



香山昆明湖架构虚拟化扩展的 设计和技术规划

裴晓坤¹ 徐泽凡²

¹中国科学院大学

²中国科学技术大学

2023年8月24日@第三届 RISC-V 中国峰会



目录

- 虚拟化技术
- RISC-V虚拟化扩展
- 香山虚拟化扩展设计思路
- 香山虚拟化扩展功能验证
- 总结与展望

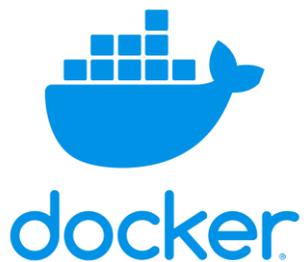


目录

- **虚拟化技术**
 - **虚拟化技术的概念**
 - **虚拟化技术的分类**
 - **虚拟机管理程序**
- RISC-V虚拟化扩展
- 香山虚拟化扩展设计思路
- 香山虚拟化扩展功能验证
- 总结与展望

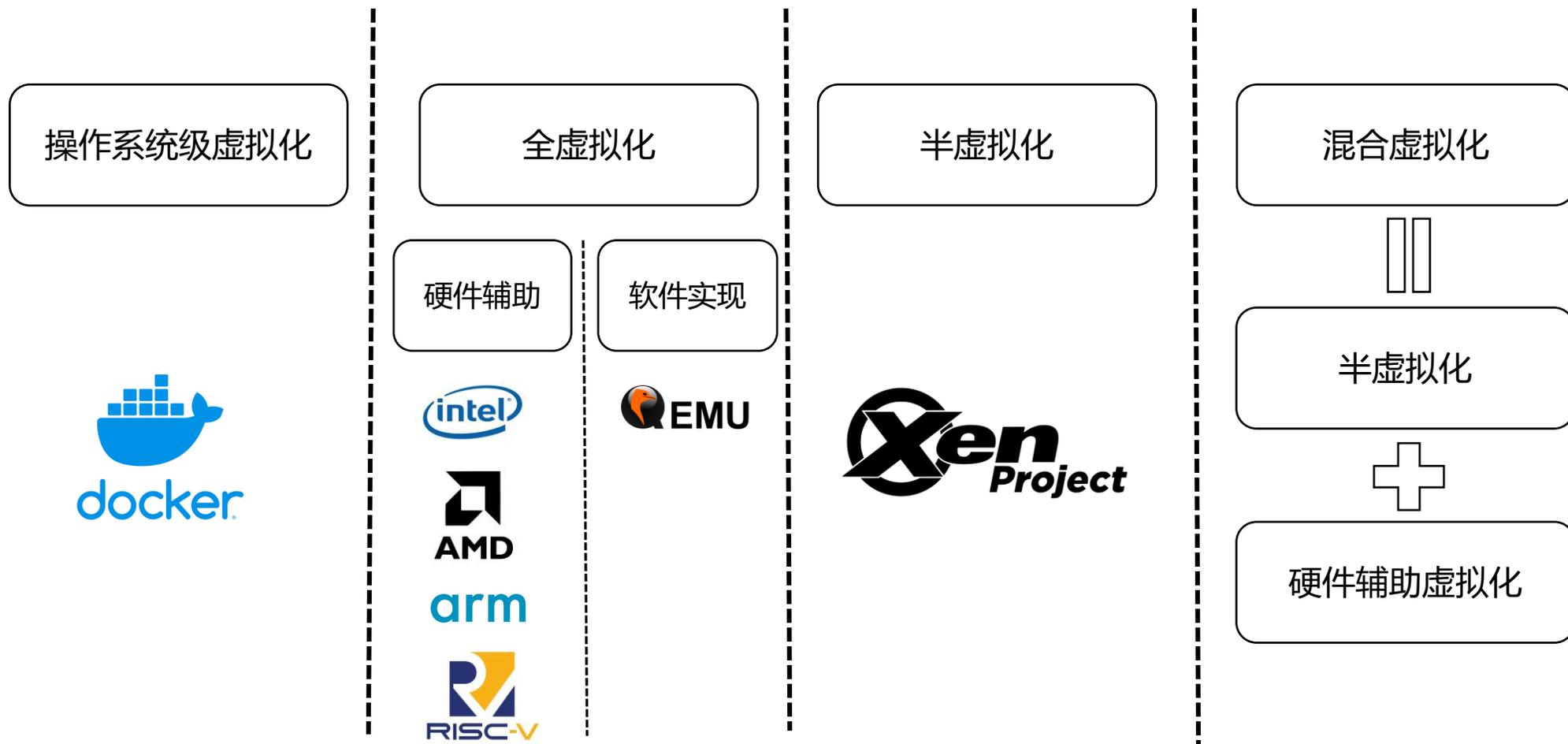
虚拟化技术·概念

- 一种资源管理技术
 - 单台计算机中的硬件资源划分为名为虚拟机 (VM) 的多个虚拟计算机
- 常见的虚拟化技术



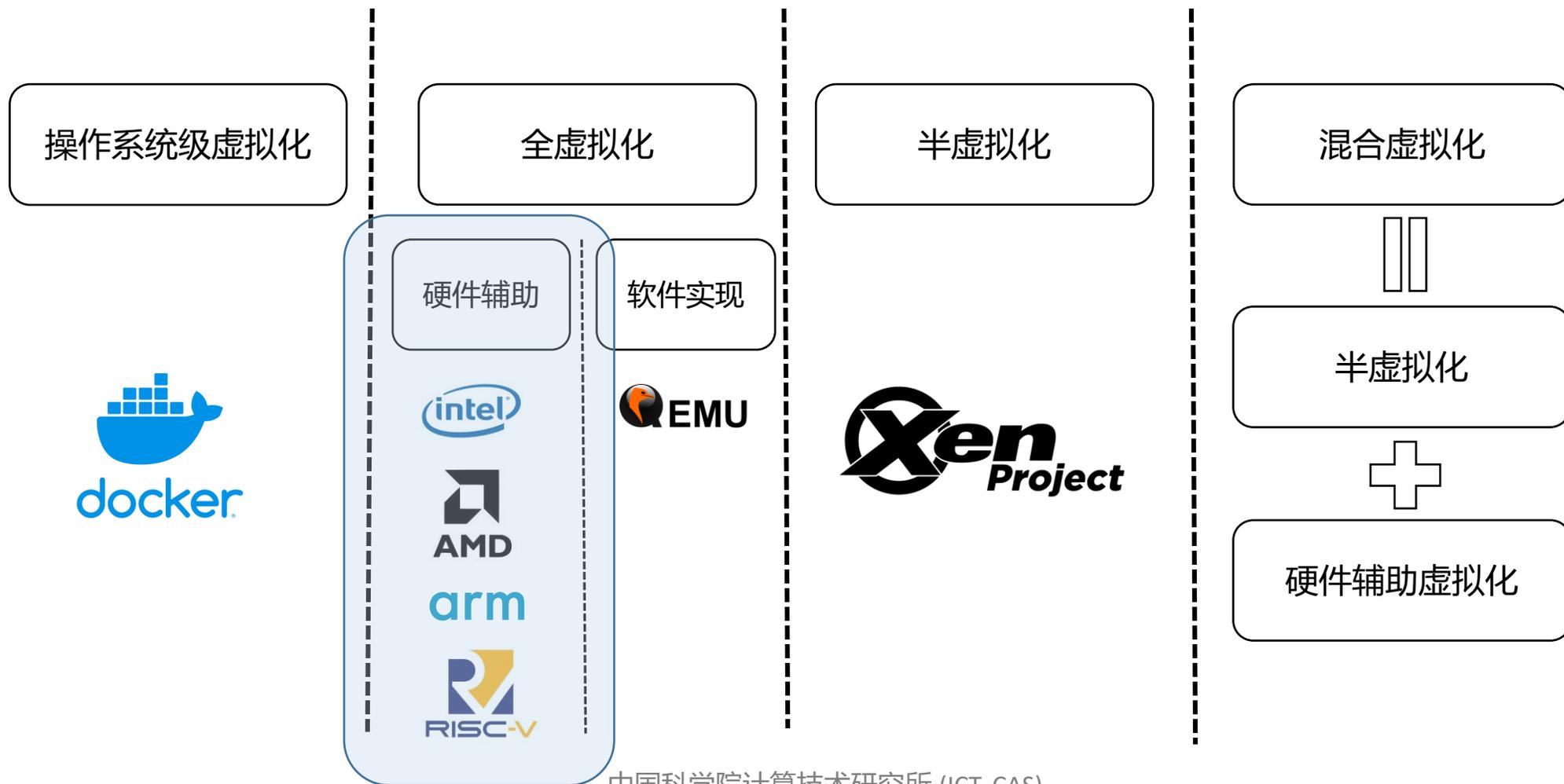
虚拟化技术·分类

- 根据虚拟化程度，可以将虚拟化分为以下几类



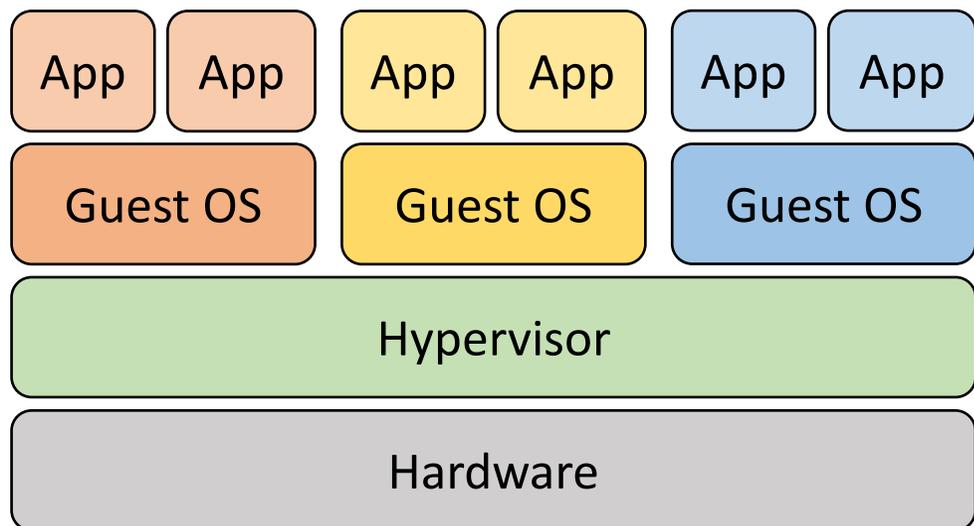
虚拟化技术·分类

- 根据虚拟化程度，可以将虚拟化分为以下几类

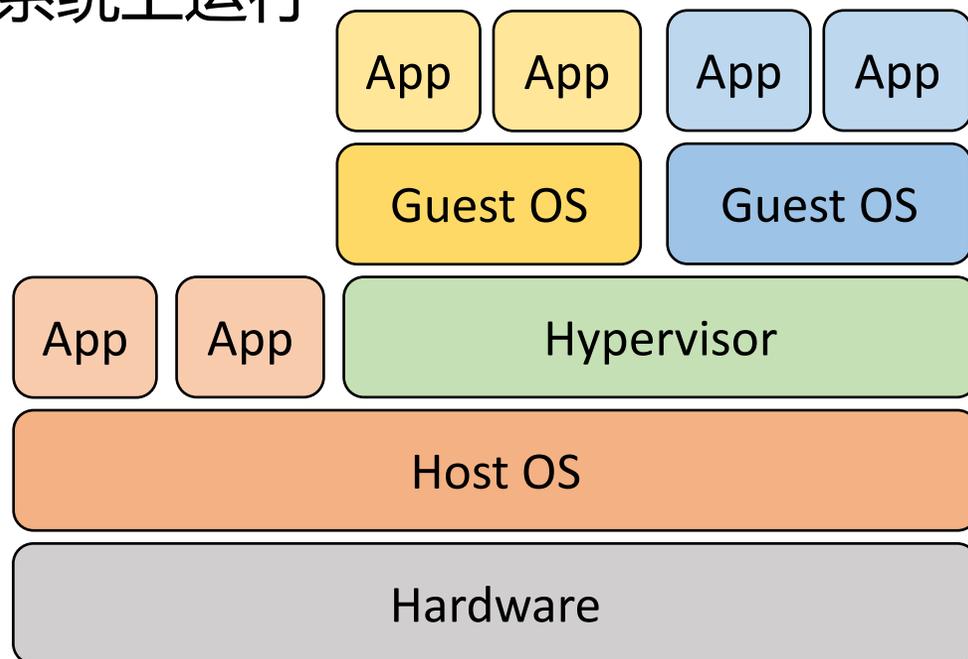


