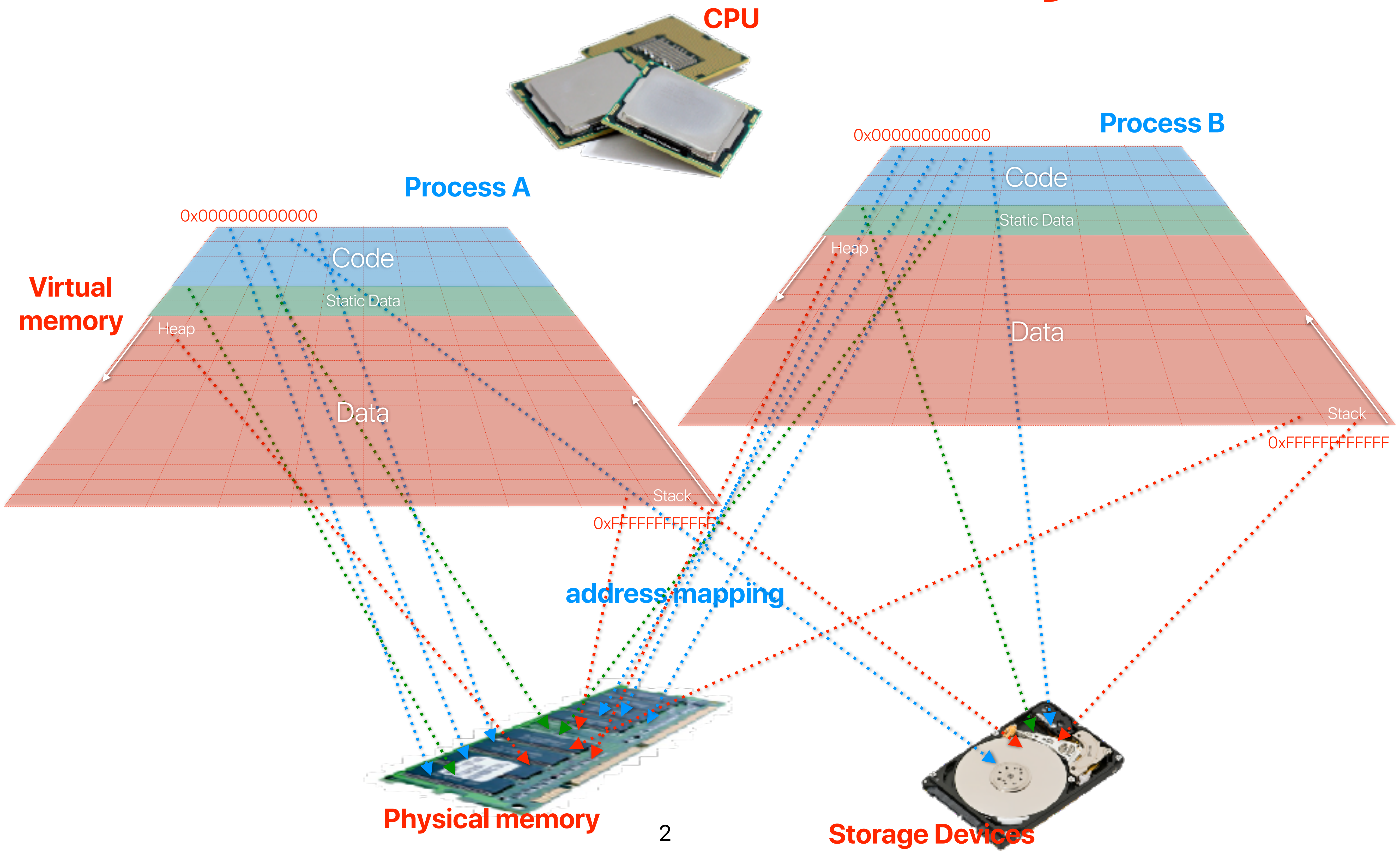
