XUANTIE

RISC-V 中国峰会 2024

QEMU RISC-V支持的最新进展

刘志伟

阿里巴巴达摩院 技术专家

XUANTIE

目录

Contents

XUANTIE

QEMU 对 RISC-V 的支持概况

9个开发板

- Virt
- Kunminghu
- Microblaze
- Spike

25个CPU

动态		厂商		裸		Profile
max	RV64	Kunminghu	Ascalon	RV32I	RV64I	RVA23S64

6个Profile

RVA22		RVB23		RVA23	
RVA22S64	RVA22U64	RVB23S64	RVB23U64	RVA23S64	RVA23U64

141个扩展

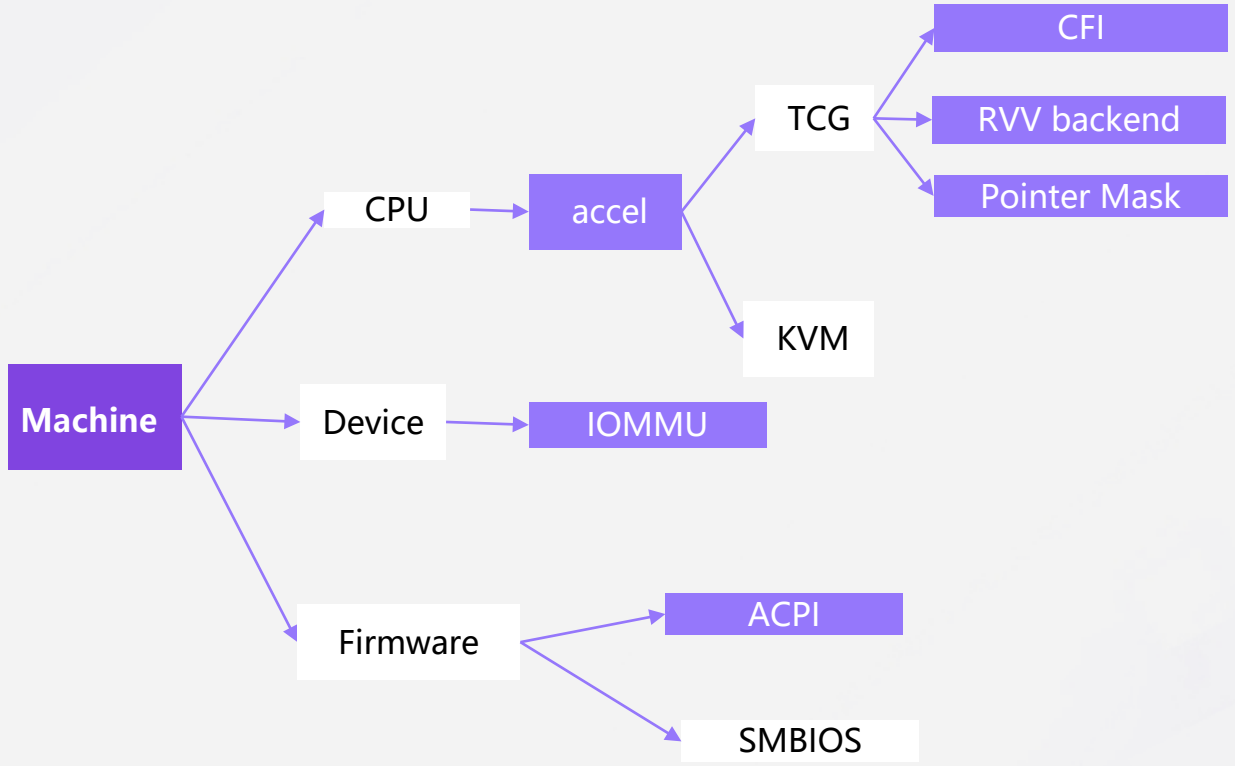
RV标准扩展				草案扩展	厂商扩展
Zicond	Zicfilp	I	A	x-svukte	Ventana
Zbb	Smmpm	M	V		XuanTie

QEMU for RISCV 进展

最近一年合入的重要特性

RISCV架构的工作进展

- 支持RVA23 profile
- 提升RVV模拟效率
- 支持IOMMU
- 支持CFI，Pointer Mask等重要扩展
- 声明式CPU
- TCG后端支持Vector
- 动态SXLEN支持

