北京开源芯片研究院
BEIJING INSTITUTE OF OPEN SOURCE CHIP

香山开源处理器社区
XiangShan Open-Source Processor Community



开源IOMMU设计和软硬件协同验证解决方案

报告人：杨媛媛
作者：杨媛媛、郭亚星
2026-10-18

目录

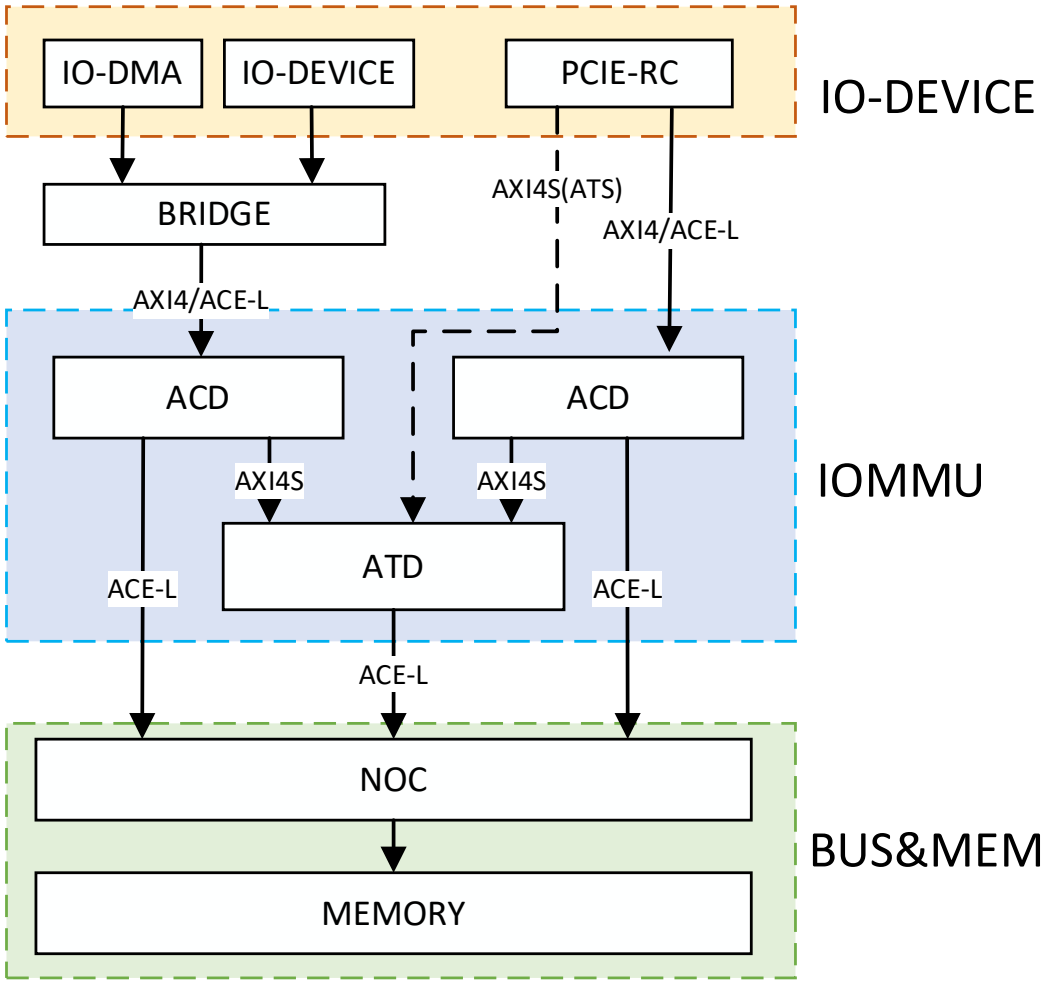
00 OpenIOMMU介绍：微架构/特性/可综合

01 OpenIOMMU软硬件协同验证方法

02 OpenIOMMU社区对比与选型优势

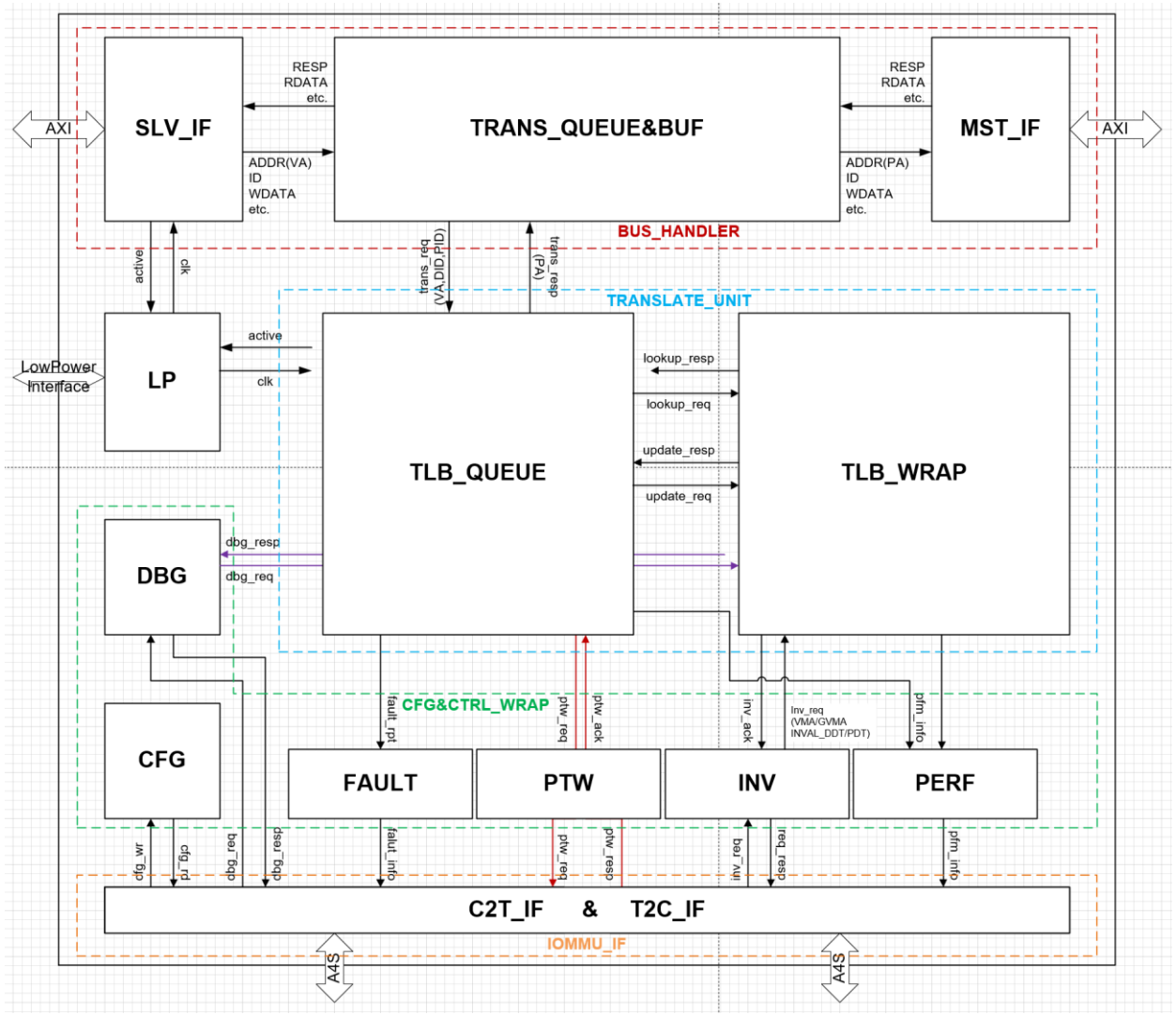
OpenIOMMU 介绍 — 微架构总览

- **ACD：地址缓存设备**
- 靠近 I/O 设备，承担数据面功能。
- 负责快速 micro TLB、main TLB 命中。
- 接收、转发 DMA 数据，降低地址转换延迟。
- 支持多条 outstanding 事务并发处理
- **ATD：地址转换设备**
- 负责设备上下文和进程上下文查找。
- 完成 Stage 1、Stage 2 页表遍历及权限检查。
- 处理 ATS、PRI 请求。
- 负责队列、中断和性能监控等软件交互功能
- **ACD 与ATD 互联**
- - 通过 AXI4-Stream 传递内部消息。- 支持多个 ACD 与 ATD 灵活互联和扩展
- **性能特点：**
- ACD 缓存：microTLB（前端）+ 分粒度 mainTLB（4K/2M/1G…）
- ATD 缓存：DDTC/PDTC 两级；S1PTC/S2PTC 独立，亦分 micro/main
- 走表并发：多槽 PTW；S1/S2 leaf 查找可 outstanding
- 预取可选：非叶子节点可配预取，降低多级 walk 长尾
- 可配裁剪：队列深度/ATC 容量/LITE 与全配，适配面积–带宽折中