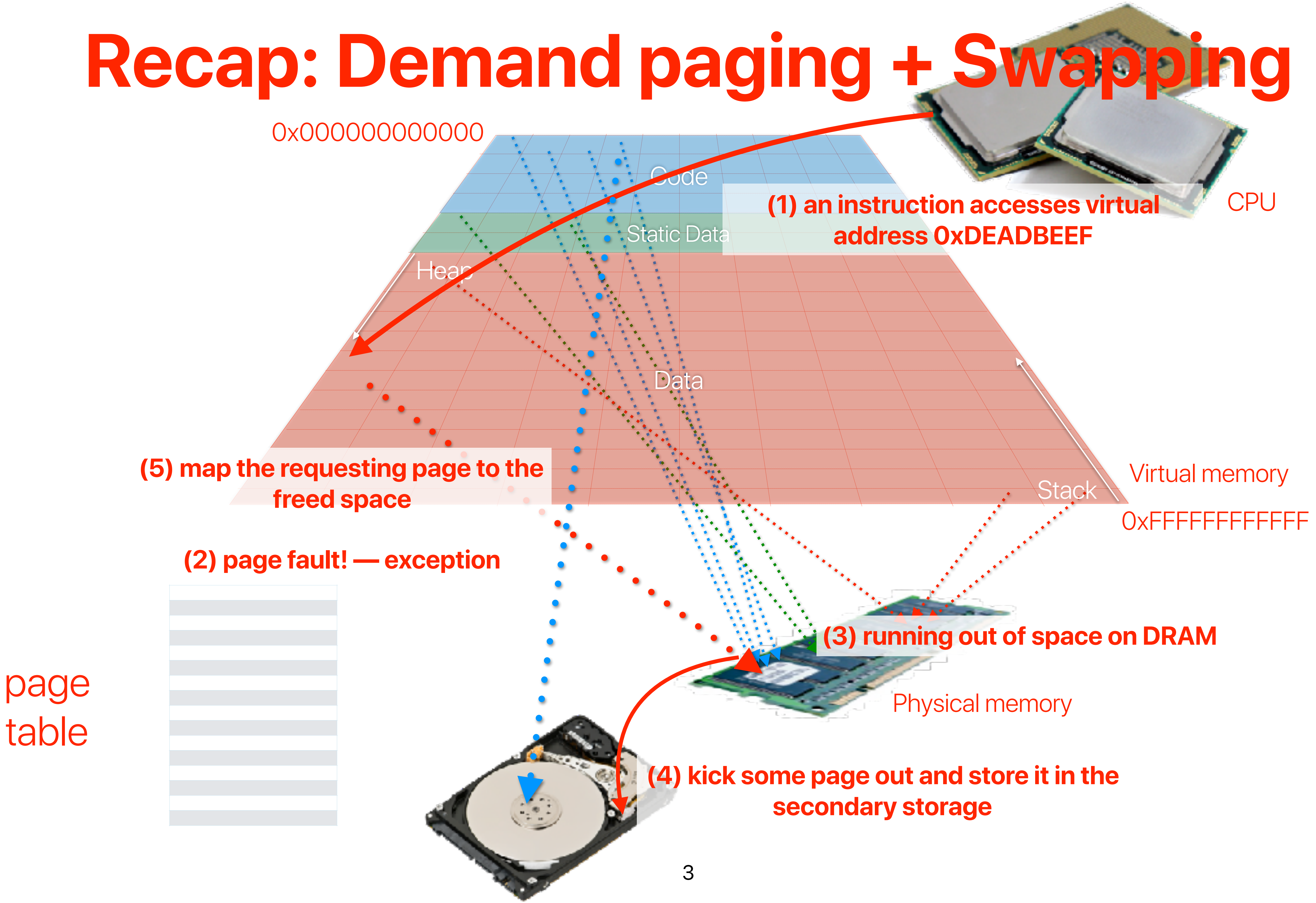
Virtual memory — policies