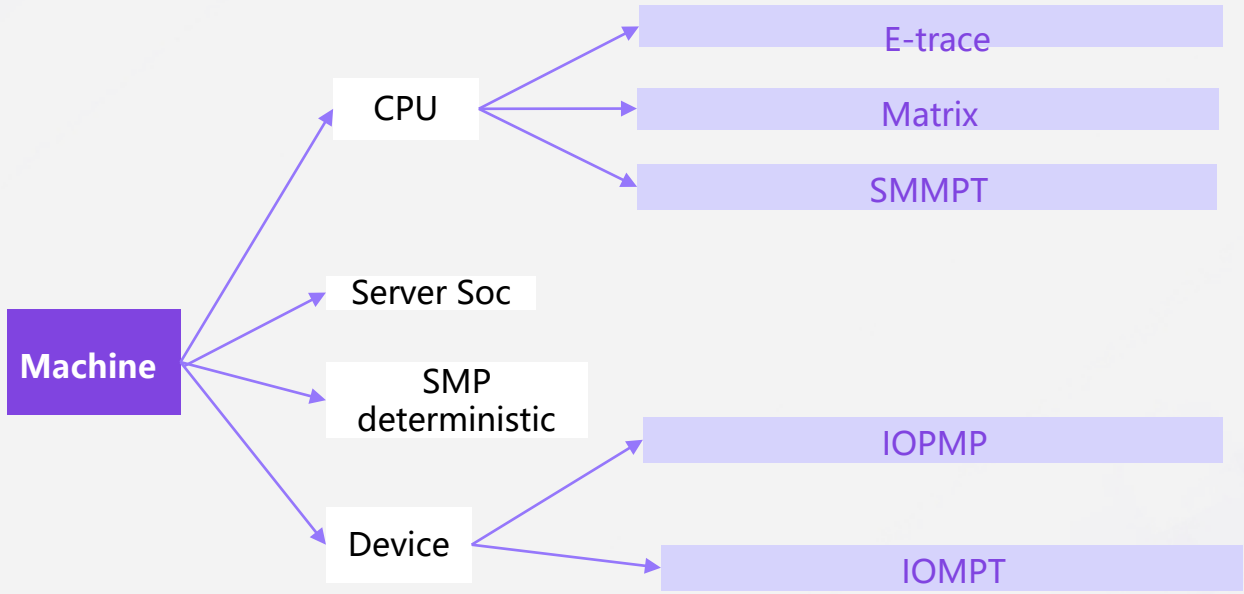


QEMU for RISCV 进展

社区在进行的工作

RISCV架构的工作进展

- 实现 IOPMP， 完善对虚拟化的支持
- 实现SMMPT， IOMPT等，支持机密计算
- 增加开发板版本支持，完善热迁移功能
- Server Soc支持
- 多核确定性执行
- 支持OCP数据类型