❄️ 虚拟化技术·虚拟机管理程序

- 虚拟机管理程序（VMM，也称 Hypervisor）分为两种：
 - Type-1, native or bare-metal hypervisors: 在裸机运行
 - Type-2 or hosted hypervisors: 在操作系统上运行



Type-1



Type-2



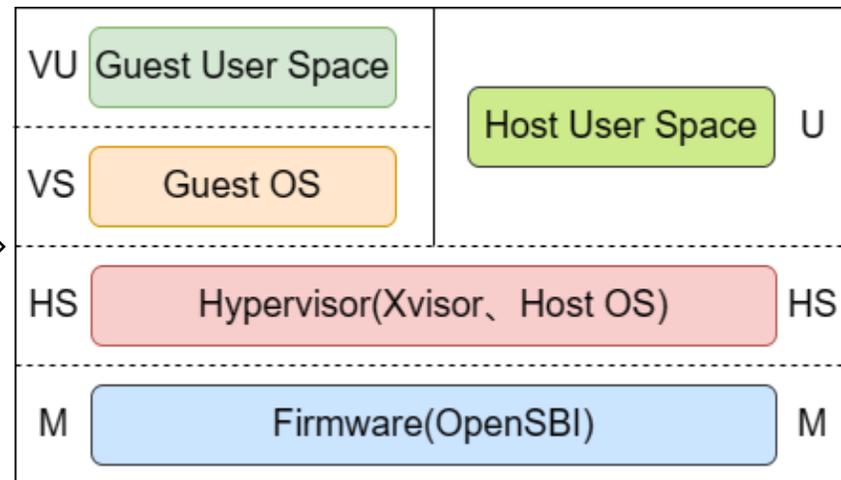
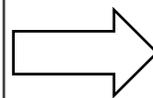
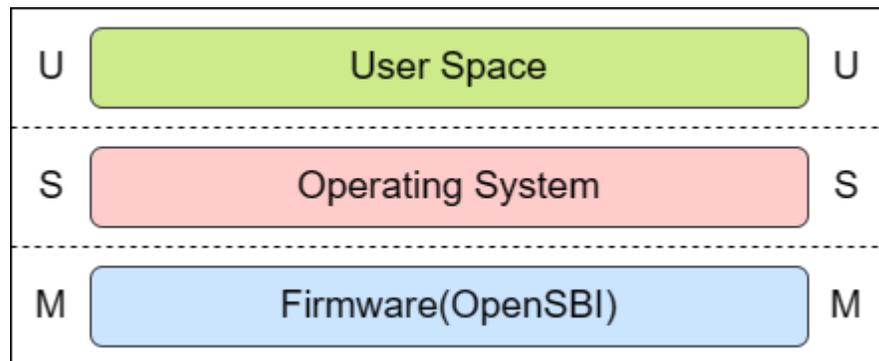
目录

