Topic V36

Virtual Memory

Reading: (Section 5.7)

Motivation 1:
Protect Data

Emily's Chat Code

```
.text
0x8000 0000     la      a0, MsgBuffer
0x8000 0004     jal     ra, WriteMsg
0x8000 0008     la      a0, MsgBuffer
0x8000 000C     jal     ra, ReadReply
…

.data
MsgBuffer:      .space 256
```

Hunter's Chat Code

```
.text
0x8000 0000     la      a0, MsgBuffer
0x8000 0004     jal     ra, WriteMsg
0x8000 0008     la      a0, MsgBuffer
0x8000 000C     jal     ra, ReadReply
…

.data
MsgBuffer:      .space 256
```

●●●●○ Optus 22:46 10%

Emily

14 Jul 2015 13:50

So Emily. I just asked siri what I should say to girl with a cute smile on Tinder...

:)

And what did she say?

She said tell Emily...

...You auto-complete me ;)

Hahah :p

●●●○○ ROGERS 1:02 AM

Back Hunter More

nothing gruesome, just like get rid of it

Apr 6, 2014, 7:42 PM

If this is going to be posted on the internet, I'd like to clear my name. I'll off the owner, not the dog. I'd never hurt a dog.

Apr 9, 2014, 12:42 PM

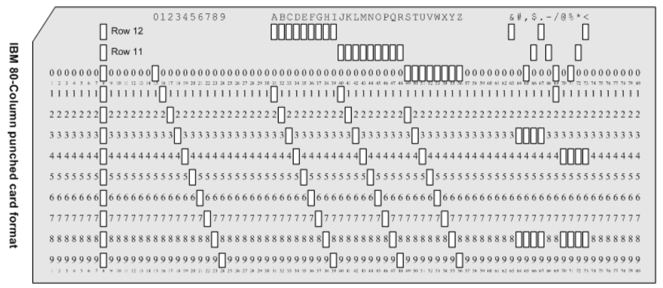
it's okay, it didn't go on the internet

It did, someone sent it to me!

stop killing dogs

Motivation 2:
Provide Illusion of
Larger Memory

IBM 1130



Punch Card

Introduced in 1965

15-bit address space

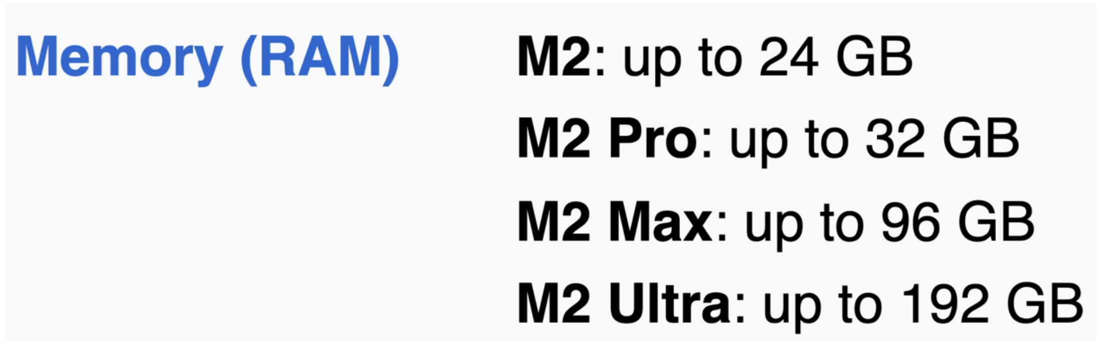
Maximum of 64 Kbytes of memory

Later could have up to 4 terminals

Fortran compiler run on machine with 8Kbytes of memory

Punch Card Perforator

48-bit address bus



Memory (RAM)	**M2**: up to 24 GB
	M2 Pro: up to 32 GB
	M2 Max: up to 96 GB
	M2 Ultra: up to 192 GB

256GB to 8TB SSD

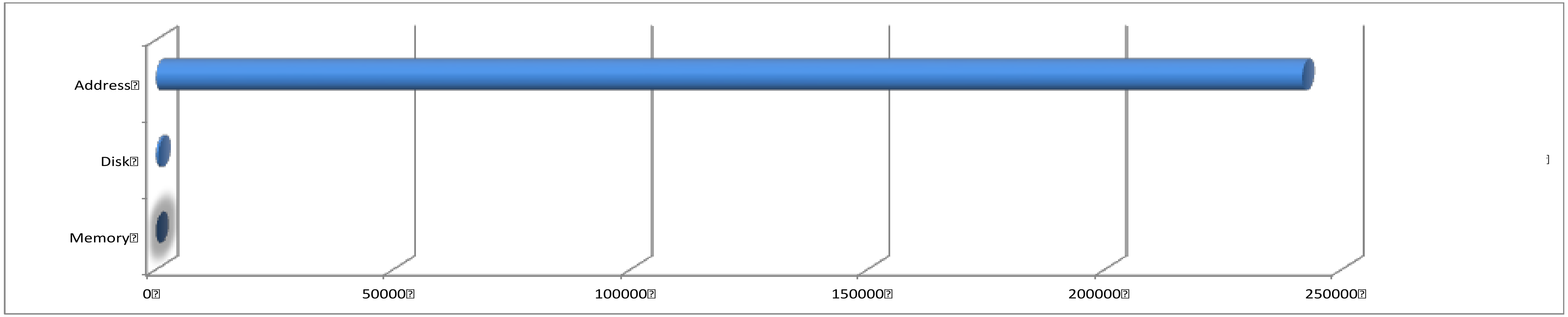
MacBook Air
Dec, 2023

48-bit address bus

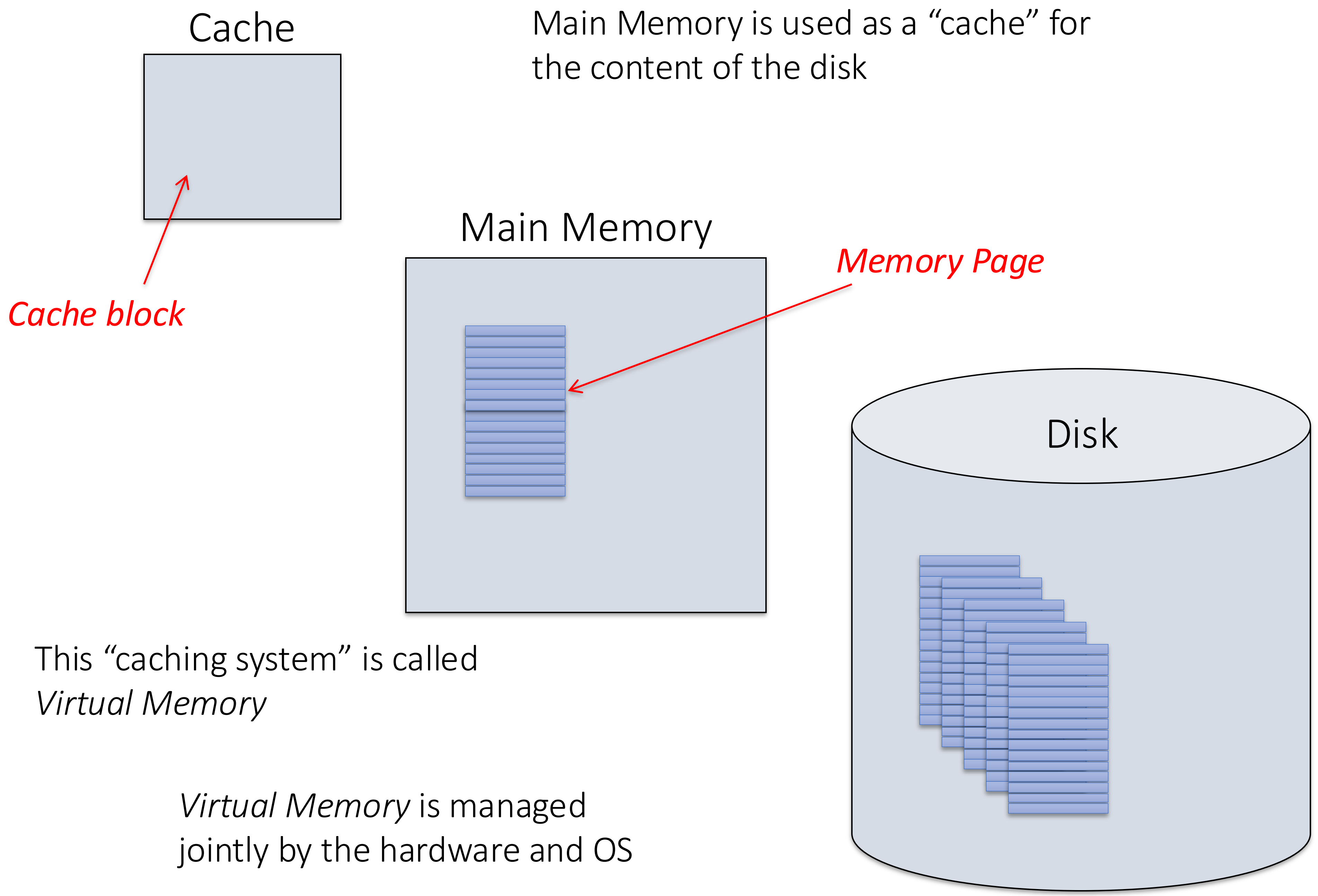
Memory (RAM)	
	M2: up to 24 GB
	M2 Pro: up to 32 GB
	M2 Max: up to 96 GB
	M2 Ultra: up to 192 GB