Hung-Wei Tseng

Recap: Virtual Memory



Recap: Demand paging + Swapping


0x000000000000
Code
Static Data
Heap
Data
Stack
CPU
(1) an instruction accesses virtual address 0xDEADBEEF
Virtual memory
0xFFFFFFFFFFFF
(5) map the requesting page to the freed space
(2) page fault! — exception
(3) running out of space on DRAM
Physical memory
page table
(4) kick some page out and store it in the secondary storage

Page replacement policy

- Goal: Identify page to remove that will avoid future page faults (i.e. utilize locality as much as possible)
- Implementation Goal: Minimize the amount of software and hardware overhead
  - Example:
    - Memory (i.e. RAM) access time: 100ns
    - Disk access time: 10ms
    - P_f: probability of a page fault
    - Effective Access Time = $10^{-7} + P_f * 10^{-3}$
  - When $P_f = 0.001$:
    Effective Access Time = 10,100ns
  - **Takeaway: Disk access tolerable only when it is extremely rare**

What VAX/VMS proposed to achieve these goals?

- Considering the optimization goals and the proposed VAX/VMS mechanisms, which of the following combinations is incorrect?

	Goal		Optimization	
[✓ A]	Process startup cost	W	Demand-zero & copy-on-refernce	
[✓ B]	Process performance interference	X	Process-local replacement	
C	Page table lookup overhead	Y	Page clustering	also helps reduce disk loads
[✓ D]	Paging load on disks	Z	Page caching	

The impact of VAX/VMS

- We’re still using their proposed techniques almost everyday!
- It’s basically the baseline UNIX VM design

Outline

- Page replacement policies
- Page replacement policy once used in UNIX: Converting a Swap-Based System to do Paging in an Architecture Lacking Page-Reference Bits
- Another popular page replacement policy: WSClcok - A Simple and Effective Algorithm for Virtual Memory Management
- Machine-Independent Virtual Memory Management for Paged Uniprocessor and Multiprocessor Architectures

Swapping policies

Page replacement policy

- We need to determine:
  - Which page(s) to remove
  - When to remove the page(s)
- Goals
  - Identify page to remove that will avoid future page faults (i.e. utilize locality as much as possible)
  - Minimize the amount of software and hardware overhead

Page replacement algorithms

- FIFO: Replace the oldest page
- LRU: Replace page that was the least recently used (longest since last use)

FIFO v.s. LRU

	FIFO	LRU
Implementation	Easy — circular queue	May require hardware support or linked list or additional timestamps in page tables
Execution overhead	Low	High — you need to manipulate the list or update every counter
Performance	Usually not as good as LRU	Usually better than FIFO

Converting a Swap-Based System to do Paging in an Architecture Lacking Page-Reference Bits

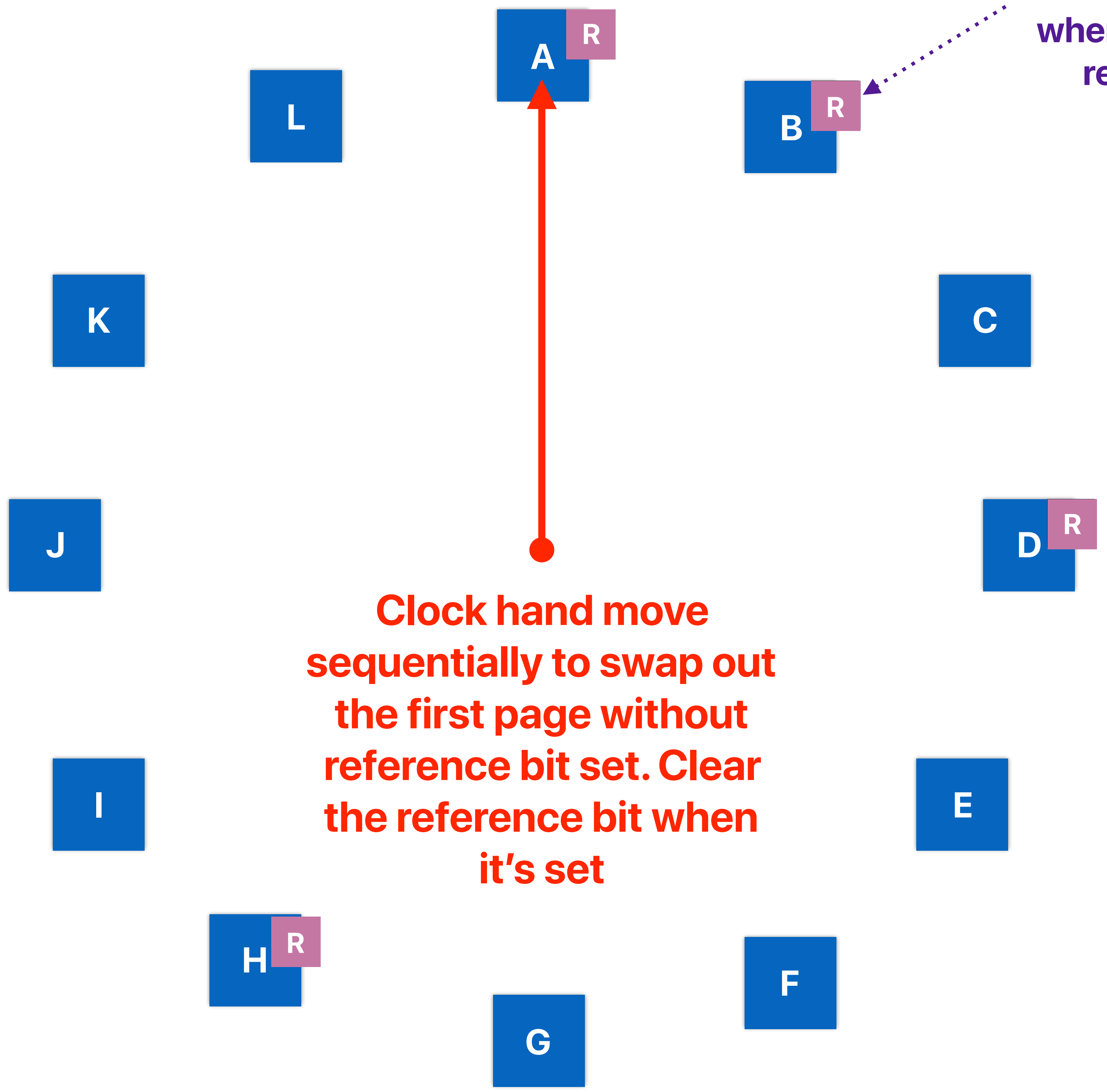
Özalp Babaoglu and William Joy*

Cornell University and University of California, Berkeley

The Why of Babaoglu new UNIX VM

- The original UNIX is a “swap-based” system
  - Whenever you have a context switch, swap the whole process out from the memory
  - Really inefficient if you have frequent context switches or if you have many applications in-fly
- Efficient page replacement policies and other virtual optimization techniques cannot be implemented easily without appropriate hardware support