- 虚拟化技术
- **RISC-V虚拟化扩展**
 - 内容介绍
 - 支持情况
- 香山虚拟化扩展设计思路
- 香山虚拟化扩展功能验证
- 总结与展望

RISCV虚拟化扩展·内容

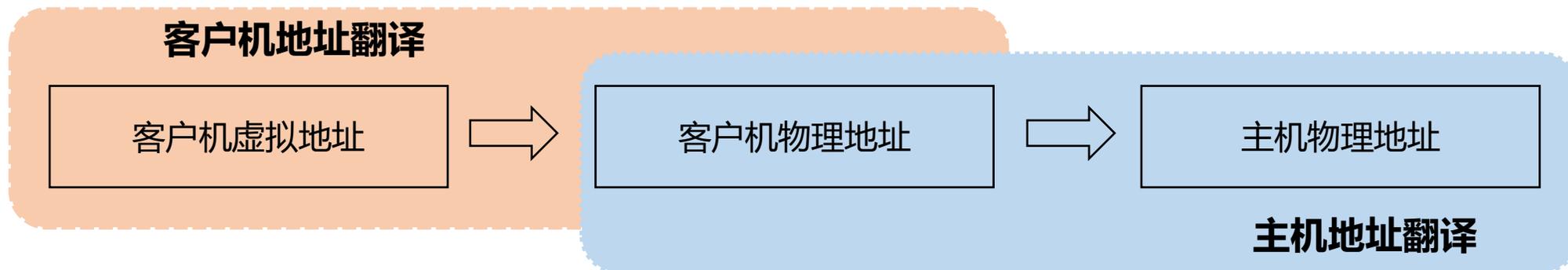
• CPU虚拟化

- 特权级拓展
- CSR拓展
- 指令拓展
- Trap拓展



• 内存虚拟化

- 两阶段地址翻译：客户机的地址翻译、主机的地址翻译



❄️ RISC-V 虚拟化拓展·支持情况

• 软件层面支持:

- 模拟器: QEMU、Spike



- Hypervisor: KVM、Xvisor、Bao等



硬件层面支持:

- 开源: Rocket chip、NOEL-V、CVA6



- 商业: 赛昉的昉·天枢、SiFive 的 P 系列等





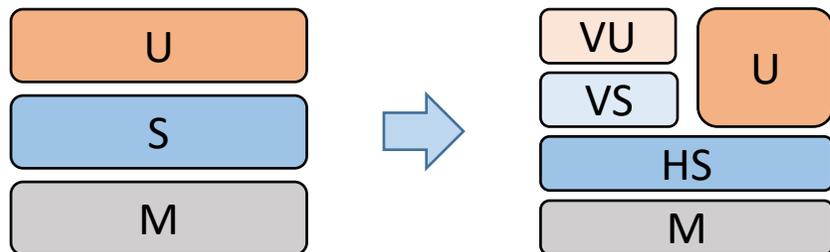
目录

- 虚拟化技术
- RISC-V虚拟化扩展
- **香山虚拟化扩展设计思路**
 - CPU虚拟化
 - 内存虚拟化
- 香山虚拟化扩展功能验证
- 总结与展望

香山虚拟化扩展设计·CPU虚拟化

• 特权级

- 新增V位, 区分VS和HS、VU和U



• CSR寄存器

Hypervisor CSR	hstatus、hedeleg、hideleg、hvip、hip、hie、hgatp等
Virtual Supervisor CSR	vsstatus、vsip、vsie、vstvec、vsepc、vsatp等
Machine CSR	mstatus、mideleg、mip、mie、mtval2 (新增)、mtinst (新增)

• Hypervisor指令

访存指令	HLV.width、HLVX.HU/WU、HSV.width
Fence指令	HFENCE.VVMA/GVMA

• Trap

- 增加VS级陷入陷出的处理

新增中断	VS software interrupt、VS timer interrupt、VS external interrupt、Supervisor guest external interrupt
新增异常	Environment call from VS-mode、Instruction guest-page fault、Load guest-page fault、Virtual instruction、Store/AMO guest-page fault

香山虚拟化扩展设计·内存虚拟化

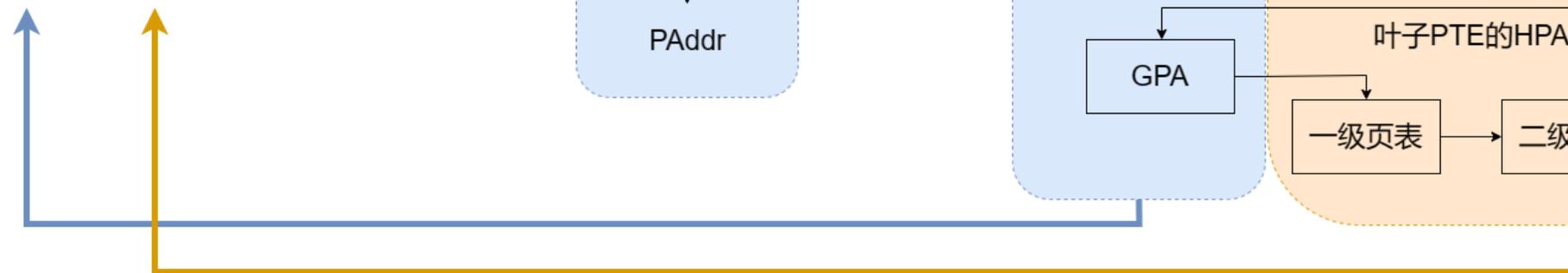
- 地址翻译模式

- Sv39 → Sv39 + Sv39x4

- 内存页表查询:

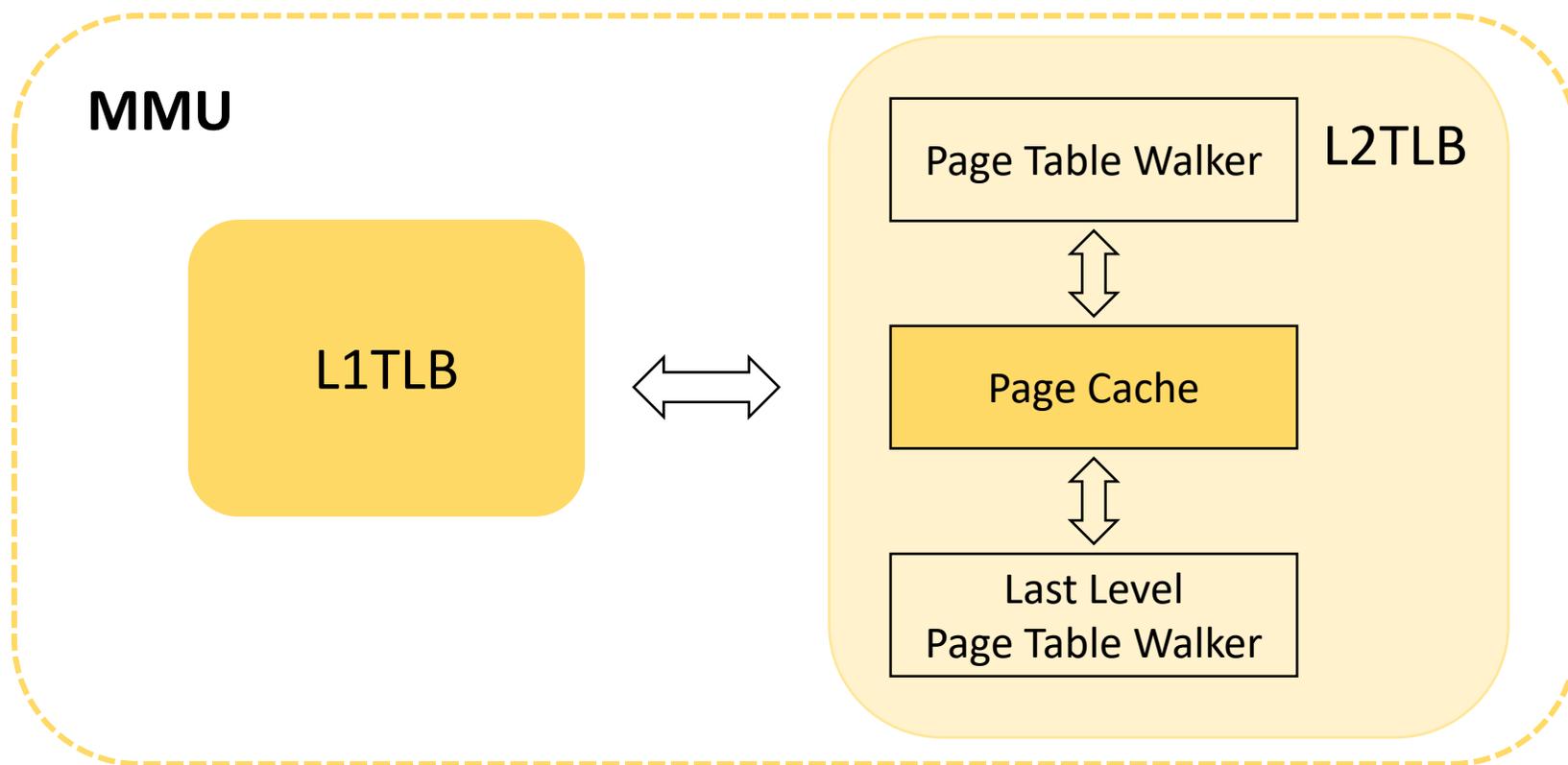
- 访存维度
- 访存次数

$$4 * 3 = 12$$



香山虚拟化扩展设计·内存虚拟化

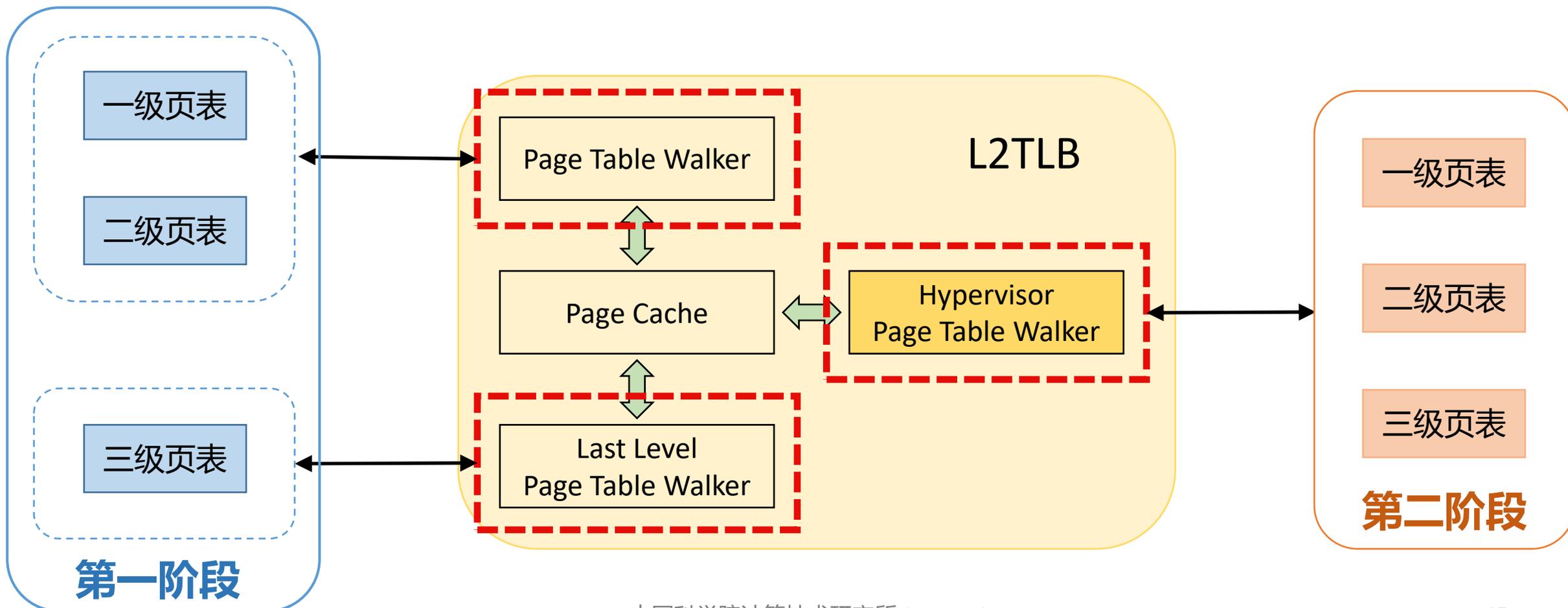
- MMU (Memory Management Unit)
 - 增加第二阶段的翻译, 客户机物理地址 → 主机物理地址
 - 存储两阶段翻译过程中的页表, L1TLB和L2TLB的Page Cache



