中国科学院计算技术研究所 INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

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



XS-GEM5：XS-RTL对齐的高性能微架构模拟器

作者：香山团队

汇报人：甄好

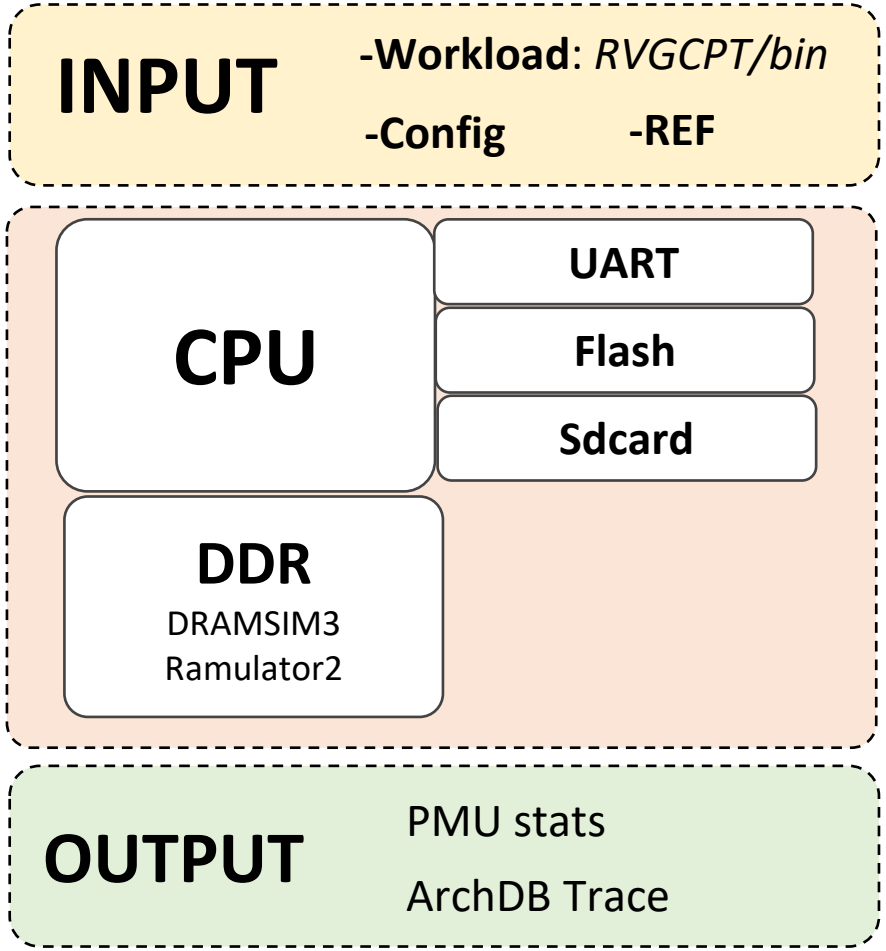
中国科学院计算技术研究所、北京开源芯片研究院

2026年10月18日

什么是XS-GEM5?

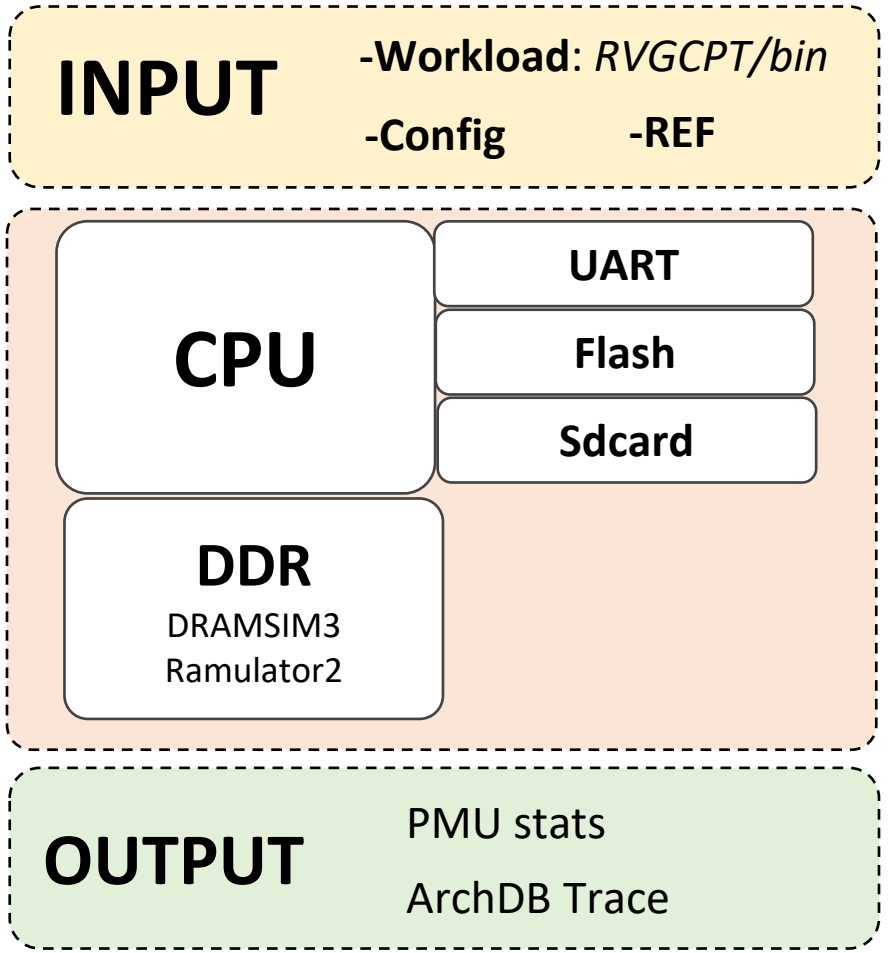
- 基于开源模拟器 GEM5 实现
- Cycle-based CPU 微架构模拟器
- 执行 RISC-V 指令
- 精确建模香山处理器的微结构