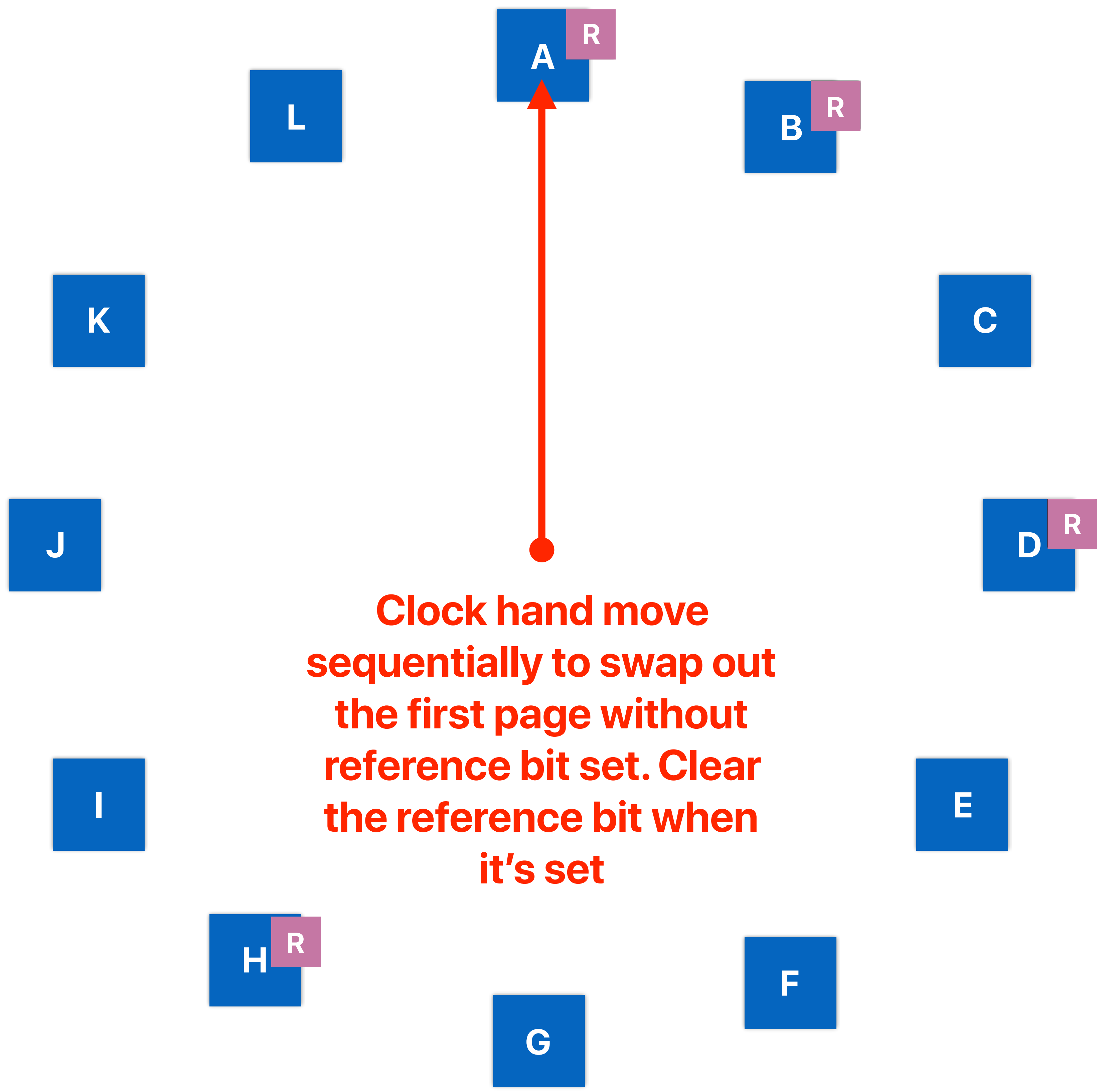
Clock algorithm



attach a "reference bit" to each PTE, set to true when the page is referenced

Clock hand move sequentially to swap out the first page without reference bit set. Clear the reference bit when it's set