香山虚拟化扩展设计·内存虚拟化

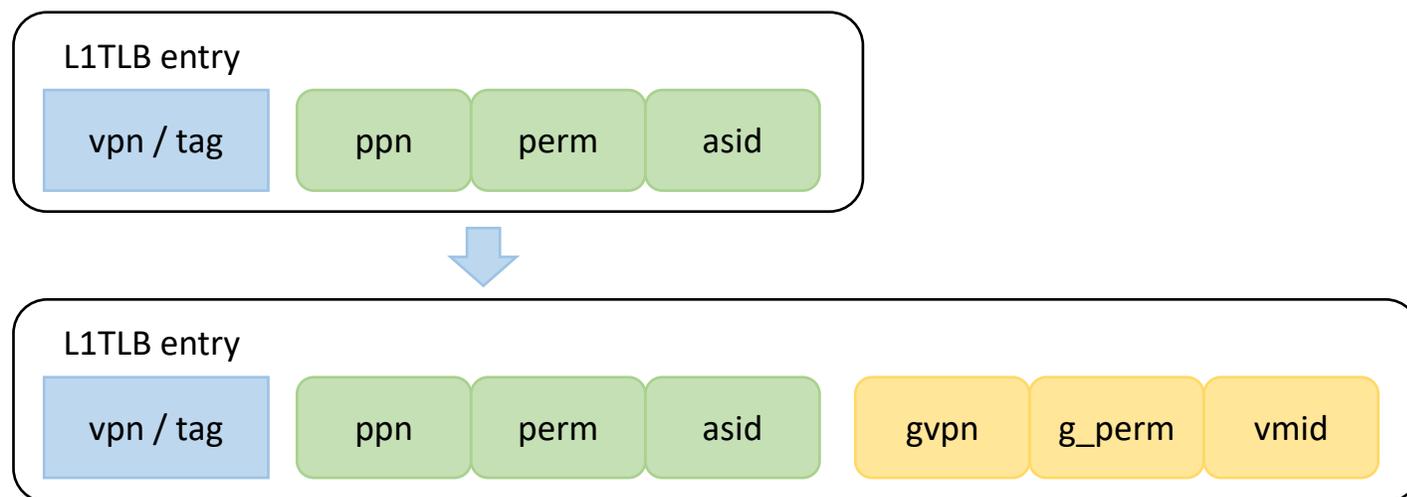
- 增加第二阶段地址翻译

- 新增Hypervisor Page Table Walker，负责客户机物理地址转换为主机物理地址



香山虚拟化扩展设计·内存虚拟化

- L1TLB存储项修改

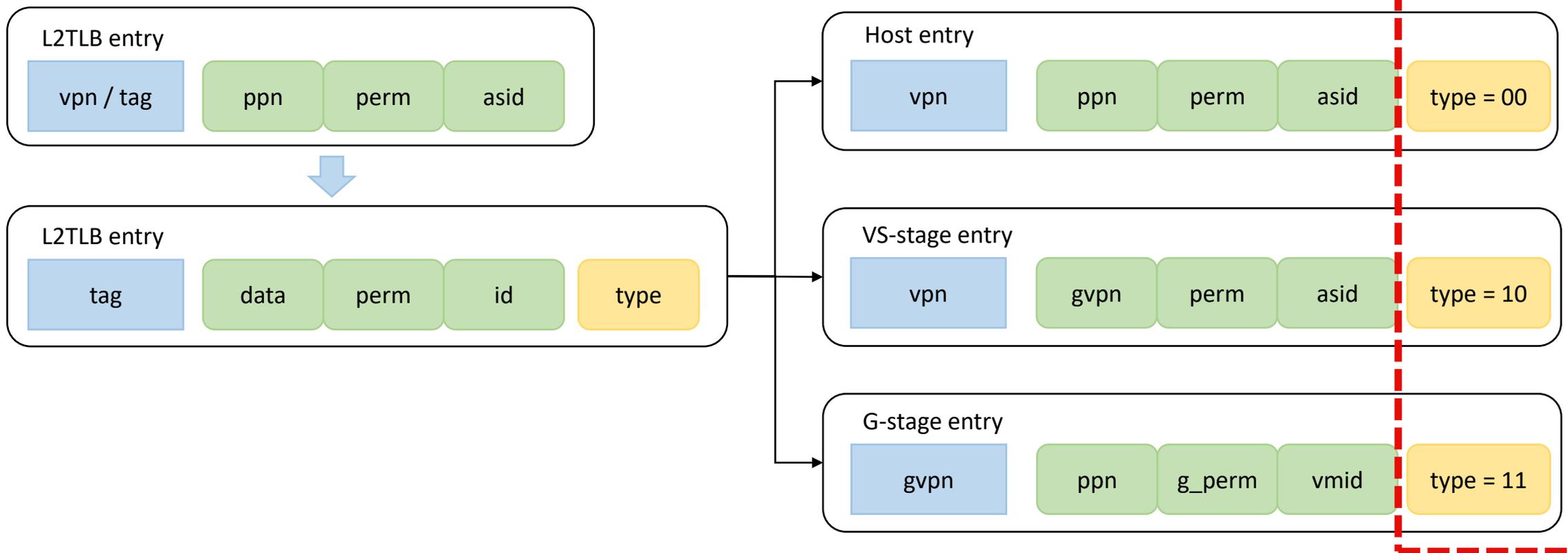


- gvpn: 第一阶段翻译的ppn, 第二阶段翻译的vpn
