



香山：迈向产业实践的 开源高性能 RISC-V 处理器

冯浩原

中国科学院计算技术研究所 北京开源芯片研究院

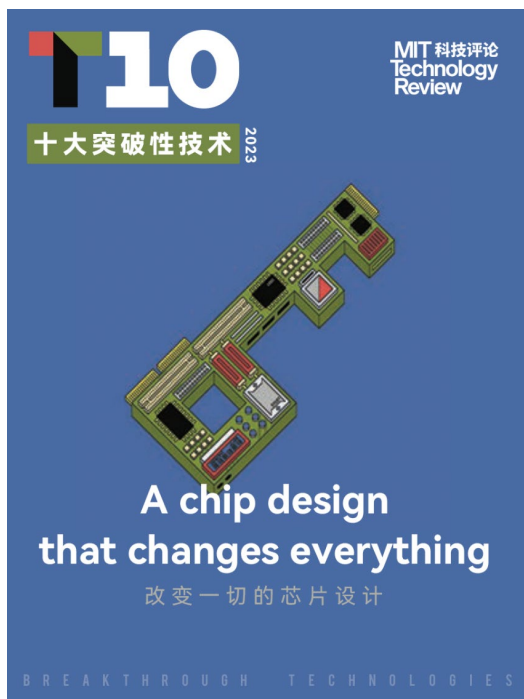
2025 年 7 月 16 日

第二届香山开源社区大会日程

时间	题目	演讲者
13:30 – 14:00	香山：迈向产业实践的开源高性能 RISC-V 处理器	冯浩原
14:00 – 14:20	XSAI：以 CPU 的编程范式支持现代 LLM 核函数	周耀阳
14:20 – 14:40	香山社区：2024 – 2025	徐泽凡
14:40 – 15:00	香山社区贡献者颁奖	
15:00 – 15:20	香山昆明湖 V3 前端架构	李衍君、徐之皓
15:20 – 15:40	香山昆明湖后端流水线的设计演进	刘泽昊
15:40 – 16:00	香山昆明湖缓存、访存的设计演进	丁昊楠
16:00 – 17:00	香山处理器与 XS-GEM5 模拟器 Tutorial	李衍君、徐之皓、 陈泱宇、杨泽辰
17:00 – 17:30	万众一芯：让每个开发者都能参与芯片验证	王俊越、李小龙

新一轮芯片设计产业变革浪潮正在来临

- RISC-V 入选《MIT 科技评论》2023 年**十大突破性技术**



《麻省理工科技评论》2023 年“全球十大突破性技术”榜单：

- 一、詹姆斯·韦伯太空望远镜 (James Webb Space Telescope)
- 二、用于高胆固醇的 CRISPR (CRISPR for high cholesterol)
- 三、制作图像的 AI (AI that makes images)
- 四、按需器官制作 (Organs on demand)
- 五、远程医疗堕胎药 (Abortion pills via telemedicine)
- 六、改变一切的芯片设计 (A chip design that changes everything)**
- 七、古代 DNA 分析 (Ancient DNA analysis)
- 八、电池回收利用 (Battery recycling)
- 九、必然到来的电动汽车 (The inevitable EV)
- 十、大规模生产的军用无人机 (Mass-market military drones)

“芯片设计正走向开放，灵活、开源的 RISC-V 有望成为改变一切的芯片设计”

迄今性能最高的开源处理器

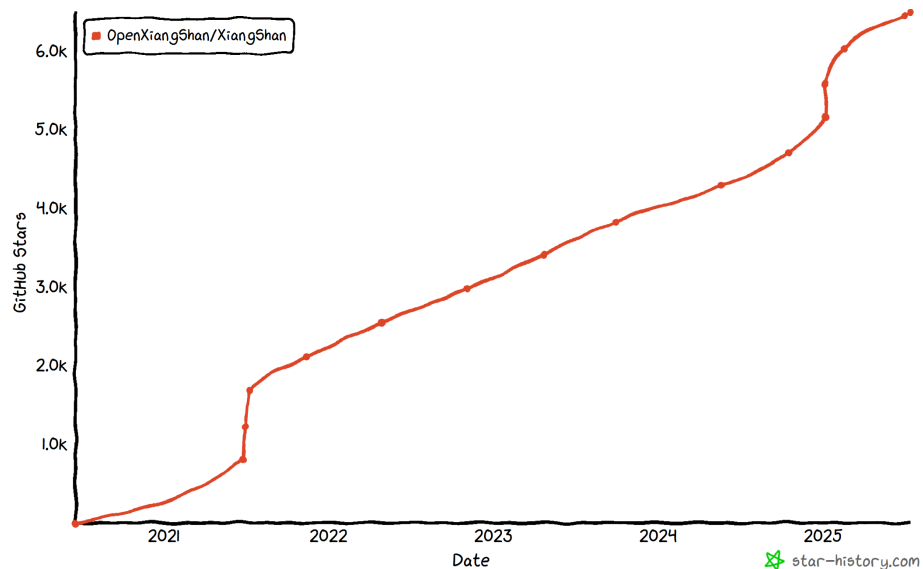
- 受到学术界、工业界的广泛关注



托管在 GitHub 上

> 6.5K stars ☆
> 780 forks 🍴

XIANGSHAN



迄今性能最高的开源处理器

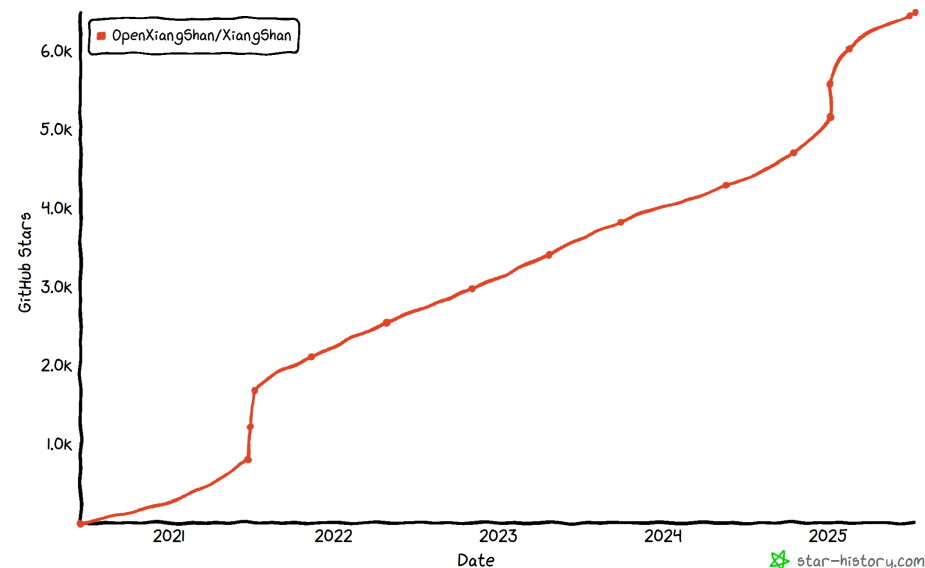
- 受到学术界、工业界的广泛关注



托管在 GitHub 上

> 6.5K stars ☆
> 780 forks 🍴

XIANGSHAN



- 愿景：建立一个像 Linux 的开源 RISC-V 核主线

- 既能支持**学术界**试验创新想法
- 也能面向**产业界**实际应用需求

“处理器界的 Linux”



V.S



XIANGSHAN

开源敏捷方法赋能高性能产业级 CPU 开发

面向产业需求的处理器

性能突出

- SPEC06 45@3GHz
- 对标 ARM 数据中心处理器 Neoverse N2
- 高性能向量 (Vector) 扩展实现

功能丰富

- 支持 RVA23 profile **全部必选项**
- 支持 AIA 标准, **Debug、Trace 扩展**
- 满足服务器场景需求

高可靠性

- UT、BT、IT、ST 多层次协同验证
- 工业级验证流程
- 开发 PLDM / FPGA **difftest 调试工具**

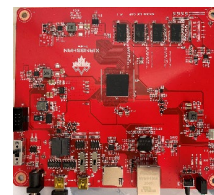
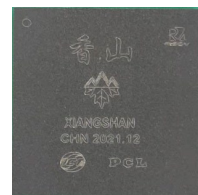
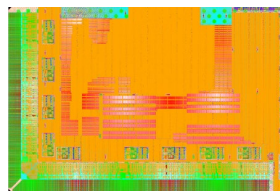
灵活可配

- 丰富的总线接口 **CHI, AXI, Tilelink**
- 高可配置的核内参数及 Cache 容量
- 满足 SoC 设计需求

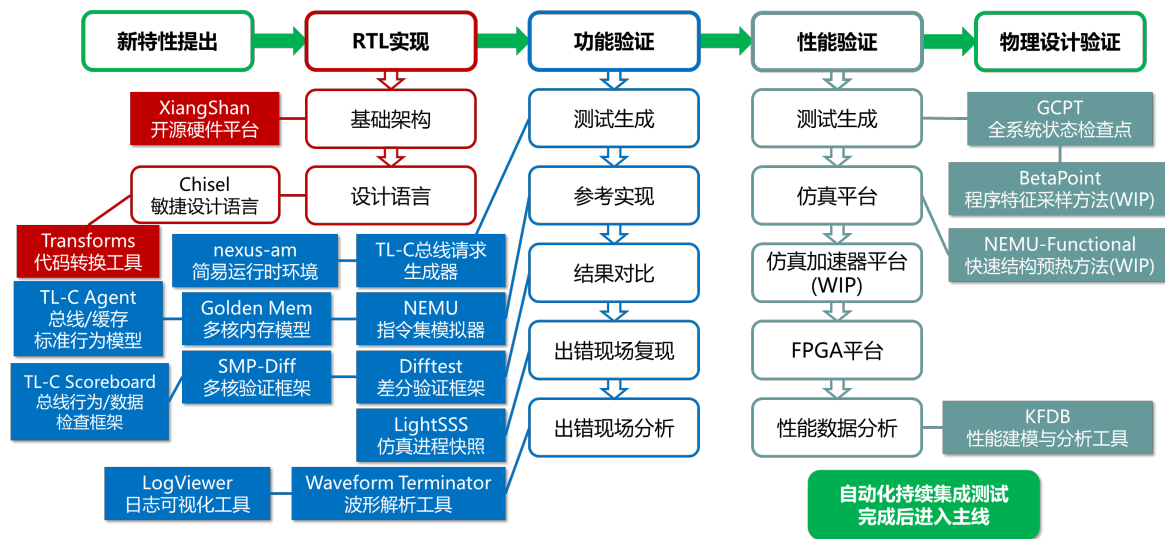
敏捷设计与验证基础设施

开源开放的处理器芯片研发能力体系

- 香山核心价值是构建一套**芯片敏捷设计基础设施**，缩短迭代优化周期
- **开源开放能力体系**，联合企业加速处理器研发节奏，支持按需快速定制芯片



成果开源



能力开放

开源敏捷方法赋能高性能产业级 CPU 开发

面向产业需求的处理器

性能突出

- SPEC06 45@3GHz
- 对标 ARM 数据中心处理器 Neoverse N2
- 高性能**向量 (RVV)** 扩展实现

功能丰富

- 支持 RVA23 profile **全部必选项**
- 支持 AIA 标准, Debug、Trace 扩展
- 满足服务器场景需求

高可靠性

- UT、BT、IT、ST 多层次协同验证
- 工业级验证流程
- 开发 PLDM / FPGA difftest 调试工具

灵活可配

- 丰富的总线接口 CHI, AXI, Tilelink
- 高可配置的核内参数及 Cache 容量
- 满足 SoC 设计需求

敏捷设计与验证基础设施

性能：开源处理器中处于领先地位

- 对标 ARM **数据中心处理器核** Neoverse N2
- **基于 SPEC06 SimPoint 的程序采样评估**
 - 编译器：GCC 12 -O3, RV64GCB, jemalloc
 - 缓存配置：64KB I/D + 1MB L2 + 16MB L3
 - 内存模型：DRAMsim3 DDR4 @ 3200MHz
 - 70ns 延迟
 - 双通道, 2 x 64

• 评估结果*

Base Score	SPECint 2006	SPECfp 2006
昆明湖V2@3GHz	44.70	44.98

SPECint 2006 est.@ 3GHz		SPECfp 2006 est.@ 3GHz	
400.perlbench	35.89	410.bwaves	67.25
401.bzip2	25.51	416.gamess	41.01
403.gcc	47.91	433.milc	44.95
429.mcf	60.21	434.zeusmp	51.71
445.gobmk	30.67	435.gromacs	33.67
456.hmmer	41.61	436.cactusADM	46.18
458.sjeng	30.65	437.leslie3d	47.77
462.libquantum	122.65	444.namd	28.87
464.h264ref	56.62	447.dealII	73.26
471.omnetpp	41.51	450.soplex	52.54
473.astar	29.34	453.povray	53.48
483.xalancbmk	72.75	454.Calculix	16.38
GEOMEAN	44.70	459.GemsFDTD	40.45
		465.tonto	36.72
		470.lbm	91.96
		481.wrf	40.64
		482.sphinx3	49.10
		GEOMEAN	44.98

