

Complex Verification Example: RISC-V MMU Verification of Virtualization and Hypervisor Operation for CPU and SOC platforms

Verification Futures Reading 2025 David Kelf, CEO, Breker Verification Systems

Agenda



- Why do we need MMUs ?
- Virtual Memory and Page Table Walks
- RISCV ISA MMU & Hypervisor Specification
- RISCV MMU & Hypervisor Test Plan
- Example MMU Test Cases
- Debug, Coverage and Deployment

Agenda

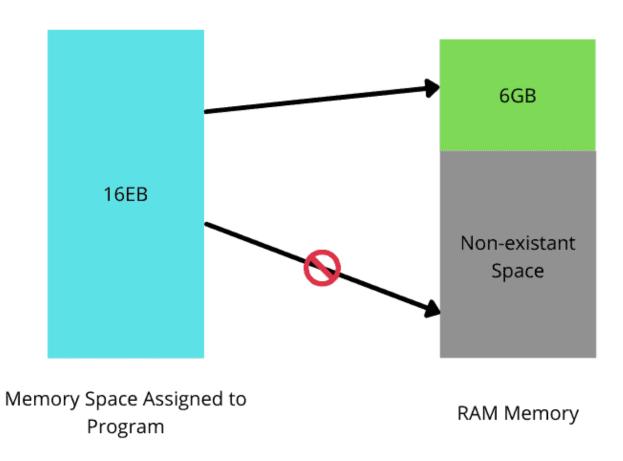


- Why do we need MMUs?
- Virtual Memory and Page Table Walks
- RISCV ISA MMU & Hypervisor Specification
- RISCV MMU & Hypervisor Test Plan
- Example MMU Test Cases
- Debug, Coverage and Deployment

Motivation for MMU (1/3)



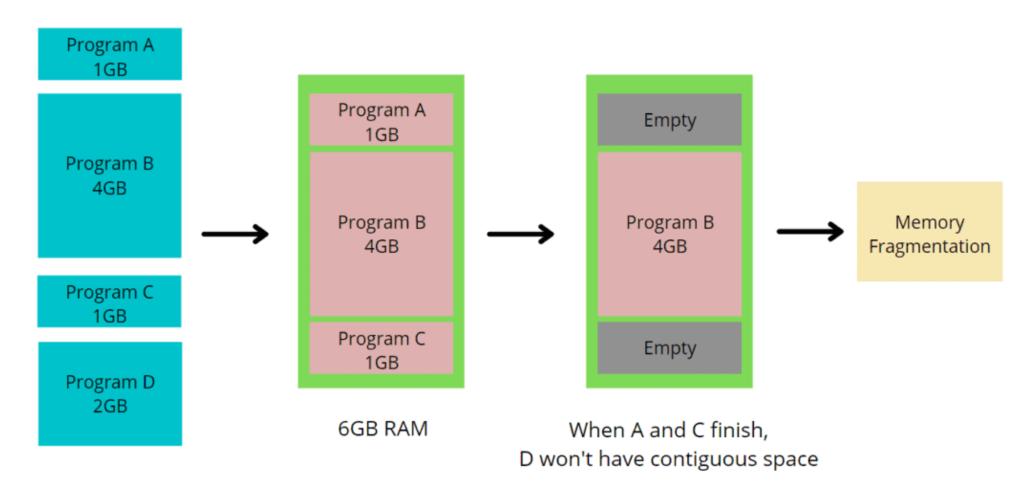
Allow software to use more memory than physically available



Motivation for MMU (2/3)



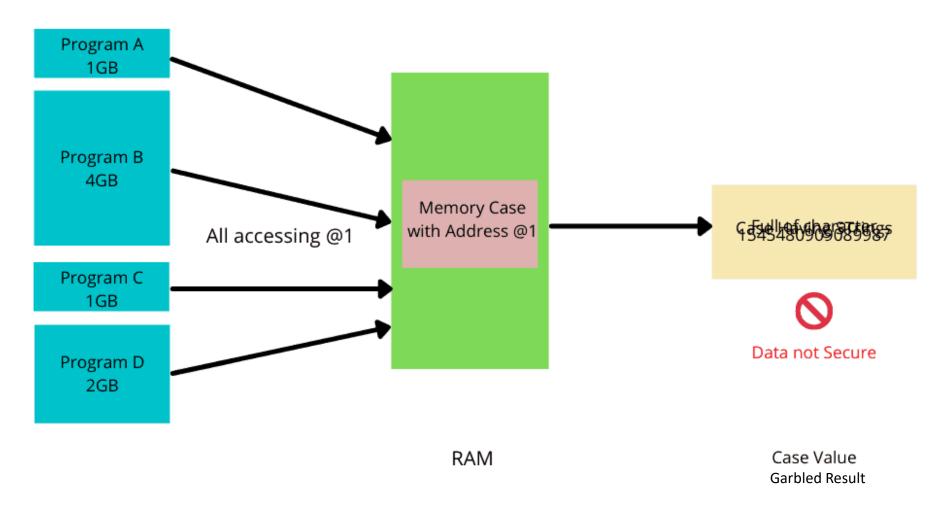
Handle Memory Fragmentation



Motivation for MMU (3/3)



Isolate memory accesses across processes



Agenda

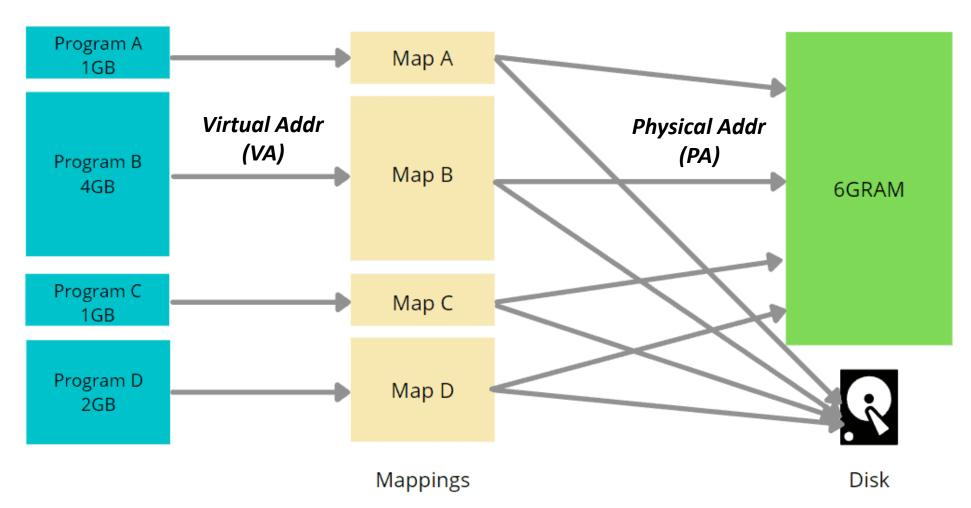


- Why do we need MMUs ?
- Virtual Memory and Page Table Walks
- RISCV ISA MMU & Hypervisor Specification
- RISCV MMU & Hypervisor Test Plan
- Example MMU Test Cases
- Debug, Coverage and Deployment