- g_perm: 第二阶段翻译得到的perm
- vmid: 所属虚拟机的id

香山虚拟化扩展设计·内存虚拟化

- L2TLB存储项修改

- 同一个结构存储三种类型的entry, 使用type区分



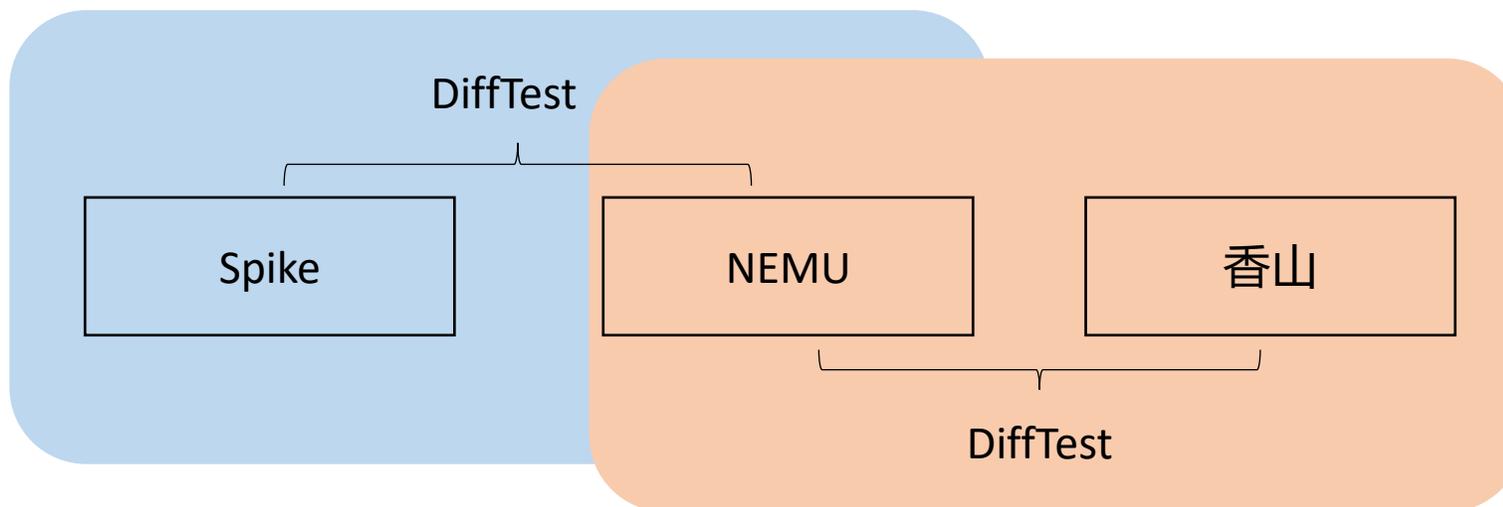


目录

- 虚拟化技术
- RISC-V虚拟化扩展
- 香山虚拟化扩展设计思路
- **香山虚拟化扩展功能验证**
 - **验证框架**
 - **单元测试**
 - **系统测试**
- 总结与展望