Clock algorithm in motion

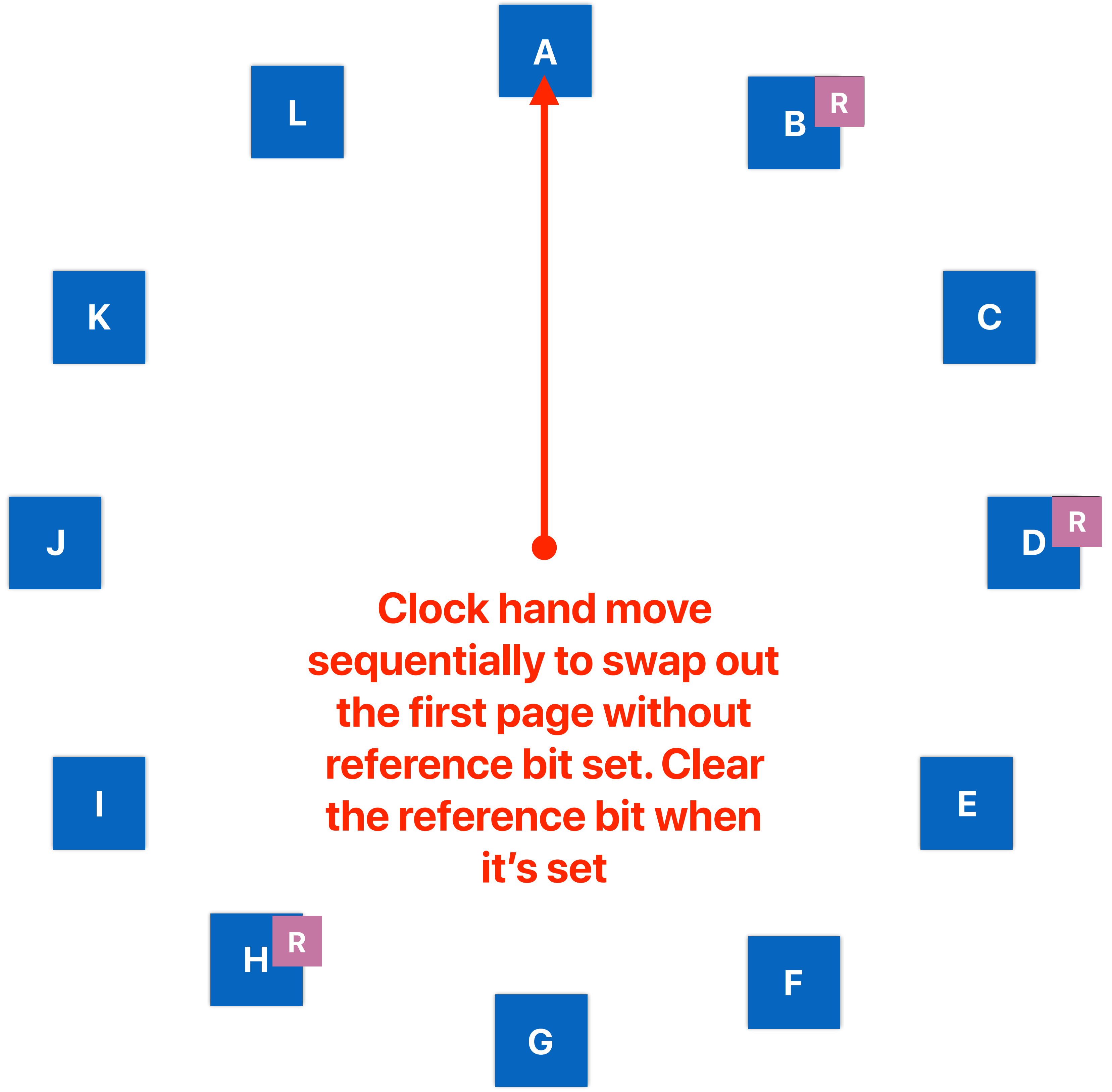


Where to put M ?

Clock hand move sequentially to swap out the first page without reference bit set. Clear the reference bit when it's set