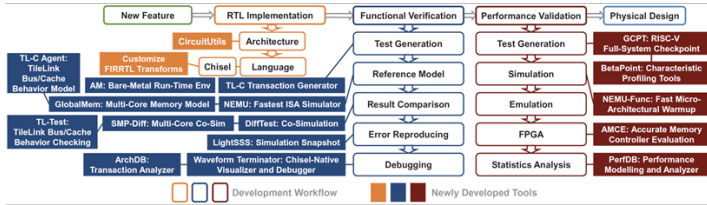
* 基于 2025 年 7 月 11 日代码版本测得的数据, commitID 6b59758, 配置为 DefaultConfig



构建芯片敏捷设计新方法与新流程

• 自研超过 17 个新工具解决敏捷开发难题，在国际上形成独特优势

芯片敏捷设计新流程



2022 年精选体系结构论文 (Top Pick)

THEME ARTICLE: TOP PICKS FROM THE 2022 COMPUTER ARCHITECTURE CONFERENCES

Towards Developing High-Performance RISC-V Processors Using Agile Methodology

Yinan Xu¹ and Zhao Yu¹, State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China

Dan Tang², Beijing Institute of Open Source Chip, Beijing, 100080, China

Ye Ca³, Shenzhen University, Shenzhen, 518060, China

Dandan Huan⁴, Beijing VCore Technology, Beijing, 100190, China

Wei He⁵, Peng Cheng Laboratory, Shenzhen, 518060, China

Ninghui Sun⁶ and Yungang Bao⁷, State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China

Agile chip design methodology has shown promise for sustaining the scaling of computing performance more efficiently. However, the practical application of this methodology has been limited by major obstacles. This article presents MinJie, an open source platform supporting agile hardware development flow. We demonstrate the usage and effectiveness of MinJie by building two generations of XiangShan, an open source RISC-V processor with industry-competitive performance. This article highlights the potential impact of MinJie and XiangShan in advancing the field of agile processor design and development as well as the potential for open source collaboration to democratize the field and drive innovation forward.

In the data-driven world, chip architects need to carefully balance conventional design metrics, such as performance, power, and area (PPA), against other metrics, such as time to market and limited budgets, when designing high-performance processors. This requires novel methods and tool flows that oversee and facilitate the entire process.

In recent years, the agile chip design methodology has attracted attention from both academia and industry. These approaches have the potential to reduce the significant engineering costs and long design cycles for chip development while maintaining reasonable PPA specifications. There is an increasing amount of research^[1-3] demonstrating the effectiveness of agile methodologies in designing CPU cores, systems on chip, and hardware accelerators.

Despite its benefits, the use of agile methodology still needs to be expanded, especially for practical modern

designing CPU cores [4], [5], [6], System-on-Chips [7], [8] and hardware accelerators [9], [10]. In addition, several agile development frameworks [11], [12], [13] are proposed to support rapid chip design, simulation and testing. Despite the benefits, agile design methodology is still of limited usage, especially for practical modern designs. To further investigate the issue, we talk to chip architects, research scientists and lead engineers from twelve hyperscalers and major chip companies across multiple

0272-1752 © 2023 IEEE
Digital Object Identifier 10.1109/MCM.2023.3273562
Date of publication 9 May 2023; date of current version 29 June 2023.

持续形成新方法和工具

面向 RISC-V CPU 的覆盖率导向模糊测试

- 原理: ①初始种子 → ②遗传算法 → ③新测试输入 → ④探索检查
- ← ⑤覆盖率反馈

面向微结构组件: 细粒度XS-TopDown性能分析模型

- 性能分析的技术挑战
- 指令集碎片化, 没有ARM工作不兼容的指令
- 微结构碎片化, 指令互斥造成数据流异构
- 流水线处理碎片化导致负载不均, 取指有性能模型分析的难度

ChiselDB: 敏捷硬件调试工具

- 使用软件工程的方法, 提高硬件化信息传递的自动化程度
- 案例: 针对关键微结构信息的调试方案

指令维度展开: TIP/TEA 指令级微观性能分析工具

- TIP/TEA性能分析方法
- 核心原理: 指令级别的运行周期和互斥状态指令, 形成指令维度颗粒度的 CPU Stack
- 分析效果: 能够细化到指令级状态指令对于处理突发性能问题的“洗刷”
- 适配 TIP/TEA 性能分析方法论驱动设计流程
- 针对流水线处理微结构优化性能数据, 精细化分析数据

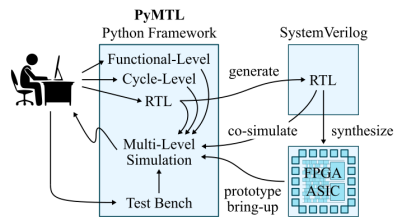
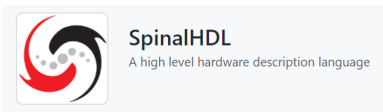
时间维度展开: XSPerRolling 微观性能分析工具

- 支持性能指标的 rolling 曲线 (滚动时间窗口) 采集与可视化工具
- 方法原理的流程图
- 性能指标分析
- 性能指标并发的性能指标
- 性能指标采集数据: Critical Region
- 性能指标采集 → 使用滚动时间窗口 ChiselDB 工具

发表于体系结构顶会 MICRO 2022

基于高层次硬件语言 Chisel 进行开发

CHISEL



bluespec

cucapra/latte21
Languages, Tools, and Techniques for Accelerator Design

Scala

python

OCaml

Verilog

SystemVerilog

敏捷硬件设计语言

新型硬件设计范式

抽象层

高抽象层次的
硬件描述语言

硬件描述语言

描述层

处理器体系结构

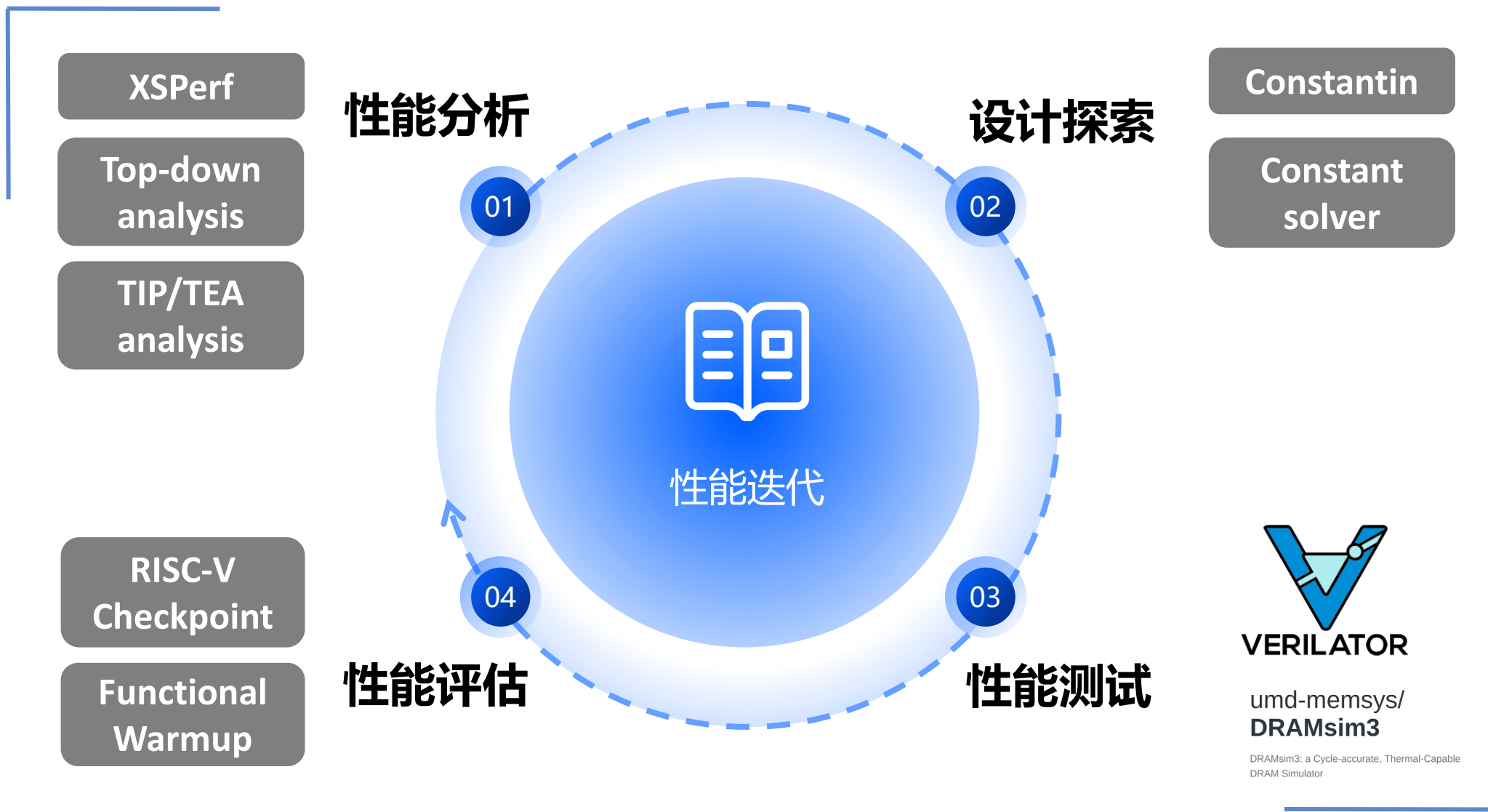
结构层

逻辑电路

电路层

敏捷硬件设计方法

敏捷性能迭代循环



持续迭代性能探索工具



Constantin



XSPerfRolling



TIP/TEA

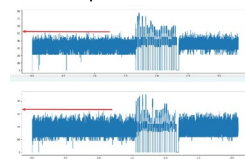
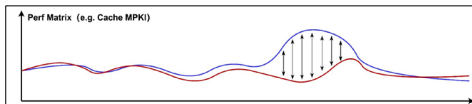
时间维度展开: XSPerfRolling 微观性能分析工具

• 支持性能指标的 Rolling 曲线 (滚动时间窗口) 采集与可视化工具

- + 方便快捷的添加流程
- + 可视化性能分析
- + 精细对比片段间的性能差异
- + 帮助锁定参数敏感的 Critical Region
- 存储开销变大 → 使用仿真时数据库 ChiseDB 缓解

```
XSPerfRolling(  
  cacheParams, "L2PrefetchAccuracy",  
  PopCount(L2prefetchMisses), PopCount(L2prefetchHits),  
  l2m, clock, reset
```

编译 + 仿真 + 脚本分析



- 运行时设计参数调控
- 敏捷设计空间探索

- 性能指标可视化
- 成功定位微架构瓶颈

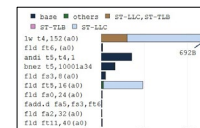
指令维度展开: TIP/TEA 指令级别微观性能分析工具

• TIP/TEA性能分析方法

- 核心思路: 将处理器的运行周期归因至静态指令, 形成指令维度细粒度的 CPI Stack
- 分析效果: 能够细化到每条静态指令对于处理器发生性能阻塞的“贡献”

• 适配 TIP/TEA 性能分析方法论至香山设计流程

- 针对香山处理器微架构优化性能计数条目, 细化分层归因



TIP/TEA分析结果

PC	weight	breakdown	id
PC: 0x112136	weight: 109244.5	breakdown: [9522.5, 99722, 0, 0]	ld a5,0(a4)
PC: 0x112166	weight: 77884.5	breakdown: [15298.5, 62586, 0, 0]	thi a5,0(s0)
PC: 0x118c08	weight: 64953.5	breakdown: [1539.5, 63414, 0, 0]	thi a4,0(a0)
PC: 0x118c08	weight: 49580.5	breakdown: [1851.5, 47729, 0, 0]	ld a4,16(a5)
PC: 0x118c0c	weight: 47718.0	breakdown: [1355.0, 46363, 0, 0]	ld a0,8(a5)
PC: 0x11210c	weight: 46610.0	breakdown: [10698.0, 35912, 0, 0]	ld ra,24(sp)
PC: 0x112112	weight: 42427.0	breakdown: [5894.0, 36533, 0, 0]	thi a0,5(a0) # 26228
PC: 0x3e3e30	weight: 42988.5	breakdown: [2627.5, 40361, 0, 0]	thi a4,0(a0)
PC: 0x118f50	weight: 36694.0	breakdown: [642.0, 36052, 0, 0]	ld s0,0(a3)
PC: 0x112164	weight: 30150.0	breakdown: [4926.0, 25224, 0, 0]	ld a5,8(a0)

