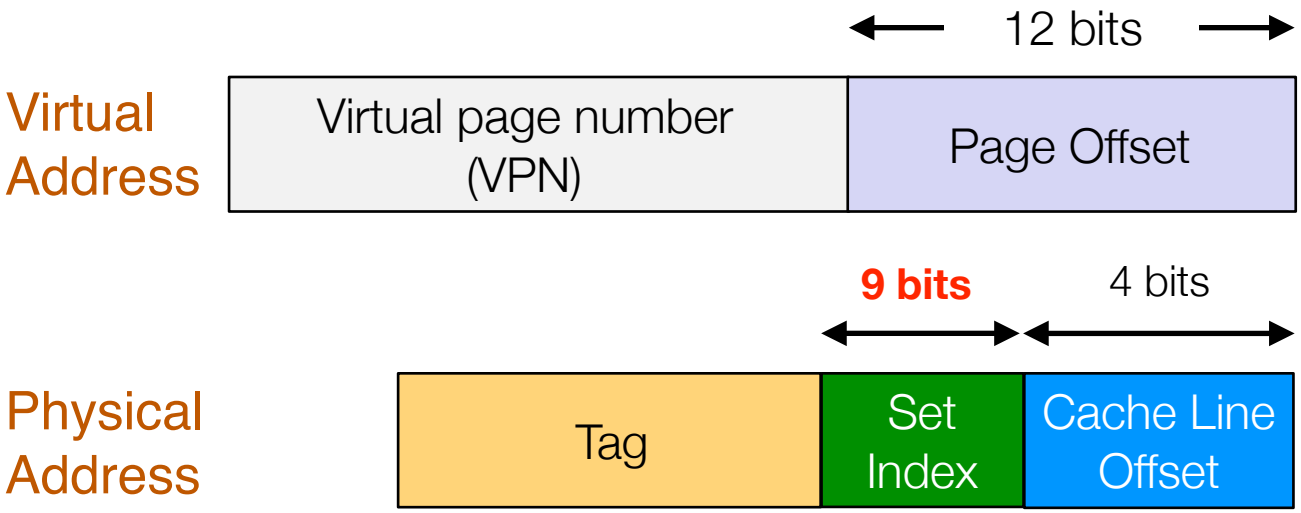
- Assuming 4K page size, cache line size is 16 bytes.
- Set Index = 8 bits. Can only have 256 Sets => Limit cache size
- Increasing cache size then requires increasing associativity
  - Not ideal because that requires comparing more tags
- Solutions?

Any Implications?



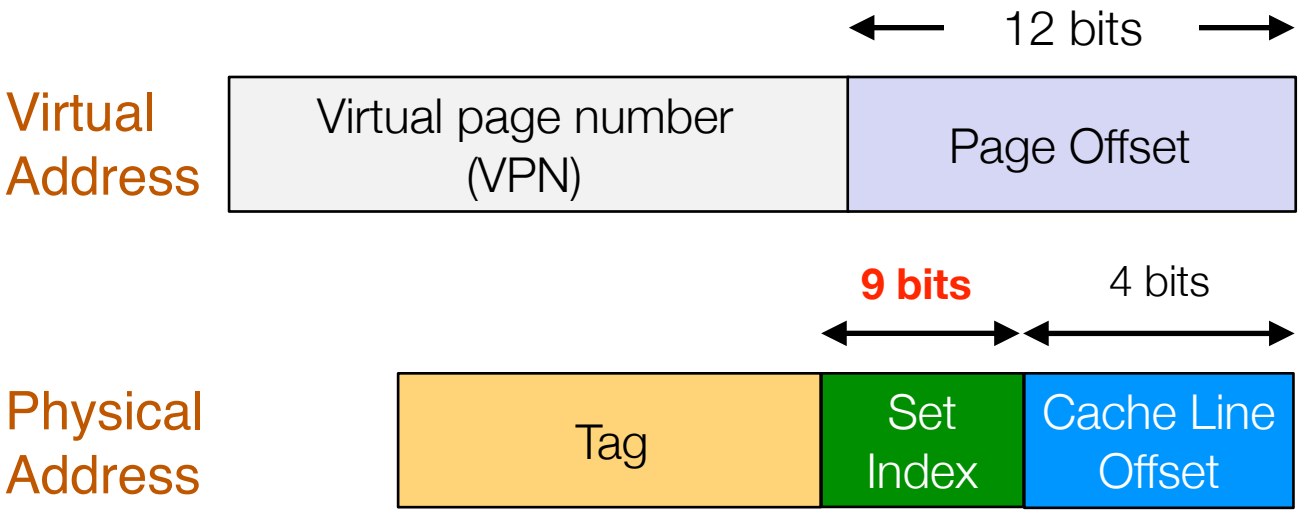
- What if we use 9 bits for Set Index? More Sets now.

Any Implications?