Clock algorithm in motion

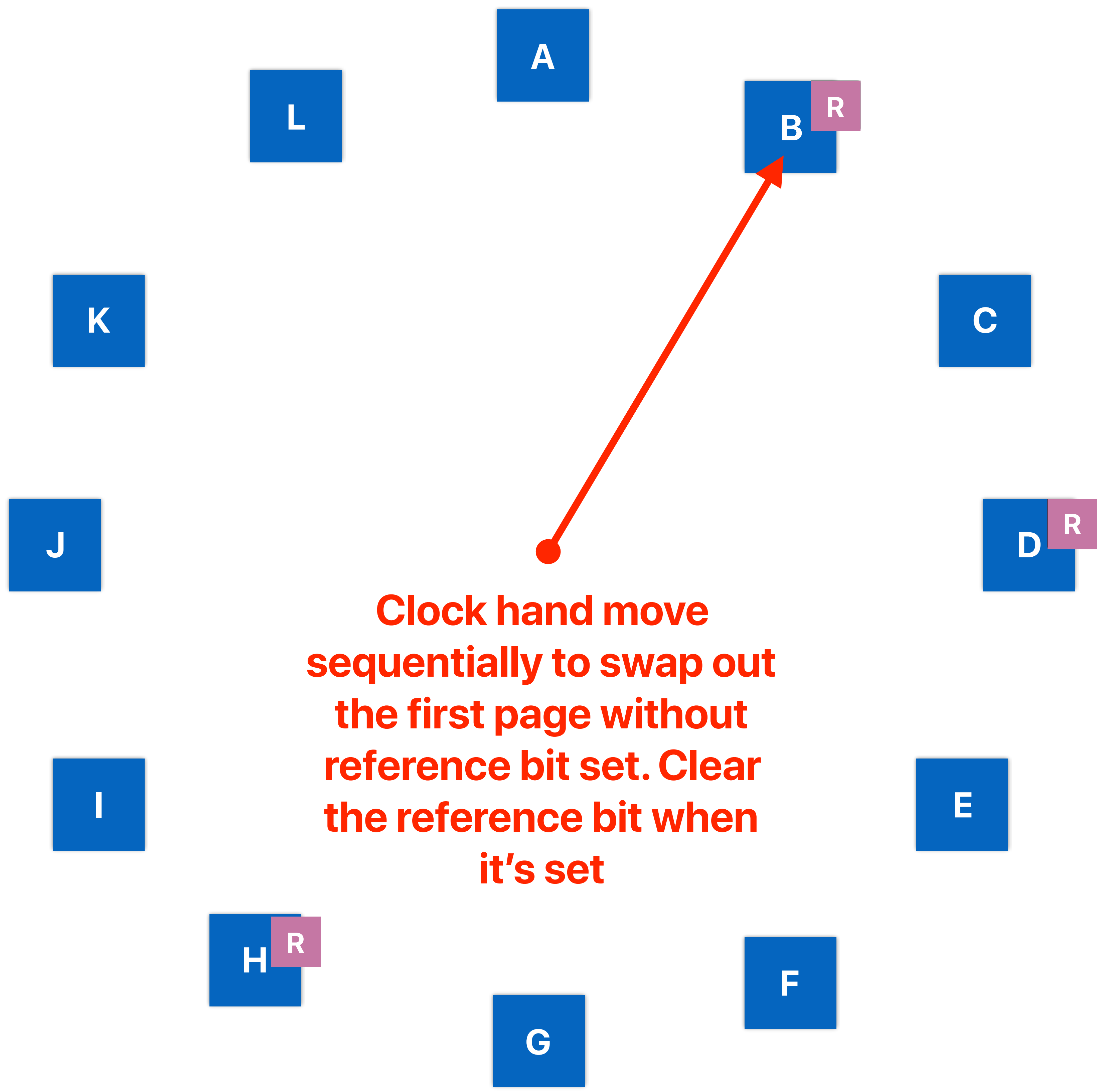
Where to put M ?



Clock hand move sequentially to swap out the first page without reference bit set. Clear the reference bit when it's set