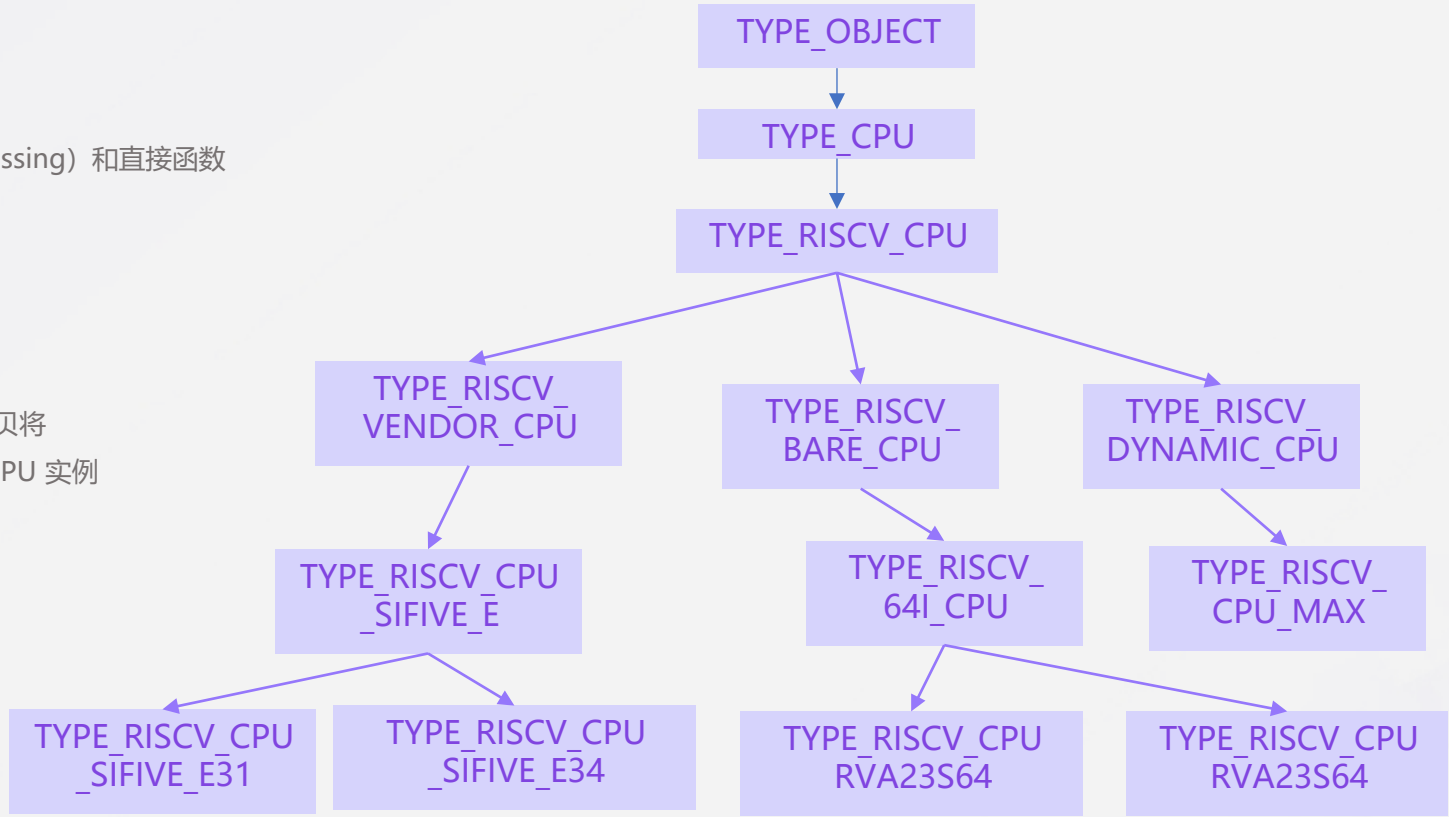


QEMU for RISCV最新特性介绍

声明式CPU

声明式CPU

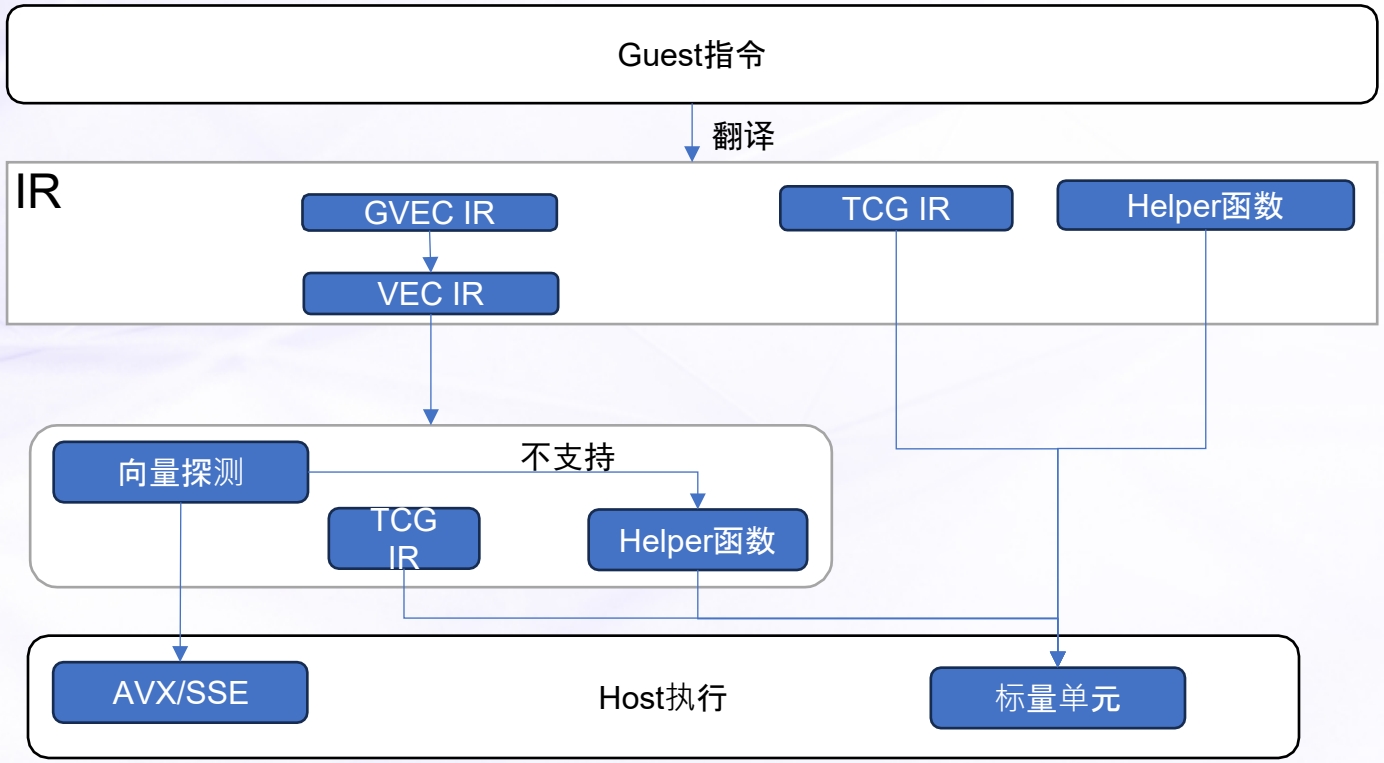
- **存在的问题**
  - 初始化逻辑碎片化
  - 混合编程模式同时使用继承（subclassing）和直接函数调用管理属性，增加代码复杂度
- **声明式机制**
  - 引入 RISCVCPUDef 结构体
  - 在 .instance_init 阶段，通过单次拷贝将 RISCVCPUDef 数据应用到 RISCVCPU 实例
  - 消除instance_post_init