Virtual Memory Mapping



Each program sees a virtual address (VA) that is mapped to a physical address (PA)

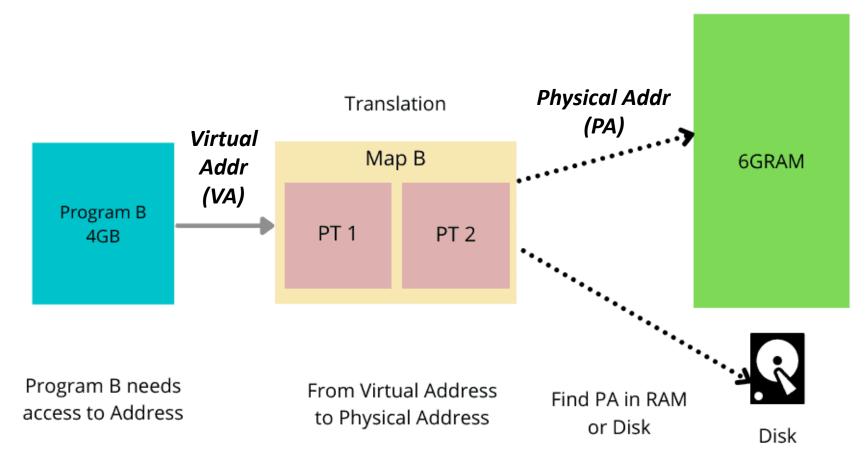


Virtual to Physical mapping via Page Tables



Each program gets a multi-stage page table lookup

Having multiple levels of page tables reduces overall memory usage

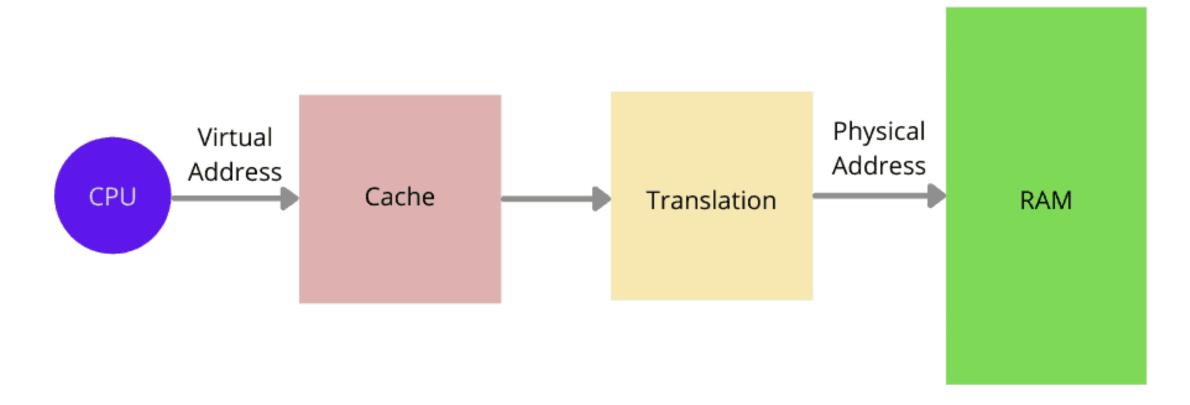


Virtual to Physical mapping via Page Tables



TLB Cache store a copy of previous table walks to improve performance

Software must flush appropriate entries from TLB when page mapping is changed



One Stage Address Translation Page Table Walks



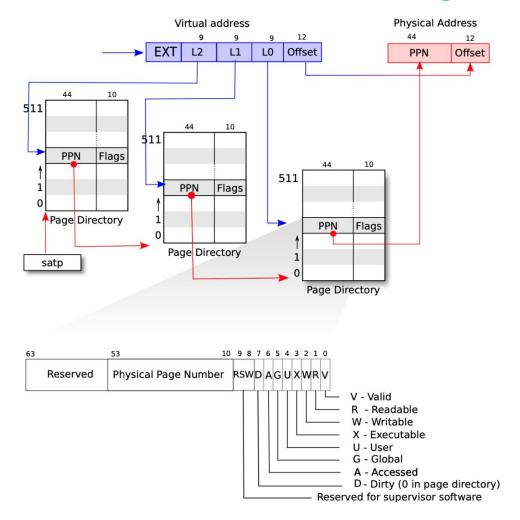


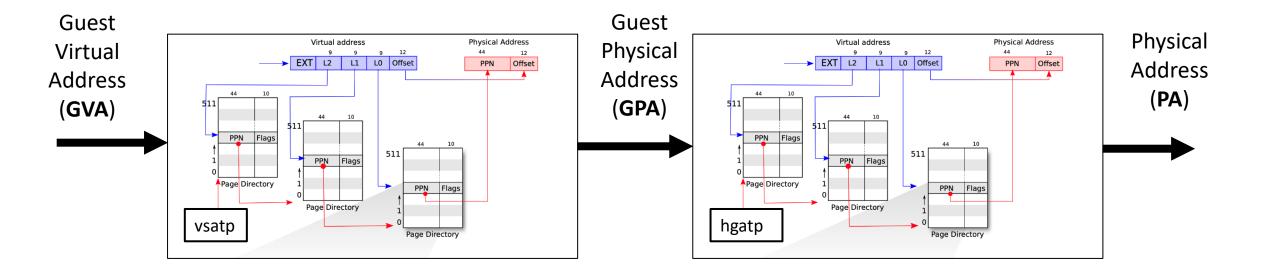
Figure 3.2: RISC-V address translation details.