Clock algorithm in motion

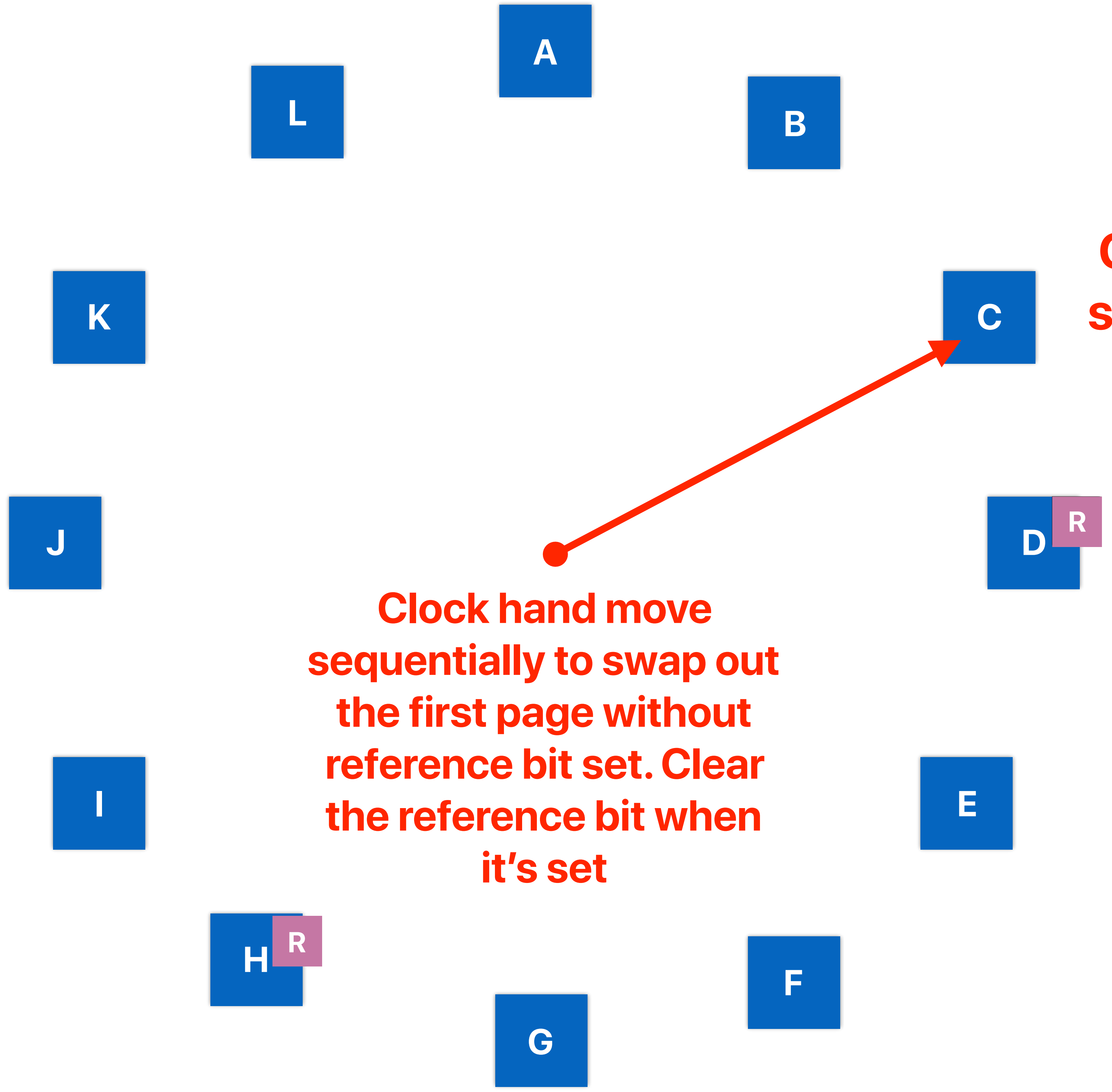
Where to put M ?



Clock hand move sequentially to swap out the first page without reference bit set. Clear the reference bit when it's set

Clock algorithm in motion

Where to put M ?



C will be selected to swap out, but Rs of A and B are cleared

Clock hand move sequentially to swap out the first page without reference bit set. Clear the reference bit when it's set

Free list

- So far, we need to trigger clock policy and swap in/out on each page fault
- Why don't we prepare more free pages each time so that we can feed page faults with pages from the list?
- Free list
  - When we need a page, take one from the free list
  - Have a daemon running the background, managing this free list — you can do this when system is not loaded
  - If size of free list gets too small, trigger the clock algorithm to add pages into the free list (by swapping out to disk)
  - Free list can be used as a disk cache