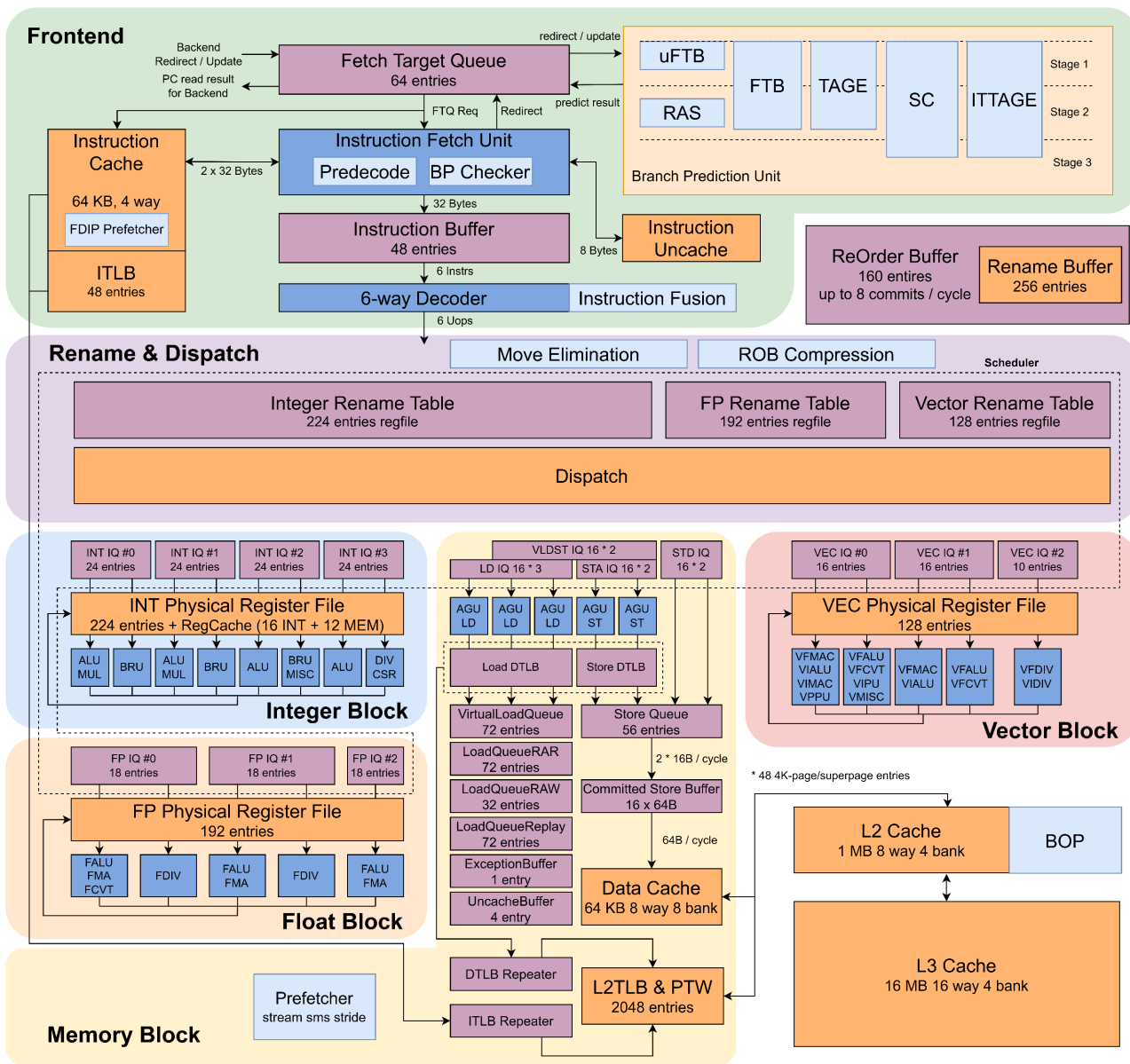
GCC 测试片断在香山上仿真得到的高权重指令信息

* B. Gottschall, L. Eeckhout, and M. Jahre, "TIP: Time-Proportional Instruction Profiling," MICRO-54
* B. Gottschall, L. Eeckhout, and M. Jahre, "TEA: Time-Proportional Event Analysis," ISCA-2023

- 性能归因至指令
- 缩小性能瓶颈范围



昆明湖微架构总览



• 解耦合的前端设计

- 减少取指空泡
- 更好地指导指令预取

• 充足的乱序执行窗口

- 更大容量的 ROB 及访存队列

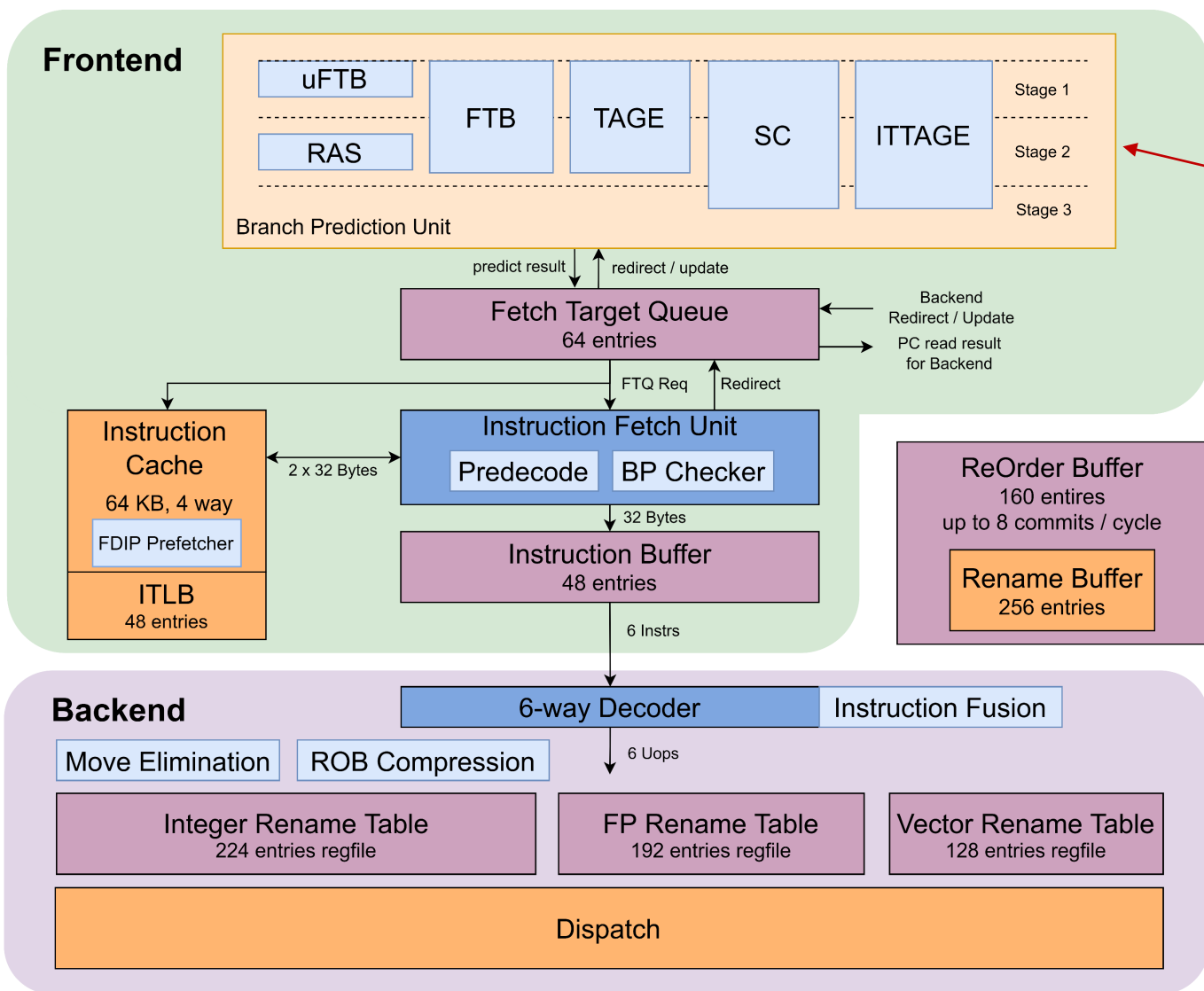
• 低延迟、高带宽的 Cache 访问

- 与 L1 / L2 Cache 紧耦合的访存流水线
- 分 Bank 的 Cache 设计
- 实现多级复合预取器

• 功能支持

- 支持全部 RISC-V RVA23 profile 必选项
- 高性能的向量 (Vector) 扩展实现
- 可配置的 CHI Issue E.b / B / C 总线协议

昆明湖微架构 — Frontend

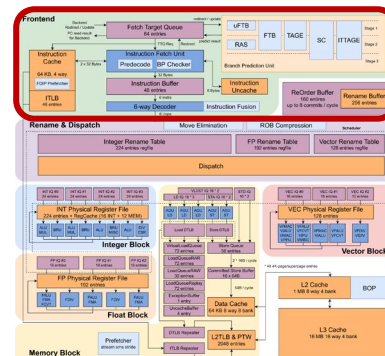


多级分支预测器

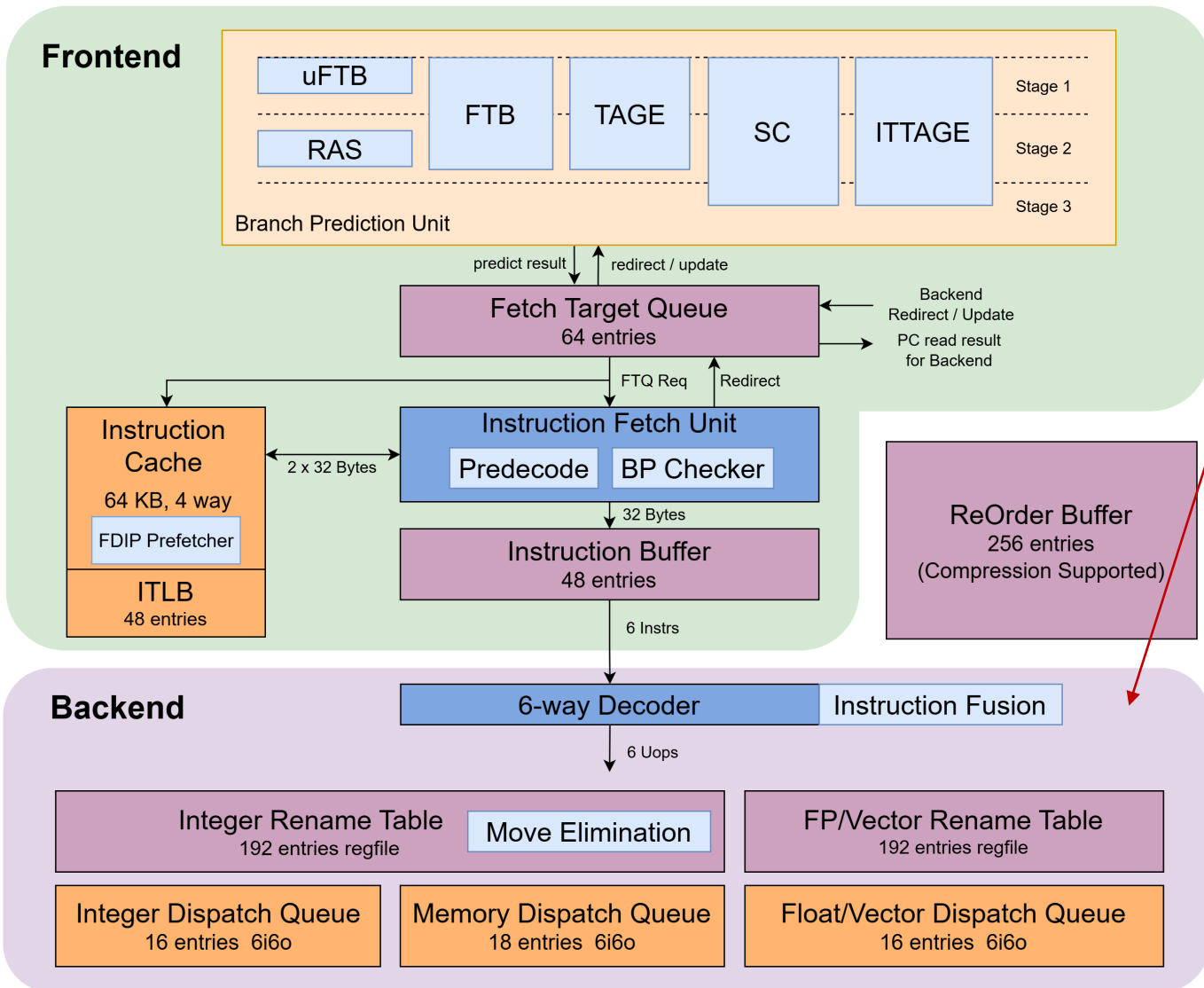
- 32 项 μ FTB
- 2K 项 FTB、以及可选的 L2FTB
- 16K 项 TAGE-SC 分支方向预测器
- 2K 项 ITTAGE 间接跳转预测器
- 48 项返回地址栈 (RAS)