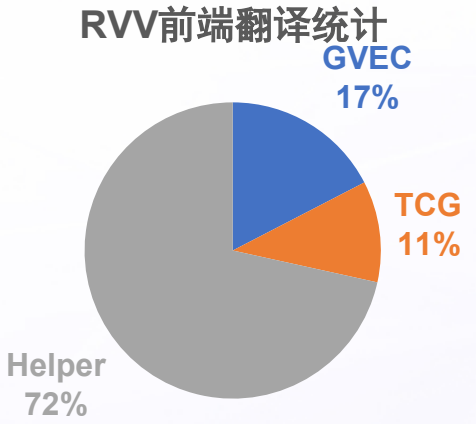


QEMU for RISCV最新特性介绍

向量优化

向量翻译为什么慢





利用主机向量能力不足，主要的原因：

1. 向量IR不支持Mask操作
2. 向量寄存器只支持4种操作宽度
3. GVEC IR的映射情况（无法直接映射）

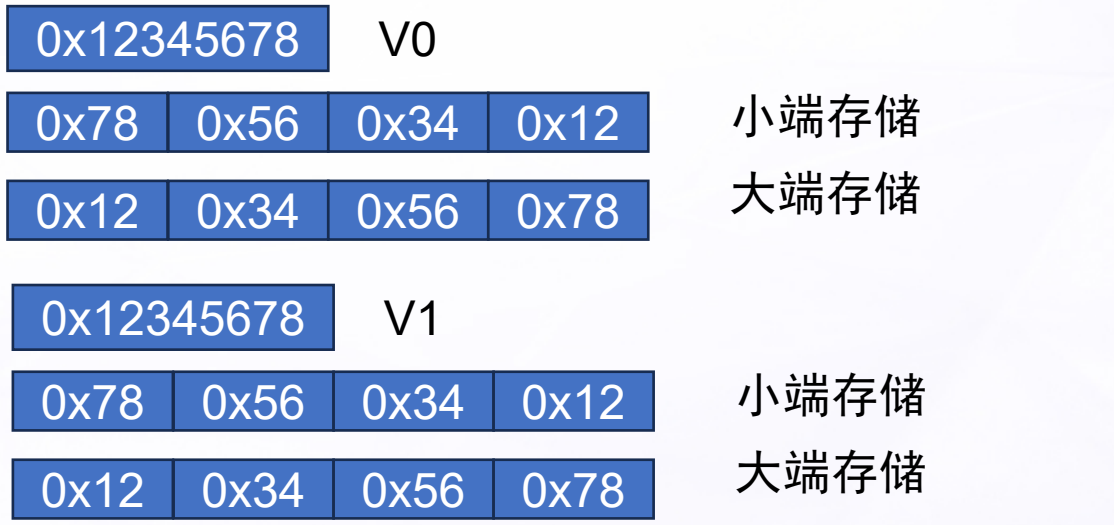
QEMU for RISCV最新特性介绍

向量优化

Helper函数为什么慢

以单精度浮点加法为例的运算类指令典型片段

```
for (i = env->vstart; i < vl; i++) {
    if (!vm && !vext_elem_mask(v0, i)) {
        /* set masked-off elements to 1s */
        vext_set_elems_1s(vd, vma, i * ESZ,
                          (i + 1) * ESZ);
        continue;
    }
    float32 s1 = *((float32 *)vs1 + H4(i));
    float32 s2 = *((float32 *)vs2 + H4(i));
    *((float32 *)vd + H4(i)) = float32_add(s2,
        s1, &env->fp_status);
}
```



V0.w * V1.w = [0x5678 * 0x5678, 0x1234 * 0x1234]

运算类指令慢的原因是无法向量化

- 边界。可以通过TB Flags解决。
- 掩码。可以通过TB Flags解决。
- 大小端 。可以通过预编译宏解决。
- 精确浮点 。只能动态检测。

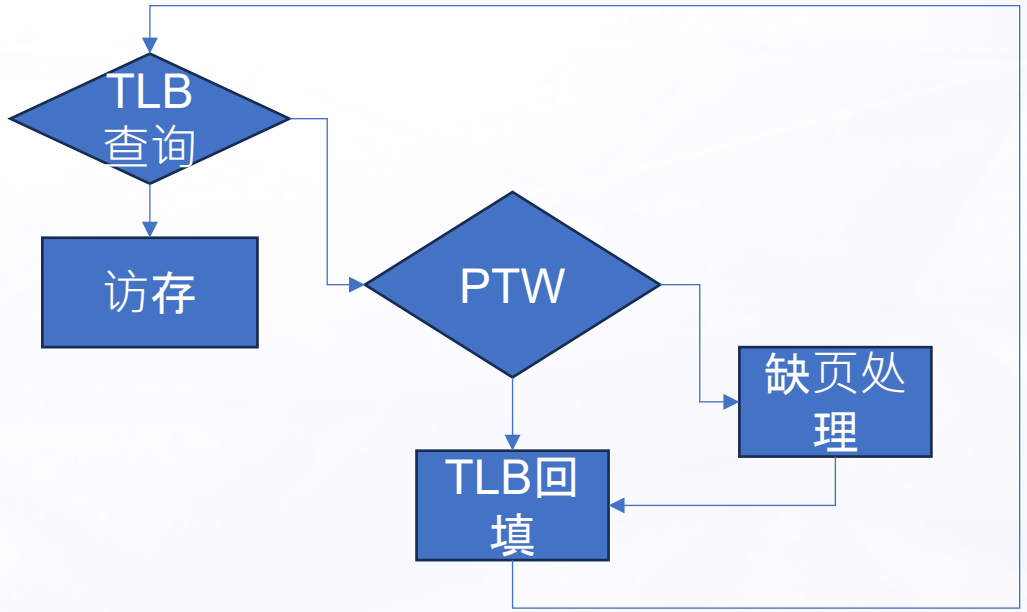
QEMU for RISCV最新特性介绍

向量优化

Helper函数为什么慢

以unit-stride为例的访存指令典型片段

```
for (i = env->vstart; i < evl; env->vstart = ++i) {
    k = 0;
    while (k < nf) {
        target_ulong addr = base + ((i * nf + k)
            << log2_esz);
        ldst_elem(env, adjust_addr(env, addr),
            i + k * max_elems, vd, ra);
        k++;
    }
}
```