- Build up sparse tree of page directories
- S/U mode

Attribution: https://clownote.github.io/2021/03/06/xv6/Xv6-page-table/

Hypervisor and Two-Stage Address Translation





virtual-VS/VU-mode

Attribution: https://clownote.github.io/2021/03/06/xv6/Xv6-page-table/

Agenda



- Why do we need MMUs ?
- Virtual Memory and Page Table Walks
- RISCV ISA MMU & Hypervisor Specification
- RISCV MMU & Hypervisor Test Plan
- Example MMU Test Cases
- Debug, Coverage and Deployment

Vignettes from the RISCV ISA: satp CSR



3.1.11. Supervisor Address Translation and Protection (satp) Register

The satp CSR is an SXLEN-bit read/write register, formatted as shown in Supervisor address translation and protection (satp) register when SXLEN=32. for SXLEN=32 and Supervisor address translation and protection (satp) register when SXLEN=64, for MODE values Bare, Sv39, Sv48, and Sv57. for SXLEN=64, which controls supervisor-mode address translation and protection. This register holds the physical page number (PPN) of the root page table, i.e., its supervisor physical address divided by 4 KiB; an address space identifier (ASID), which facilitates address-translation fences on a per-address-space basis; and the MODE field, which selects the current address-translation scheme. Further details on the access to this register are described in [virt-control].

Encoding of satp MODE field. shows the encodings of the MODE field when SXLEN=32 and SXLEN=64. MODE When MODE=Bare, supervisor virtual addresses are equal to supervisor physical addresses, and there is no additional memory protection beyond the physical memory protection scheme described in [pmp]. To select MODE=Bare, software must write zero to the remaining fields of satp (bits 30-0 when SXLEN=32, or bits 59-0 when SXLEN=64). Attempting to select MODE=Bare with a nonzero pattern in the remaining fields has an UNSPECIFIED effect on the value that the remaining fields assume and an UNSPECIFIED effect on address translation and protection behavior.

> When SXLEN=64, three paged virtual-memory schemes are defined: Sv39, Sv48, and Sv57, described in Sv39: Page-Based 39-bit Virtual-Memory System, Sv48: Page-Based 48-bit Virtual-Memory System, and Sv57: Page-Based 57-bit Virtual-Memory System, respectively. One additional scheme, Sv64, will be defined in a later version of this specification. The remaining MODE settings are reserved for future

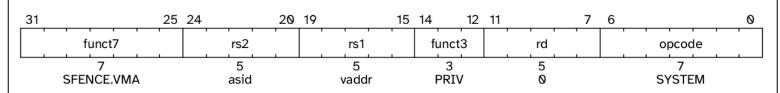
use and may defir

The satp CSR is considered active when the effective privilege mode is S-mode or U-mode. Executions of the address-translation algorithm may only begin using a given value of satp when satp is active.

Vignettes from the RISCV ISA: sfence.vma instruction



3.2.1. Supervisor Memory-Management Fence Instruction



The supervisor memory-management fence instruction SFENCE.VMA is used to synchronize updates to in-memory memory-management data structures with current execution. Instruction execution causes implicit reads and writes to these data structures; however, these implicit references are ordinarily not ordered with respect to explicit loads and stores. Executing an SFENCE.VMA instruction guarantees that any previous stores already visible to the current RISC-V hart are ordered before certain implicit references by subsequent instructions in that hart to the memory-management data structures. The specific set of operations ordered by SFENCE.VMA is determined by rs1 and rs2, as

associated v instruction a

described b For the common case that the translation data structures have only been modified for a single address mapping (i.e., one page or superpage), rs1 can specify a virtual address within that mapping to effect a

translatio • If $rs1 \neq x0$ and rs2 = x0, the fence orders only reads and writes made to leaf page table entries corresponding to the virtual address in rs1 for all address snaces. The fence also invalidates all

address-trar address in r

An implicit read of the memory-management data structures may return any translation for an address that was valid at any time since the most recent SFENCE.VMA that subsumes that address. The ordering implied by SFENCE.VMA does not place implicit reads and writes to the memorymanagement data structures into the global memory order in a way that interacts cleanly with the standard RVWMO ordering rules. In particular, even though an SFENCE.VMA orders prior explicit accesses before subsequent implicit accesses, and those implicit accesses are ordered before their associated explicit accesses, SFENCE.VMA does not necessarily place prior explicit accesses before subsequent explicit accesses in the global memory order. These implicit loads also need not otherwise obey normal program order semantics with respect to prior loads or stores to the same address.

structures

space. Th

Vignettes from the RISCV ISA: Page Table Entries



31	20	19 10 9 8 7 6 5 4 3 2 1 0				
	PPN[1]	PPN[0] RSW D A G U X W R V				
	12	10 2 1 1 1 1 1 1 1				
	Fig	ure 19. Sv32 page table entry.				
The PTE for valid; if it is page-fault exception. Attempting to execute a load or load-reserved instruction whose effective address lies within a page without read permissions raises a load page-fault exception. Attempting to execute a store, store-conditional, or AMO instruction whose effective address lies within a page without write pecase, supervisor code will fault on accesses to user-mode pages. Irrespective of SUM, the supervisor may not execute code on pages with U=1.						
bits.		An alternative PTE format would support different permissions for supervisor and user. We omitted this feature because it would be largely redundant with the SUM The RSW field is reserved for use by supervisor software; the implementation shall ignore this field.				
	The U bit indicathe page when access pages when	end Each leaf PTE contains an accessed (A) and dirty (D) bit. The A bit indicates the virtual page has been designaread, written, or fetched from since the last time the A bit was cleared. The D bit indicates the virtual				

table are global.
whereas marking Two schemes to manage the A and D bits are defined:

For non-leaf PTEspage has been written since the last time the D bit was cleared.