- 支持：
  - 性能分析（定位性能瓶颈）
  - Feature评估



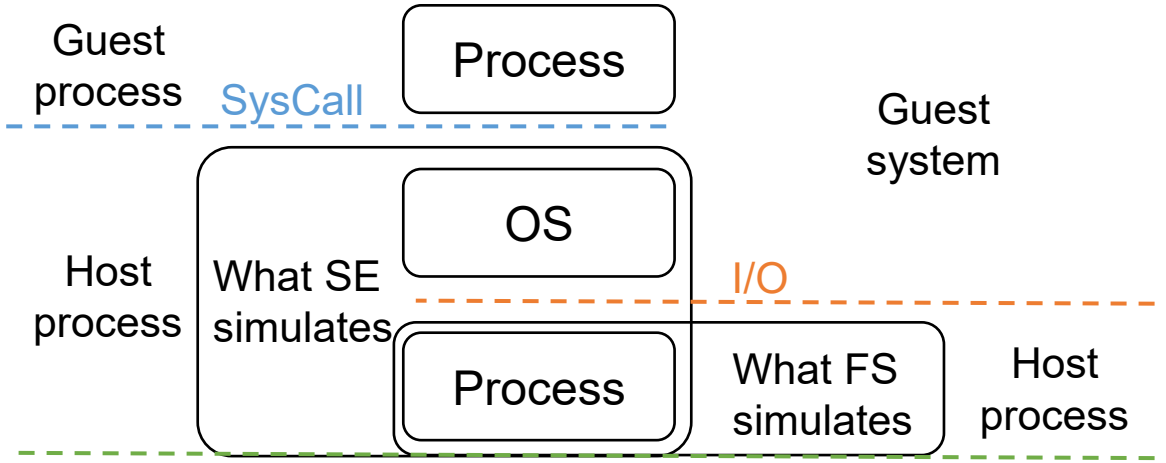
什么是XS-GEM5?

- SE simulation
- FS simulation
  - RVGCpt：基于 SimPoint 估算 SPEC CPU 2006 和 2017 性能
  - 支持多样化TRACE: CBP/Google Trace
- 前端：精确对齐
- 后端：
  - 延迟与吞吐量对齐
  - 流水线、发射队列和寄存器堆
- 访存&缓存：
  - 预取器 && L2 (2MB) && L3 (32MB)
  - MMU：两级TLB && TLB预取 && 并行PTW
  - 支持DRAMSIM3 及 Ramulator 多种DDR仿真器



支持 SE 及 FS 模式

- SE 模式：
  - 支持部分microbenchmark
- FS 模式：
  - 虚拟地址到物理地址的映射
  - 精确模拟系统调用
  - 模拟串行化指令的性能开销
  - 精确模拟虚拟内存行为
  - 精确建模 TLB 和页表遍历器（PTW）
- 推荐使用全系统模式（FS）



为什么需要架构模拟器?