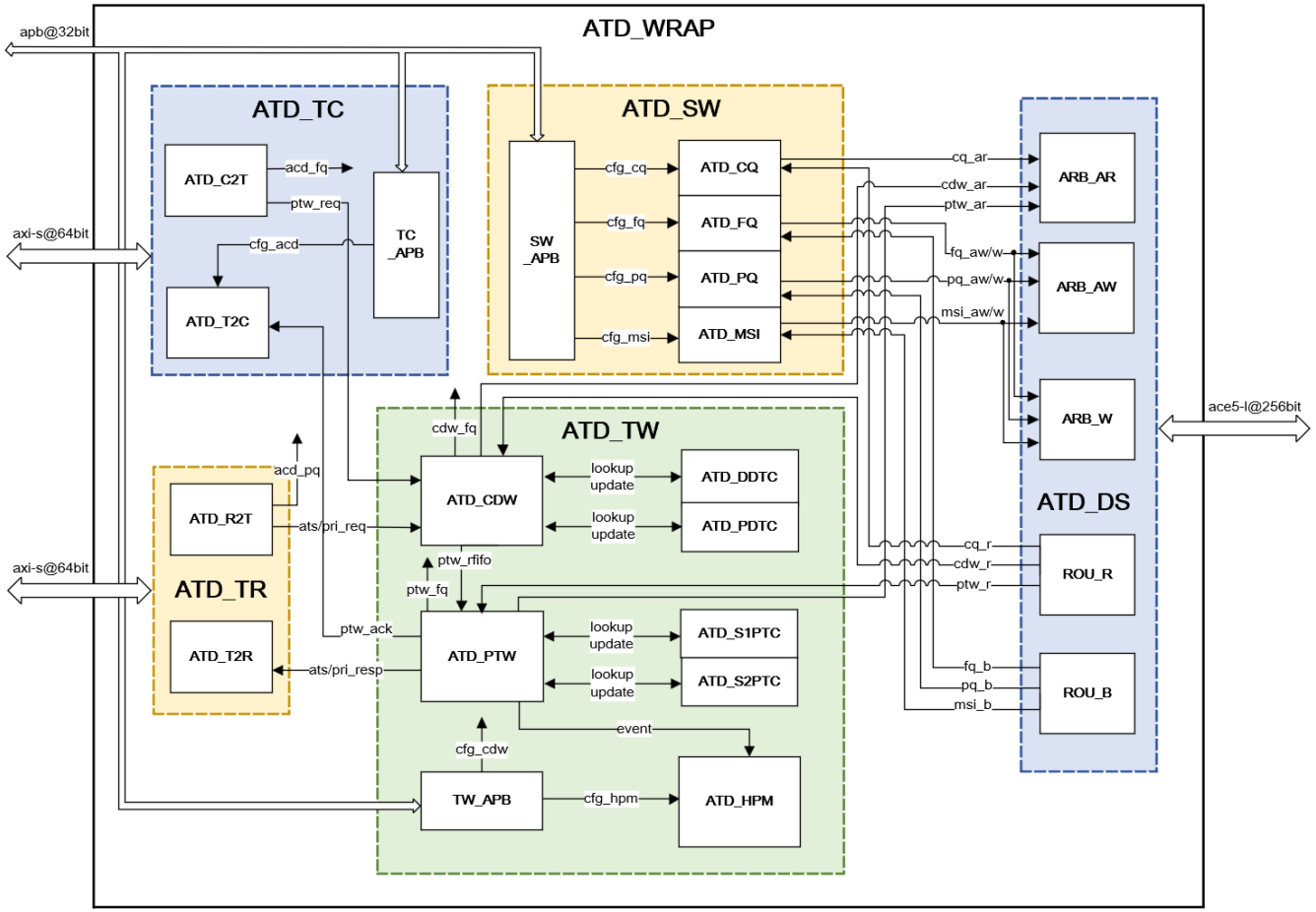
OpenIOMMU 介绍 — ACD 微架构

- **ACD：设备侧数据面与快速翻译缓存**
- **BUS_HANDLER**
- IO-Device ↔ SoC Memory 的 ACE5-Lite 接口与总线特性
- 多 outstanding（BUS_INFLY）、写缓冲与旁路通路控制
- **TRANSLATE_UNIT**
- ACD 翻译前端：microTLB + 分粒度 mainTLB（4K/2M/1G）
- 命中后修改地址输出到下游；缺失经 TMI 请求 ATD 走表
- **CFG & CTRL_WRAP**
- APB配置、CACHE 失效、故障上报、HPM、DEBUG 等控制面
- ACD 控制流 ↔ TMI 协议转换，协同 ATD
- **IOMMU_IF**
- 对接 ATD：收发 TMI Message（翻译请求/应答/失效等）