访存类指令慢的原因是

- 慢速路径访存。
- 按照XLEN和Pointer mask的要求，计算有效地址。
- 大小端 。
- NF的处理。

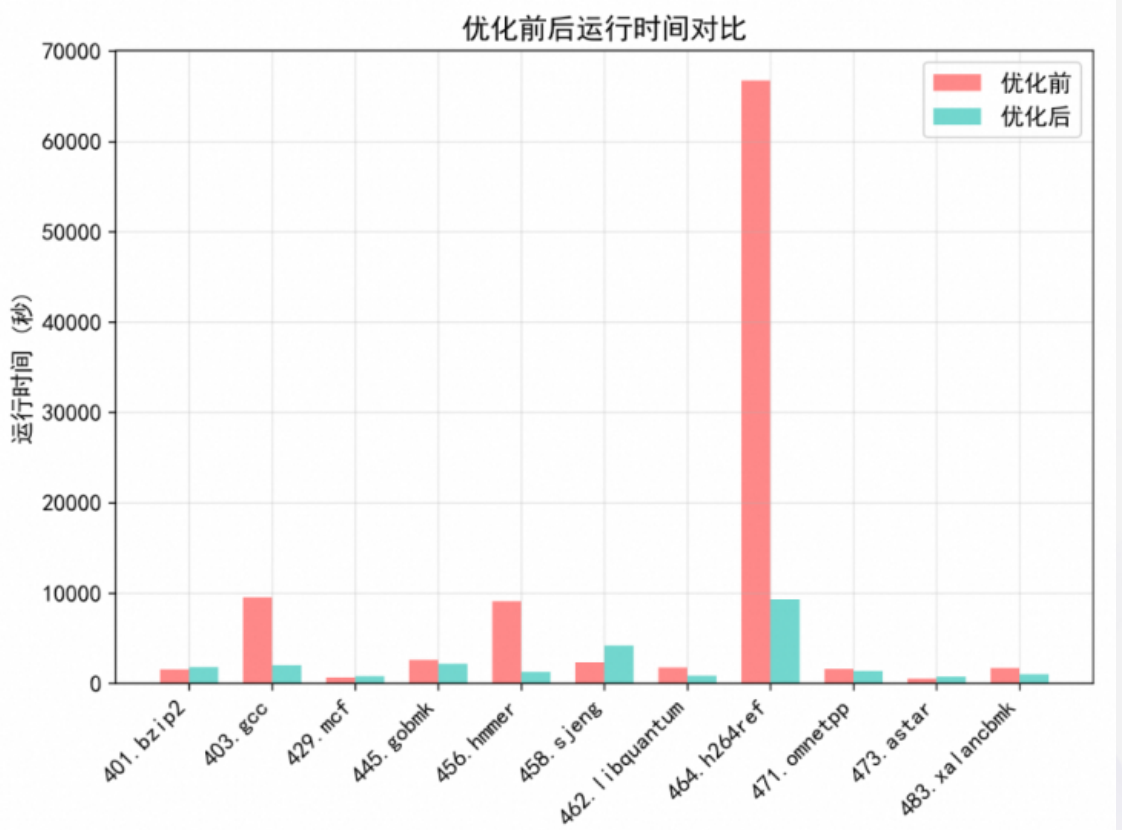
QEMU for RISCV最新特性介绍

向量优化

向量翻译优化

- **意义**
  - 满足快速运行spec等大型程序的需要
- **对Helper函数的关键优化**
  - 探测页面可访问性并解析宿主机物理地址
  - 在满足NF=1，小端，无掩码等条件下，优先快速路径，利用直接访问宿主机内存的快速路径（Fast Path）提升效率
  - 慢速路径兜底
- **探索方向**
  - GVEC IR增加掩码处理机制。
  - 优化stride或index访存的特殊形式。