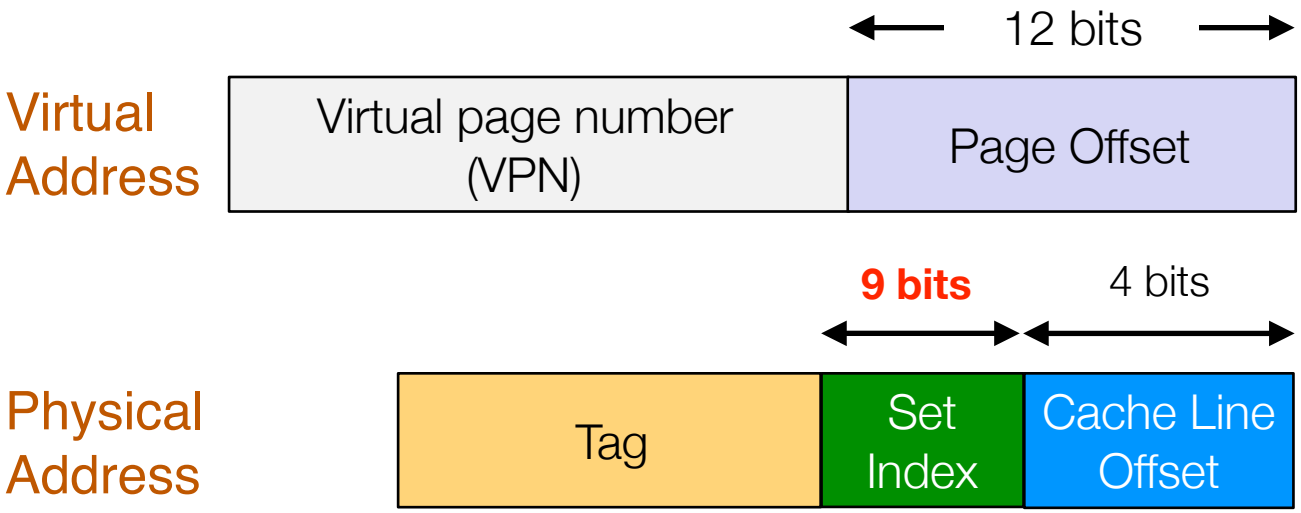
- What if we use 9 bits for Set Index? More Sets now.
- How can this still work???

Any Implications?



- What if we use 9 bits for Set Index? More Sets now.
- How can this still work???
- The least significant bit in VPN and PPN must be the same

Any Implications?