256GB to 8TB SSD

$2^{48} = 2^{8}$ TB = 256 TB = 241,664 GB



Gigabytes

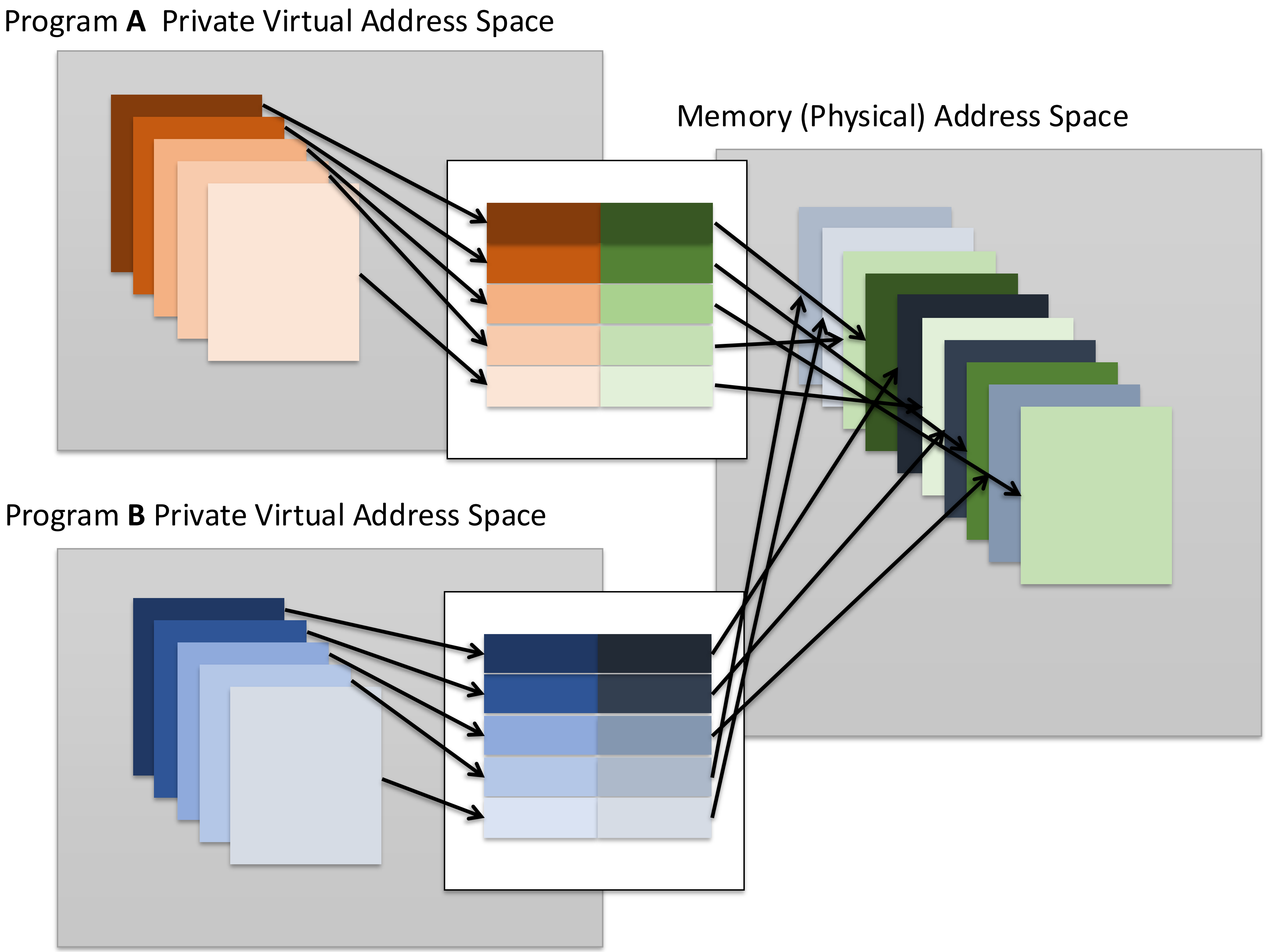


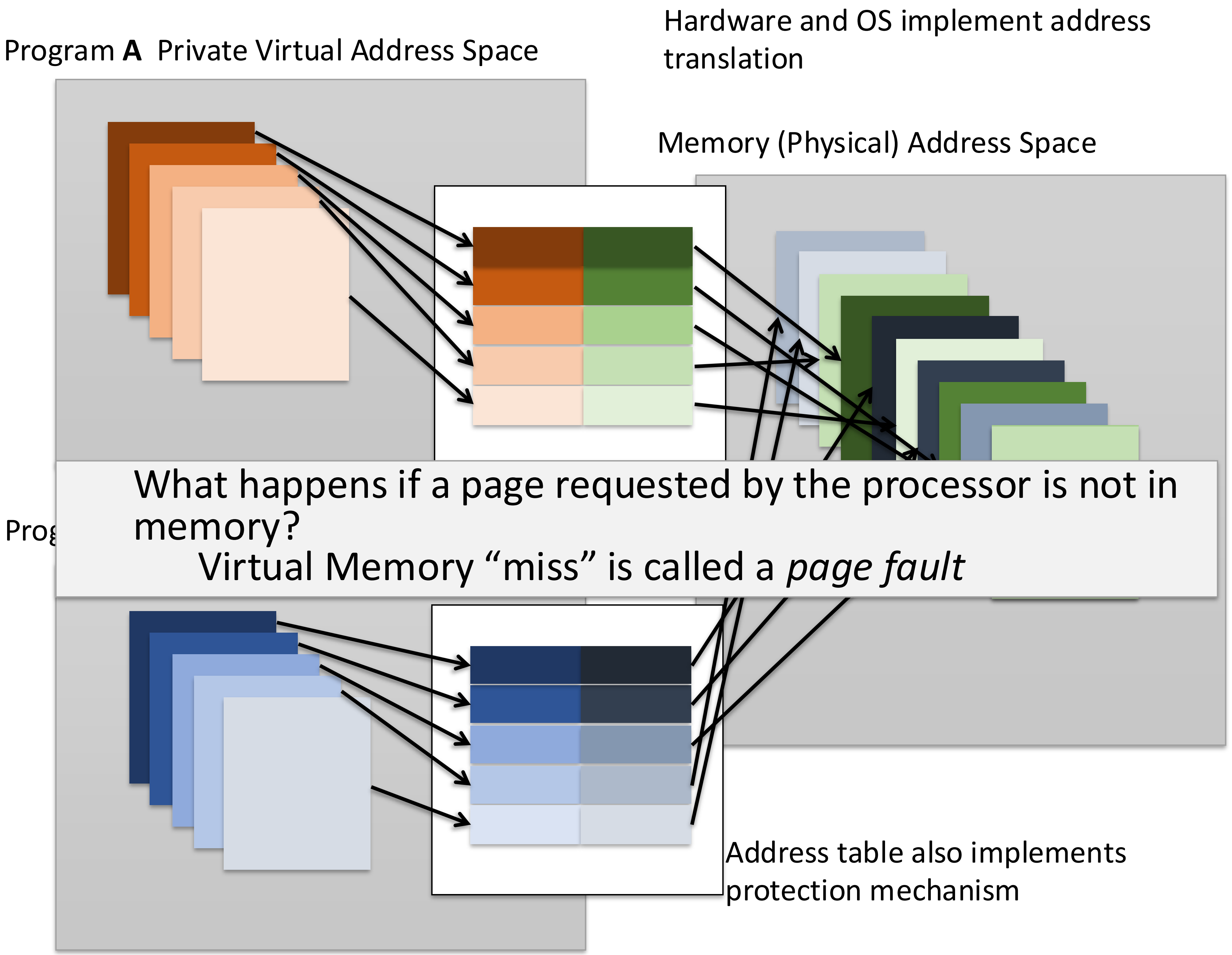
Main Memory is used as a "cache" for the content of the disk

This "caching system" is called *Virtual Memory*

Virtual Memory is managed jointly by the hardware and OS

Virtual Memory


Program **A** Private Virtual Address Space
Memory (Physical) Address Space
Program **B** Private Virtual Address Space

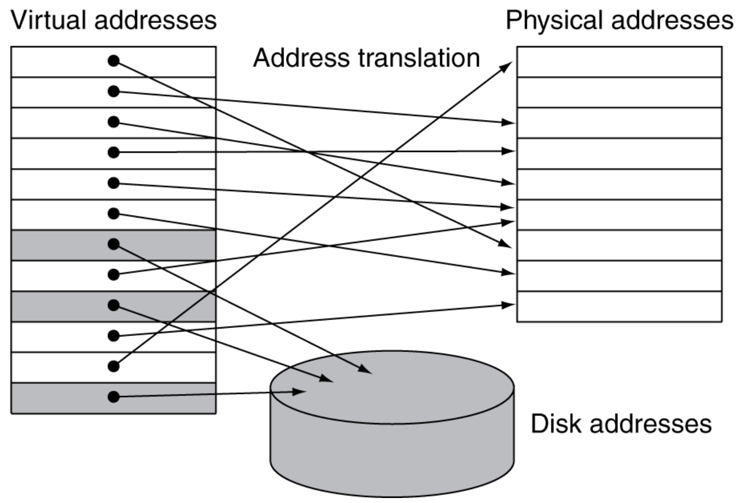

Program A Private Virtual Address Space
Hardware and OS implement address translation
Memory (Physical) Address Space
Prog
Address table also implements protection mechanism

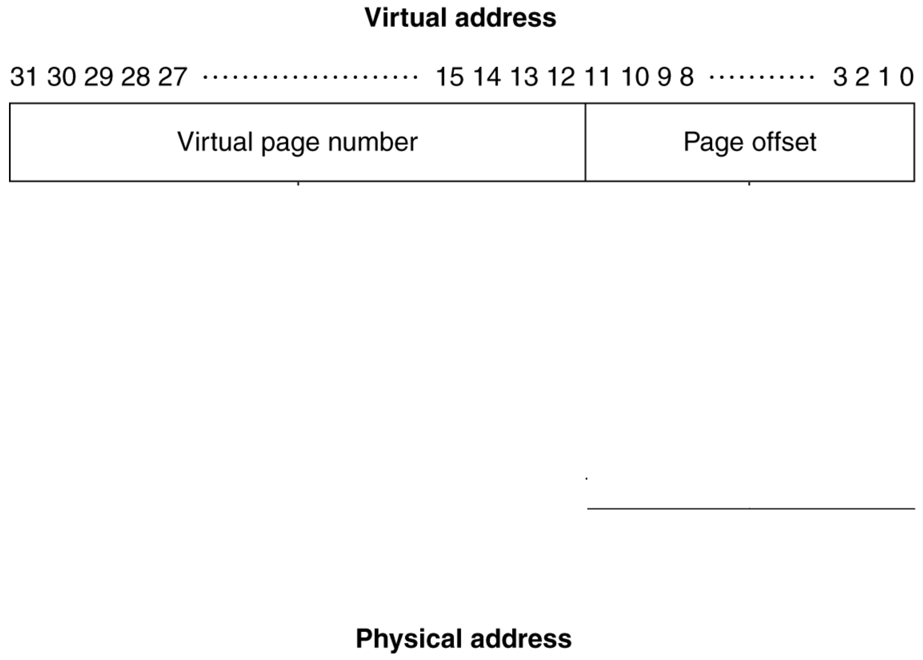
What happens if a page requested by the processor is not in memory?

Virtual Memory "miss" is called a *page fault*

Address Translation

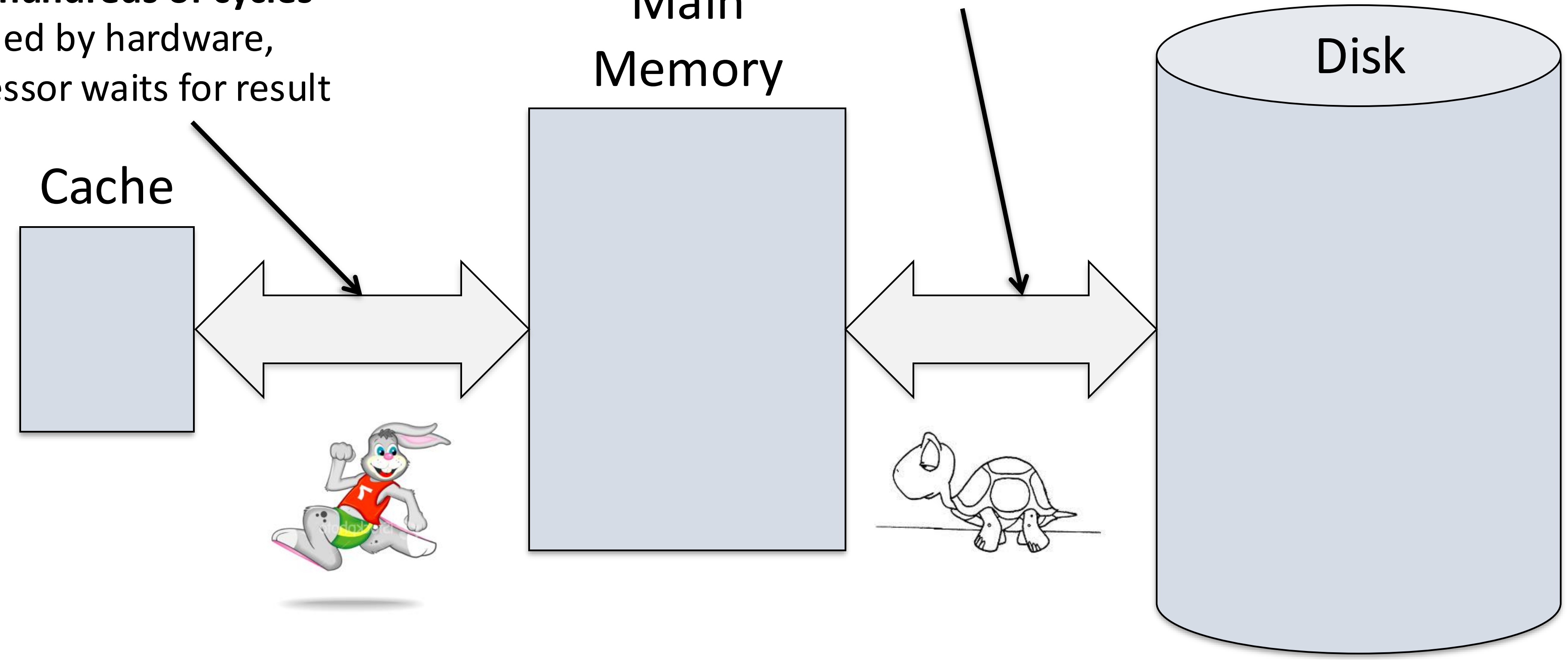
Fixed-size pages (e.g., 4K)