指令缓存与 ITLB

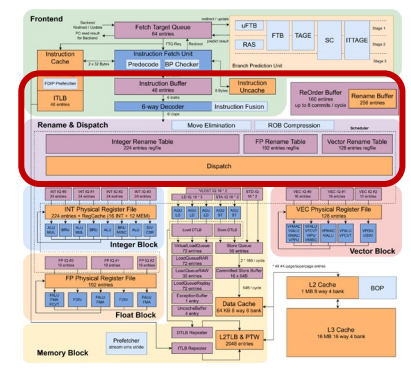
- 64KB 4 路组相联 ICache
- 48 项 ITLB
- 实现 FDIP 预取器



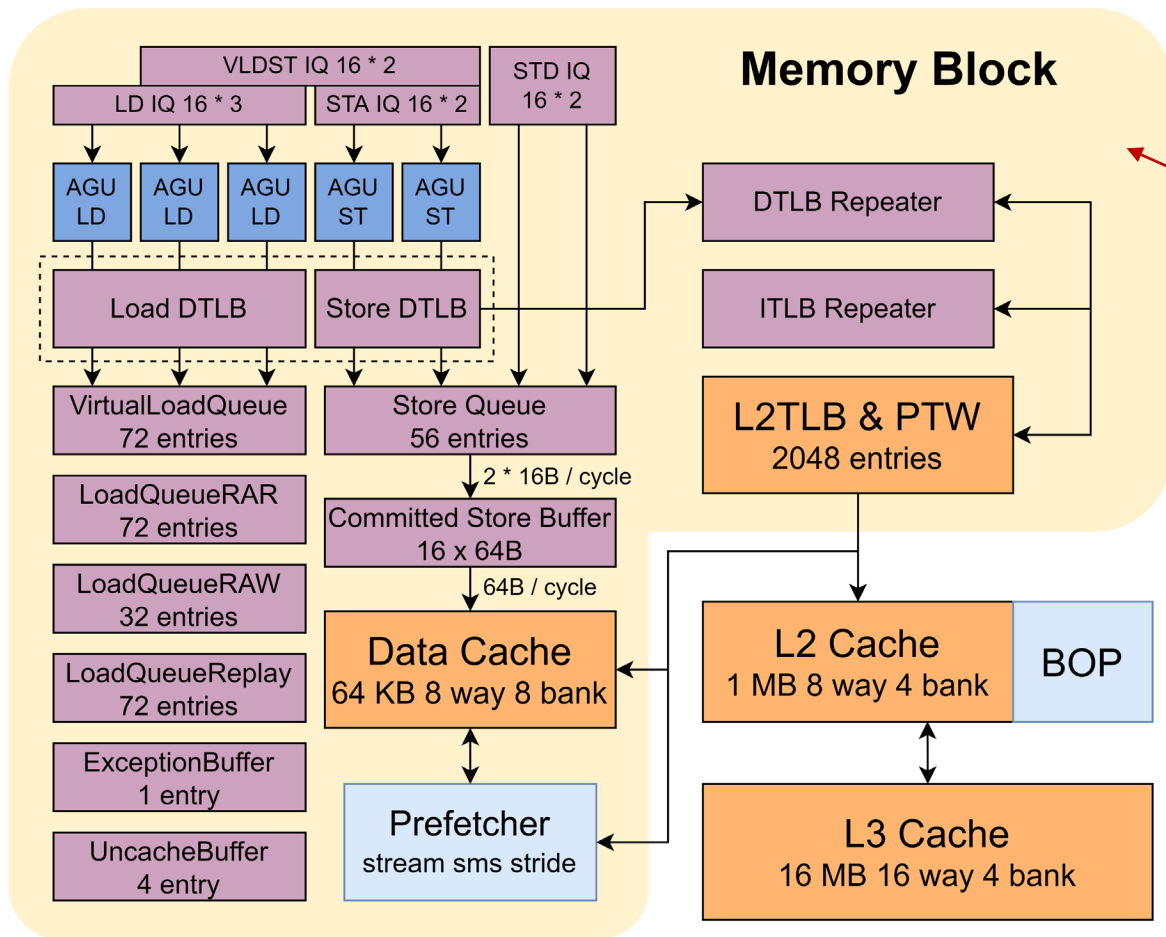
昆明湖微架构 — Backend



- ## 6 发射译码/重命名/分派
- ### 寄存器重命名
- 224 项整数寄存器堆
 - 192 项浮点寄存器堆
 - 128 项向量寄存器堆
 - 支持基于 RAT 的 Move 消除、及指令融合
- ### 160 项重排序缓冲 (ROB)
- 支持 ROB 压缩, 最多压缩 6 uop / 项
 - 每周最多提交 8 条指令
 - 通过检查点回滚的方式实现重定向恢复
- ### 256 项 Rename Buffer
- 将寄存器映射信息从 ROB 中独立存储
 - 解耦指令提交与重命名表更新过程



昆明湖微架构 — MemBlock



访存流水线与队列

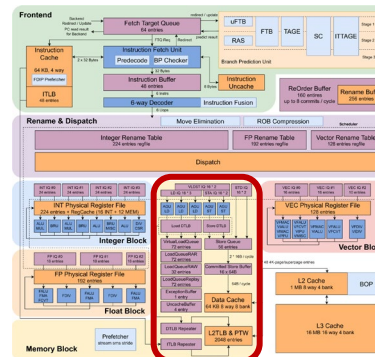
- 3 条 load 流水线、72 项 LoadQueue
- 2 条 store 流水线、56 项 StoreQueue

内存管理单元

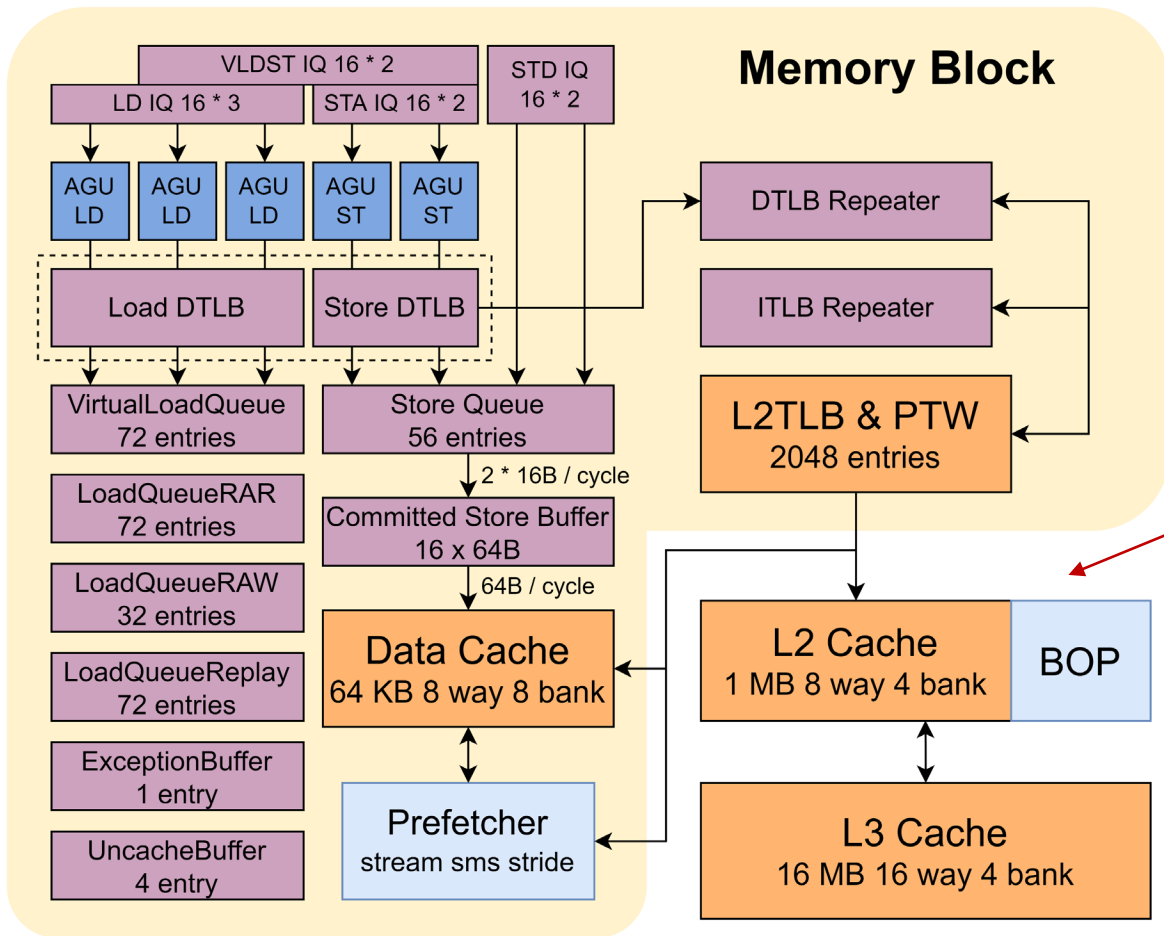
- 支持 Sv48 地址翻译模式 / 48 位物理地址
- 48 项 DTLB, 2K 项 L2TLB
- 支持至多 7 条并行的 Page Walk 请求
- 支持 RVH (Hypervisor) 扩展

数据缓存

- 64KB 8-way VIPT DCache (硬件处理别名)
- 复合预取器
 - L1: Stream & Stride & SMS 预取器
 - L2: BOP 预取器
 - L3: Temporal 预取器



昆明湖微架构 — L2 / L3 Cache

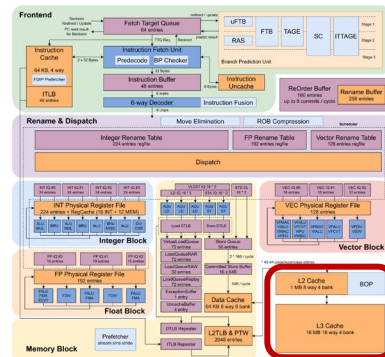


私有 L2 缓存

- 8 路组相联，最大容量 **1MB** / core
- 支持最多 **64+** 个并行请求 (可配置)
- Load-to-Use 延迟约 **10** 周期
- 支持 PLRU / RRIP 替换策略

共享 L3 缓存

- 16 路组相联，最大容量为 **16MB**
- 支持对 L2 配置为 Inclusive 或 Non-inclusive
- 支持最多 **150+** 个并行请求 (可配置)
- Load-to-Use 延迟约 **30** 周期
- 支持 PLRU / RRIP 替换策略



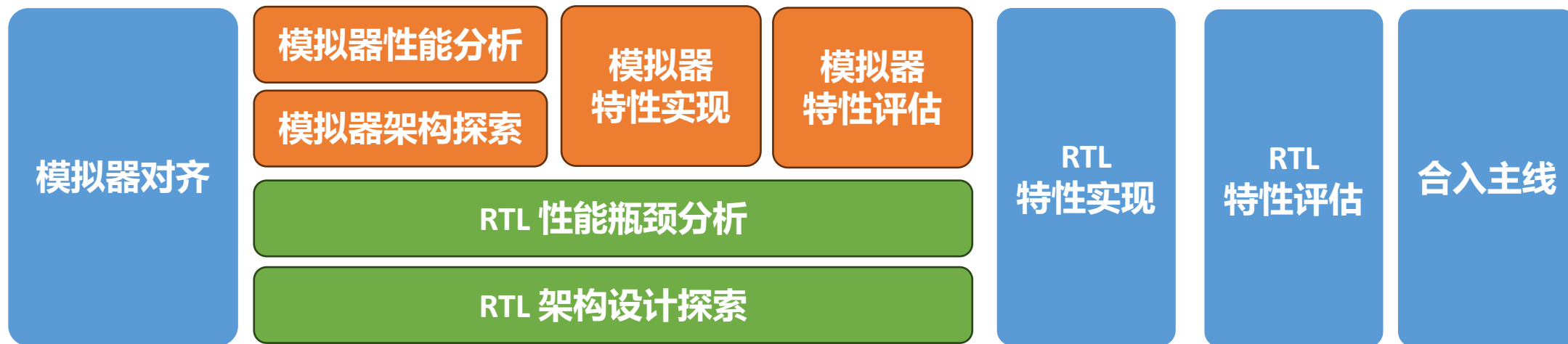
XS-GEM5 微结构模拟器

- 基于开源模拟器 GEM5 实现
- **与香山微架构对齐**，使得其探索结果对香山的性能迭代有指导意义

	香山	XS-GEM5
编译速度	45 min	1 min (增量)
运行速度	3 - 4 KHz	100 KHz

赋能快速设计空间探索

■ 模拟器探索流程 ■ RTL 探索流程



性能迭代 workflow

🏔️ 模拟器与 RTL 的性能对齐

- 原理：基于**外部观测**获取**处理器微结构**的关键特征参数，实现架构对齐
- 工作：设计并实现了一套面向**模拟器-RTL 微结构对齐**的 microbenchmarks