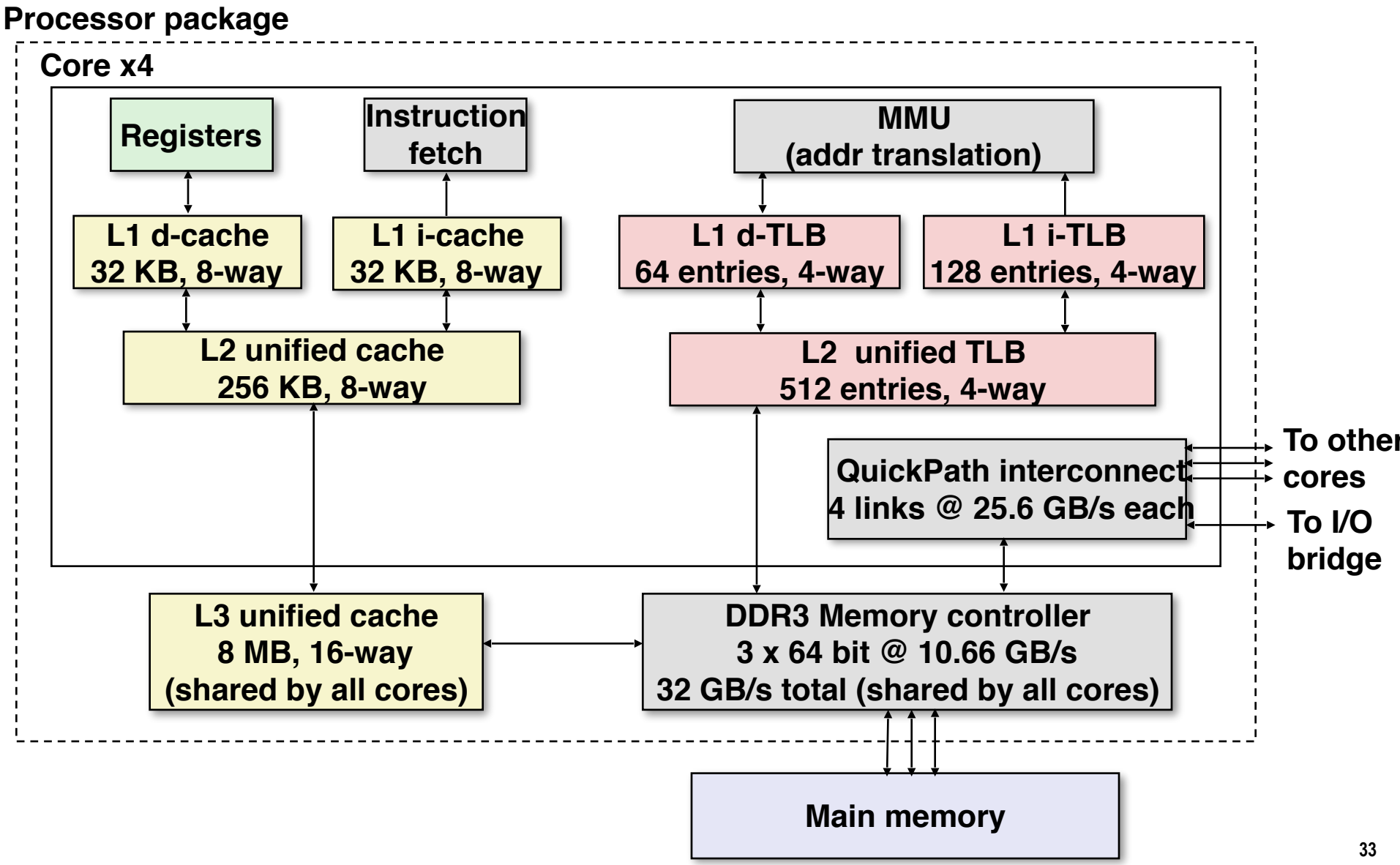


- What if we use 9 bits for Set Index? More Sets now.
- How can this still work???
- The least significant bit in VPN and PPN must be the same
- That is: an even VA must be mapped to an even PA, and an odd VA must be mapped to an odd PA

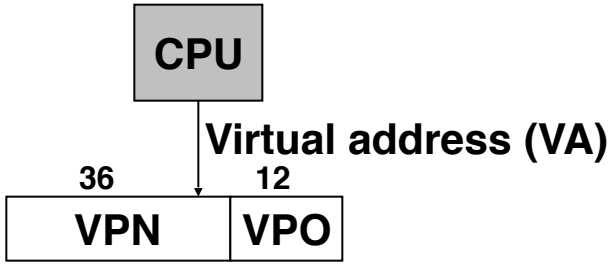
Today

- Three Virtual Memory Optimizations
  - TLB
  - Page the page table (a.k.a., multi-level page table)
  - Virtually-indexed, physically-tagged cache
- Case-study: Intel Core i7/Linux example