Try to **minimize page fault rate**:

- Fully associative page placement
- Smart replacement algorithms



Fast: hundreds of cycles
handled by hardware,
processor waits for result

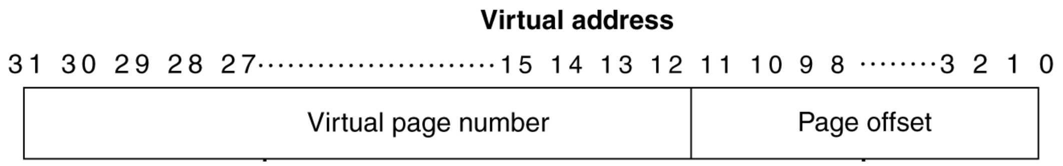
Slow: millions of cycles
exception to OS,
context switch

Unlike caches, virtual memory is not a place where data is stored

Virtual memory is:

- a mapping between two address spaces
- a mechanism that allows main memory to serve as a cache to disk storage

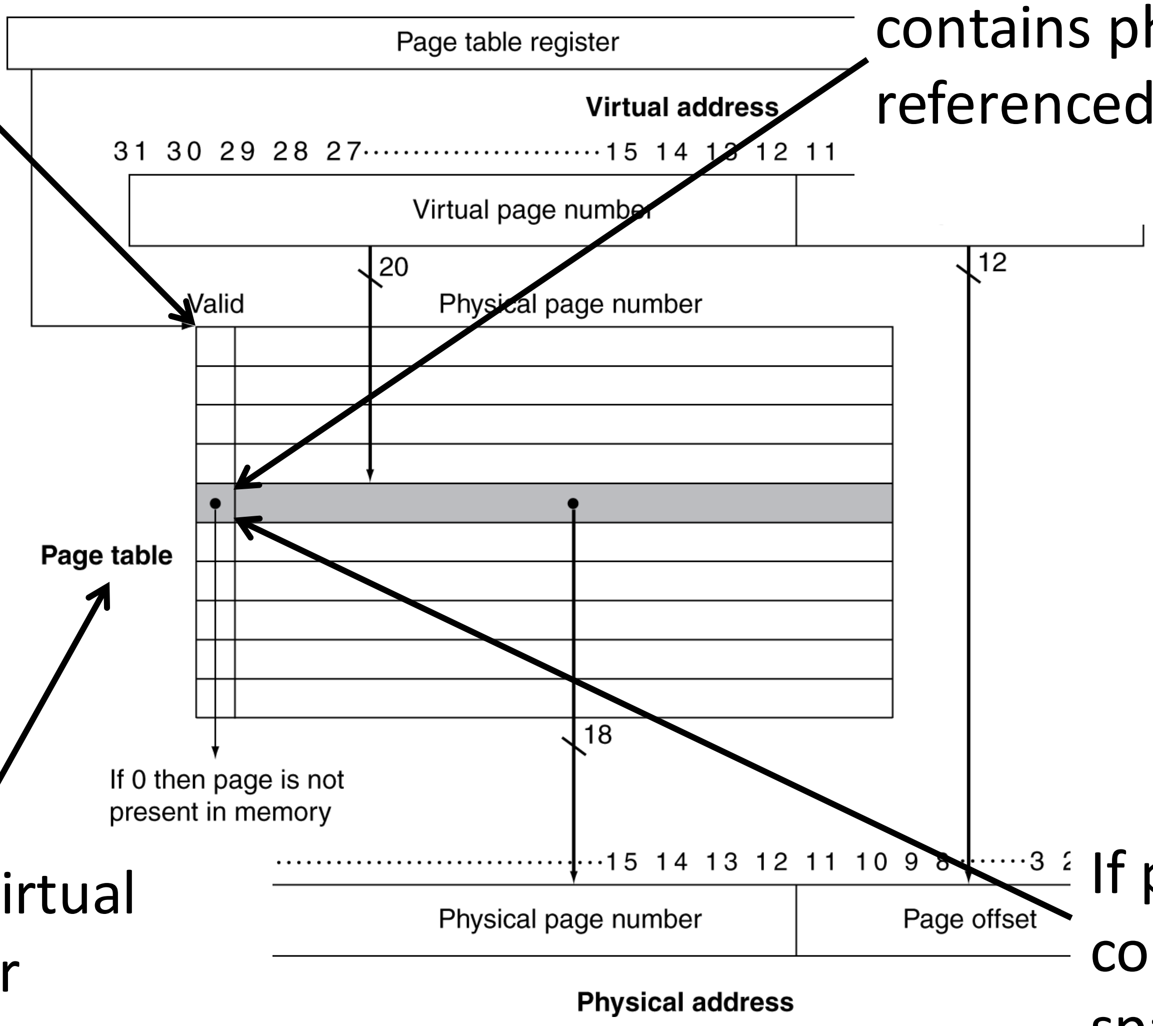
Translation Using a Page Table



า Using a Page Table

Page Table Register in CPU points to start address of page table in memory

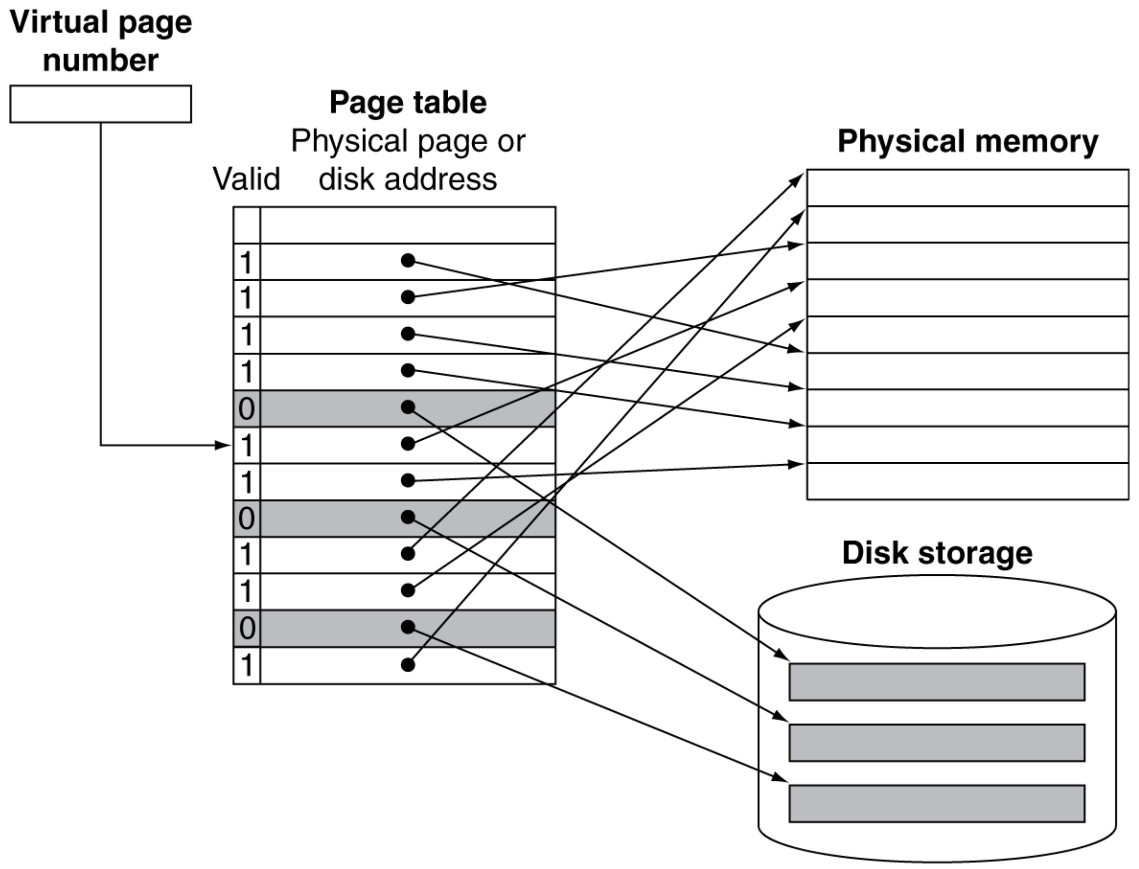
If page is in memory: contains physical address, referenced bit, dirty, protection info



Indexed by virtual page number

If page is <u>not</u> in memory: contains location in swap space on disk

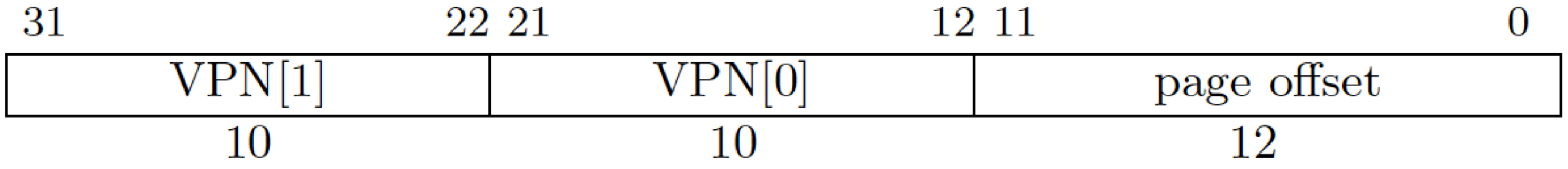
Mapping Pages to Storage



Address Translation in RISC-V

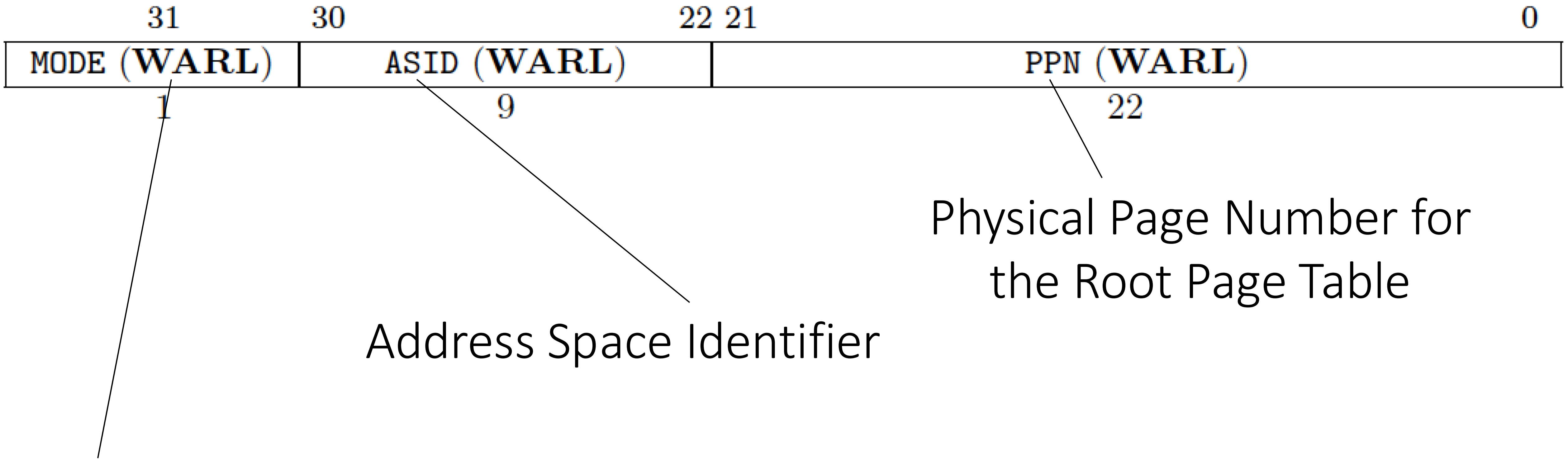
Radix-Based Page Table
And SATP register

Address Translation in RISC-V (32 bits)



Sv32 virtual address.