观测角度

流水线和调度

- 指令片段运算延迟
- 操作数依赖的除法运算、调度延迟
- 指令间调度延迟
- Decode 宽度
- Rename 宽度
- Dispatch 宽度

Cache & TLB

- Load-to-use 延迟
- Load-Load 调度延迟
- L1/L2/L3 访问延迟
- L1/L2/L3 访问带宽
- L2 TLB 访问延迟
- L2 TLB 对各级缓存、内存的访问延迟

分支预测和预取

- 分支误预测后冲刷指令的时间代价
- 不包含跳转的长指令序列测试
- Stride 预取器
- Stream 预取器
- PBOP / VBOP 预取器

难对齐项

- 公版 GEM5 缺少对流水线的细节建模
- 较理想化的分派算法和调度策略
- 访存流水线和队列的重发机制
- L2Cache 主流流水线
- 持续对齐 RTL 细节

基础设施：自动化编译 + 运行 + 对比分析流程

🏔️ 模拟器与 RTL 的性能对齐

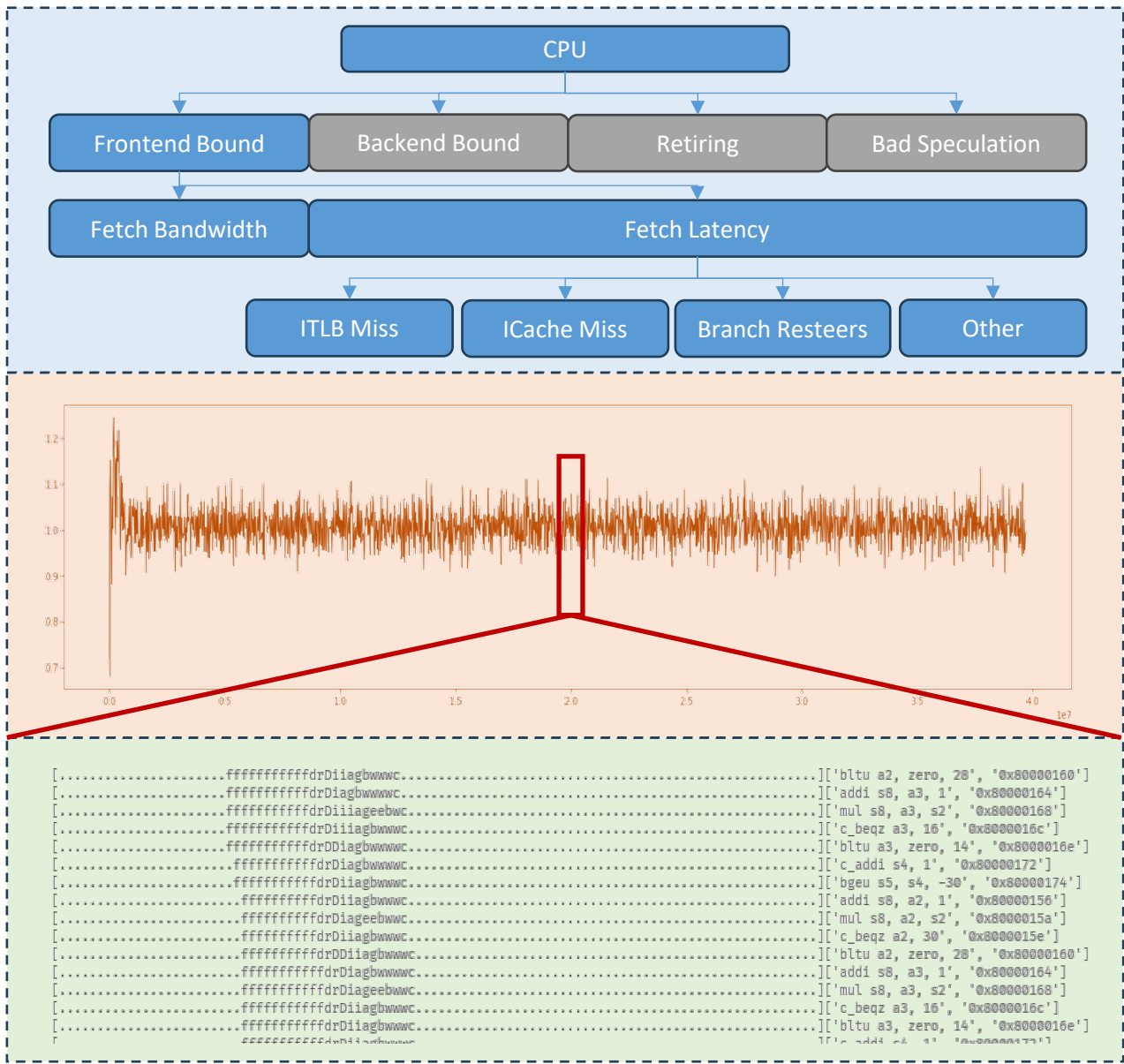
TopDown



IPCRolling



PerfCCT



模拟器与 RTL 对齐效果

平均误差 < 1.84%

SPEC06 int / GHz, 未开启预取			
	香山	GEM5	DIFF
400.perlbench	9.93	9.99	0.55%
401.bzip2	6.61	6.72	1.66%
403.gcc	10.63	10.61	-0.15%
429.mcf	7.82	7.91	1.18%
445.gobmk	9.57	10.26	7.22%
456.hmmmer	13.56	13.73	1.20%
458.sjeng	10.34	10.70	3.51%
462.libquantum	9.86	10.00	1.46%
464.h264ref	18.47	18.31	-0.87%
471.omnetpp	7.15	7.45	4.21%
473.astar	8.89	9.09	2.32%
483.xalancbmk	7.46	7.45	-0.05%
GEOMEAN	9.62	9.80	1.84%

欢迎大家加入 XS-GEM5 的开发

- XS-GEM5 仓库: <https://github.com/OpenXiangShan/GEM5/>
- 对应文档: <https://xs-gem5.readthedocs.io/zh-cn/latest/>

XS-GEM5 仓库



XS-GEM5 文档



Thanks

In the development of the Kunming Lake v3 performance model, developers Zhang Qianlong, Zhang Lutong, Hu Kai, Yan Yuwei, Liu Shuquan, and Chen Dewei from Tencent Penglai Laboratory jointly made the following contributions:

1. Contributed and debugged the MGSC predictor model, which reduced MPKI by approximately 10% and improved overall performance by 0.3-0.4 points/GHz in the SpecInt 2006 checkpoint tests.
2. Contributed and debugged the L2 Next Prefetcher model, which improved MCF performance by approximately 23% and overall performance by 1.7% in the SpecInt 2006 checkpoint tests.
3. Contributed the model codes for the L2 Adjacent Block Prefetcher and Sector Cache.

We are grateful for their contributions and hope to continue to work together to improve the performance of XS-GEM5.

开源敏捷方法赋能高性能产业级 CPU 开发

面向产业需求的处理器

性能突出

- SPEC06 45@3GHz
- 对标 ARM 数据中心处理器 Neoverse N2
- 高性能向量 (RVV) 扩展实现

功能丰富

- 支持 RVA23 profile **全部必选项**
- 支持 AIA 标准, **Debug、Trace 扩展**
- 满足服务器场景需求

高可靠性

- UT、BT、IT、ST 多层次协同验证
- 工业级验证流程
- 开发 PLDM / FPGA difftest 调试工具

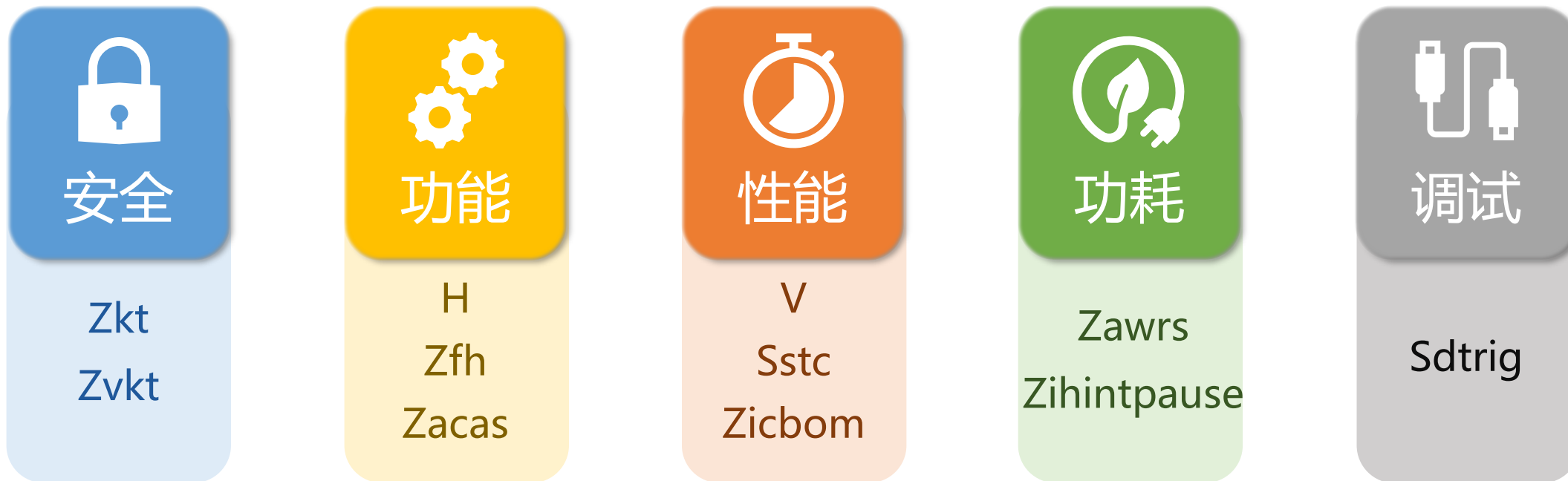
灵活可配

- 丰富的总线接口 CHI, AXI, Tilelink
- 高可配置的核内参数及 Cache 容量
- 满足 SoC 设计需求

敏捷设计与验证基础设施

功能：支持最新的 RVA23 Profile

- RVA23：一组标准化指令集扩展
 - 2024年10月由RISC-V国际基金会通过
 - 支持**更丰富的软件生态**，提高软件兼容性，对于服务器生态至关重要



昆明湖支持 RVA23 Profile 全部必选扩展和部分可选特性

产业级处理器的功能支持

- AIA: 支持虚拟化和优先级的高级中断架构
- Trace: 支持压缩记录指令执行路径



香山昆明湖后端流水线的 设计演进

刘泽昊^{3,1}张紫飞¹胡轩¹张梓悦¹唐浩晋¹贾志杰¹林志达¹
吉骏雄¹曾锦鸿¹肖飞豹²程光辉²游朝阳²张铭旭⁴

¹中国科学院计算技术研究所 ²北京开源芯片研究院

³中国科学院大学 ⁴西安交通大学

2025年7月16日@第五届 RISC-V 中国峰会

产业级处理器的功能支持

- AIA: 支持虚拟化和优先级的高级中断架构
- Trace: 支持压缩记录指令执行路径
- Svpbmt: 支持页表覆盖 PMA 标记的原始物理内存类型
- Zicclsm: 支持硬件非对齐 load / store 访问