- 香山自发布以来，完成了多代架构升级：
  - Yanqihu →Nanhu → Kunminghu → KunminghuV3
  - SPEC CPU 2006: 7 /GHz → 10 /GHz → 15 /GHz → 20 /GHz

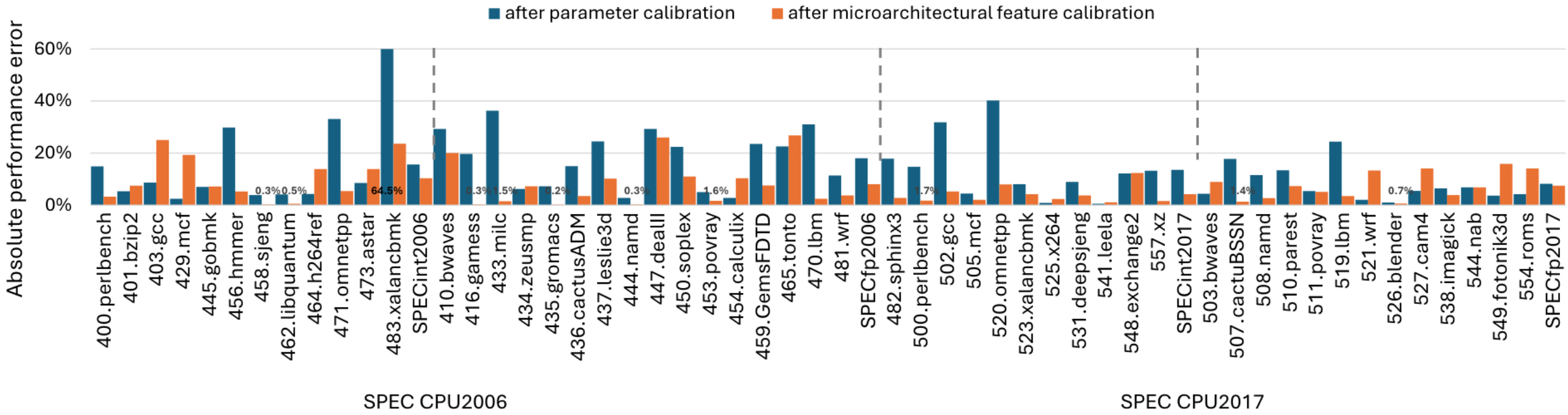
- 在设计前期探索多种设计方案，并针对广泛的应用开展性能分析。
- 需求：
  - 详细的微架构建模
  - 正确性保障
  - 高效的开发流程
  - 快速的仿真运行
  - 高质量的代码

XS-GEM5：精准、高效、可靠的微架构模拟器

- 详细的微架构建模
  - Decoupled Frontend、乱序后端、分支预测器和预取器等
- 正确性保障
  - 与 XS-RTL 对齐，结合 SPEC 与 Microbenchmark 多角度评估微架构对齐度
- 开发效率
  - 提供强大的正确性 (DIFFTEST) 与性能分析基础设施 (TopDown)
- 仿真速度
  - 使用 C/C++ 编写高效代码，仿真速度可达100KHz
- 代码质量
  - CI、回归测试，对所有PR进行代码审查