SPEC 2006整数基准测试总体性能提升: 74.02%
平均运行时间: 从8913.70秒降至2316.19秒

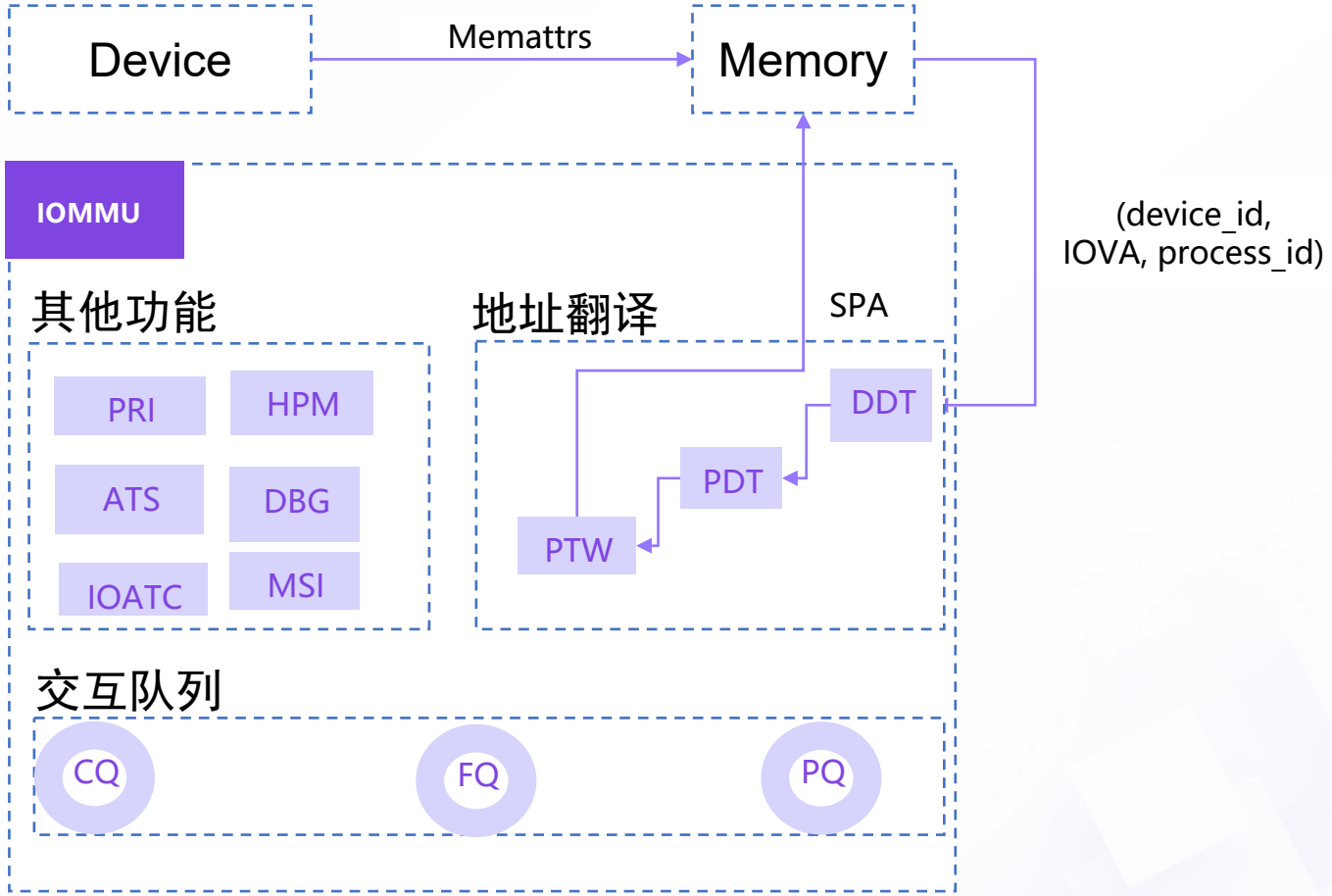


QEMU for RISCV最新特性介绍

虚拟化关键扩展

IOMMU

- **功能**
  - 实现iova到spa的地址翻译
  - 实现中断的重映射
- **机制**
  - 复用IOMMUMemoryRegion
  - 扩展了PASID到Memattrs
  - 扩展了PCIIOMMUOps，满足和PCIHost集成
  - 动态指定BDF到Memattrs
- **用法**
  - 平台设备
  - PCI设备