 中国科学院计算技术研究所  中国开放指令生态 (RISC-V) 联盟  北京开源芯片研究院



香山昆明湖后端流水线的设计演进

刘泽昊^{3,1}张紫飞¹胡轩¹张梓悦¹唐浩晋¹贾志杰¹林志达¹
吉骏雄¹曾锦鸿¹肖飞豹²程光辉²游朝阳²张铭旭⁴

¹中国科学院计算技术研究所 ²北京开源芯片研究院

³中国科学院大学 ⁴西安交通大学

2025 年 7 月 16 日 @ 第五届 RISC-V 中国峰会

 中国科学院计算技术研究所  中国开放指令生态 (RISC-V) 联盟  北京开源芯片研究院



香山处理器访存、缓存架构 解决方案与演进

丁昊楠¹刘威丁²

¹中国科学院计算技术研究所 ²北京开源芯片研究院

2025 年 7 月 16 日

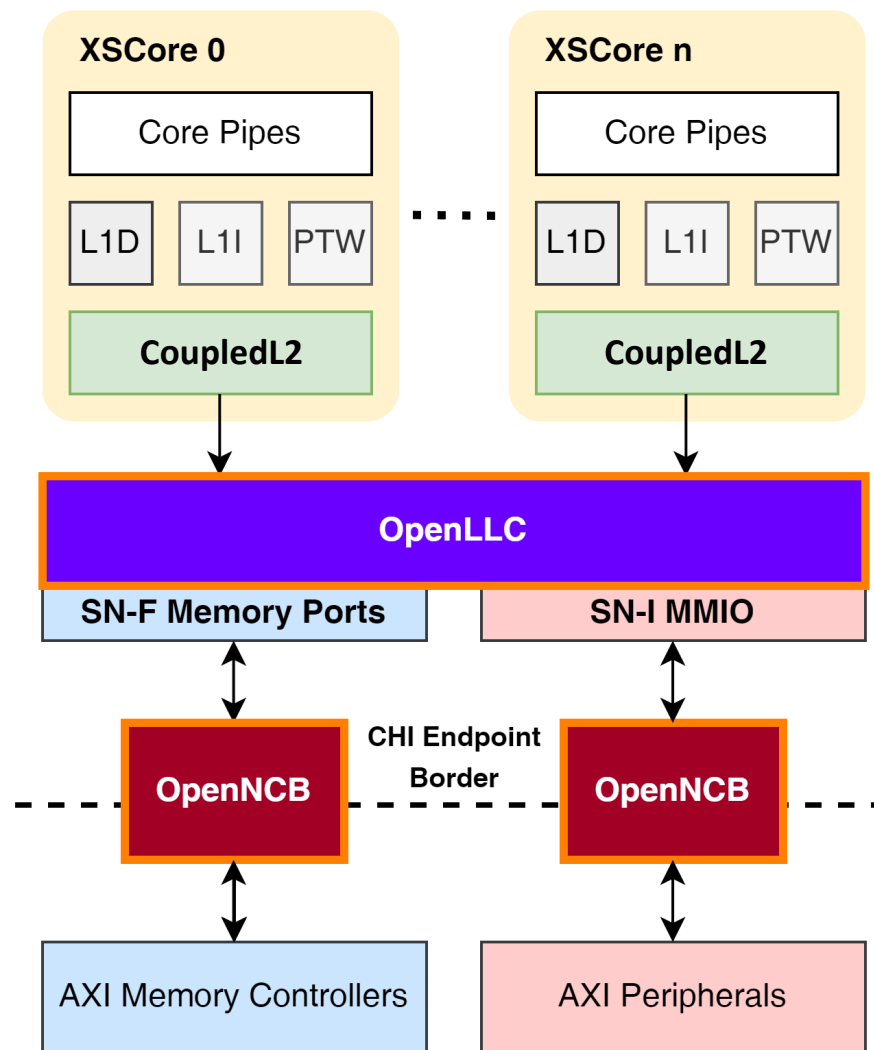
🔥 开源 CHI SoC IP 实践

- **CoupledL2: 高可扩展一致性 L2 缓存**
 - 支持 TileLink、CHI Issue B / C / E.b 总线协议
- **CHI SoC IP 和基础设施**
 - OpenNCB: 开源高性能 CHI to AXI4 协议桥
 - OpenLLC: 开源高性能 CHI HN-F 实现
 - CHIron: 开源 AMBA CHI 基础设施
- **敬请期待:**

ICIT 中国科学院计算技术研究所
CRA 中国开放指令生态 (RISC-V) 联盟
IS 北京开源芯片研究院

**香山处理器访存、缓存架构
解决方案与演进**

丁昊楠¹ 刘威丁²
¹中国科学院计算技术研究所 ²北京开源芯片研究院
2025 年 7 月 16 日



基于香山的机密虚拟机合作成果

- 中科院信工所团队为昆明湖架构贡献“虚拟机级可信执行环境”
 - 基于香山开放的敏捷设计方法和流程
 - 保护使用中的数据 and 代码，免受未经授权的访问或篡改

CVM Development #3980

 Open

yxtx1994 wants to merge 22 commits into `OpenXiangShan:master` from `yxtx1994:dev-cvm` 

```
/*
 * Copyright (c) 2024 Institute of Information Engineering Chinese Academy of Sciences
 *
 * XiangShan is licensed under Mulan PSL v2.
 * You can use this software according to the terms and conditions of the Mulan PSL v2.
 * You may obtain a copy of Mulan PSL v2 at:
 *     http://license.coscl.org.cn/MulanPSL2
 */
```