XS-GEM5 的准确度

- SPEC CPU 2006及SPEC CPU2017绝对性能误差均在5%以内。

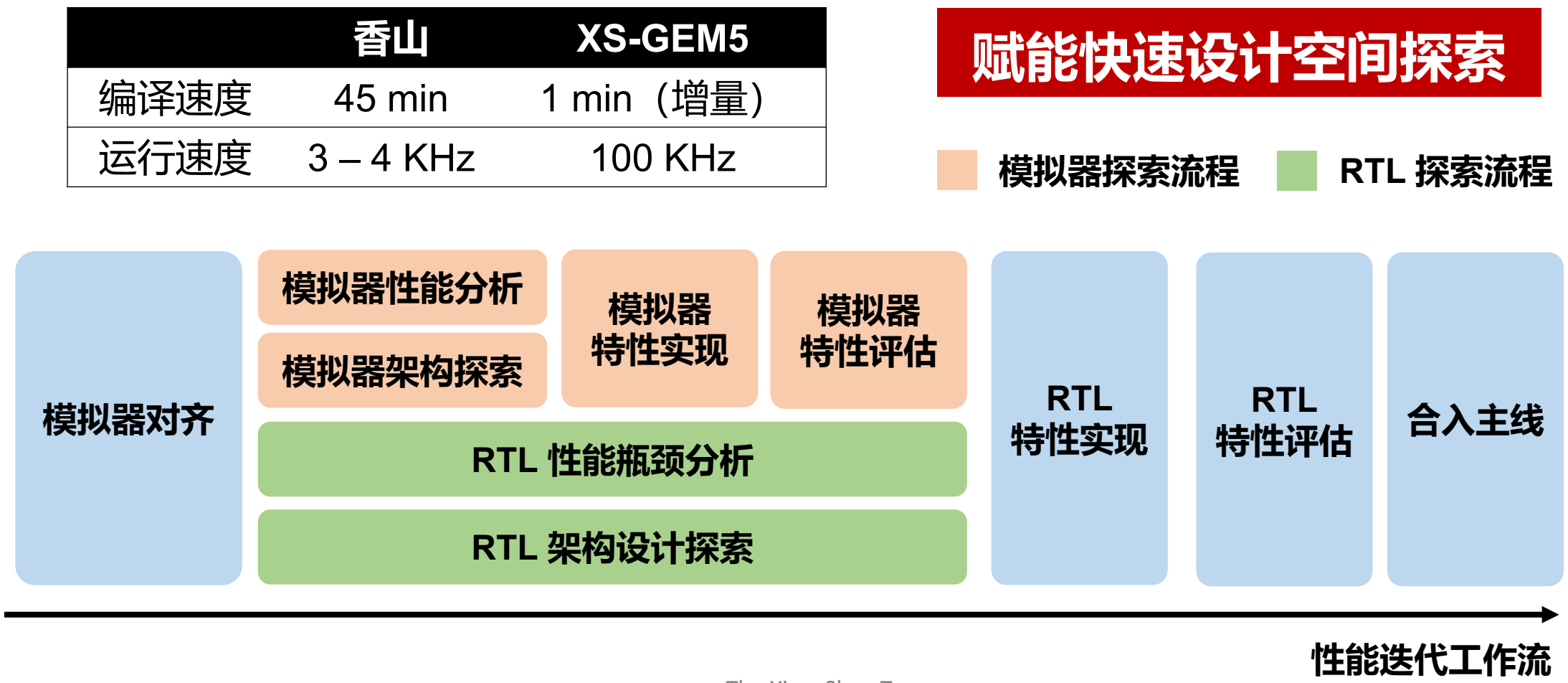


加速性能迭代流程

- **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



•XS-GEM5平台支撑了多个feature的探索工作

前端

- Two Fetch
- Two Taken
- Victim BTB
- TAGE
- 指令预取
- uBTB预取
- LLBP
- Ahead prediction
- APF
- …

访存 & 缓存

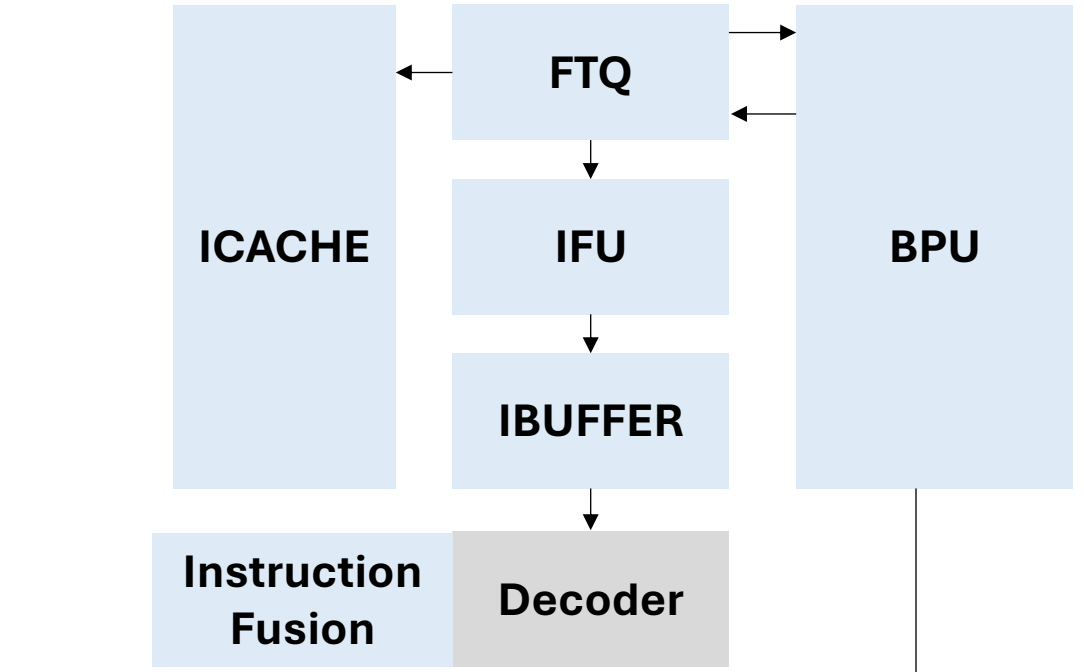
- BOP预取限流
- LLDP
- MDP算法
- SMS分层预取优化
- Coherency 优化
- Dcache bank冲突优化
- Berti 预取优化
- 多预取调控
- L2分bank优化
- L3配置优化
- …