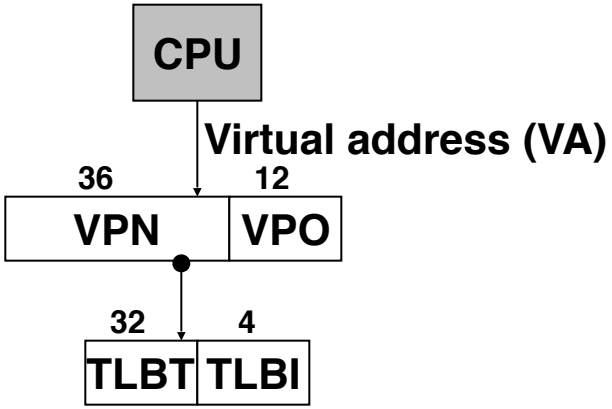
Intel Core i7 Memory System



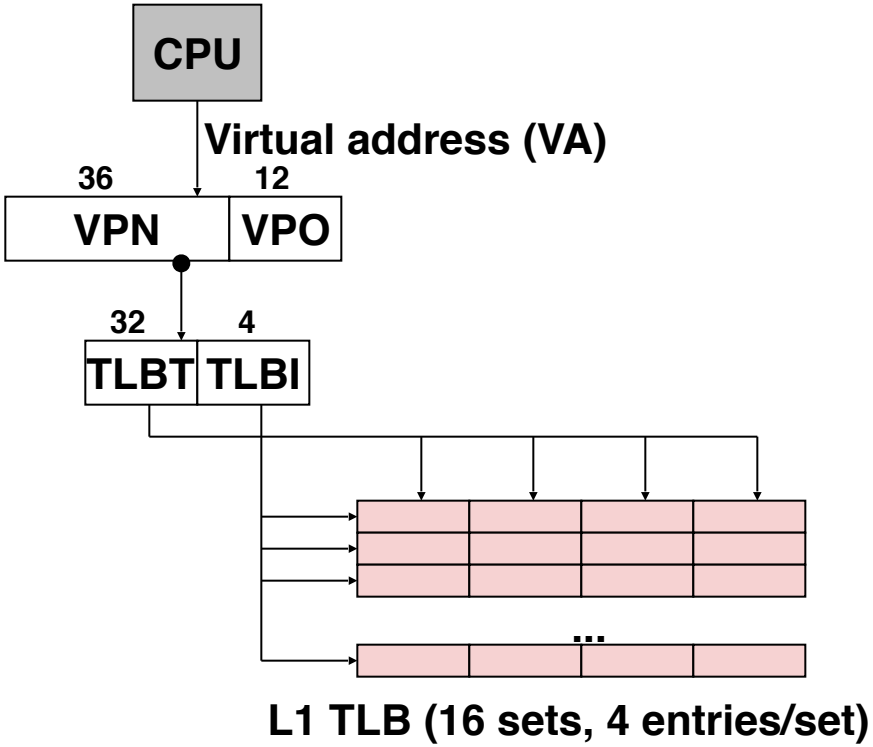
End-to-End Core i7 Address Translation



End-to-End Core i7 Address Translation