• The Svade extension: when a virtual page is accessed and the A bit is clear, or is written and the D bit is clear, a page-fault exception is raised.

space with a diffe

mapping being us

Vignettes from the RISCV ISA: Virtual Address Translation **Process**



3.3.2. Virtual Address Translation Process

acc

oric

stai

3. If

4. Oth

page-fau

Shadow S

and raise

7. Determin

8. Determin

next level of the

to the original ac

A virtual address va is translated into a physical address pa as follows:

- 5. A leaf PTE has been reached. If i>0 and $pte.ppn[i-1:0] \neq 0$, this is a misaligned superpage; stop and raise a page-fault exception corresponding to the original access type. sati
 - 6. Determine if the requested memory access is allowed by the *pte.u* bit, given the current privilege mode an 9. If *pte.a*=0, or if the original memory access is a store and *pte.d*=0:
 - - If the Svade extension is implemented, stop and raise a page-fault exception corresponding to the original access type.
 - The results of implicit address-translation reads in step 2 may be held in a read-only, incoherent • If a address-translation cache but not shared with other harts. The address-translation cache may hold an corres arbitrary number of entries, including an arbitrary number of entries for the same address and ASID.
 - Perfor Entries in the address-translation cache may then satisfy subsequent step 2 reads if the ASID
 - Implementations may also execute the address-translation algorithm speculatively at any time, for any associated with 1
 - virtual address, as long as satp is active (as defined in Supervisor Address Translation and Protection mapping. To ens ■ If instruction must (satp) Register). Such speculative executions have the effect of pre-populating the address-translation
 - cache.
 - The address-tran ■ If in memory direct
 - 10. The translation is successite
 - pa.pgoff = va.pgoff.

Speculative executions of the address-translation algorithm behave as non-speculative executions of the algorithm do, except that they must not set the dirty bit for a PTE, they must not trigger an exception, and they must not create address-translation cache entries if those entries would have been invalidated by any SFENCE.VMA instruction executed by the hart since the speculative execution of the algorithm began.

Agenda



- Why do we need MMUs ?
- Virtual Memory and Page Table Walks
- RISCV ISA MMU & Hypervisor Specification
- RISCV MMU & Hypervisor Test Plan
- Example MMU Test Cases
- Debug, Coverage and Deployment

MMU & Hypervisor Test Plan: Select Privilege Level



1.1. MMU & Hypervisor Page Translation

Tests Plan for Paging and Hypervisor functionality.

Start: Select privilege level

1.1.1. Select privilege level

IsaRef: link

CvgRef: gotoPrv

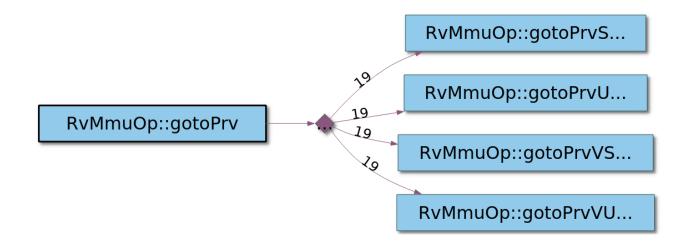
Pick a random privilege level to use for the scenario.

Select one of:

- User mode
- Hypervisor-extended supervisor mode
- Virtual user mode
- Virtual supervisor mode

NOTE: cannot run MMU operations in Machine mode.

TODO: support disabling of misa. H for S and U mode even if H extension is supported



MMU & Hypervisor Test Plan: User Mode

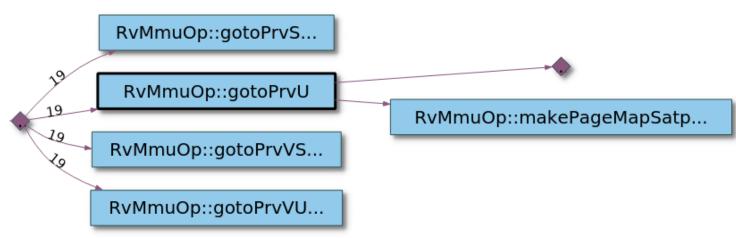


1.1.1.1. User mode

CvgRef: gotoPrvU

Do

- require PTE.U flag
- Setup one-stage address translation
- Switch to U-mode
- Select operation



MMU & Hypervisor Test Plan: Setup one-stage address translation



RvMmuOp::satpModeBare

1.1.3.1. Setup one-stage address translation

IsaRef: link

CvgRef: makePageMapSatp

RvMmuOp::satpModeSv39

RvMmuOp::satpMode

RvMmuOp::satpModeSv48

RvMmuOp::makePageMapSatp

RvMmuOp::makePageMapSatp

RvMmuOp::makePageMapSatp

RvMmuOp::makePageMapSatp

RvMmuOp::makePageMapSatp

Do

- Select one-stage paging mode
- Allocate 4KB naturally aligned root page
- Page map code stack and code addresses
- Write satp with root page table address and mode

TODO: randomize AISD mapping

MMU & Hypervisor Test Plan: Select operation

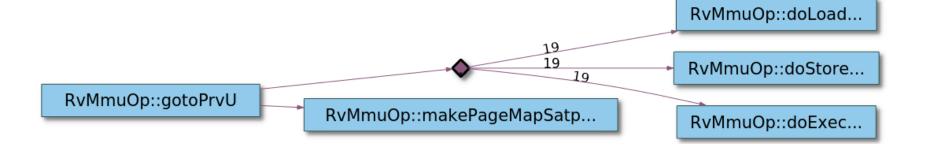


1.1.2. Select operation

Pick an operation to exercise page translation

Select one of

- Do load operation
- Do store operation
- Do execute operation



TODO: need to add LR, SC and AMO operations.

MMU & Hypervisor Test Plan: Do store operation



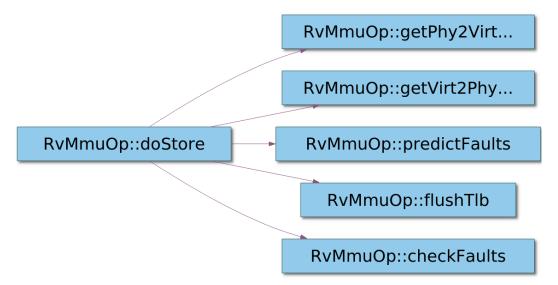
1.1.2.2. Do store operation

IsaRef: link

CvgRef: doStore

- Allocate virtual address
 - require PTE flags D, A, W, R and V set
 - if (U or VU mode) require PTE flag U set
- Store to allocated virtual address and update memory scoreboard
- Check expected page faults for mcause CAUSE_STORE_[GUEST_]PAGE_FAULT

