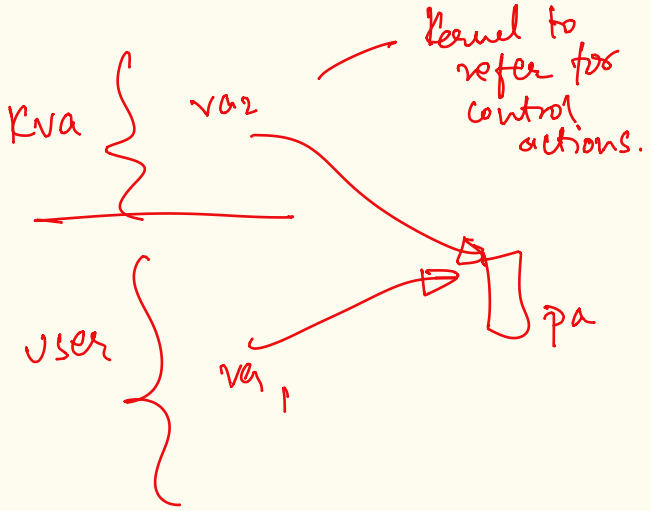
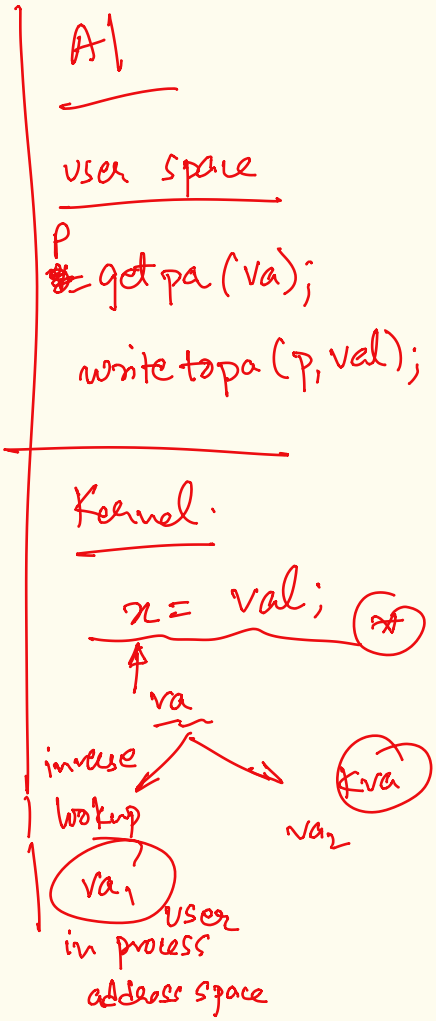
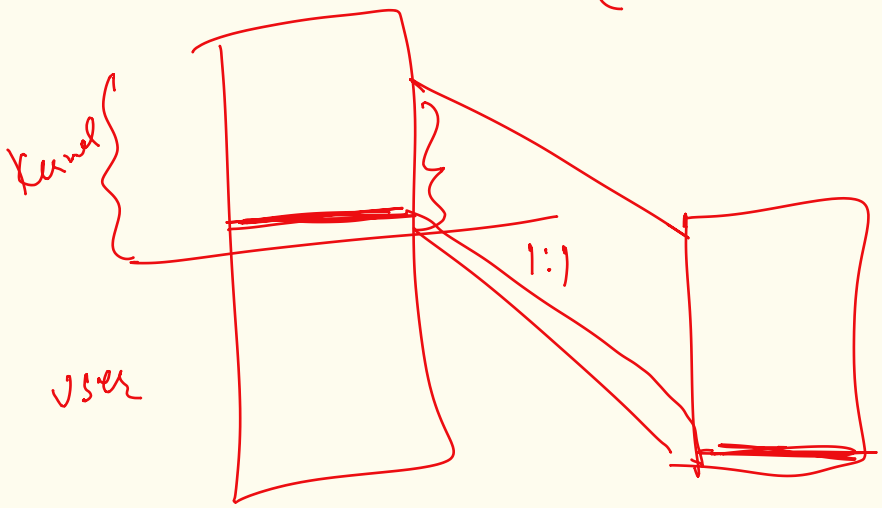
Lecture 5 CS 695

CPU virtualization for VMs. with

recap: trap-and-emulate
binary translation / trap-and-emulate.







~~(Kva - OFFSET) : pa~~

Kva : Kva - OFFSET
VA PA

virt 2 phys Kva → pa
pa → Kva

Design 3: paravirtualized approach for CPU virt.

(*) Guest OS knows about the VMM/VM abstraction.

⇒ not issue problem instructions!

⇒ enabled via a mechanism: hypercalls.

hypercall: similar to the system call interface

⇒ interface between Guest OS the hypervisor

(abstraction) VM ↑ provider of abs.

⇒ explicit invocation.

int 0x80 — x86 syscall

int 0x82 — hypercall (Xen

idea: on any critical work

anything that needs hypervisor intervention

explicitly make a hypercall.

Guest va → pa ?