香山虚拟化扩展功能验证·验证框架

- 以DiffTest框架为核心，进行验证
 - DiffTest：在线差分验证框架
 - NEMU：高性能指令级解释器
- 验证NEMU：NEMU \leftrightarrow Spike（已完成）
- 验证香山：香山 \leftrightarrow NEMU（调试中）



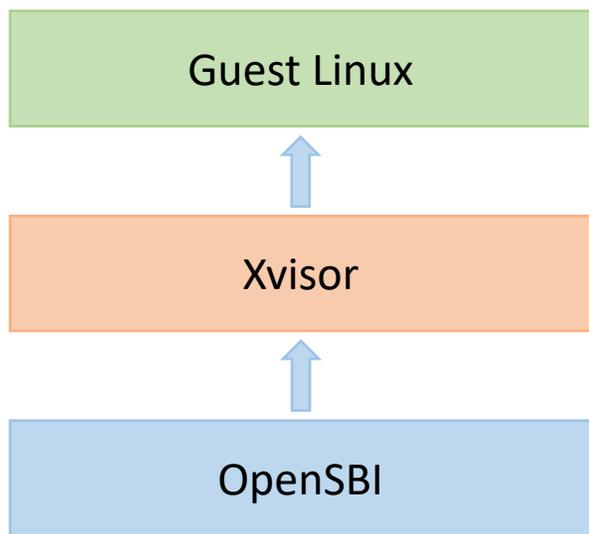
香山虚拟化扩展功能验证·单元测试

- 虚拟化验证程序集： riscv-hyp-tests
 - 开源项目， <https://github.com/josecm/riscv-hyp-tests>
 - 9个测试程序，共108个测试点，包含CPU虚拟化和内存虚拟化
 - 用于测试虚拟化扩展的基本功能

测试用例名	测试内容
tinst_tests	触发各种情况的 pagefault，检查 mtinst
wfi_exception_tests	不同特权级下，wfi 指令可能引发的异常
hfence_test	虚拟化 fence 指令的功能
virtual_instruction	虚拟化指令异常的触发
interrupt_tests	VS 级软件中断的触发及其代理机制
check_xip_regs	中断相关的 CSR 寄存器的读写
m_and_hs_using_vs_access	测试 mprv 位的功能、虚拟化访存指令
second_stage_only_translation	测试只有第二阶段地址翻译的情况
two_stage_translation	测试 VS 级下的地址翻译情况

香山虚拟化扩展功能验证·系统测试

- Xvisor
 - 开源 type-1 hypervisor
 - OpenSBI + Xvisor + Guest Linux



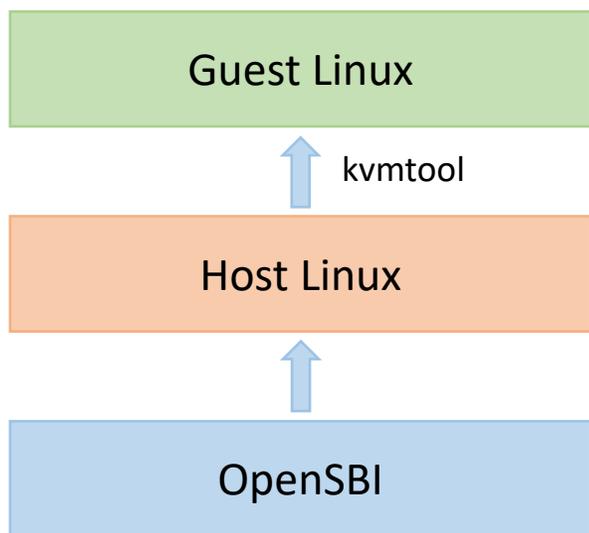
- NEMU运行Xvisor效果

```
~/xs-env/NEMU master = 12 > ./build/riscv64-nemu-interpretor -b ${XS_PROJECT_ROOT}/build/payload.bin[] xuzefan@open17
```

香山虚拟化扩展功能验证·系统测试

- KVM

- Type-2 hypervisor (有分类认为type-1)
- OpenSBI + Linux + kvmtool + Guest Linux
- kvmtool: 轻量级的虚拟机管理工具



- NEMU运行KVM效果

```
~ cd NEMU_DEBUG
~ NEMU_DEBUG cd NEMU
~ NEMU git:(master) X ./build/riscv64-nemu-interpreter -b -d ../riscv-isa-sim/difftest/build/riscv64-spike-so ../kvm-riscv/opensbi_xiangshan/build/platform/generic/firmware/fw_payload.bin
```

I



目录

- 虚拟化技术
- RISC-V虚拟化扩展
- 虚拟化扩展设计思路
- 虚拟化扩展功能验证
- **总结与展望**

总结与展望

- 总结

- 虚拟化技术、RISC-V虚拟化拓展的内容
- CPU虚拟化和内存虚拟化在香山上的实现
- 香山功能验证

- 展望

- RISC-V AIA (Advanced Interrupt Architecture)
- RISC-V IOMMU

敬请批评指正!