NOTE: need both PTE.W and PTE.R — no write-only PTE flag configuration



MMU & Hypervisor Test Plan: Allocate virtual address



1.1.4. Allocate virtual address

- allocate physical memory
- Page map pysical address to virtual address
- predict expected page faults
- **if** (two-stage translation)
 - Generate hypervisor memory-management fence instructions
- Generate supervisor memory-management fence instruction

MMU & Hypervisor Test Plan: Page Map physical to virtual address



1.1.6. Page map pysical address to virtual address

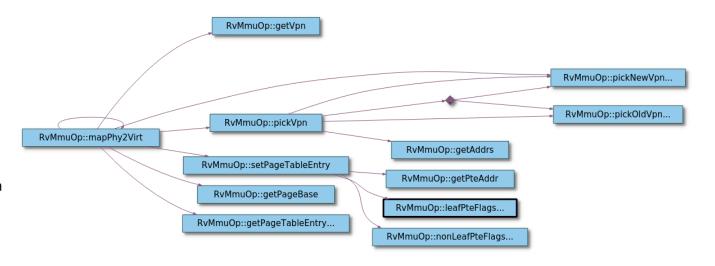
IsaRef: link

CvgRef: getPhy2Virt

- **if** (two-state translation)
 - Page map pysical address to virtual address to get gpa
- for each PTE level
 - Pick VPN for PTE
 - Pick PTE flags

TODO: clarify if sign extension (e.g. must have bits 63-39 all equal to bit 38) is used

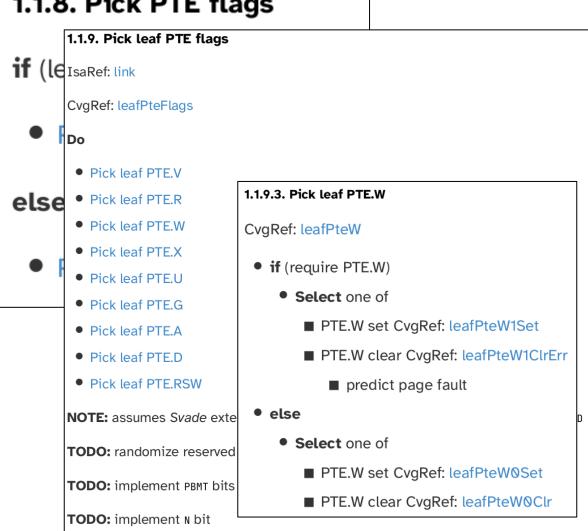
TODO: clarify usage of *megapages*, *gigapages*, *terapages* and *pentapages*