RISC-V Supervisor Address Translation and Protection (SATP) register

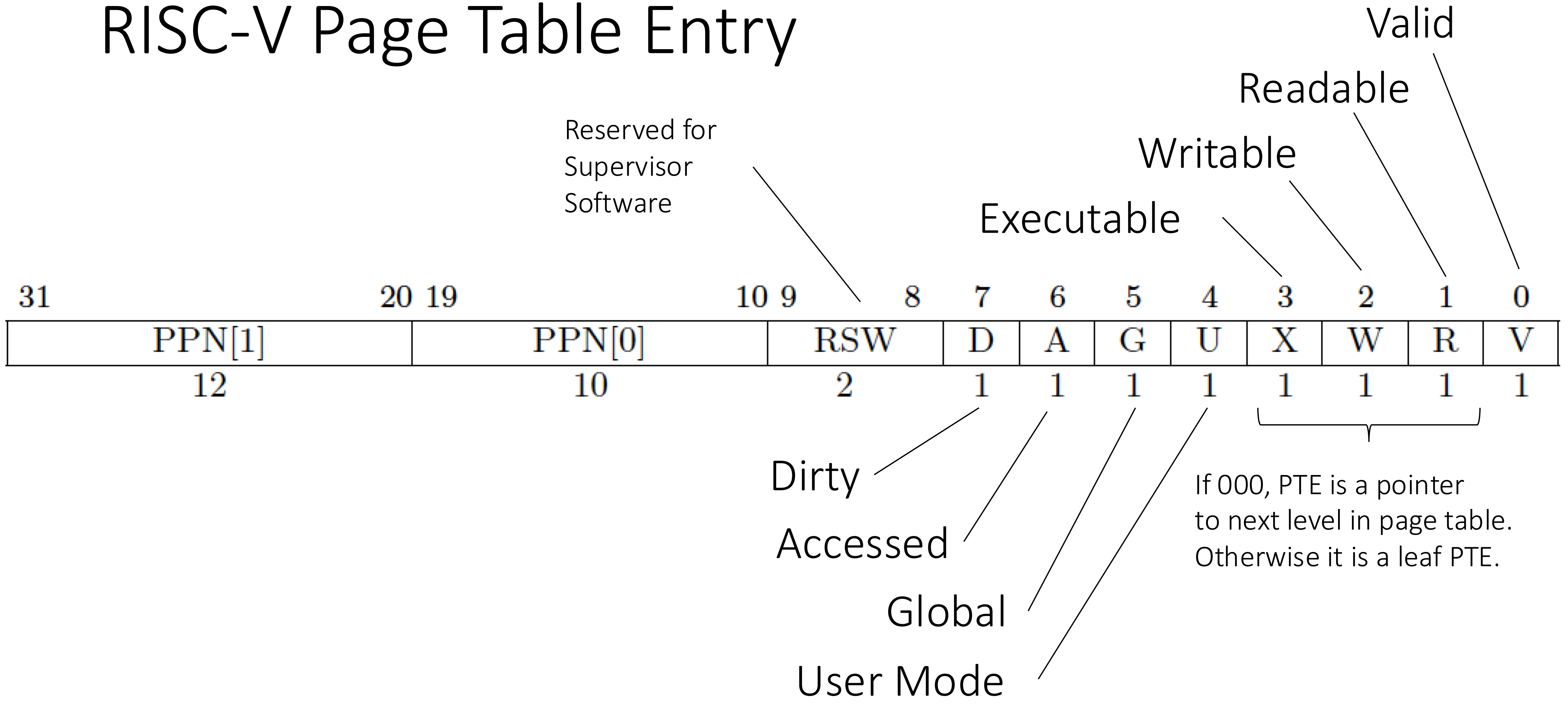


Physical Page Number for the Root Page Table

Address Space Identifier

WARL: Write Any Values, Reads Legal Values

RISC-V Page Table Entry



If 000, PTE is a pointer to next level in page table. Otherwise it is a leaf PTE.

Radix Page Table

64-bit Virtual Address

63 48	47 39	38 30	29 21	20 12	11 0
sign ext.	idx 4	idx 3	idx 2	idx 1	offset

Radix Page Table only makes new pages when accessing a new page

Each idx corresponds to a page table entry in its associated level

idx 4 -> which page table entry to access in the base page table

idx 3 -> which page table to access in that new page table from the previous page table entry

...

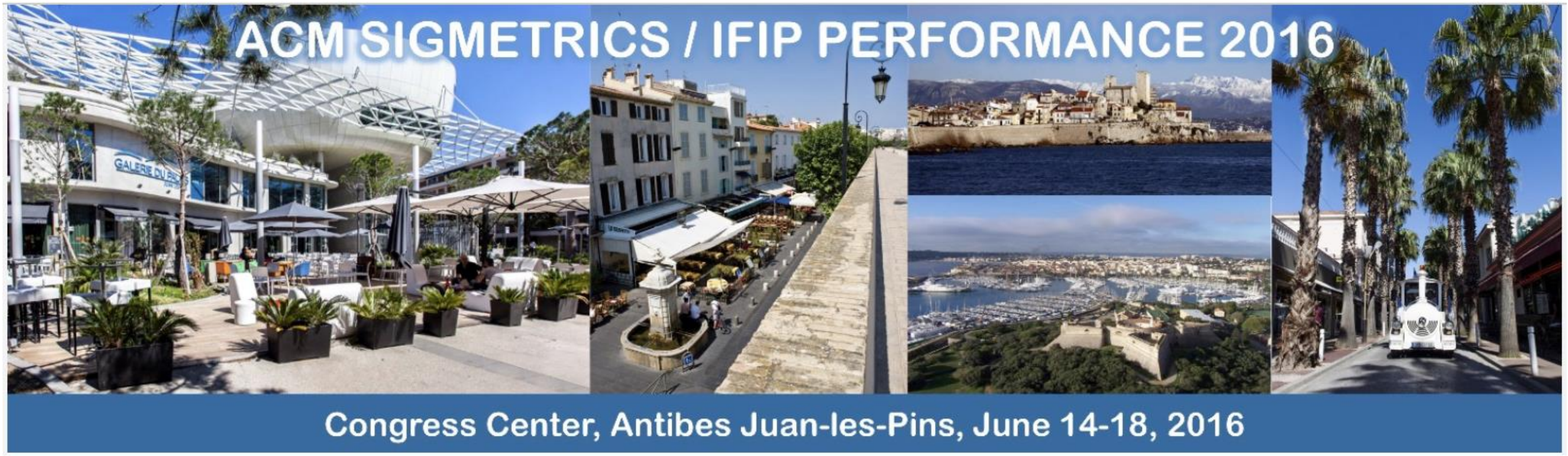
idx 1 -> which page table to access in memory from the last page table

offset -> the data to be accessed on that page

Hash, Don’t Cache (the Page Table)

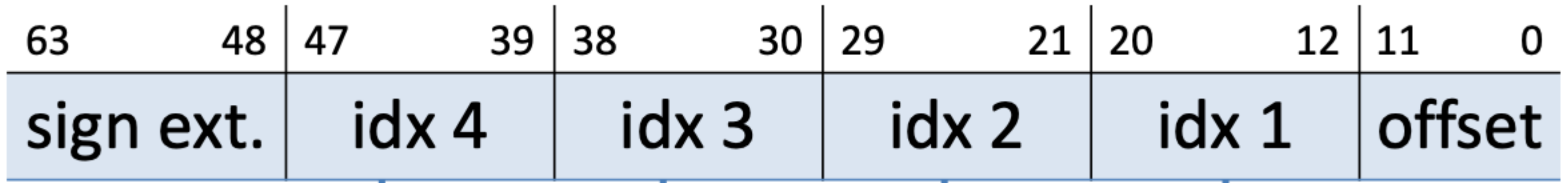
Idan Yaniv
Technion – Israel Institute of Technology
idanyani@cs.technion.ac.il

Dan Tsafrir
Technion – Israel Institute of Technology
dan@cs.technion.ac.il


ACM SIGMETRICS / IFIP PERFORMANCE 2016
Congress Center, Antibes Juan-les-Pins, June 14-18, 2016

Virtual Address -> Physical Address

64-bit Virtual Address



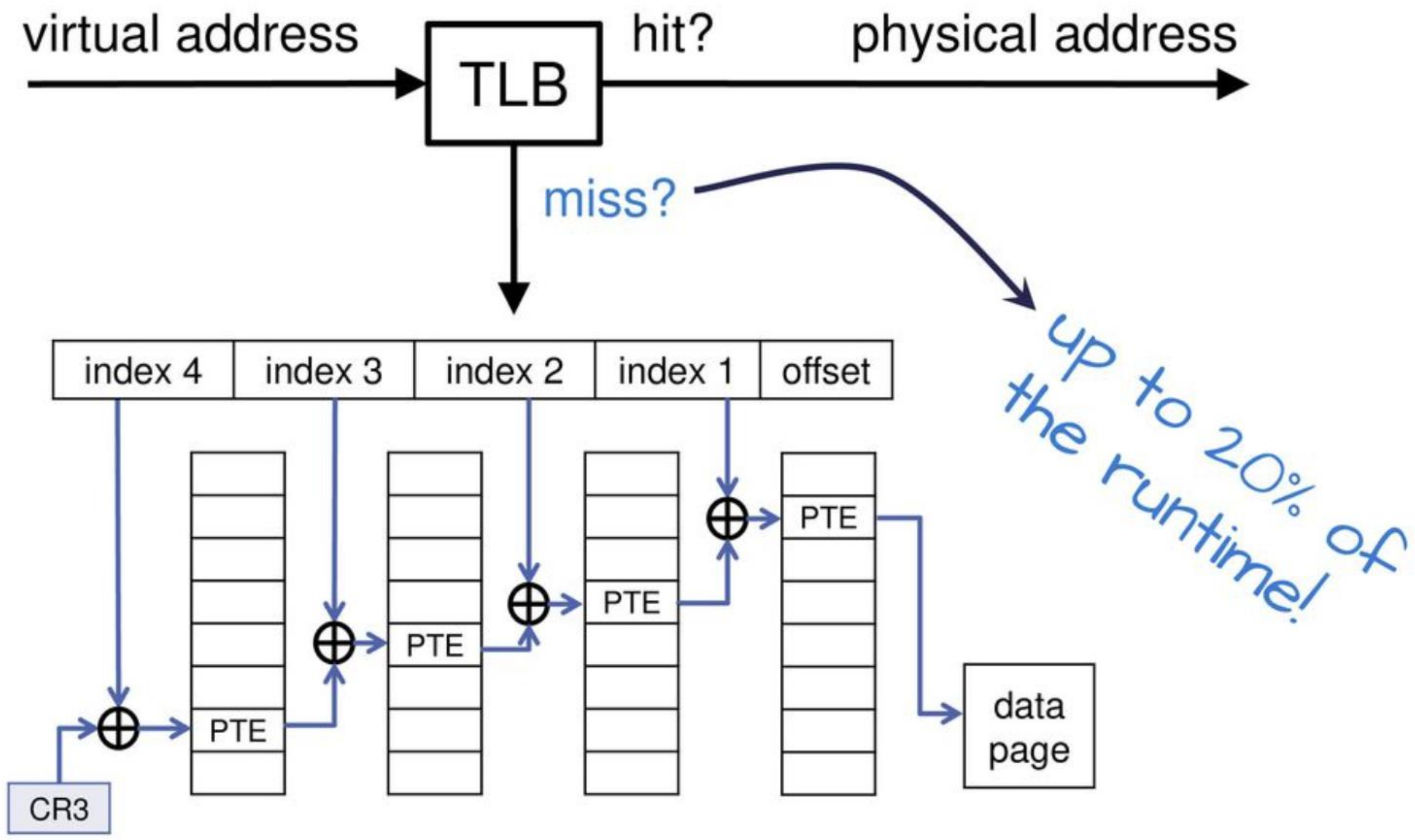
63 48	47 39	38 30	29 21	20 12	11 0
sign ext.	idx 4	idx 3	idx 2	idx 1	offset