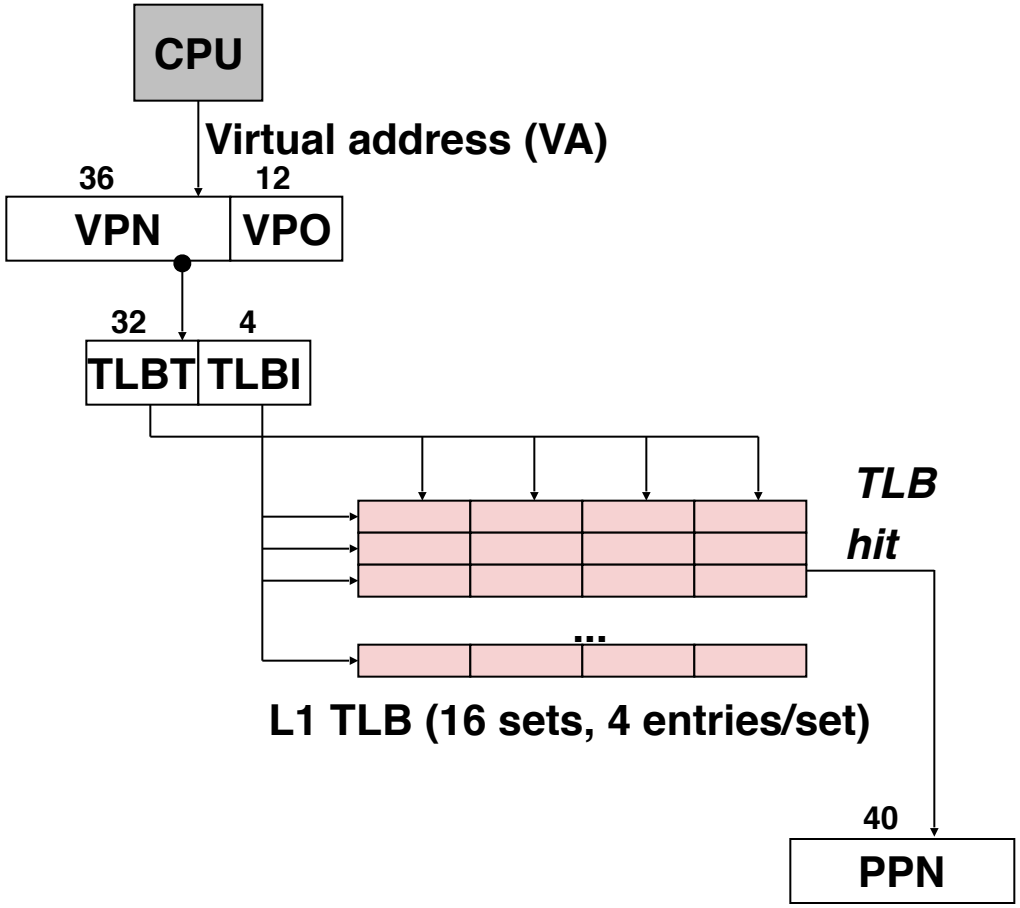


End-to-End Core i7 Address Translation

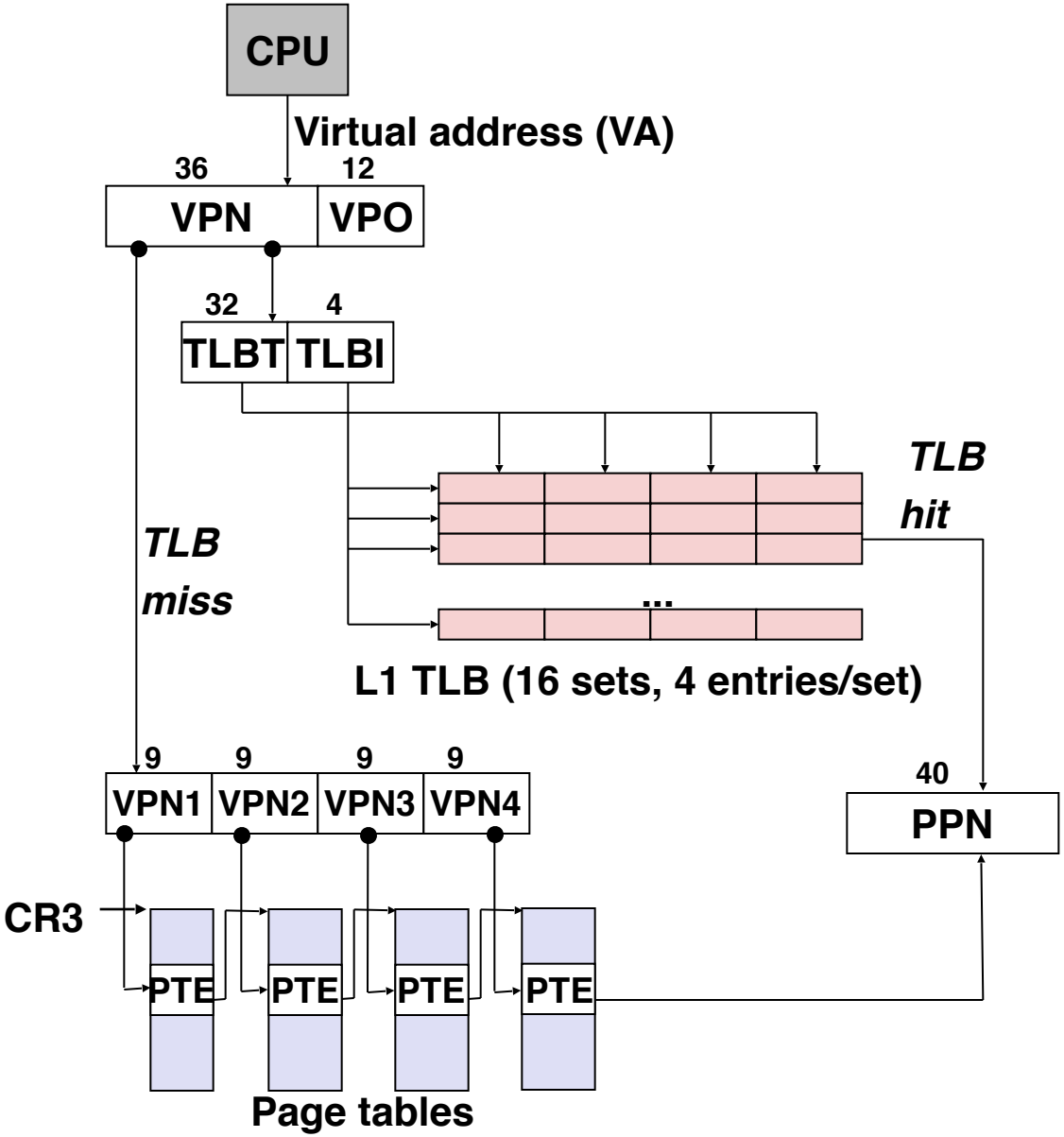


L1 TLB (16 sets, 4 entries/set)