OpenIOMMU 介绍 — ATD 微架构

- **ATD：上下文获取、PTW与软件队列**
- **ATD_TC**
- 对接 ACD：收发 TMI Message（翻译请求/应答/失效等）
- **ATD_TR**
- 对接 PCIe RC：ATS 翻译请求与 PRI请求路径
- **ATD_SW**
- 软件队列 CQ/FQ/PQ 与 MSI 中断协同
- IOTINVAL/IODIR/IOFENCE、故障与缺页上报
- **ATD_TW**
- DDT/PDT 与 S1/S2 页表遍历（CDW/PTW）
- 配合 DDTC/PDTC/S1PTC/S2PTC 实现分层 cache
- **ATD_DS**
- 走表/隐式访存的数据请求路由与仲裁通道



OpenIOMMU 介绍 — 支持特性

设备/进程上下文

- DDT/PDT walk + DDTC/PDTC cache
- DC支持 1LVL/2LVL/3LVL走表
- PC支持PD8 / PD17 / PD20
- SOC集成：PRIV/DID/PID 经ACE-lite user 传输
- 支持GSCID / PSCID

页表与翻译

- Sv39 / Sv48 及 x4 两级
- Superpage（4K / 2M / 1G…）
- Bare / S1-only / S2-only / S1+S2 可配
- Svnapot/Svpbmt 虚拟化扩展
- 不支持Sv32，尚未支持sv57

PCIe / 中断

- ATS + T2GPA（capabilities）
- PRI / PQ 路径
- MSI_FLAT + MRIF
- •AMO_MRIF / AMO_HWAD

软件队列与可观测

- CQ：IOTINVAL / IOFENCE / IODIR / ATS
- FQ 故障上报
- HPM + DBG 接口
- capabilities RO = 0x1F8C7E68610（Spec 1.0）

OpenIOMMU 介绍 — 可综合现状

组件	RTL（line）	Cell count	Cell Area	Net Area	Total Area
ACD	25959	456552	319570	55454	375025
ATD	33262	405845	264065	47371	311436
Monitor+traffic	1338+585	5375	75052	817	75870
WRAP（all）	**63291**	**885271**	664620	105675	**770295**

口径说明：

- 顶层 iommu_wrap = ACD + ATD + Monitor/traffic；XiangShan 常用 CONFIG_RISCV_IOMMU_BOSC_V2_LITE
- Area/Cell 为综合结果摘录， Total Area，单位与工艺角以开芯院通算项目FIX版本原始报告一致
- LITE 下 ACD microTLB 32 项 + mainTLB 合计约 536 项，支撑高带宽多 DMA 命中路径

ACD mainTLB 深度（bank×set×way）

配置	microTLB	4K	2M	1G	512G/0T	main 合计
Default	32	2048	64	8	8	2128
LITE	32	512	8	8	8	536

另：MAIN_TLB_INFIFO = 4

ATD-cache 深度

结构	Default	LITE
DDTC	micro 16 + main 552 (512+32+8)	micro 8 + main 88 (64+16+8)
PDTC	默认复用 DDTC 几何	同 DDTC LITE
S2PTC	micro 32 + main 9368 (8192+1024+128+16+8)	micro 32 + main 296 (256+16+8+8+8)
S1PTC	默认复用 S2PTC 几何	同 S2PTC LITE

OpenIOMMU软硬件协同验证方法

QEMU + Picker 验证 OpenIOMMU 的软硬件协同方案

QEMU TCG
- Linux/Driver
- PCIe/NVMe/Mem

→

切分边界
- 仅IOMMU→Verilator
- 其余留软件

→

Picker
- DUT→.so
- C/C++ API

→

统一 iommu_api
- MMIO/DMA/step
- CModel↔RTL

→

收益
- 速度≈软件
- 观测≈RTL

相对全量 Verilator

- CPU+外设+IOMMU 全进 RTL → Linux 启动漫长
- 难覆盖真实驱动 / 实际 DMA 负载
- **本方案：仅 IOMMU 采用RTL，其余在 QEMU**
- 分钟量级起动 Linux，仍保留周期精确波形