Control Register contains base address of root table

hysical
ddress

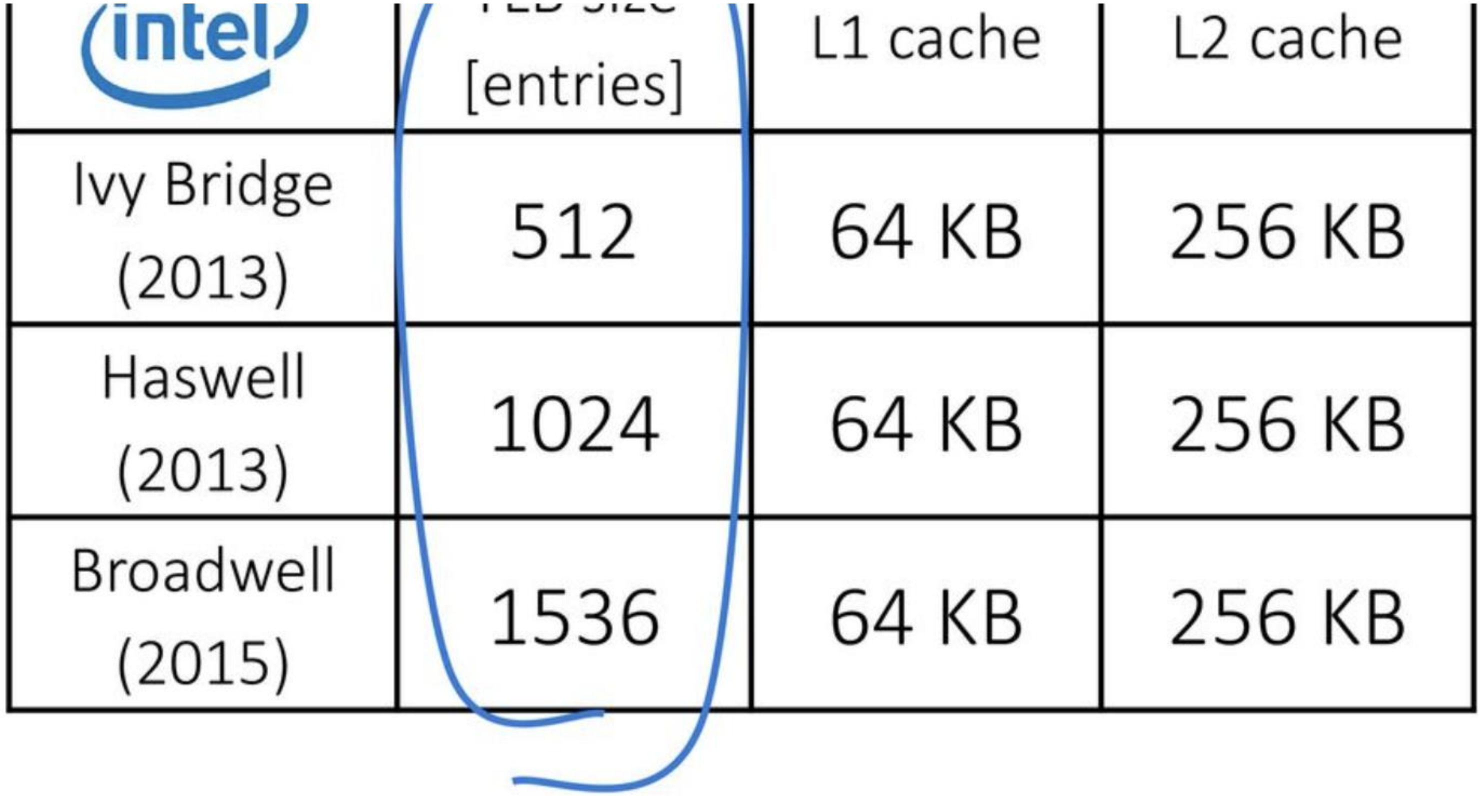
SATP: Supervisor Address Translation and Protection
PPN: Physical Page Number

Radix page tables



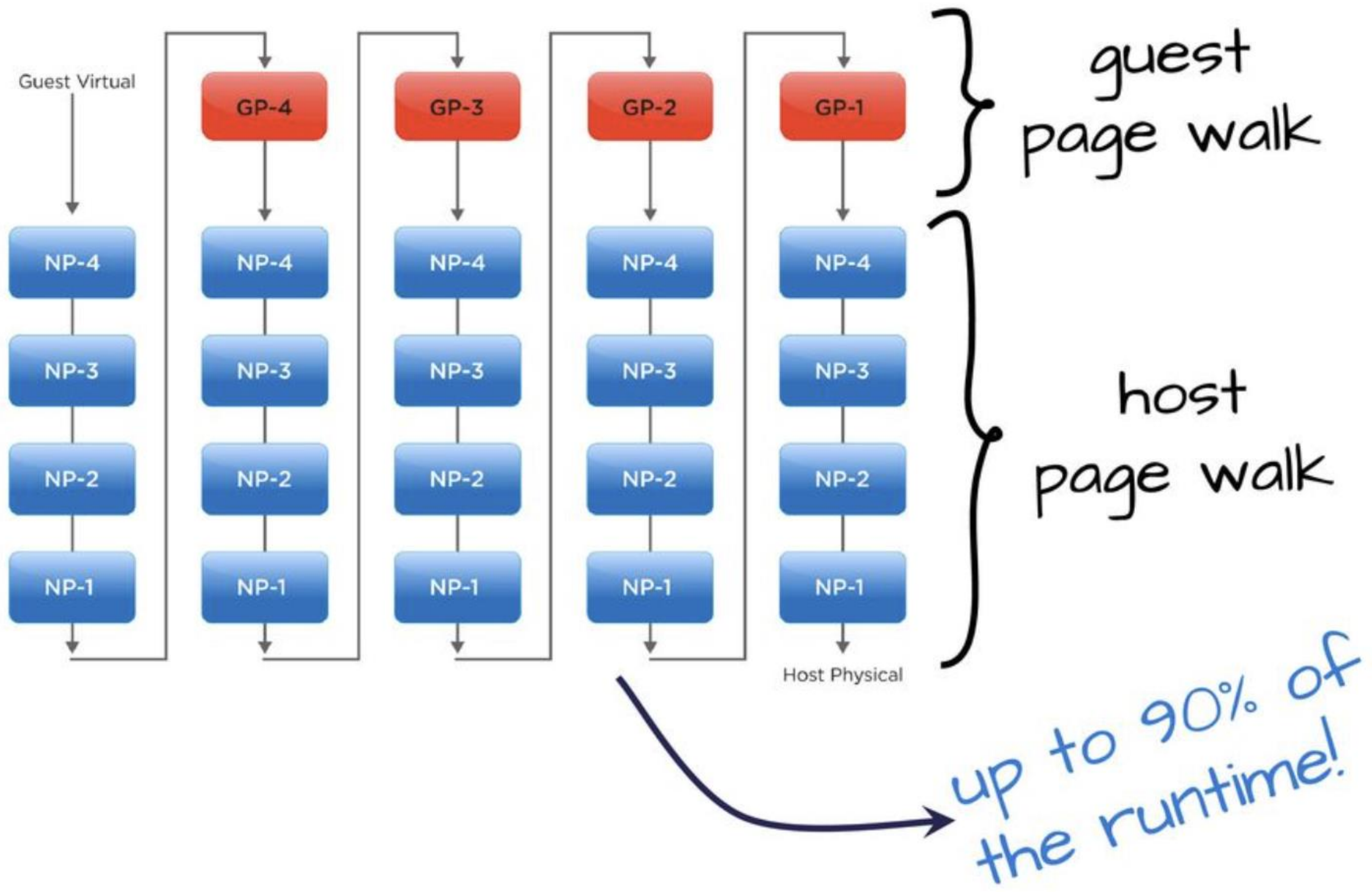
Data TLB: 4-KB Pages, 4-way set associative, 64 entries
Instruction TLB: 4-KByte pages, 8-way set associative, 64 entries

Shared 2nd-Level TLB: 4-KB / 2-MB pages, 6-way associative, 1536 entries.