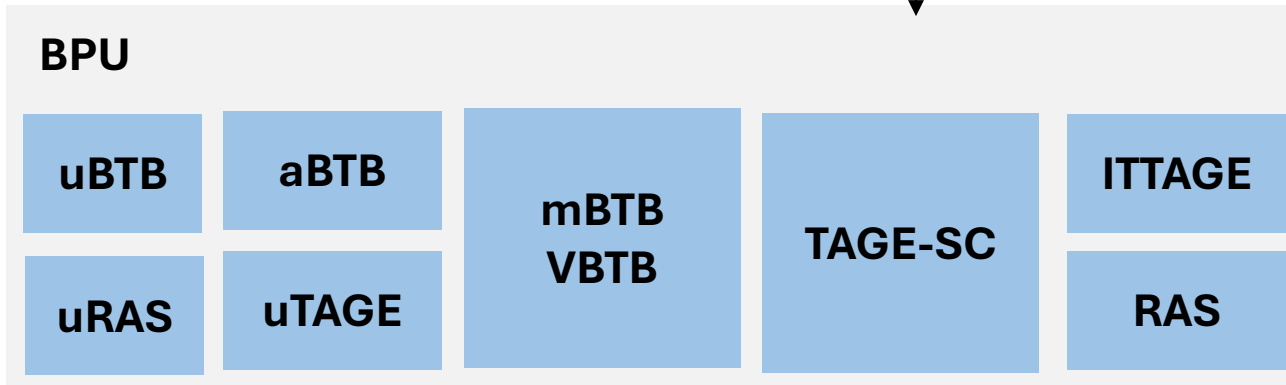
后端

- IQ redirect入队减一拍
- Regfile Prefetch 值预测部分刷新
- LVP
- ESTRIDE
- 多值预测框架
- …

Frontend

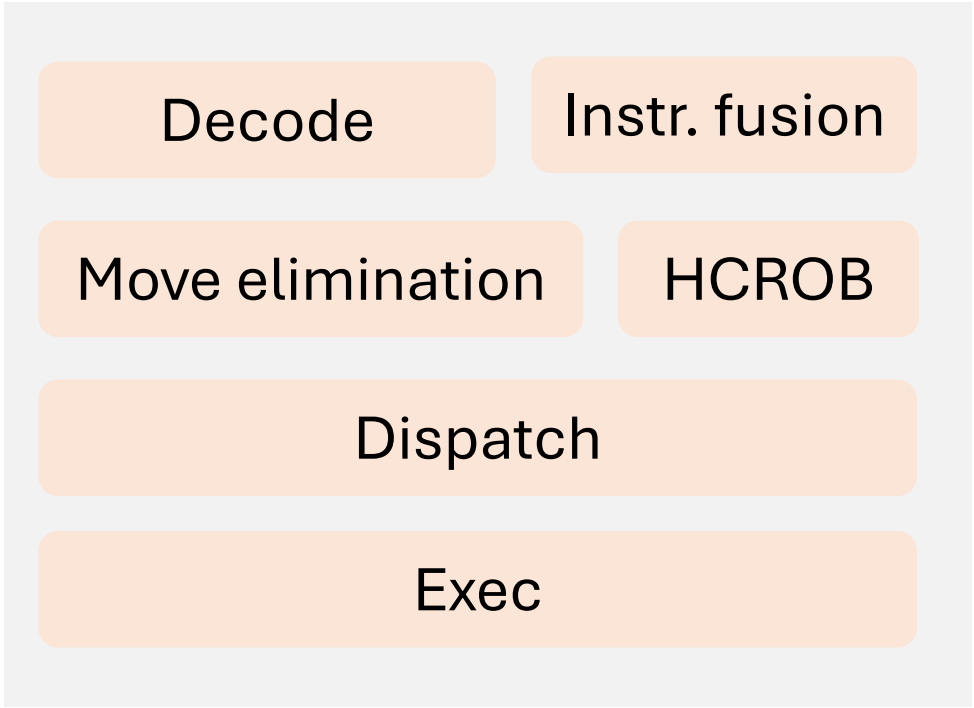
- **Decoupled Frontend**
  - 隐藏延迟，提高流水线效率
- 分支预测器组合：
- 集成 TAGE、SC 等预测器
  - TAGE
    - 8张预测表
  - SC
    - 6 张校正表
- 高度可扩展
  - 支持多个解耦的分支预测器
  - 延迟可配置