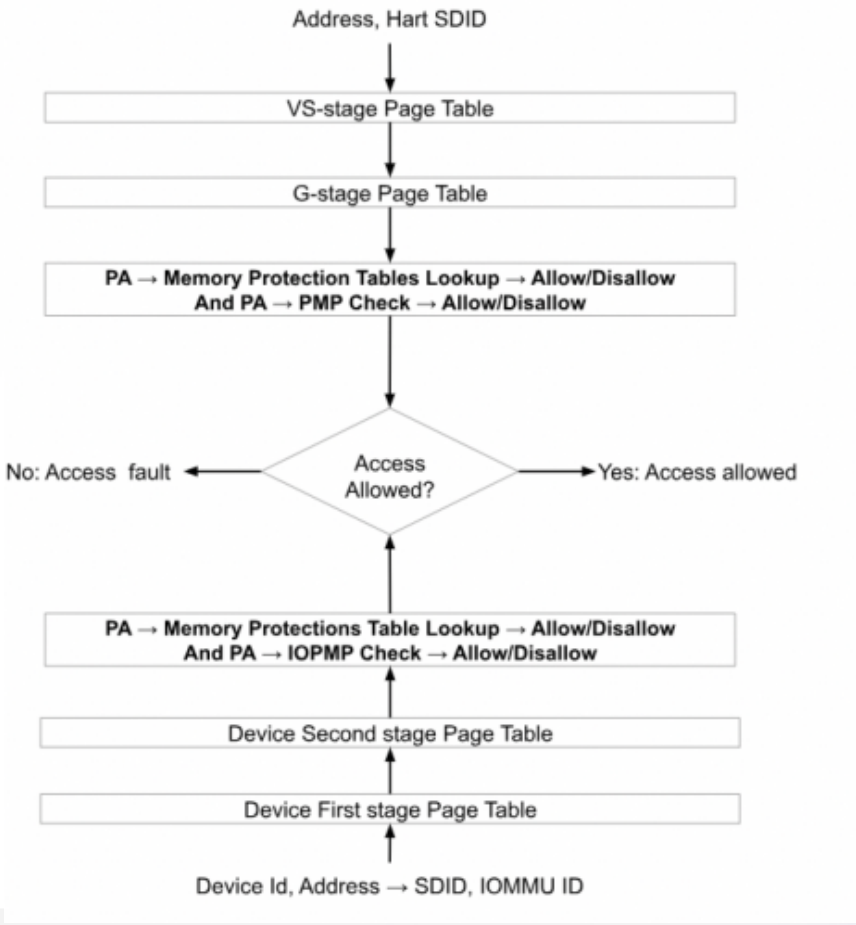


QEMU for RISCV最新特性介绍

虚拟化关键扩展

SMMPT

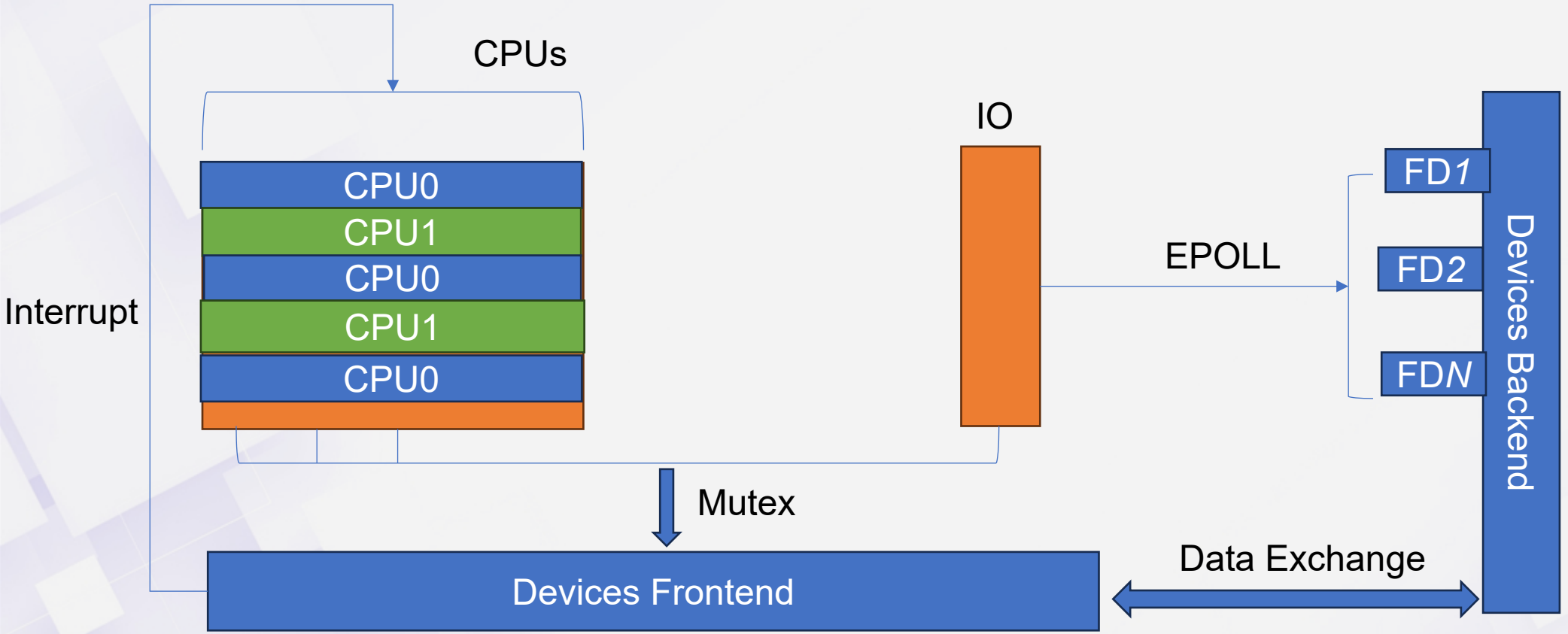
- **意义**
  - 内存按域访问，为机密计算提供支撑
- **机制**
  - 每个域通过SDID标记
  - 每个域有MPT的页目录基地址
  - 工作在M态，但通过页表方式维护，可以灵活配置
- **开发状态**
  - 每个SDID和SMMPT完成开发，upstream中
  - Iompt和smsdia开发中



QEMU for RISCV最新特性介绍

调试增强

多核确定型执行(QEMU单线程多核轮询架构)



QEMU for RISCV最新特性介绍

调试增强

多核确定型执行

- **单核确定性执行**
  - 保持IO线程和CPU线程之间的确定性
- **多核轮询模式确定性执行**
  - IO线程和CPU线程之间的非确定性
  - CPU核心之间指令数的非确定性
- **多线程模式确定性执行**
  - IO线程和CPU线程之间的非确定性
  - CPU核心之间指令数的非确定性
  - 多线程访存的非确定性