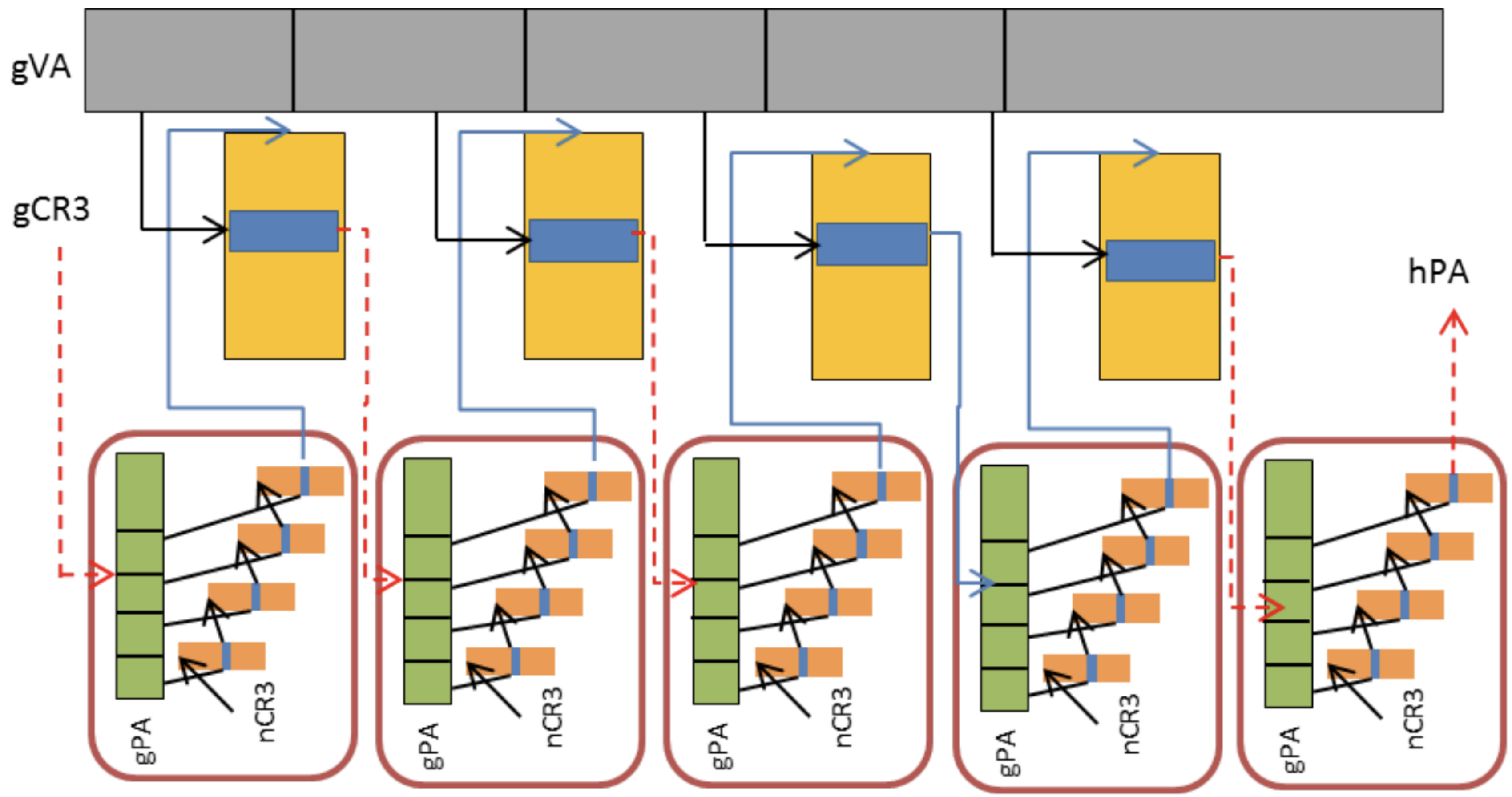
	[entries]	L1 cache	L2 cache
Ivy Bridge (2013)	512	64 KB	256 KB
Haswell (2013)	1024	64 KB	256 KB
Broadwell (2015)	1536	64 KB	256 KB

Virtualization requires 24 steps



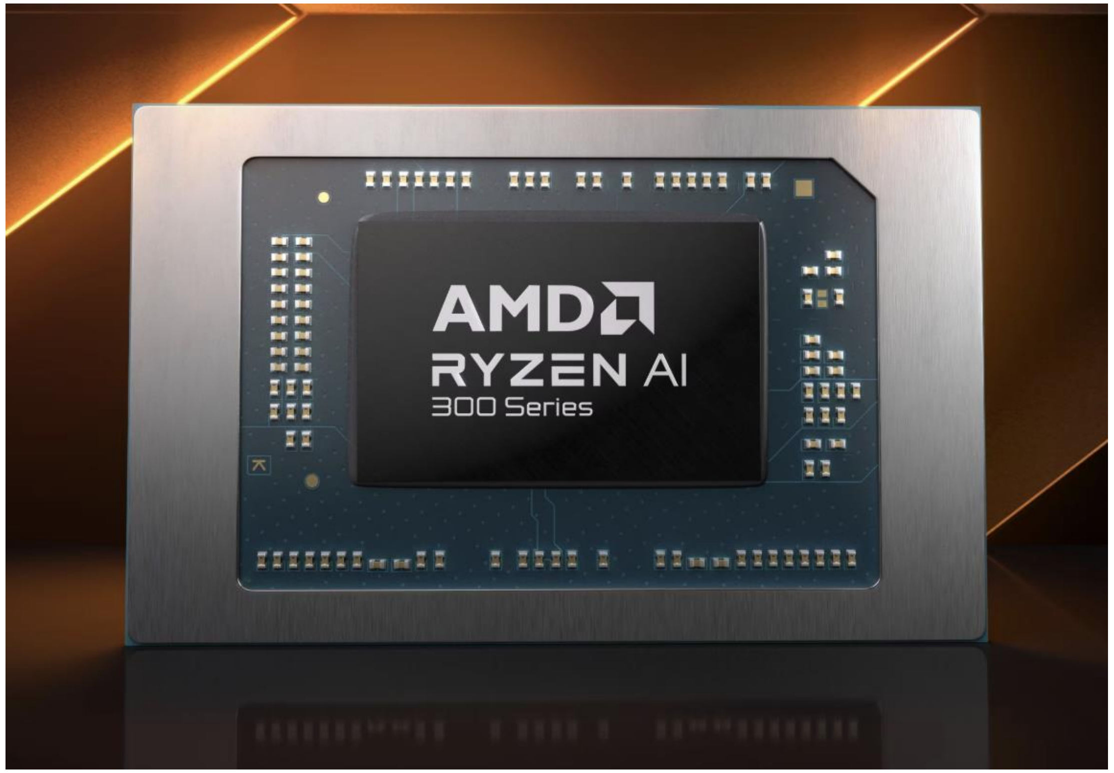
TLB and Pagewalk Performance in Multicore Architectures with Large Die-Stacked DRAM Cache

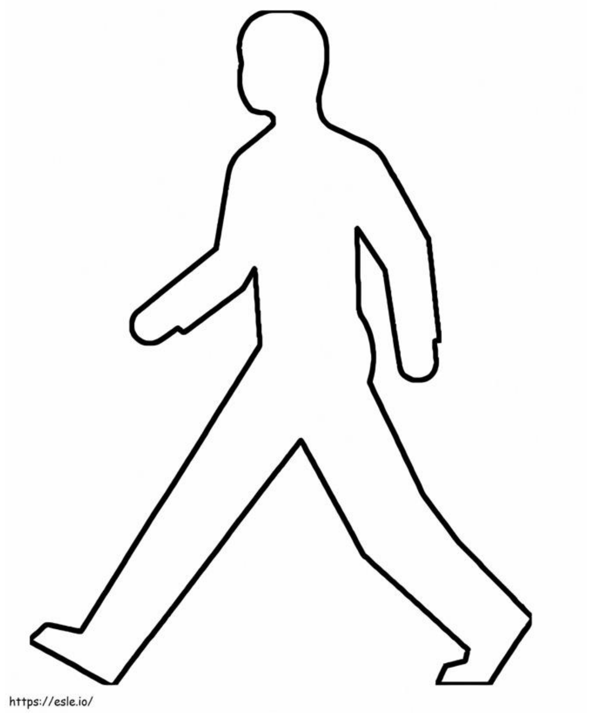
Adarsh Patil
Computer Science and Automation
Indian Institute of Sciences, Bangalore, IN
adarsh.patil@csa.iisc.ernet.in



(b) 2D page walk in Virtualization

Page Walking Caches (PWC)





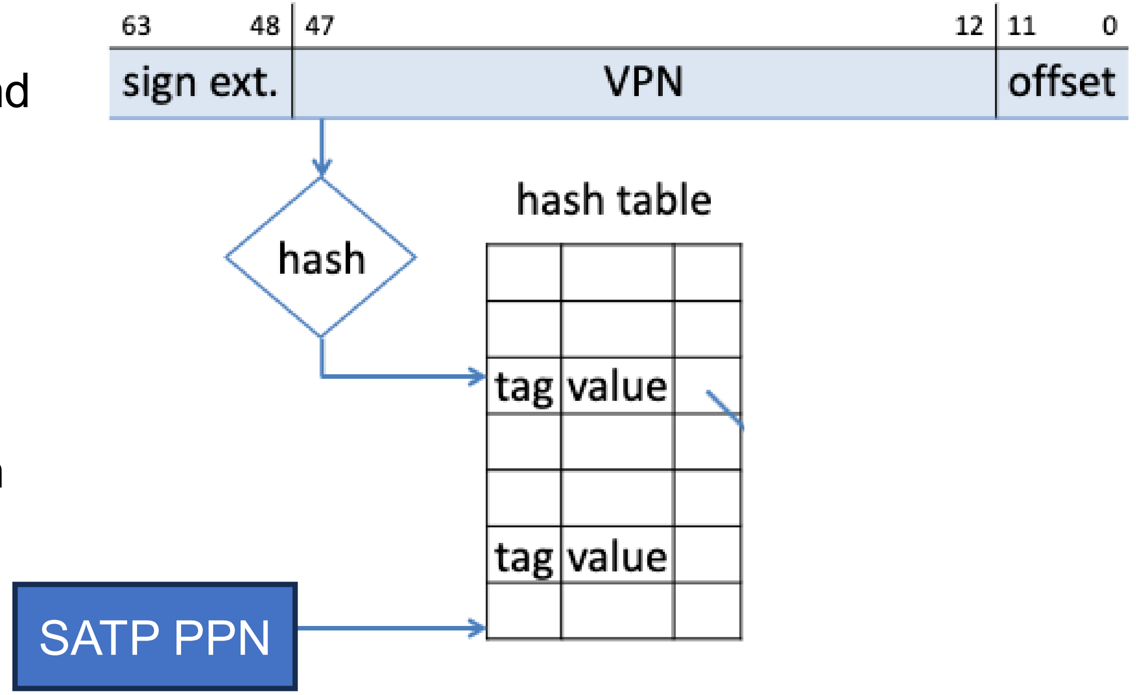
PWC is 16KB

Alternatives to Radix Structure

Hash Page Table

- The Virtual Page Number (VPN) is hashed to find an entry in the hash table

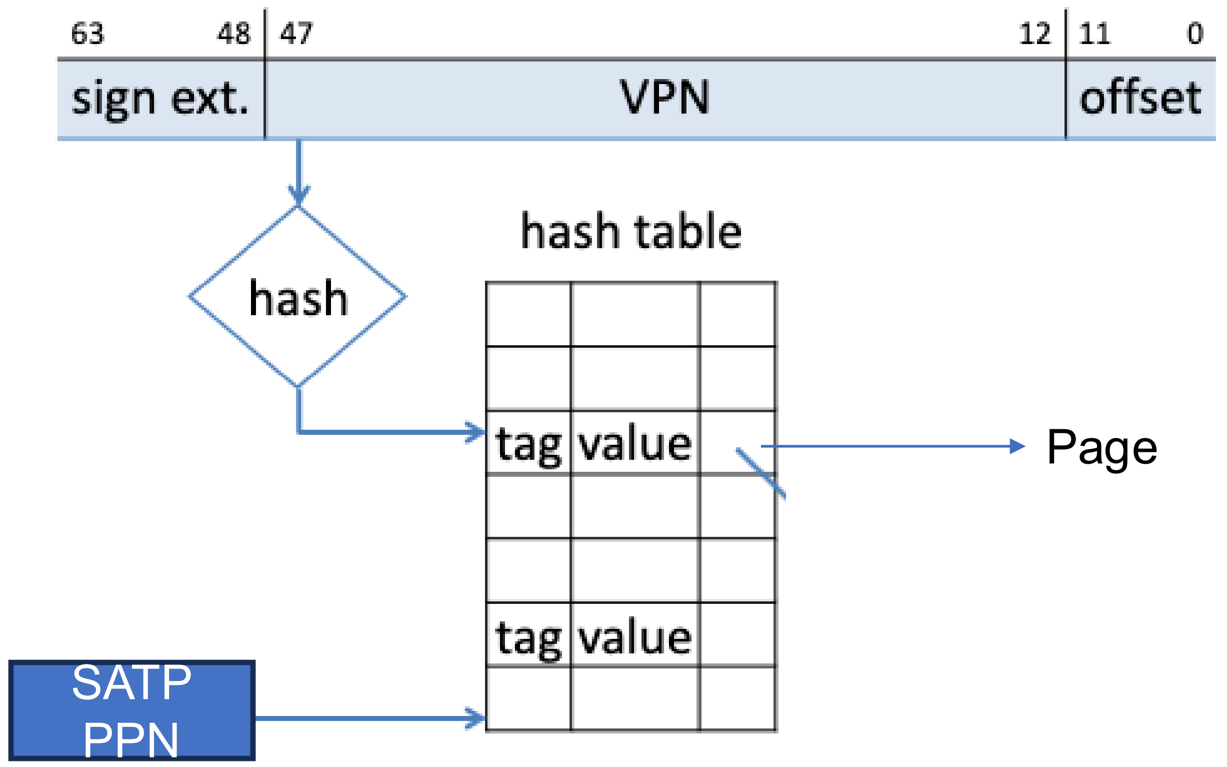
- One large page table is shared among all processors, and the Hash is used as an index in that table for the translation.