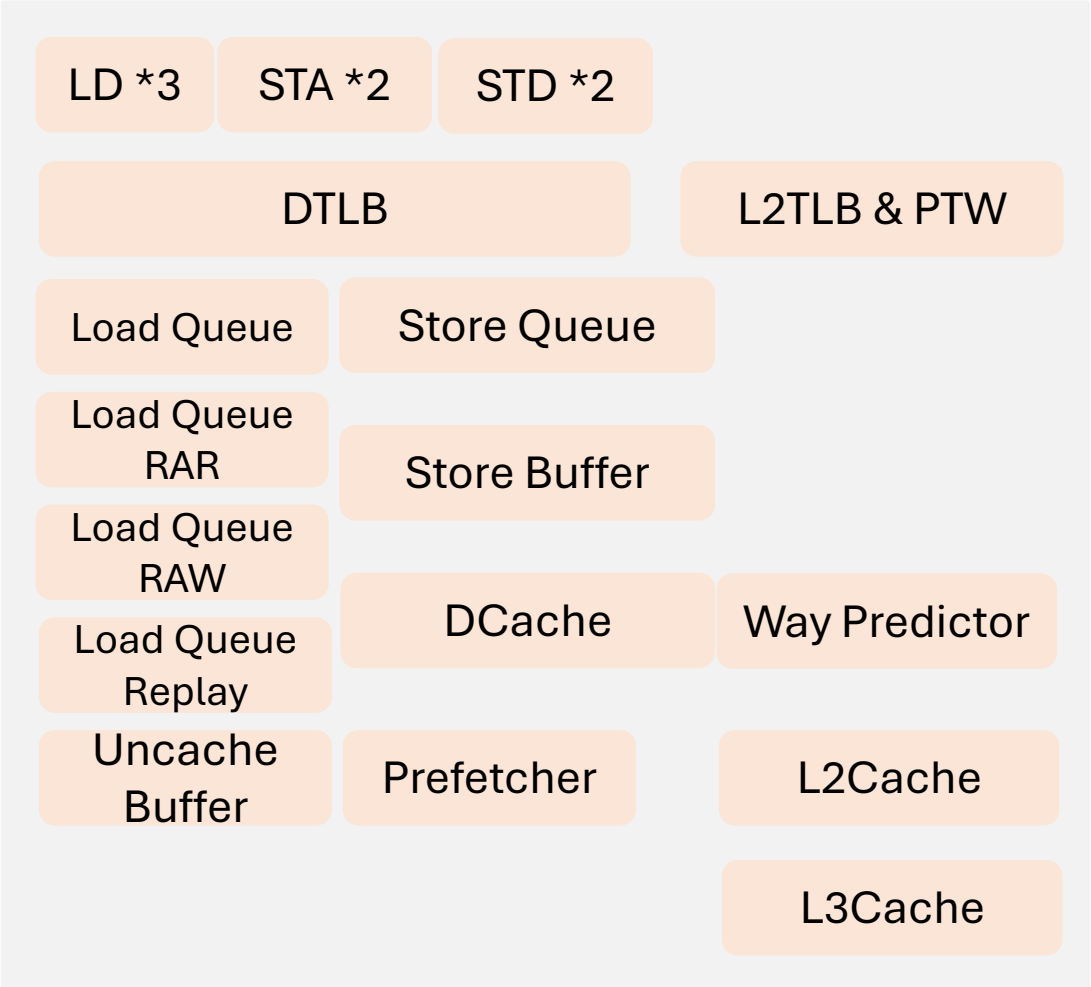
Backend

- 指令融合
- Move 消除
- 可配置的调度器
- **精确对齐：**
  - Dispatch
  - Issue
  - Replay
  - ROB压缩
- 合理抽象&参数化
  - 执行单元
  - 快速/推测唤醒通道
  - 寄存器文件读写端口