相对 FPGA 板级

- 信号可见性弱，抓波困难、复现慢
- 改 RTL / 迭代周期长，调试成本高
- **本方案：Picker/Verilator 可 dump 全波形**
- 软件侧改配置即切 CModel/RTL 后端

相对 TLM 协同

- 依赖 SystemC / TLM-2.0 协议栈
- 需厚重 RTL↔TLM wrapper
- **本方案：Picker 直出可调用 .so / C API**
- UnityChip 工具链可规模化复用

先进性：真实软件负载 × 周期精确 DUT × CModel 逐笔对拍；定位闭环短，可直接服务香山 SoC 上芯验证。

三要点：① 软硬协同加速 ② Golden/RTL 同接口对比 ③ Picker（UnityChip）快速把 RTL 变成可 dlopen 对象

开源 RISC-V IOMMU 社区对比

维度分类	riscv-non-isa	riscv-iommu（ZDL ）	OpenIOMMU (BOSC)
【定位】角色	规范 + C 金标模型	教学/研究向 RTL 子集	可上芯高性能 RTL
【微结构】取向	无总线流水 / 带宽	串行 PTW；ATC 小	多队列 + 分层 ATC + 多 outstanding
【接口】总线	抽象 transaction	AXI4 配置/数据/走表	APB + ACE5-Lite
【功能】完备度	语义全集金标	无： ATS/PRI/PQ/Sv48/	功能齐全
【交付】产物	Spec PDF + ref model	SystemVerilog IP+ demo / tests	SV IP + 香山集成+ 协同验证路径
【验证】闭环	不对芯片周期	CVA6 / FPGA demo+ arch tests	XiangShan + QEMU / Picker / CModel
【开源】许可/仓/更新	CC-BY-4.0 riscv-non-isa/riscv-iommu 最近推送 2026-10	Apache-2.0 + SHL zero-day-labs/riscv-iommu 主仓未更新 2025-09	Mulan PSL v2 OpenXiangShan/OpenIOMMU 最近推送 2026-10

相关开源补充（非并列 RTL 主实现，但同属社区生态）：

- ZDL（ zero-day-labs ）： riscv-iommu-demo（Linux/Bao FPGA）、riscv-iommu-tests（架构测试）
- 软件栈：Linux RISC-V IOMMU driver、QEMU riscv-iommu、EDK2/UEFI IOMMU driver（演进中）
- 互补：riscv-non-isa/riscv-iopmp（物理侧防护，非 IOMMU 翻译）

OpenIOMMU 社区作用与选型优势

社区作用（Community Role）

补齐“规范金标 → 可上芯 RTL”缺口

- 相对 riscv-non-isa：规范+C 参考模型只保证语义；
- OpenIOMMU 提供可综合、可查看波形、可集成香山的硬件实现

降低开源 SoC 采用门槛

- 相对仅商业 IP：源码可审计、可改、可与香山/UnityChip 同仓演进，避免黑盒与授权门槛。

形成可复用社区资产

- RTL + 能力矩阵 + QEMU/Picker 协同路径，可供教学、研究、产品预研共同演进。

生态锚点

- 对接 OpenXiangShan/OpenIOMMU、Linux/QEMU/EDK2 等软件栈，推动 RISC-V IOMMU 从纸面到系统。

先进性（Advancement）

功能面更完整

- 相对 ZDL：Sv48/x4、ATS/PRI、HPM/DBG、MSI/MRIF 等能力更全，面向服务器/高性能 SoC。

微架构面向高带宽

- 多层 banked ATC + micro/main TLB；翻译队列/多 PTW；BUS_INFLY 多 outstanding，避免串行翻译成为瓶颈。

验证方法论先进

- Picker 将 DUT→.so，QEMU 跑真实 Linux：速度≈软件、观测≈RTL；并可与 CModel 逐笔对拍。

平台扩展可落地

- APB + ACE5-Lite、hw_bypass、LITE/全配裁剪，直接服务 XiangShan WITH_IOMMU 集成。

OpenIOMMU开源网址：

https://github.com/OpenXiangShan/OpenIOMMU

THANKS