Traversing Hash Page Table

1) Hash VPN -> Position in Hash Table

2) Compare original VPN to tag
    a) VPN == tag
    b) VPN != tag



Radix vs Hash

	Radix	**Hash**

Overall, the literature mostly agrees that Radix is the winner!

Radix vs Hash Page Indexing

Perfect Hashing

One memory access

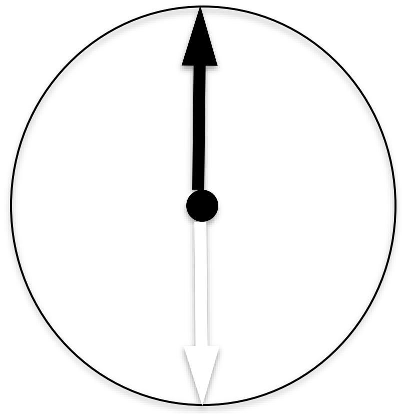
Radix Structure

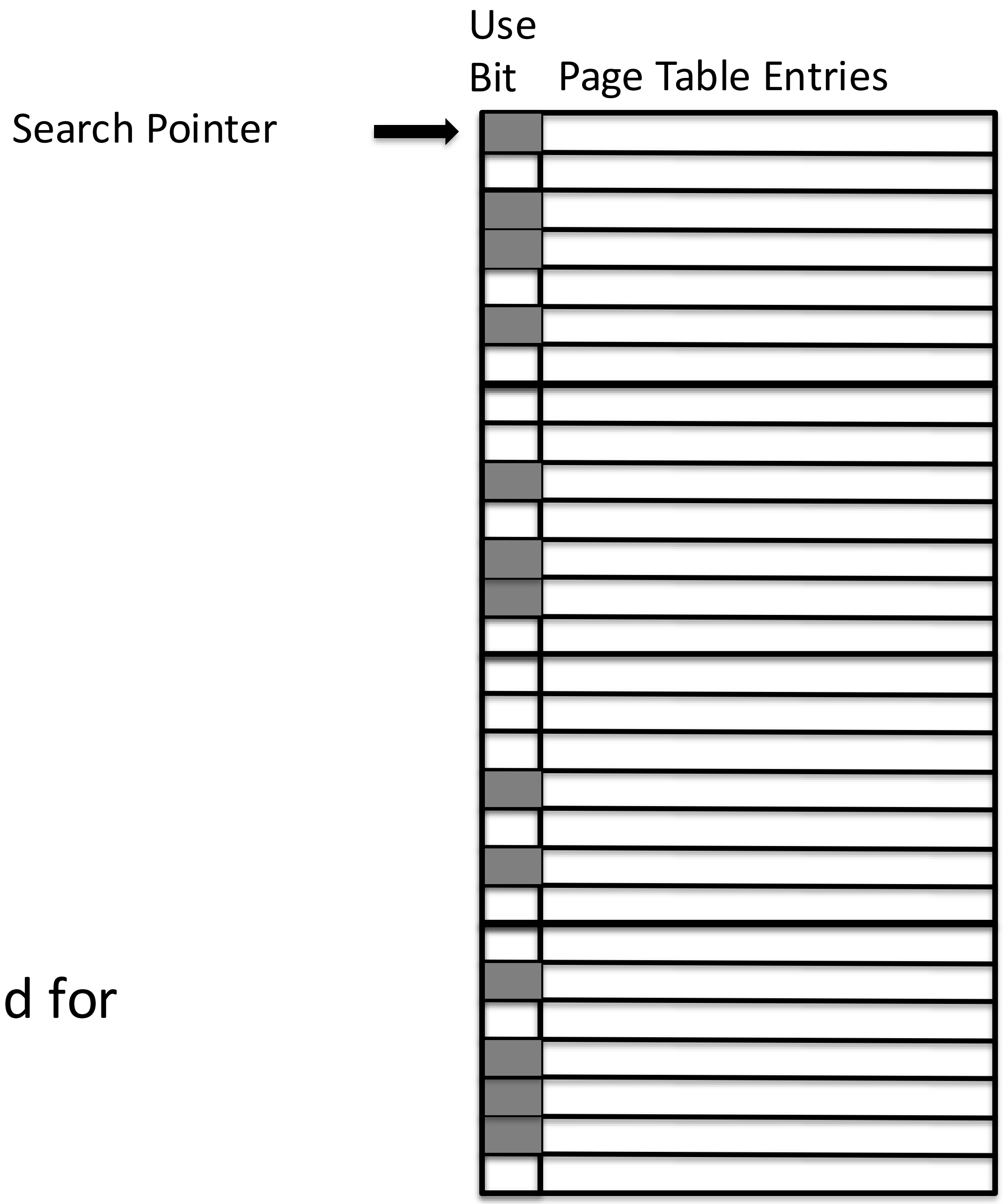
Guaranteed at least four memory accesses

Page Replacement

To reduce page fault rate, prefer least-recently used (LRU) replacement

- Reference bit (aka use bit) in PTE set to 1 on access to page
- Periodically cleared to 0 by OS
- A page with reference bit = 0 has not been used recently



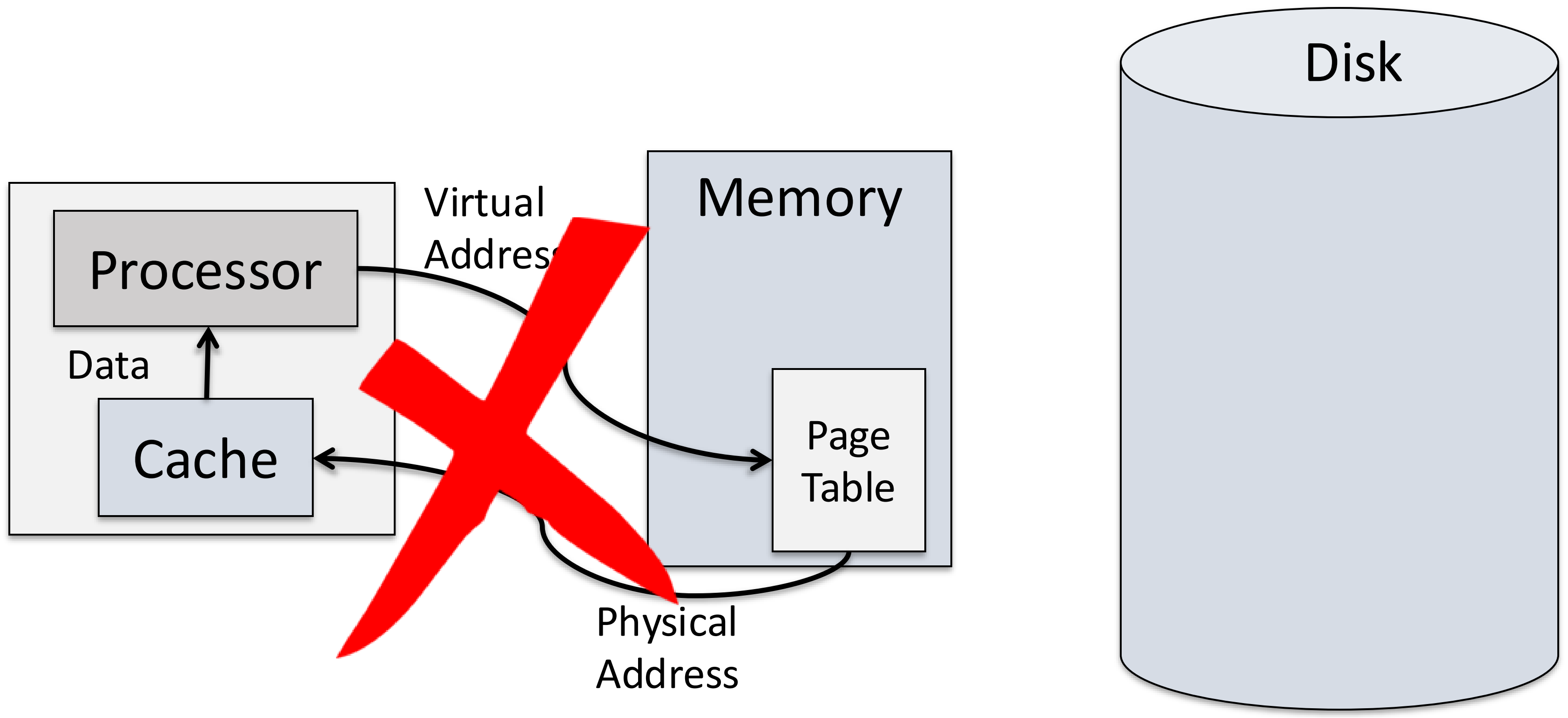


An LRU approximation used for page replacement

Writes

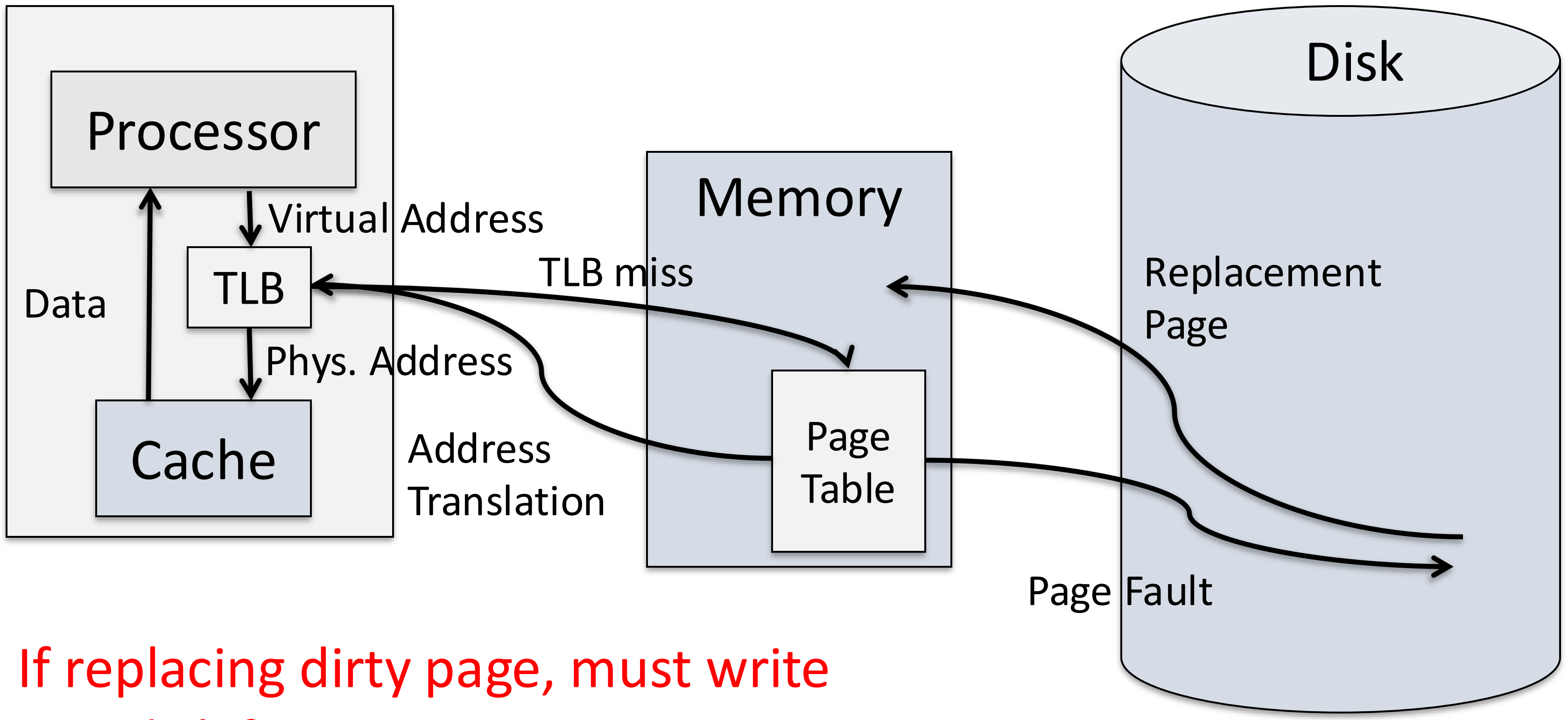
- Disk writes take millions of cycles
  - Block at once, not on individual locations
  - Write through is impractical
    - Use write-back