WSClcok - A Simple and Effective Algorithm for Virtual Memory Management

Richard Carr and John Hennessy

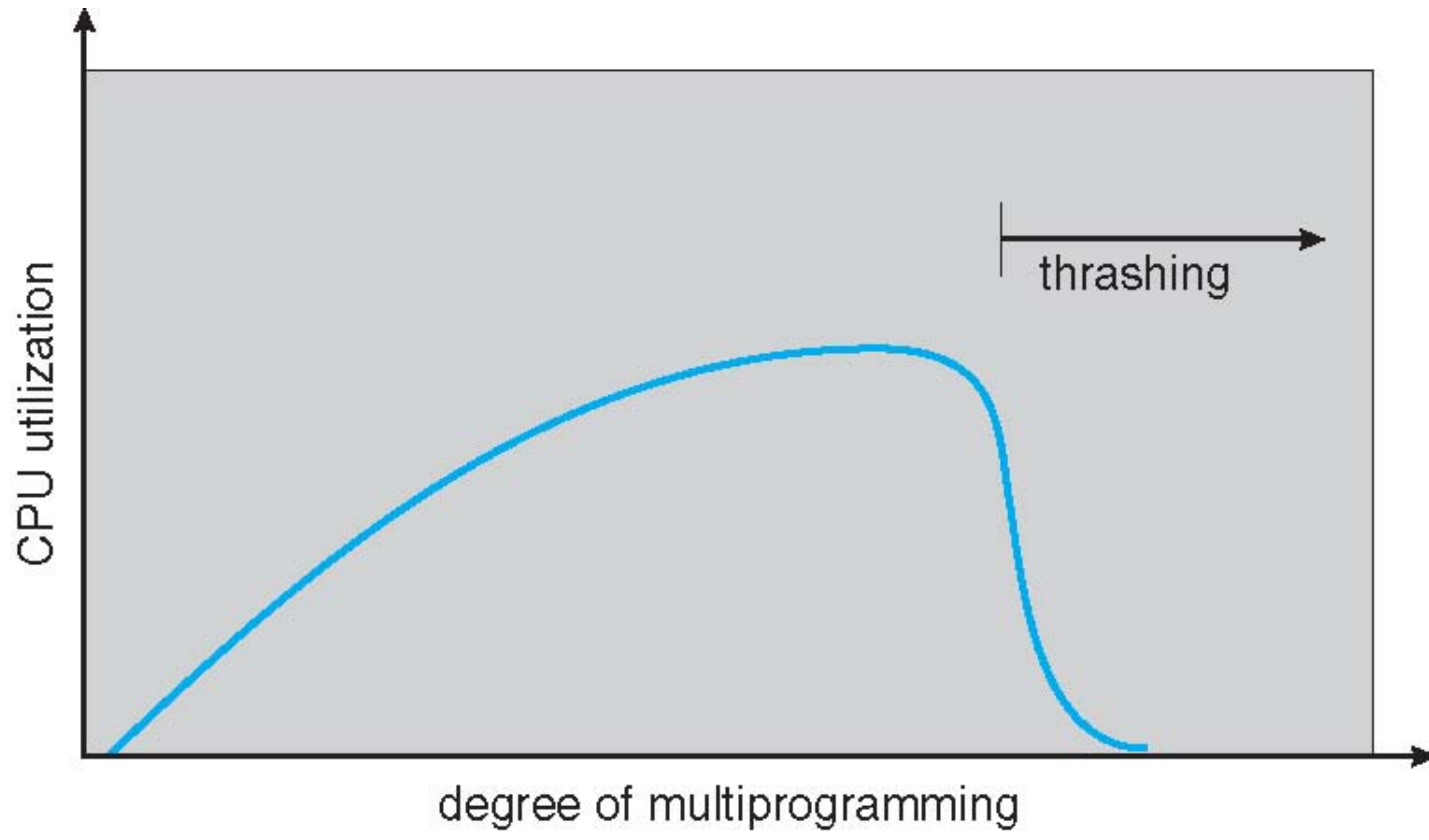
What policies are used?

- Local: select one page from the same process’ physical pages for storing the demanding page when swapping is necessary
  - VAX/VMS
  - Original UNIX
- Global: select any page that was previously belong to any process when swapping is necessary
  - UNIX after Babaoglu
  - Mach

Thrashing!

- The system overcommitted memory to tasks
- The system spends most time in paging, instead of making meaningful progress

Previously, we have seen how scheduling policies can help improving "saturation". Now, let's see how page replacement policies can address this "thrashing"



Degree of parallelism and performance

- How many of the following would happen in Babaoglu’s UNIX VM if we keep increase the amount of concurrent processes in the system?
  ① The CPU utilization will keep increasing and stay at 100%
  ② The system may spend more time in context switching than real computation
  ③ The system may spend more time in swap in/out than real computation
  ④ Some process may not respond due to the high page overhead
  A. 0
  B. 1
  C. 2
  D. 3
  [✓ E. 4]

Why WS-Clock

- Take advantages from both local and global page replacement policies
  - Global — simplicity, adaptive to process demands
  - Local — prevent thrashing

Working Set Algorithm

- Working set: the set of pages used in a certain number of recent accesses
- Assume these recently referenced pages are likely to be referenced again soon (temporal locality)
- Evict pages that are not referenced in a certain period of time
  - Swap out may occur even if there is no page faults
- A process is allowed to be executed only if the working set size fits in the physical memory

WSClock

- Use **working set** policy to decide how many pages can a process use
  - Return a page to the free list if there exists a page in the process' working set that hasn't been access for a certain period of time
- If the free list is lower than a threshold
  - Trigger the **clock** policy to select pages from any process
- On a page fault
  - Take a page from the free list

WSClock

- Wherever you need to reclaim a page —
    1. Examine the PTE pointed to by clock hand.
    2. If reference bit is set
        1. Clear reference bit;
        2. Advance clock hand;
        3. Goto Done.
    3. If reference bit is not set
        1. If the timestamp of the PTE is older than a threshold
            1. Write the page to disk if it’s dirty and use this page
            2. Goto Done
        2. Otherwise
            1. Advance clock hand
            2. Goto 1.
    4. Done
    5. If no victim page is chosen, randomly pick one

The impact of WSClock

- One of the most important page replacement policies in practice

Announcement

- Reading quiz due next Tuesday
- Project due 3/3
  - We highly recommend you to fresh install a Ubuntu 16.04.6 Desktop version within a VirtualBox
    - Virtual box is free
    - If you crash the kernel, just terminate the instance and restart virtual box
  - Use office hours to discuss projects