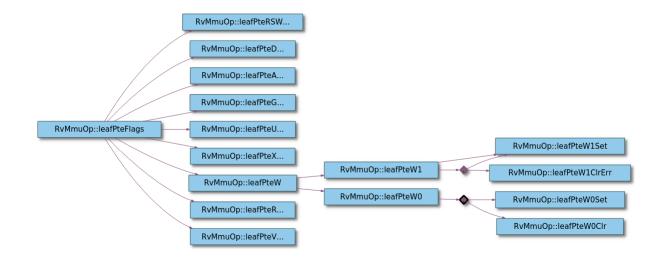


MMU & Hypervisor Test Plan: Pick PTE Flags



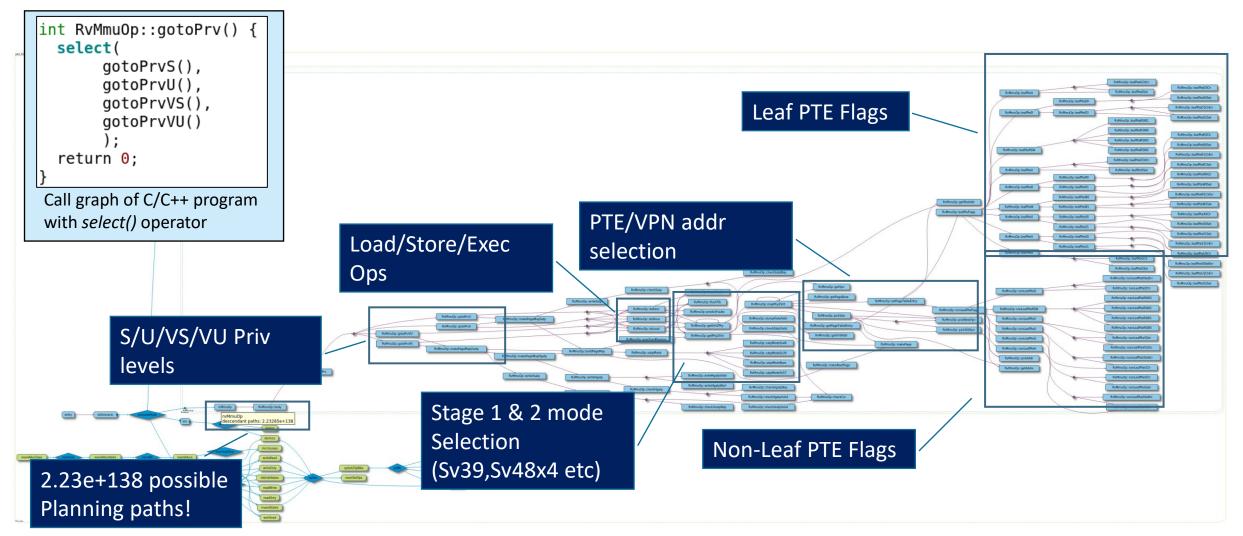
1.1.8. Pick PTE flags





MMU & Hypervisor Test Plan: Complete Scenario Model





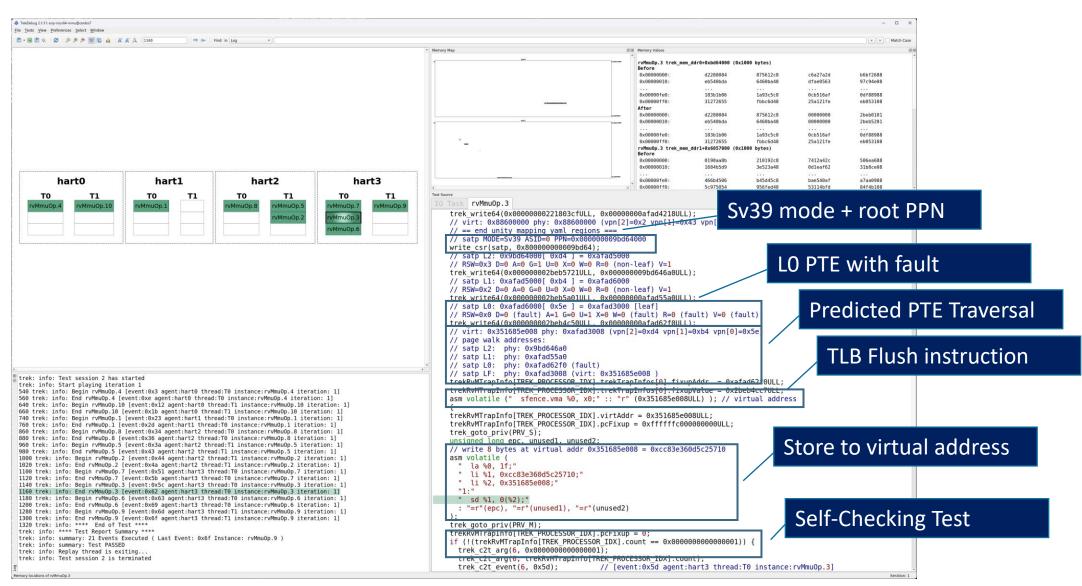
Agenda



- Why do we need MMUs ?
- Virtual Memory and Page Table Walks
- RISCV ISA MMU & Hypervisor Specification
- RISCV MMU & Hypervisor Test Plan
- Example MMU Test Cases
- Debug, Coverage and Deployment

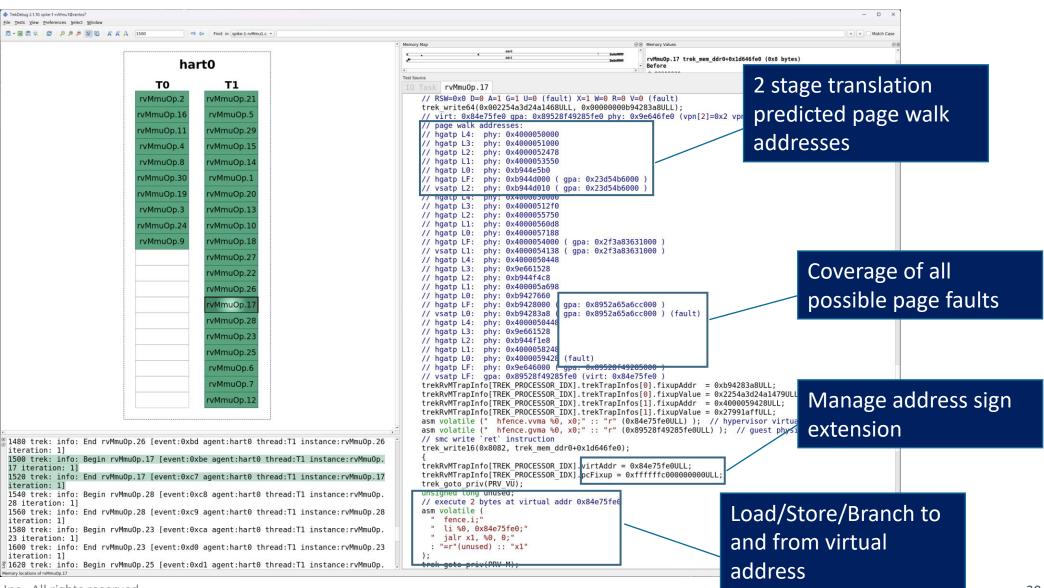
MMU One-Stage Address Translation Example





MMU Hypervisor Two-Stage Address Translation Example





Example Customer Bugs and ISA Misinterpretations



- Guest Page Fault vs (Host) Page Fault mcause for Hypervisor
- Code Prefetch past end of mapped page
- Request to support Sv39 for G-stage translation
- Failing to take page fault for bad PTE Flags

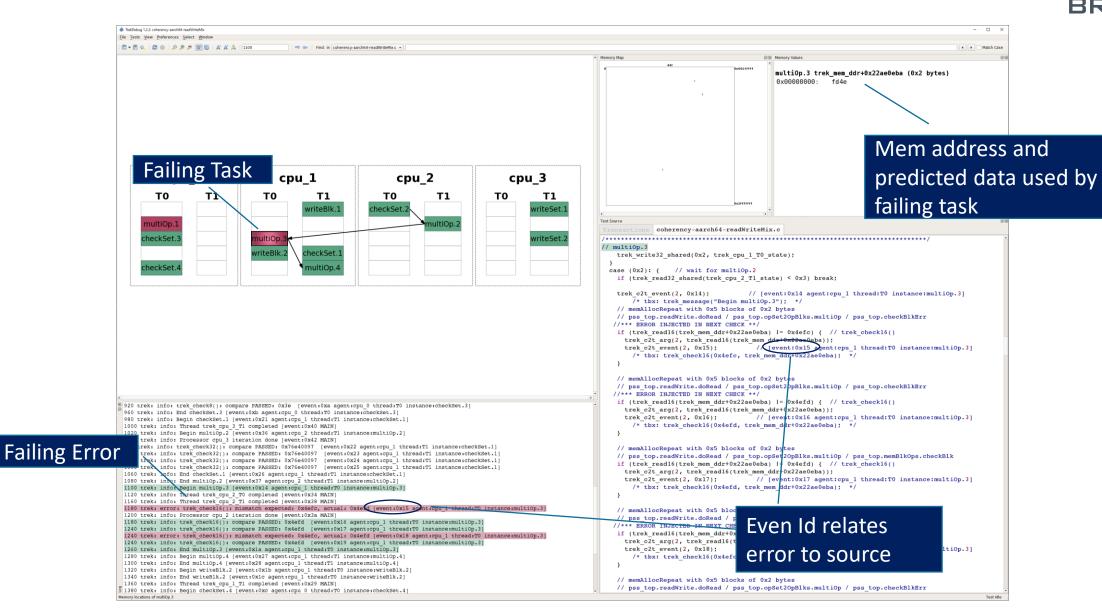
Agenda



- Why do we need MMUs ?
- Virtual Memory and Page Table Walks
- RISCV ISA MMU & Hypervisor Specification
- RISCV MMU & Hypervisor Test Plan
- Example MMU Test Cases
- Debug, Coverage and Deployment