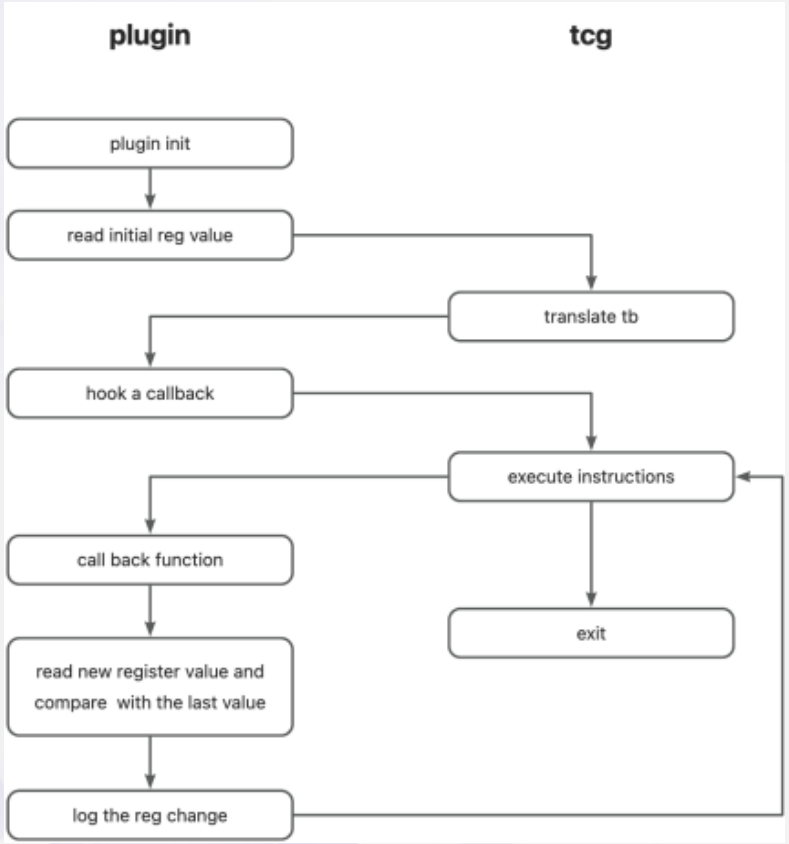
qemu-system-riscv64 -M virt -smp 2 -kernel Image –bios fw_dynamic.bin -nographic -initrdrootfs.cpio -icount shift=0,rr=replay,rrfile=rr.bin -m 4G

```
processor       : 0
hart            : 0
isa             : rv64imafdch_zicbom_zicboz_zicntr_zicsr_zifencei_zihintpaus
mmu             : sv57
mvendorid       : 0x0
marchid         : 0x0
mimpid          : 0x0

processor       : 1
hart            : 1
isa             : rv64imafdch_zicbom_zicboz_zicntr_zicsr_zifencei_zihintpaus
mmu             : sv57
mvendorid       : 0x0
marchid         : 0x0
mimpid          : 0x0

               total        used        free      shared  buff/cache   avail
Mem:         4023184       31632     3790220      180268      201332     372
Swap:              0           0           0

Linux version 6.6.36 (light@404540-6a6e1858-default) (riscv64-unknown-linux-
2.42.0.20240618) #1 SMP Wed Jan  8 15:07:45 UTC 2025

Welcome to Buildroot
buildroot login: root
root@qemu:~# ls
collect_gcda.sh  sit.sh
root@qemu:~# QEMU: Terminated
```

QEMU for RISCV最新特性介绍

调试增强

寄存器插件



追踪RA寄存器变化的插件

```
Hart  PC                 ASM                      REGS
0     0x17d8c  0xe09fe0ef  jal ra,-4600 # 0x16b94    ra -> 0x0000000000017d90
0     0x16be2  0x9782      jalr ra,a5,0             ra -> 0x0000000000016be4
0     0x17630  0xe5090ef   jal ra,39140 # 0x20f14   ra -> 0x0000000000017634
0     0x17658  0x70a2      ld ra,40(sp)             ra -> 0x0000000000016be4
```

QEMU for RISCV最新特性介绍

AI相关

OCP数据格式

- **简介**
  - 定义了bfloat16, fp8(e4m3/e5m2), fp6(e2m3/e3m2), e8m0, fp4(e2m1)等格式
  - 不完全遵守IEEE-754浮点规范
  - 在AI中广泛应用
- **机制**
  - 软浮点单元增加对应的数据类型
  - 对canonicalize和uncanonicalize做特殊处理
  - 建立默认规则来处理特殊值转换
- **社区进展**
  - BF16已合入
  - 其他数据类型正在upstream中

	E2M3	E3M2
Exponent bias	1	3
Infinities	N/A	N/A
NaN	N/A	N/A
Zeros	S 00 000_2	S 000 00_2
Max normal	S 11 111_2 = ± 2^2 × 1.875 = ± 7.5	S 111 11_2 = ± 2^4 × 1.75 = ± 28.0
Min normal	S 01 000_2 = ± 2^0 × 1.0 = ± 1.0	S 001 00_2 = ± 2^{-2} × 1.0 = ± 0.25
Max subnorm	S 00 111_2 = ± 2^0 × 0.875 = ± 0.875	S 000 11_2 = ± 2^{-2} × 0.75 = ± 0.1875
Min subnorm	S 00 001_2 = ± 2^0 × 0.125 = ± 0.125	S 000 01_2 = ± 2^{-2} × 0.25 = ± 0.0625

QEMU for RISCV最新特性介绍

AI相关

RISCV社区对OCP的需求

- **类似格式**
  - IEEE-P3109 , Arithmetic Formats for Machine Learning，进行中
  - AGQ
  - TSL

Format	P3109			OCP		AGQ		TSL	
Subformat	P3	P4	P5	E5	E4	E5	E4	E4	E5
Special values shared by all subformats	Y			N		Y		N	
Exactly one NaN	Y			N		Y		Y	
Positive and negative infinity	Y			Y	N	N		N	
Include negative zero	N			Y		N		N	
Max exponent emax	15	7	3	15	8	15	7	N/A	N/A

- **RISCV扩展**
  - 正式发布了zvfbfmin,zvfbfmin,zvfbfwma
  - 制定中Zvfbfa,Zvfofp8min
- **RISCV社区讨论**
  - ALTFMT机制
  - 未来会兼容 P3109 或类似被AI广泛采用的数据类型

XUANTIE

未来规划和展望

未来一年社区的主要工作

扩展支持

- OCP
- Memory Tagging
- Matrix
- P
- 64位指令宽度
- E-trace
- IOPMP
- SMMPT
- SMSDIA
- IOMPT

支持Server Soc specification

- PCIE
- Qos
- RAS

XUANTIE

RISC-V 中国峰会 2024

Thank you

玄铁公众号

玄铁中文站

玄铁海外站