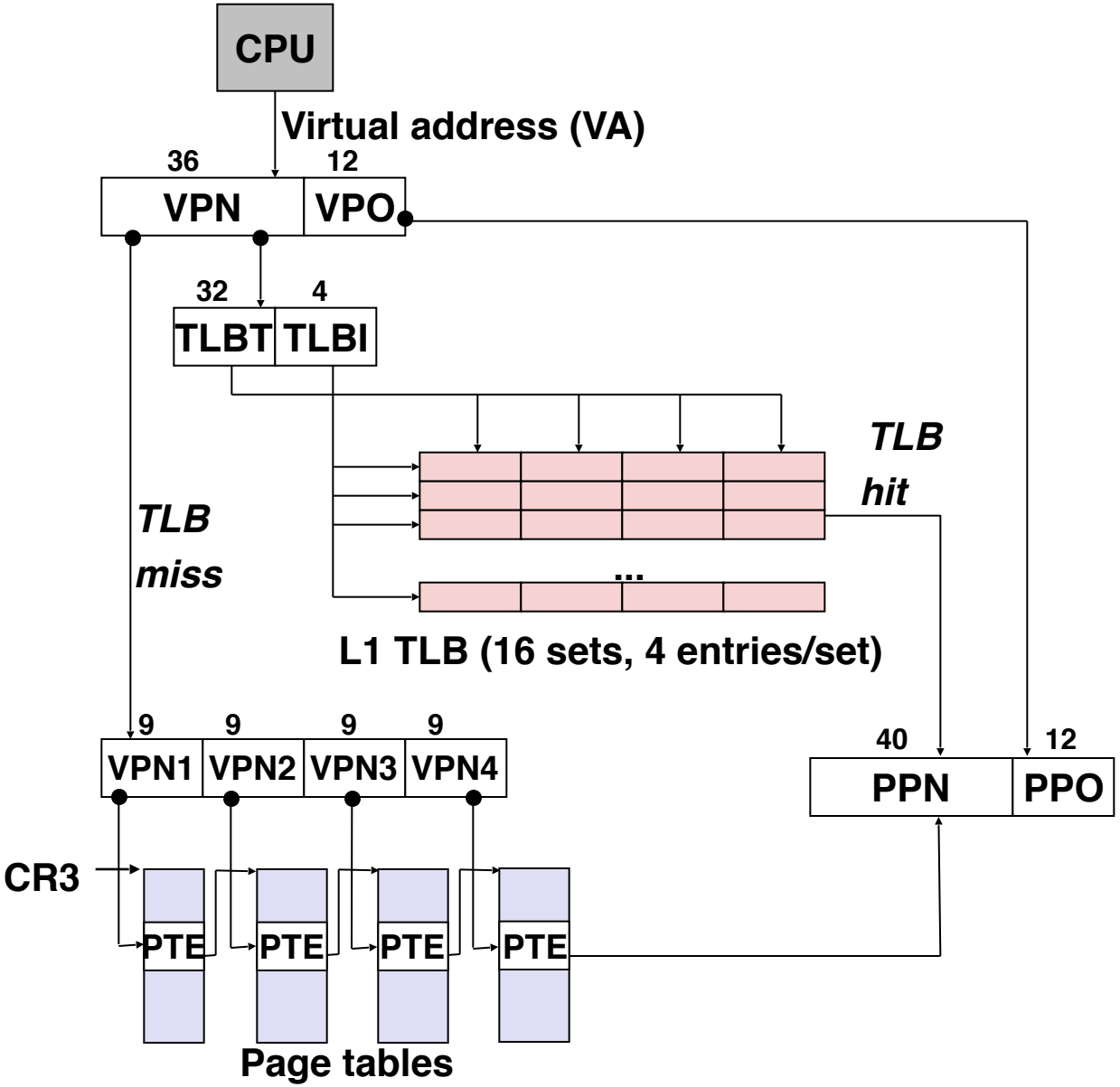
End-to-End Core i7 Address Translation



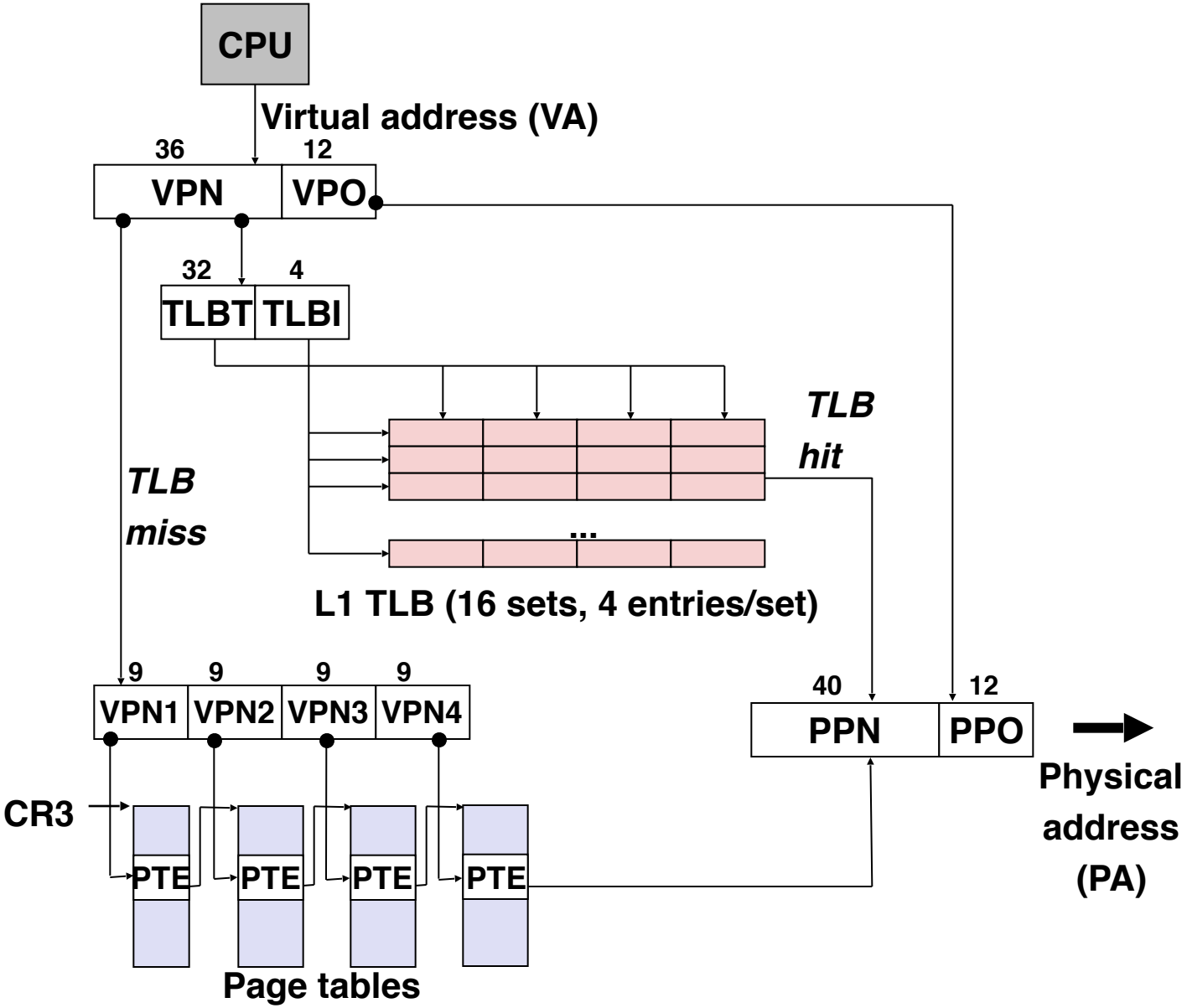
End-to-End Core i7 Address Translation