Memory Block

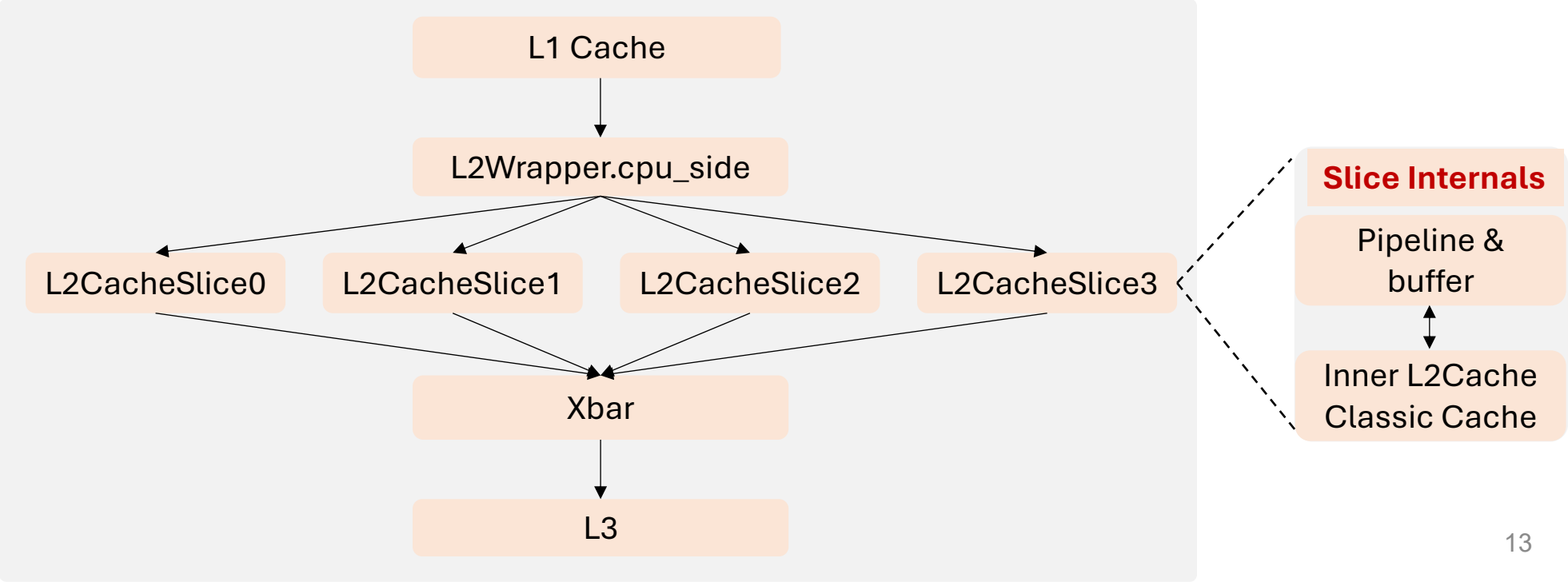
- **精确对齐**
  - Load流水线
  - Load/Store Queue
  - STA / STD 分离
  - Load Replay 场景
  - 预取器 & 路预测器
  - Store Buffer 建模
  - 访存违例
  - MMU 建模
- **合理抽象**
  - L2/L3 Cache 硬件逻辑细节



Code details
*[1] src/cpu/o3/lsq**
*[2] src/arch/riscv/tlb**
*[3] src/arch/riscv/pagetable_walker**
*[4] src/mem/cache/prefetch/**
*[5] src/mem/cache/way_prediction_policies/**

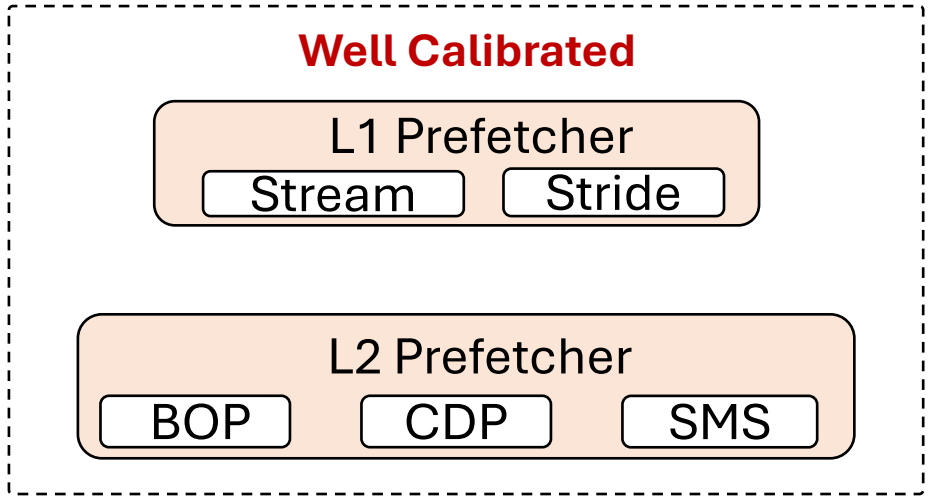
L2 Cache

- 基于 GEM5 Classic Cache，增加 Wrapper 层以模拟硬件行为
- 精确对齐：
  - 细粒度建模流水线仲裁、停顿、调度与控制逻辑
  - Slice / Prefetch