0x2000 → 0xAfffabcd

?

e.g.: mmu_update (va, pa, pte details)

representative hypercall.

pa ✗

① trying to update page table.

② setup the IDT.

G&P 1974

needs Guest OS changes.

— PV'ed OS

efficiency — ↑

control — ✓

equivalence — ??

Design 4 : hardware-assisted virtualization support for the CPU.

(*) h/w architecture redesign/upgrade to support system-level virtualization.

Key idea: programmatic control/config. of traps || exits

- x86 ISA

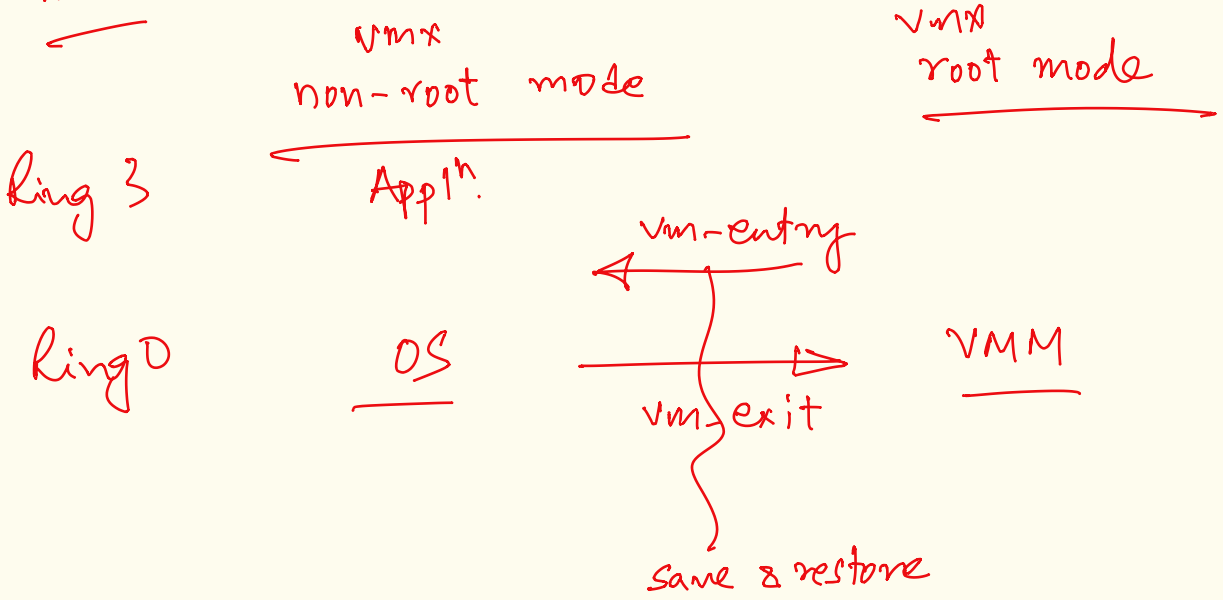
(i) vmx modes of operation

(ii) vmx instructions

(iii) vmx state (context & config.)

- vmx modes.

(i)



VMCS : virtual machine control structure

4KB Object in-memory

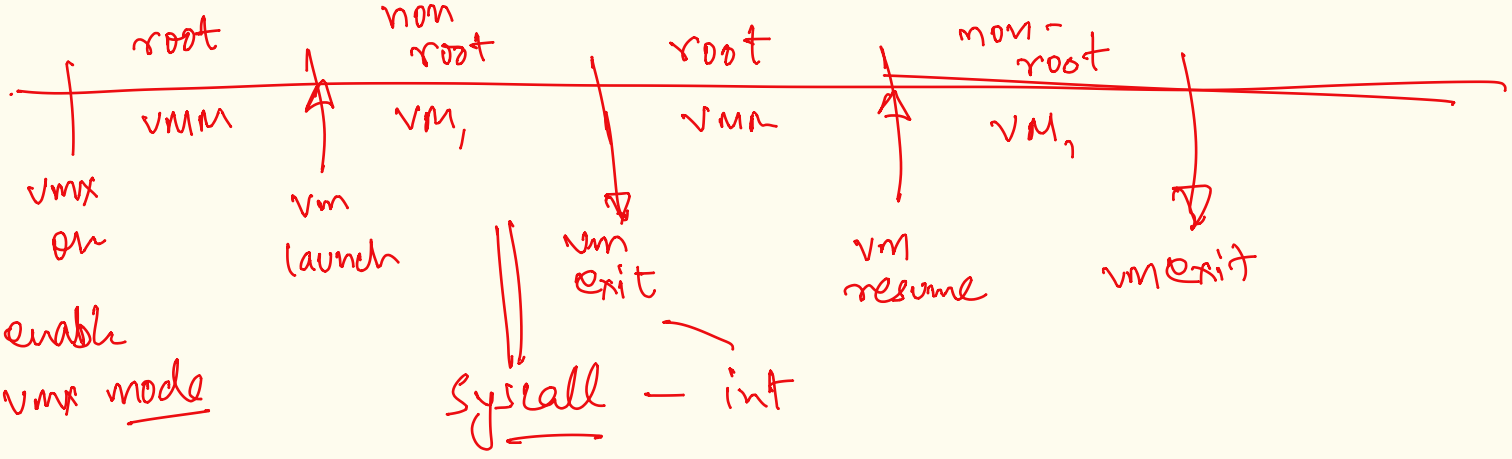
- guest state area — all guest execution state, CPU regs. saved on exit, restored on entry
- host area — save VMM's execution status — saved on entry to VM, used on exit from VM
- vm execution control — behaviour in non-root mode (exit conditions)
- vm entry control fields — what state to restore on vm entry — interrupt injection.
- vm exit control fields — what state to restore on exit
- vm exit information. — exit status, exit details. (?)