开源敏捷方法赋能高性能产业级 CPU 开发

面向产业需求的处理器

性能突出

- SPEC06 45@3GHz
- 对标 ARM 数据中心处理器 Neoverse N2
- 高性能向量 (RVV) 扩展实现

功能丰富

- 支持 RVA23 profile **全部必选项**
- 支持 AIA 标准, Debug、Trace 扩展
- 满足服务器场景需求

高可靠性

- UT、BT、IT、ST 多层次协同验证
- 工业级验证流程
- 开发 PLDM / FPGA difftest 调试工具

灵活可配

- 丰富的总线接口 CHI, AXI, Tilelink
- 高可配置的核内参数及 Cache 容量
- 满足 SoC 设计需求

敏捷设计与验证基础设施

处理器验证是芯片开发的瓶颈环节之一

206%

2007年至今
验证工程师增长

5:1

验证：设计
工程师数量比

86%

首次流片未成功的
芯片项目占比

75%

超过预期时间的
芯片项目占比

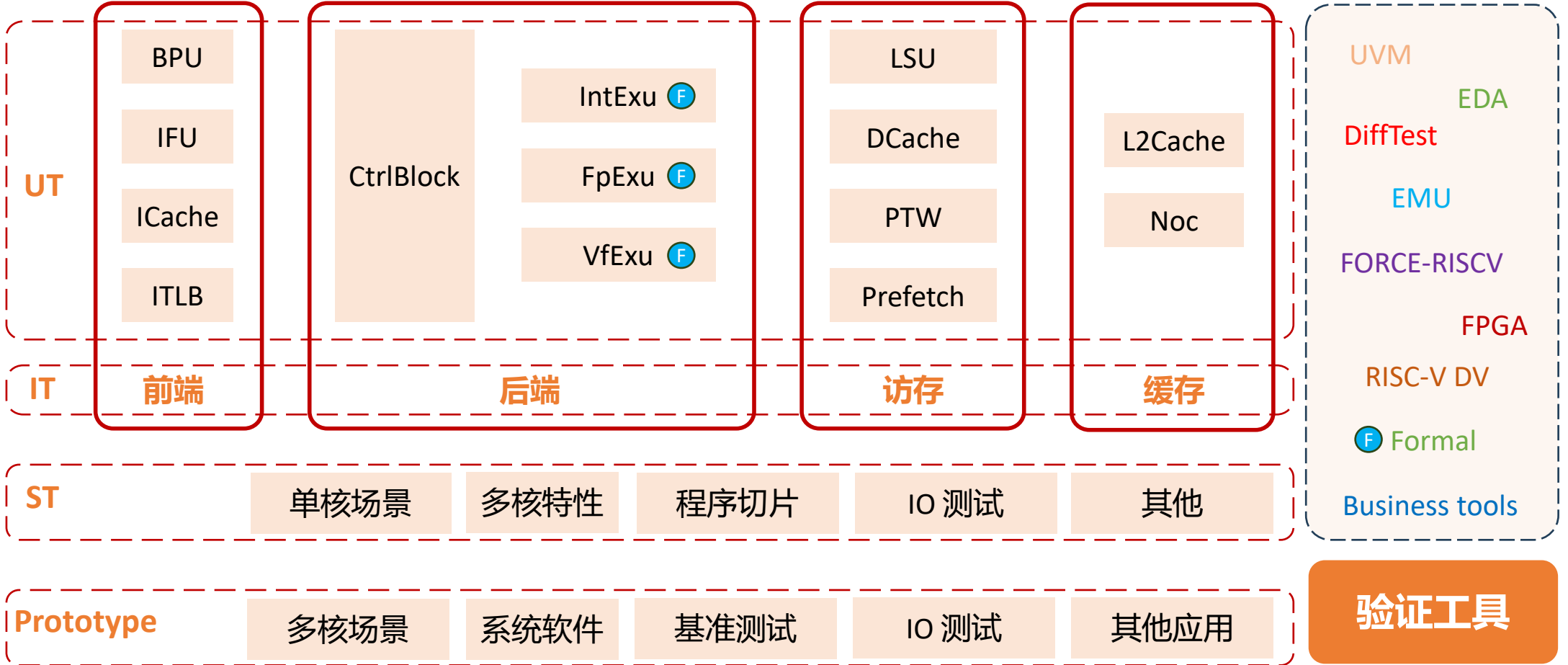
尽管投入大量资源，芯片验证质量和效率仍未达预期

产业级验证

验证流程

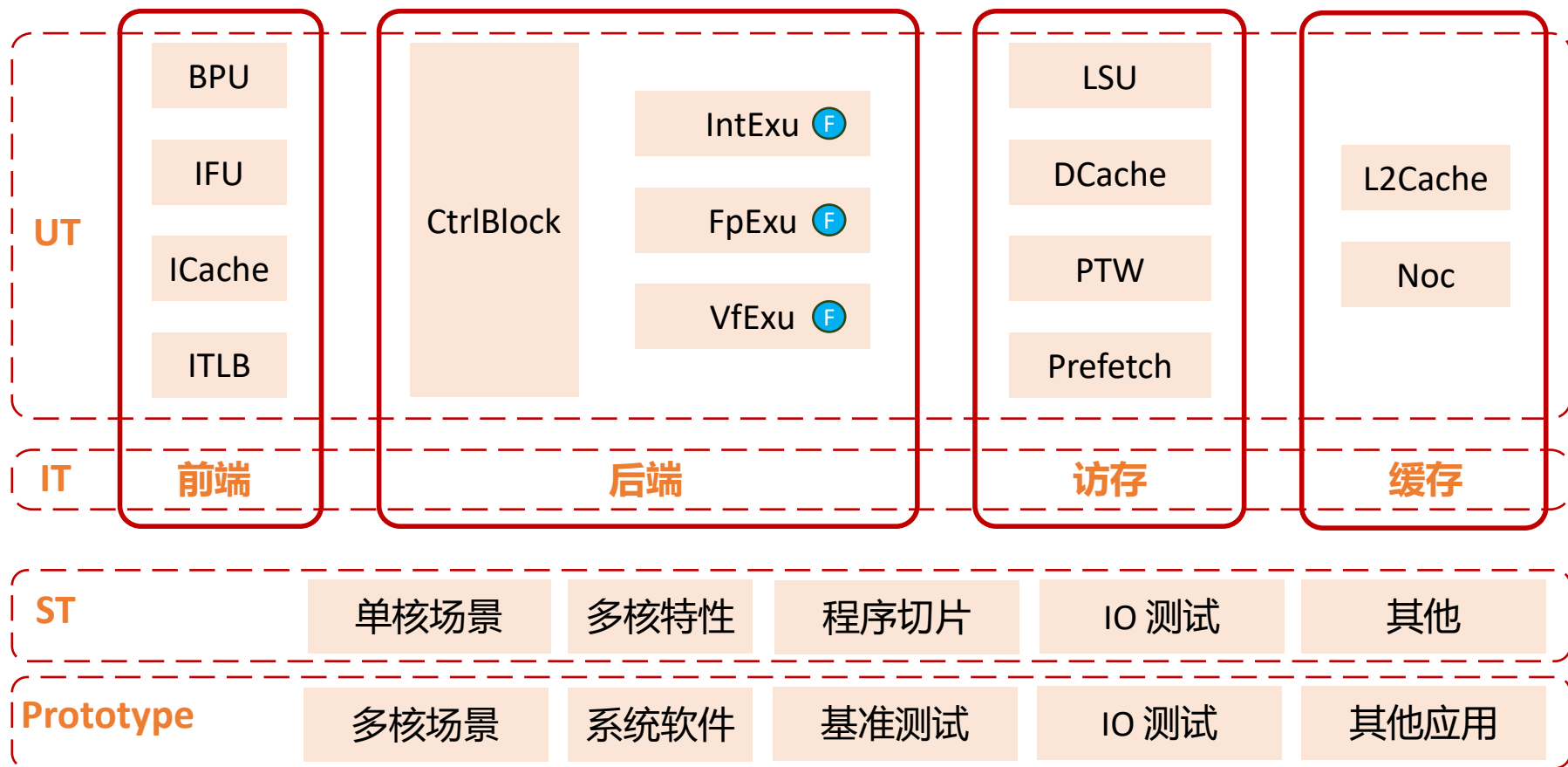


验证层级



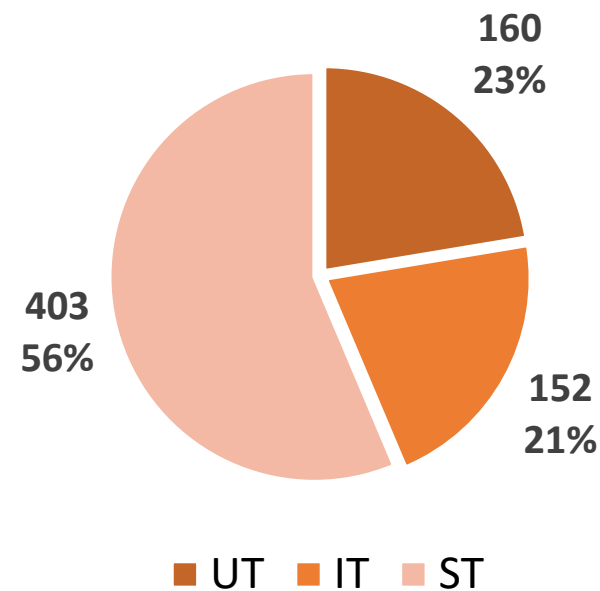
产业级验证流程

- 面向实际应用场景，发现**真实系统**中的 Bug
- 保证模块级**覆盖率**水平，提升验证完备性



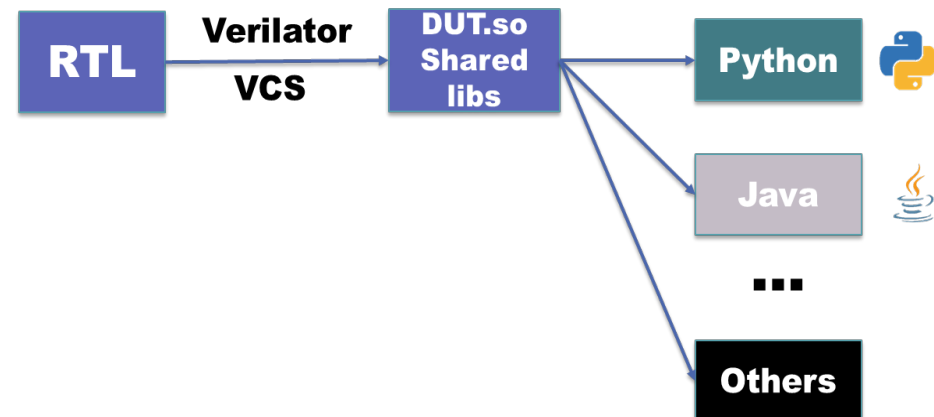
各验证层级发现

Bug 数量



🌟 万众一芯：芯片开放验证平台

- **愿景: 让软件背景开发者参与芯片验证**
 - 采用众包验证模式
 - 支持多种软件编程语言, 无需学习新语言
 - 复用 UVM 验证环境
- **高效定位**
 - 5 位本科生在两个月内发现了 **10 个 Bug**
- **易于使用**
 - 对 Python 和 Linux 没有基础的同学
 - 在 5 天内学习使用万众一芯工具链
 - 在 15 天内拥有添加基本测试用例的能力
- **香山模块级 (UT) 验证任务**
 - 任务: 对 IFU, ICache, LSQ 等模块进行单元验证
 - 扫描右侧二维码查看更多内容



万众一芯官网



GitHub 仓库链接

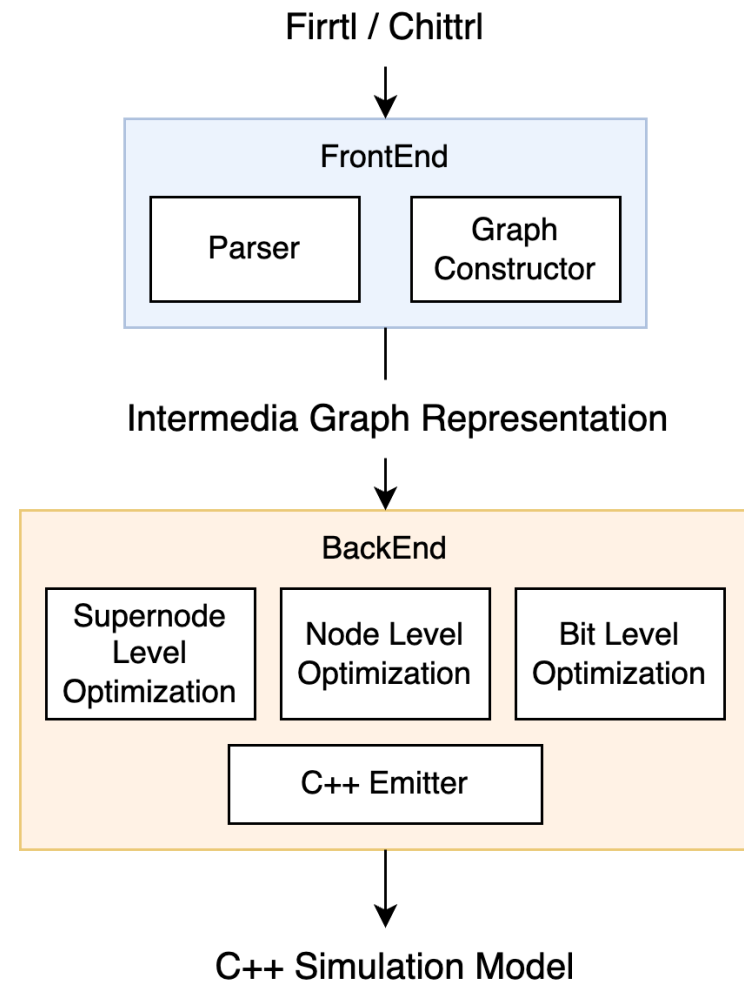
GSIM: 面向大规模设计的高速 RTL 软件仿真器

- **GSIM 采用了基于 ESSENT 的惰性仿真框架**

- 每一个节点绑定一个激活位，根据激活位设置情况判断当前节点和下游节点是否需要更新
- 仿真器接收 FIRRTL 或 CHIRRTL 的输入，输出 C++ 仿真文件
- 支持仿真 Rocket、BOOM、香山在内的多个开源处理器核

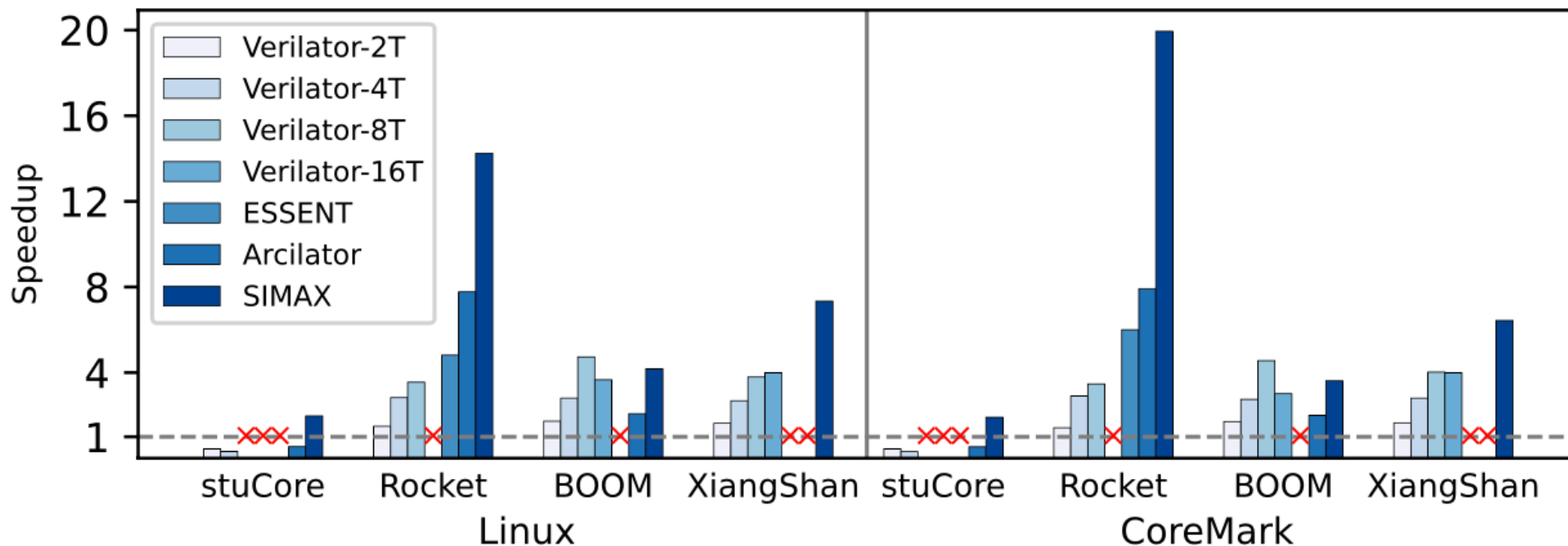
- **通过对软件仿真计算过程进行分解，GSIM 实现了多层次的优化策略**

- **超节点级优化**：基于增强型图划分算法划分超节点，以超节点粒度激活
- **节点级优化**：精准剪枝，消除别名节点、死节点、短路节点、无用寄存器四类冗余节点
- **比特级优化**：比特级拆分节点，细粒度激活受影响的后继节点，降低激活节点的比例



GSIM: 面向大规模设计的高速 RTL 软件仿真器

- 性能远超其他单线程 RTL 仿真器
 - 仿真香山启动 Linux 性能为 verilator 的 **7.34** 倍
 - 仿真 Rocket 运行 CoreMark 性能为 verilator 的 **19.94** 倍
- 在大多数处理器核上性能超过多线程 verilator



DiffTest-H: 基于硬件加速器的协同仿真框架

- 香山开发流程采用硬件加速器加速电路仿真，提升验证效率
 - 软件仿真器：速度 < 10 kHz，难以满足数十亿仿真周期的充分验证
 - 硬件加速器：Emulator (e.g. 帕拉丁, Zebu) 约 1MHz, FPGA 约 50MHz
- 基于硬件加速器的协同仿真验证框架
 - 硬件 (RTL 侧)：利用 Emulator、FPGA 硬件仿真平台加速待测处理器电路仿真
 - 软件 (Host 侧)：使用 x86、ARM 上位机运行参考模型软件逻辑
 - RTL-Host 通信：利用 PCIe、以太网等高速通信手段在软硬件间传递数据