Do We Need an Extra Load for Each Memory Reference?



Translation-Lookaside Buffer (TLB)

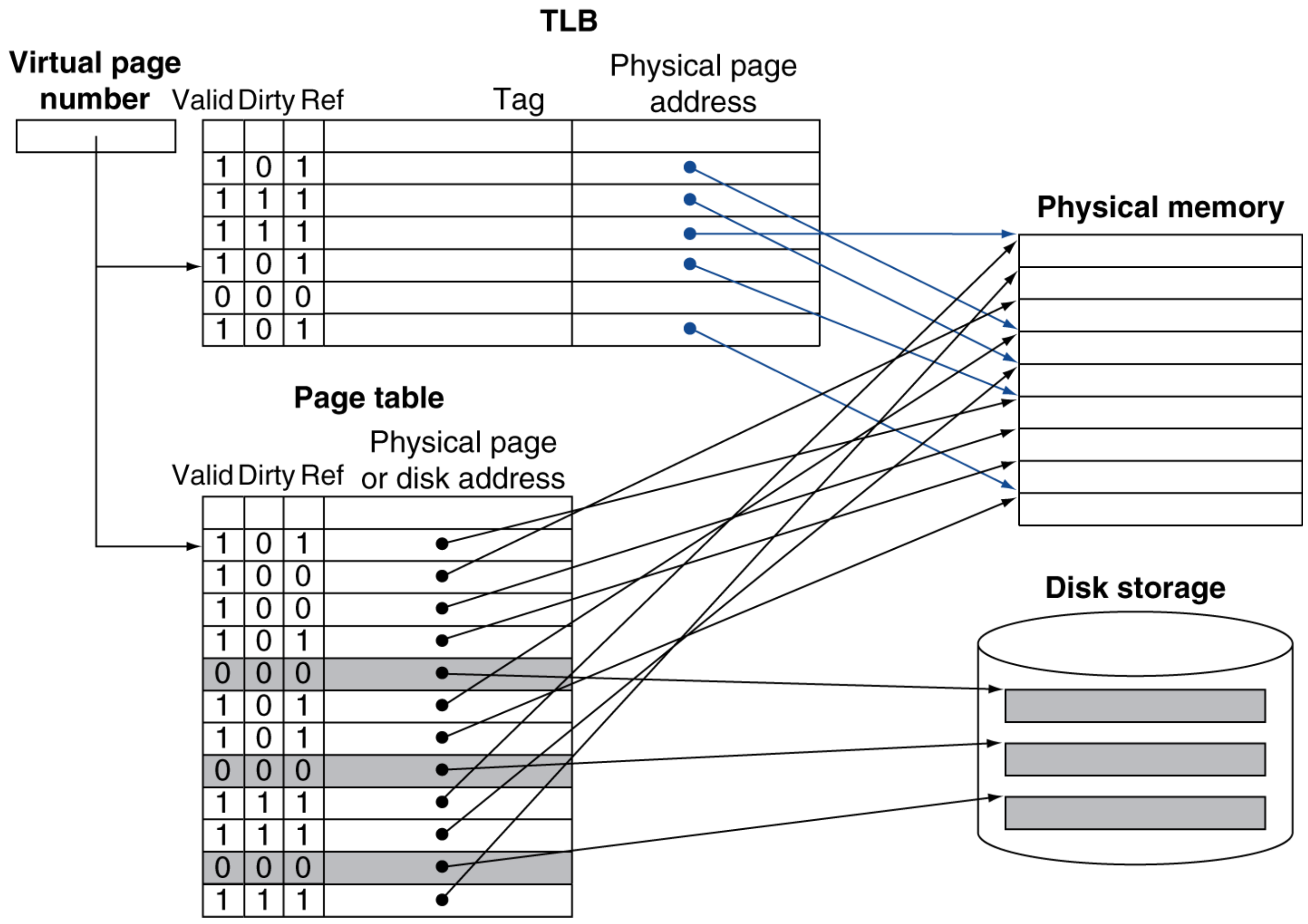
Page Fault is handled by OS



If replacing dirty page, must write it to disk first

Typical TLB: 16–512 Page Table Entries, 0.5–1 cycle for hit, 10–100 cycles for miss, 0.01%–1% miss rate

Fast Translation Using a TLB



TLB Misses

- If page is in memory
  - Load the Page Table Entry (PTE) from memory and retry
  - Could be handled in hardware
    - Can get complex for more complicated page table structures
  - Or in software
    - Raise a special exception, with optimized handler
- If page is not in memory (page fault)
  - OS handles fetching the page and updating the page table

TLB Miss Handler

TLB miss indicates either:

- Page present, but PTE not in TLB
- Page not present

Must recognize TLB miss before destination register overwritten

- Raise exception

Exception Handler copies PTE from memory to TLB

- Then restarts instruction

Page Fault Handler

Use faulting virtual address to find PTE

Locate page on disk

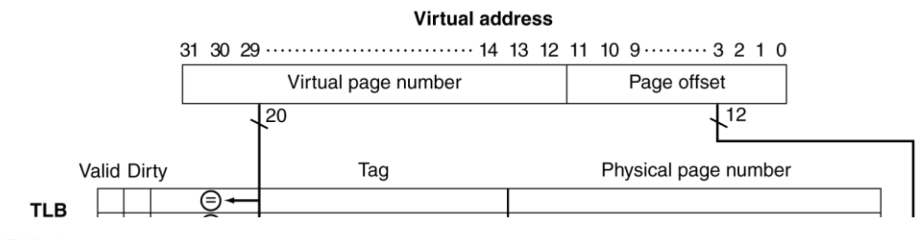
Choose page to replace

- If dirty, write to disk first

Read page into memory and update page table

Make process runnable again

TLB and Cache Interaction



If cache tag uses physical address

- Need to translate before cache lookup

Alternative: use virtual address tag

- Complications due to aliasing

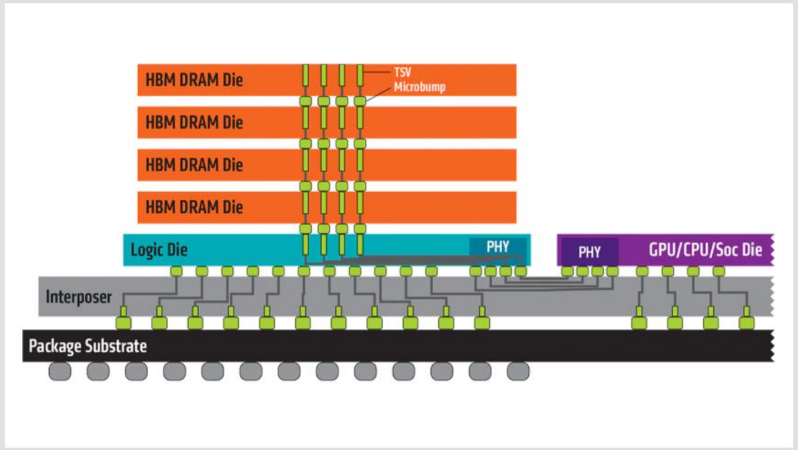
Memory Protection

Different tasks can share parts of their virtual address spaces

- But need to protect against errant access
- Requires OS assistance

Hardware support for OS protection

- Privileged supervisor mode (aka kernel mode)
- Privileged instructions
- Page tables and other state information only accessible in supervisor mode
- System call exception (e.g., ecall

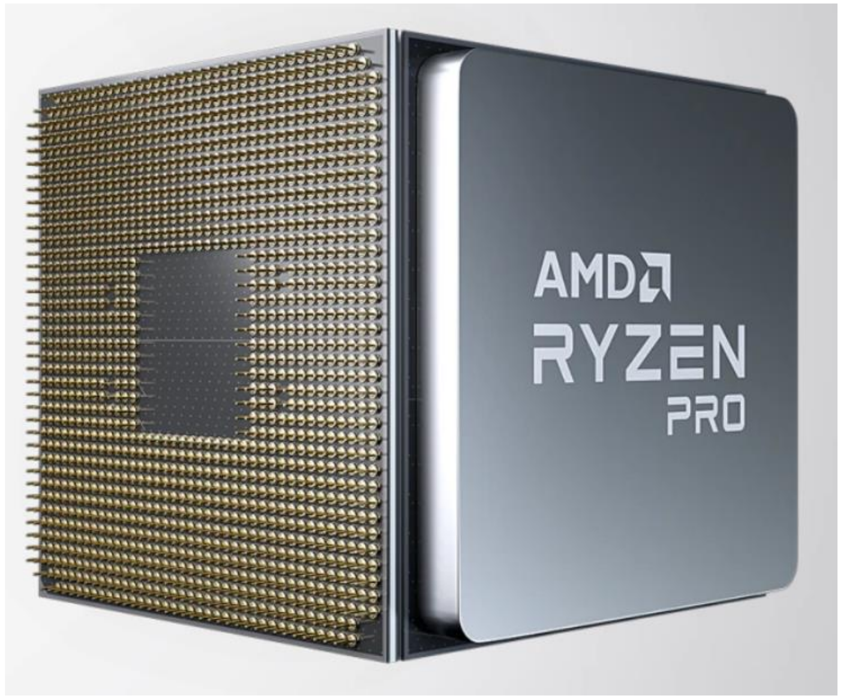


Revolutionary HBM breaks the processing bottleneck

HBM is a new type of CPU/GPU memory ("RAM") that vertically stacks memory chips, like floors in a skyscraper. In doing so, it shortens your information commute. Those towers connect to the CPU or GPU through an ultra-fast interconnect called the "interposer." Several stacks of HBM are plugged into the interposer alongside a CPU or GPU, and that assembled module connects to a circuit board.

Though these HBM stacks are not physically integrated with the CPU or GPU, they are so closely and quickly connected via the interposer that HBM's characteristics are nearly indistinguishable from on-chip integrated RAM.

2021



High Bandwidth Memory