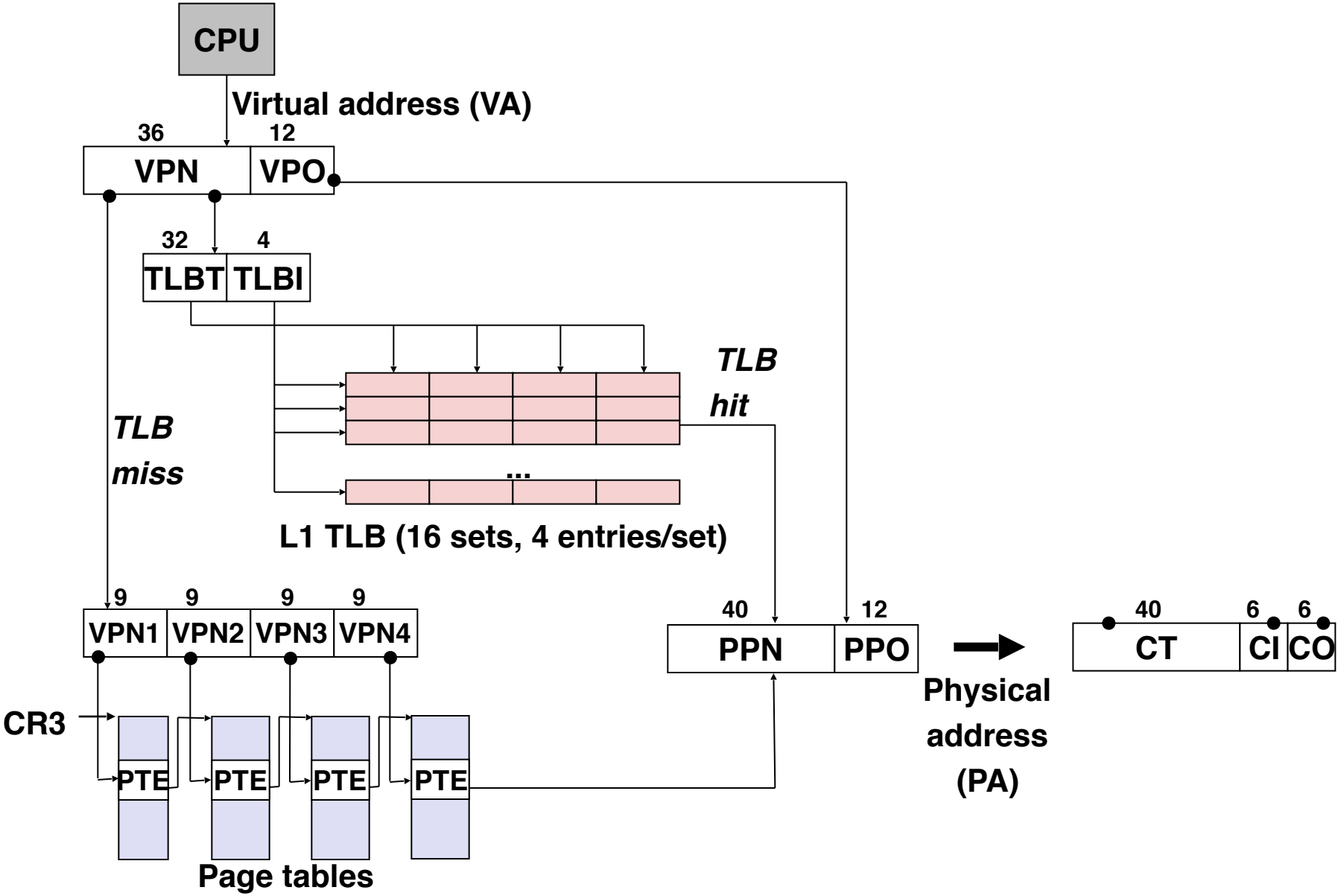
End-to-End Core i7 Address Translation



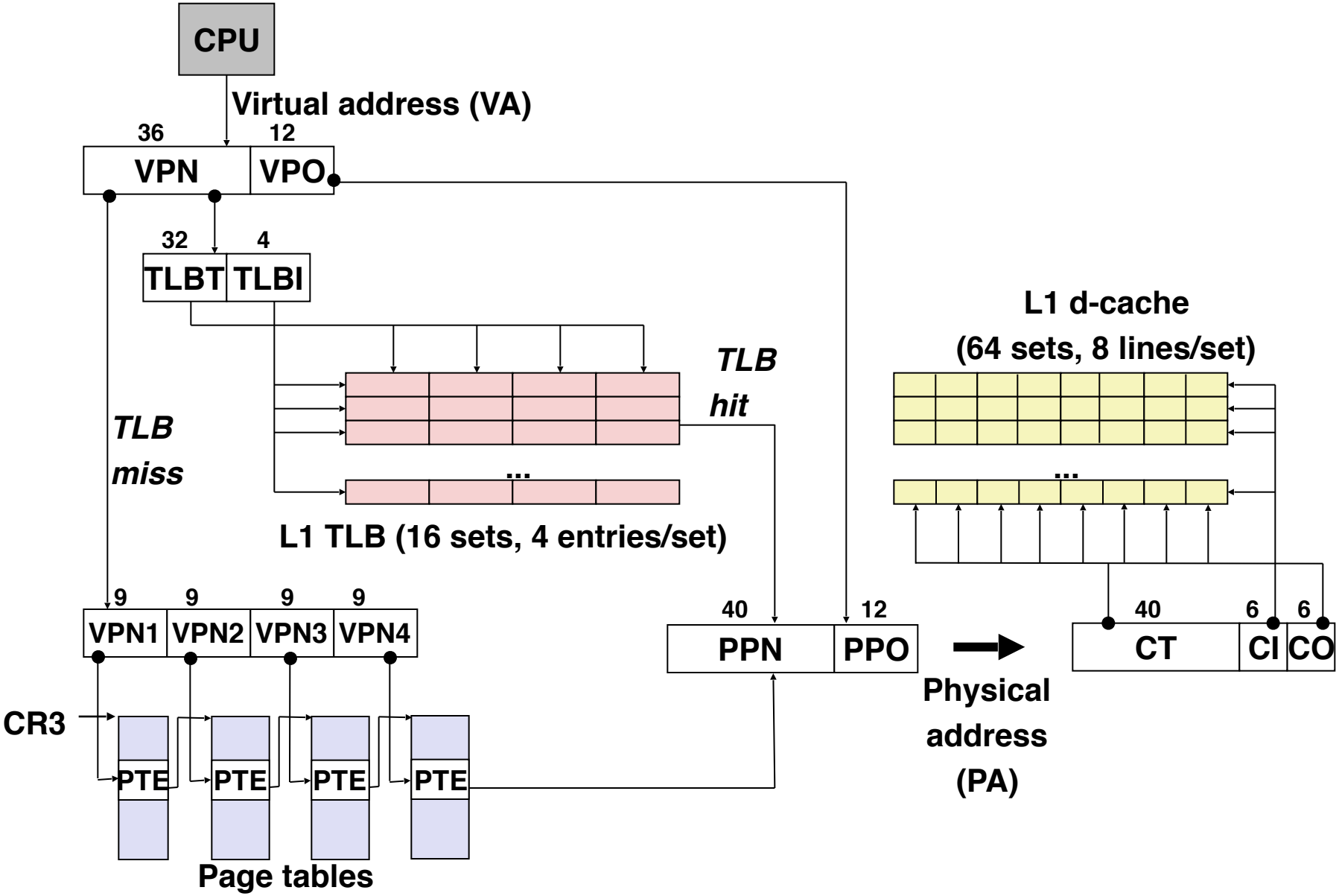
End-to-End Core i7 Address Translation



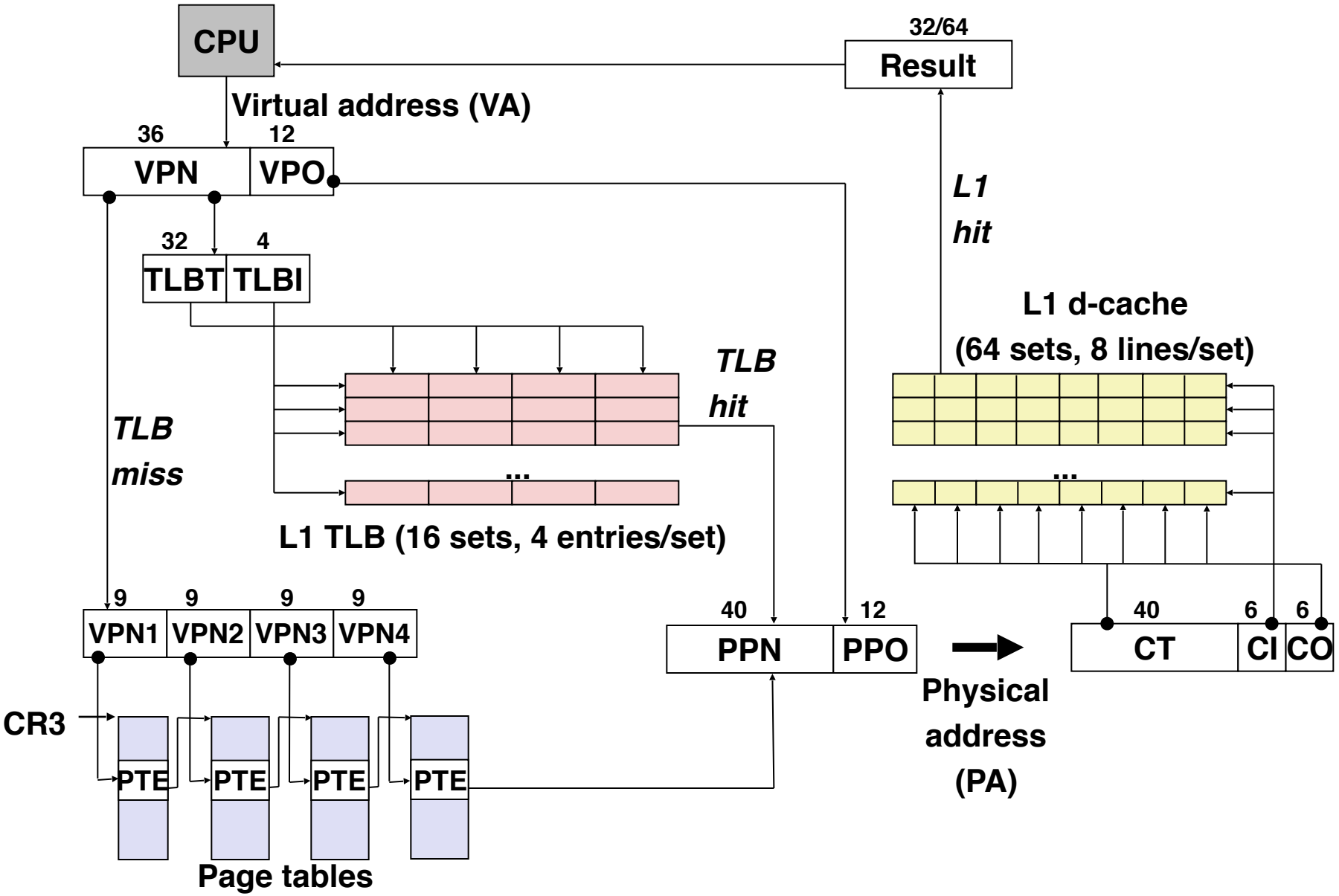
End-to-End Core i7 Address Translation



End-to-End Core i7 Address Translation



End-to-End Core i7 Address Translation



End-to-End Core i7 Address Translation