How new privileged inst^n.

vmread vmwrite vmclear, vmptrld

vmx on vmx off vm launch { vm resume, vmexit } — ISA

— per-CPU vmcs ptr. (regs).



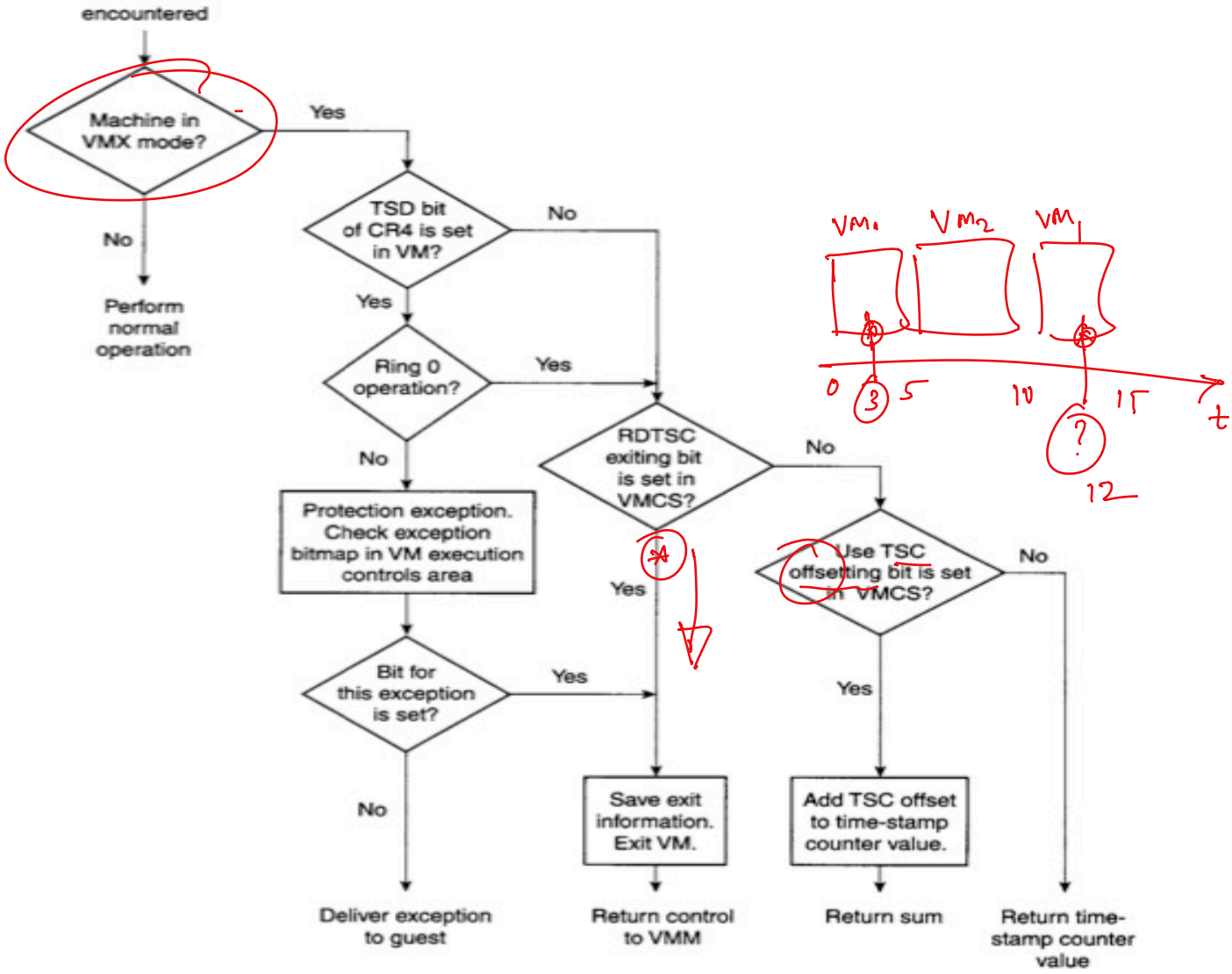


Figure 8.27 Actions Taken by Hardware When a Read Time-Stamp Counter (rdtsc) Instruction Is Encountered.