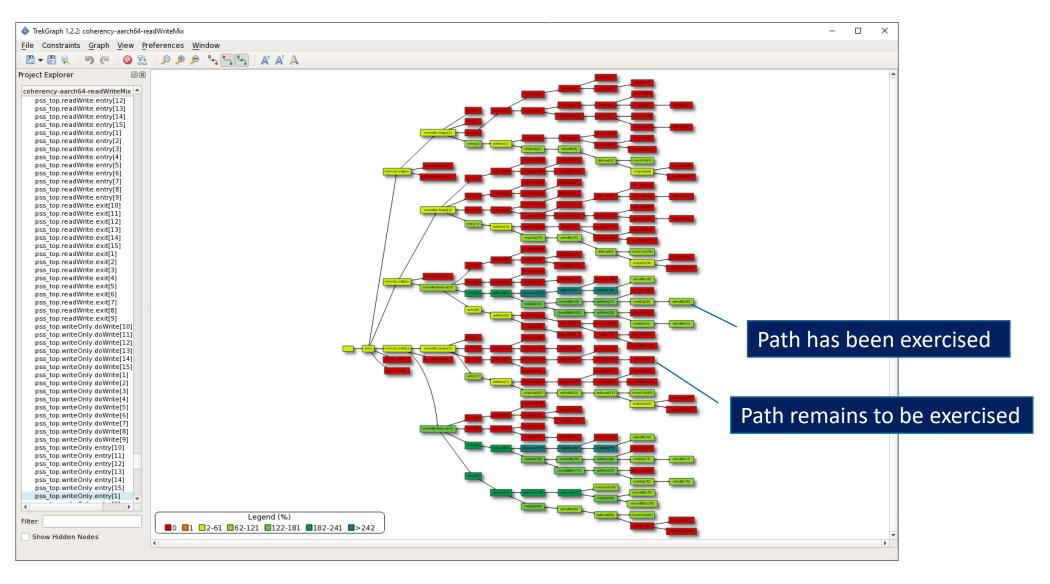
Debug Failing Test





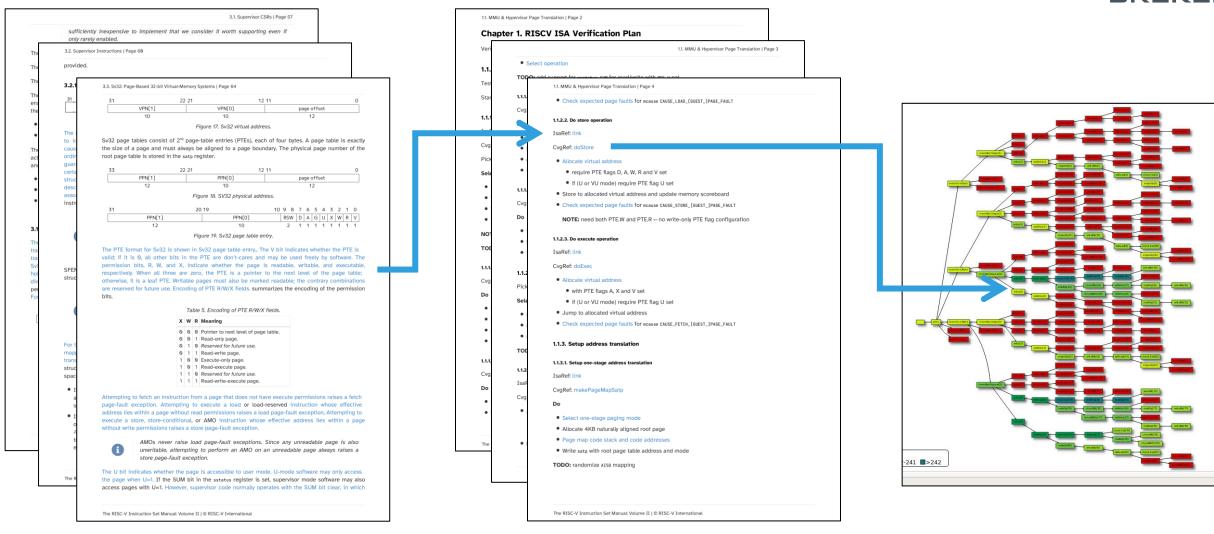
Analyze & Close Scenario Model Coverage





ISA Specs, Test Plans and Coverage





ISA Specification

RISCV Test Plan

Coverage Reports

RVA23 Profile Coverage Roadmap



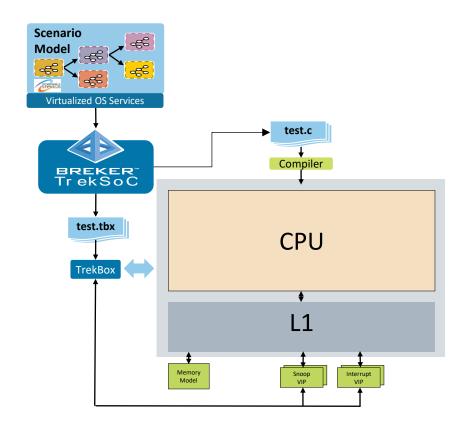
Extension	Description	Available
1	RV64I is the mandatory base ISA for RVA23U64.	Yes
М	Integer multiplication and division.	Yes
Α	Atomic instructions.	Yes
F	Single-precision floating-point instructions.	4Q24
D	Double-precision floating-point instructions.	4Q24
С	Compressed instructions.	Yes
Zicsr	CSR instructions.	Yes
Zicntr	Base counters and timers.	1Q25
Zihpm	Hardware performance counters.	3Q25
Ziccif	Coherence PMAs must support instruction fetch.	Yes
Ziccrse	Coherence PMAs must support RsrvEventual.	4Q24
Ziccamoa	Coherence PMAs must support all atomics.	Yes
Zicclsm	Misaligned loads and stores.	Yes