问题：通信成为速度瓶颈

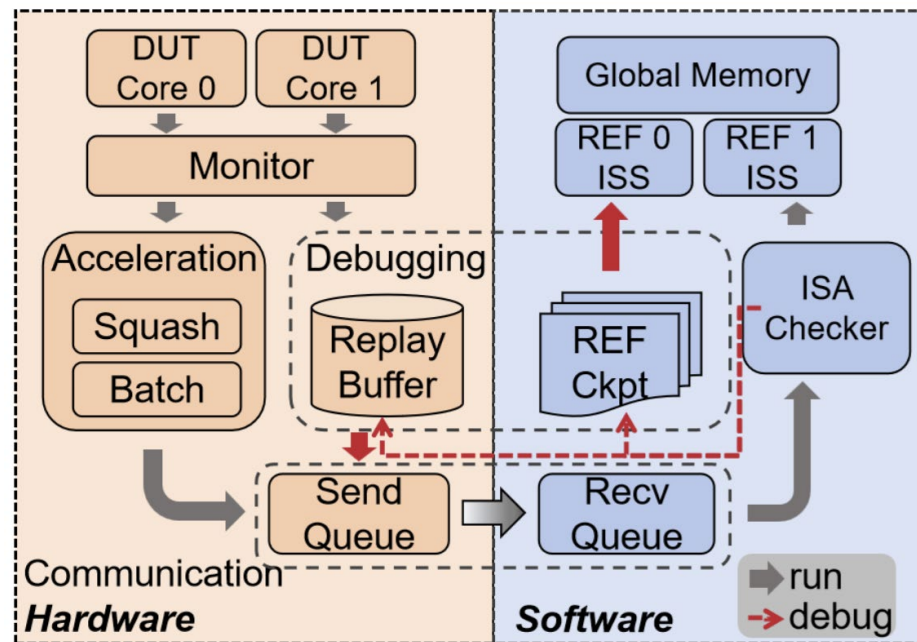
- 软硬件通信：>99%
- RTL仿真：<1%

整体协同仿真速度仅 KHz 量级

DiffTest-H: 基于硬件加速器的协同仿真框架

- **特点：对验证数据包的通信优化**

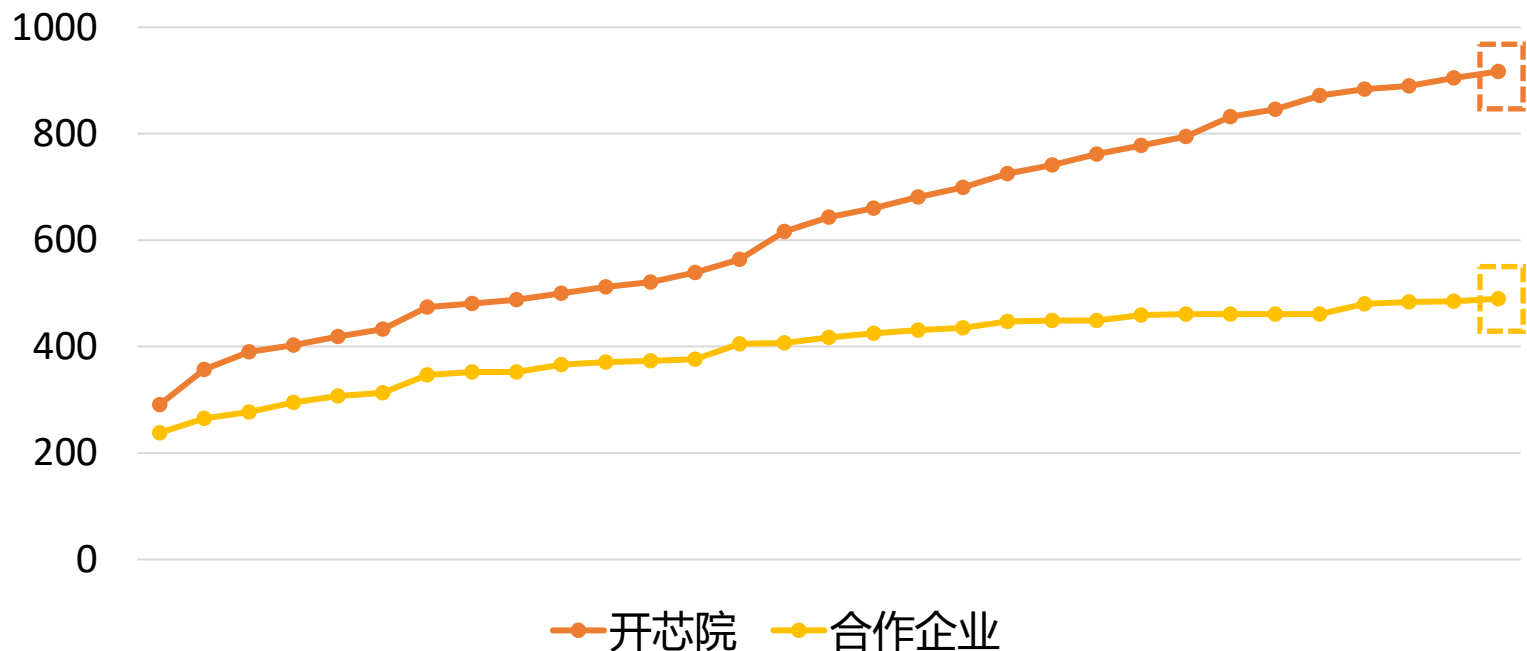
- **Batch**：降低通信频次
- **Squash**：减少通信数据量
- **Replay**：保持指令级别的调试粒度
- 仿真验证加速比达 **57~109** 倍
- DiffTest-H 帮助香山发现了 **151** 个硬件 bug, 涉及到 **780** 行和 **19** 个 PR 的代码修改



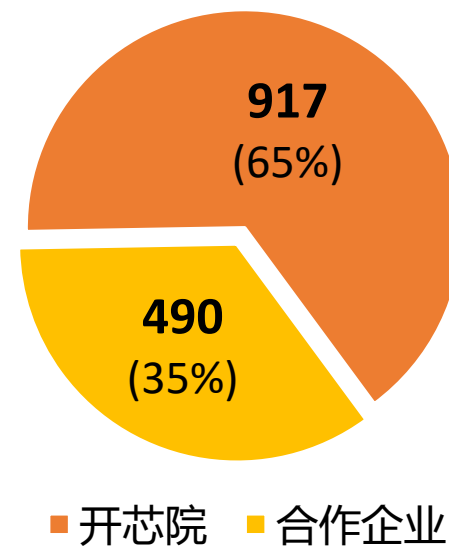
配置	NutShell (帕拉丁)	香山 (帕拉丁)	香山 (FPGA)
无优化	14KHz	4KHz	0.1MHz
+Batch	102 KHz (7×)	24 KHz (6×)	1.3 MHz (13×)
+NonBlock	398 KHz (28×)	71 KHz (12×)	2.2 MHz (22×)
+Squash	1080 KHz (77×)	478 KHz (109×)	5.7 MHz (57×)

🏔️ 开源模式下的验证协作

- 依托于香山的开源开放属性
 - 企业可使用**非公开内部测试集**共同对香山进行验证
 - 昆明湖开发过程中，合作企业帮助发现了约 **35%** 的 bug



截至 7 月 14 日的累计缺陷数量



🚀 开源敏捷方法赋能高性能产业级 CPU 开发

面向产业需求的处理器

性能突出

- SPEC06 45@3GHz
- 对标 ARM 数据中心处理器 Neoverse N2
- 高性能向量 (RVV) 扩展实现

功能丰富

- 支持 RVA23 profile **全部必选项**
- 支持 AIA 标准, Debug、Trace 扩展
- 满足服务器场景需求

高可靠性

- UT、BT、IT、ST 多层次协同验证
- 工业级验证流程
- 开发 PLDM / FPGA difftest 调试工具

灵活可配

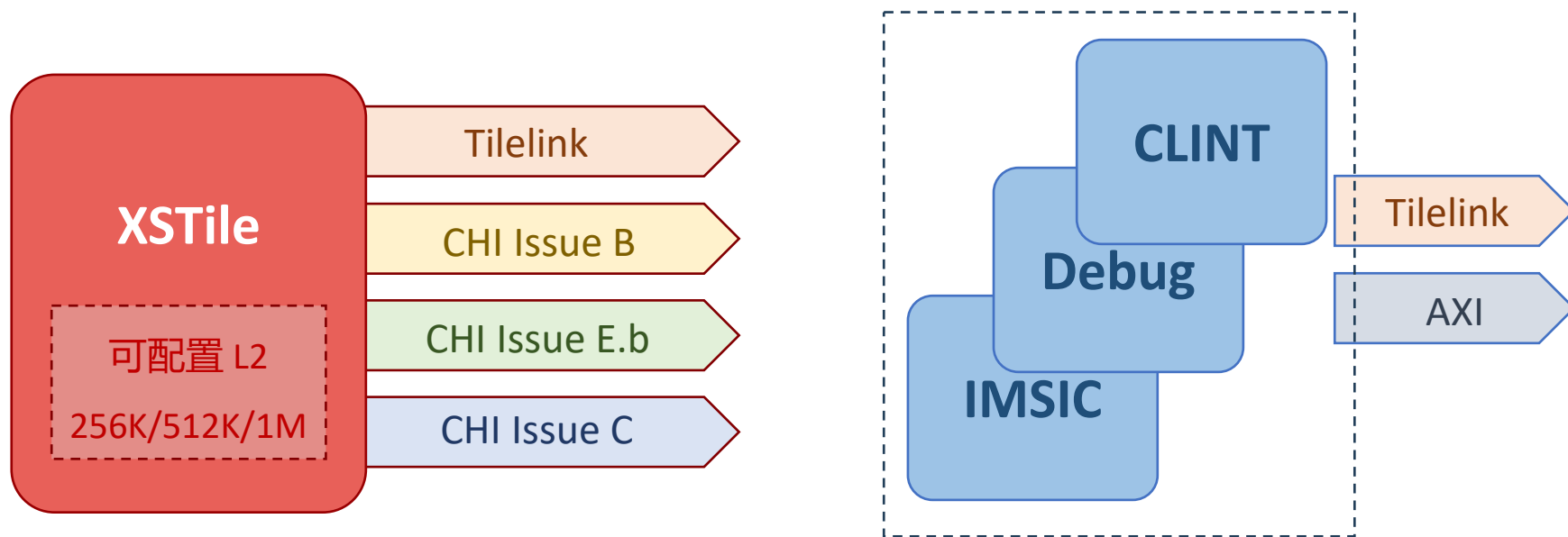
- 丰富的总线接口 **CHI, AXI, Tilelink**
- 高可配置的核内参数及 Cache 容量
- 满足 SoC 设计需求

敏捷设计与验证基础设施

🌟 丰富可配置的交付选项

- **灵活可配置**的交付选项可以满足不同的定制化需求

丰富的总线接口



高度可配置

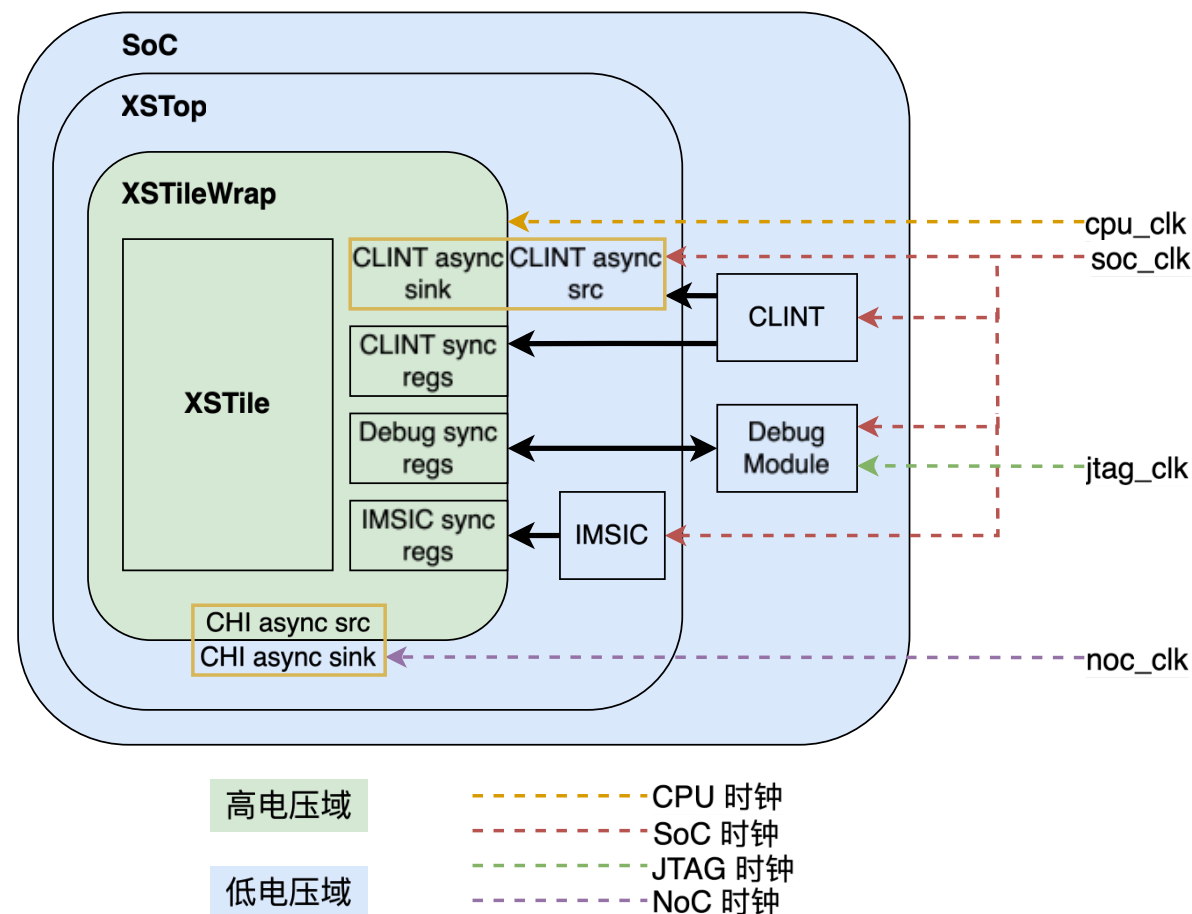


丰富可配置的交付选项

- **灵活可配置**的交付选项可以满足不同的定制化需求

- **满足多样的 SoC 设计需求**

- 高低电压域分离
- 跨模块信号异步处理
- 支持 CLINT 模块灵活部署, 允许共享或私有



第二届香山开源社区大会日程

时间	题目	演讲者
13:30 – 14:00	香山：迈向产业实践的开源高性能 RISC-V 处理器	冯浩原
14:00 – 14:20	XSAI：以 CPU 的编程范式支持现代 LLM 核函数	周耀阳
14:20 – 14:40	香山社区：2024 – 2025	徐泽凡
14:40 – 15:00	香山社区贡献者颁奖	
15:00 – 15:20	香山昆明湖 V3 前端架构	李衍君、徐之皓
15:20 – 15:40	香山昆明湖后端流水线的设计演进	刘泽昊
15:40 – 16:00	香山昆明湖缓存、访存的设计演进	丁昊楠
16:00 – 17:00	香山处理器与 XS-GEM5 模拟器 Tutorial	李衍君、徐之皓、 陈泱宇、杨泽辰
17:00 – 17:30	万众一芯：让每个开发者都能参与芯片验证	王俊越、李小龙

敬请批评指正!