Extension	Description	Available
Za64rs	Reservation sets 64B,contiguous,aligned.	4Q24
Zihintpause	Pause hint.	1Q25
Zba	Address generation.	Yes
Zbb	Basic bit-manipulation.	Yes
Zbs	Single-bit instructions.	Yes
Zic64b	Cache blocks must be 64 bytes.	Yes
Zicbom	Cache-block management instructions.	Yes
Zicbop	Cache-block prefetch instructions.	1Q25
Zicboz	Cache-Block Zero Instructions.	1Q25
Zfhmin	Half-precision floating-point.	4Q24
Zkt	Data-independent execution latency.	3Q25
V	Vector extension.	4Q24
Zvfhmin	Vector minimal half-precision floating-point.	4Q24
Zvbb	Vector basic bit-manipulation instructions.	4Q24
Zvkt	Vector data-independent execution latency.	4Q24
Zihintntl	Non-temporal locality hints.	1Q25
Zicond	Integer conditional operations.	4Q24
Zimop	may-be-operations.	2Q25
Zcmop	Compressed may-be-operations.	2Q25
Zcb	Additional compressed instructions.	Yes
Zfa	Additional floating-Point instructions.	4Q24
Zawrs	Wait-on-reservation-set instructions.	3Q25
Supm	Pointer masking.	3Q25
Zvkng	Vector crypto NIST algorithms with GCM.	3Q25
Zvksg	Vector crypto ShangMi algorithms with GCM.	3Q25
Zabha	Byte and halfword atomic memory operations.	2Q25
Zacas	Compare-and-Swap instructions.	2Q25
Ziccamoc	Coherence PMAs must provide AMOCASQ.	2Q25
Zvbc	Vector carryless multiplication.	2Q25
Zama16b	Misaligned loads, stores, and AMOs are atomic.	Yes
Zfh	Scalar half-precision floating-point.	2Q25
Zbc	Scalar carryless multiply.	Yes
Zicfilp	Landing Pads.	3Q25
Zicfiss	Shadow Stack.	3Q25
Zvfh	Vector half-precision floating-point.	3Q25
Zfbfmin	Scalar BF16 converts.	3Q25
Zvfbfmin	Vector BF16 converts.	3Q25
Zvfbfwma	Vector BF16 widening mul-add.	3Q25
Zifencei	Instruction-Fetch Fence.	Yes
Ss1p13	Supervisor architecture version 1.13.	Yes
Svbare	The satp mode Bare must be supported.	Yes

Extension	Description	Available
Sv39	Page-based 39-bit virtual-Memory system.	Yes
Svade	Page-fault exceptions for A and D bits.	Yes
Ssccptr	Coherence PMAs must support page-tables.	Yes
Sstvecd	stvec.MODE=Direct is supported.	1Q25
Sstvala	stval gets faulting virtual address.	1Q25
Sscounterenw	hpmcounter and scounteren.	3Q25
Svpbmt	Page-based memory types.	1Q25
Svinval	Fine-grained TLB invalidation.	1Q25
Svnapot	NAPOT translation contiguity.	2Q25
Sstc	supervisor-mode timer interrupts.	1Q25
Sscofpmf	count overflow and mode-based filtering.	3Q25
Ssnpm	Pointer masking for senvcfg.PME, henvcfg.PME.	3Q25
Ssu64xl	sstatus.UXL UXLEN=64.	Yes
Н	The hypervisor extension.	Yes
Ssstateen	Supervisor-mode view of the state-enable extension.	3Q25
Shcounterenw	hpmcounter and hcounteren.	3Q25
Shvstvala	vstval gets faulting virtual address.	1Q25
Shtvala	htval gets faulting guest physical address.	1Q25
Shvstvecd	vstvec.MODE=Direct is supported.	1Q25
Shvsatpa	All translation modes in vsatp.	Yes
Shgatpa	hgatp SvNNx4 modes supported.	Yes
Sv48	Page-based 48-bit virtual-memory system.	Yes
Sv57	Page-based 57-bit virtual-memory system.	Yes
Zkr	Entropy CSR.	3Q25
Svadu	Hardware A/D bit updates.	1Q25
Sdtrig	Debug triggers.	3Q25
Ssstrict	No non-conforming extensions.	3Q25
Svvptc	Invalid to valid PTEs w/o fence.	1Q25
Sspm	Supervisor-mode pointer masking.	3Q25

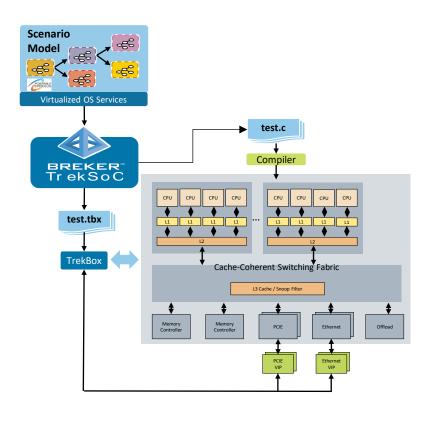
Single Hart vs Multi-Hart + SoC





Testbench control needed for

- External Interrupts,
- IOMMU,
- Debug Extension



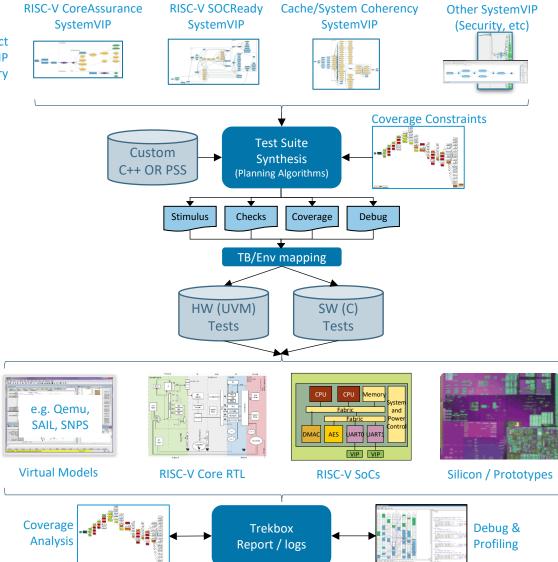
Multi-core tests needed for

- AMO/atomics
- CMO/cache flush, invalidate
- RVWMO Memory Ordering

Breker Technology Overview for RISC-V



Abstract SystemVIP Library



- EDA Test Generation Tool company
- RISCV Core and SoC System VIP products
- 17+ Commercial RISCV Deployments
 - 4+ open-source RISCV deployments
- Support organization of RISCV SME's
 - Help customers understand spec
 - Issue tracking, escalation, etc.
- Track Record:
 - 100% compliance + verification of x86_64
 - 10+ years on ARM SoC verification

Thanks for Listening! Any Questions?