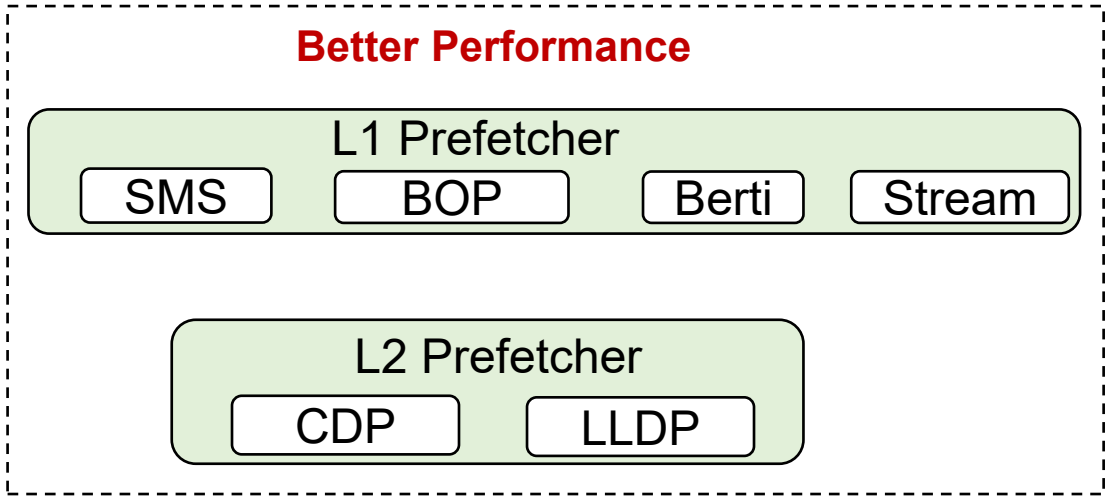


Prefetcher

- **RTL 对齐配置**
  - **L1：** Stream + Stride
  - **L2：** SMS + BOP + CDP



- **性能探索配置**
  - **L1：** SMS + BOP + Berti + Stream
  - **L2：** CDP + LLDP
- 支持 **Prefetch Request Offloading**
- **根据置信度进行L1/L2/L3预取的调控**



SMT

- 基于XS-GEM5平台，**达摩院・中兴微电子・北京开源芯片研究院**联合展开SMT的建模和探索工作
- **SMT吞吐提升约 30%**
- 敬请关注：20日香山如意开发者大会
  - 来自三家单位的开发者会进行SMT的详细分享



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，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 主流水线
- 持续对齐 RTL 细节

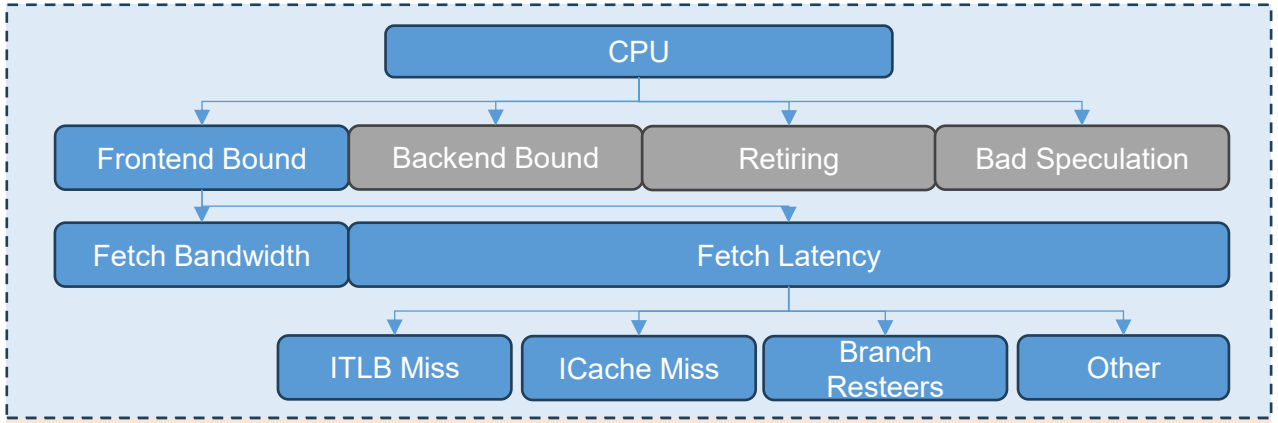
基础设施：LLM辅助microbenchmark生成 + 自动化编译 + 运行 + 对比分析流程

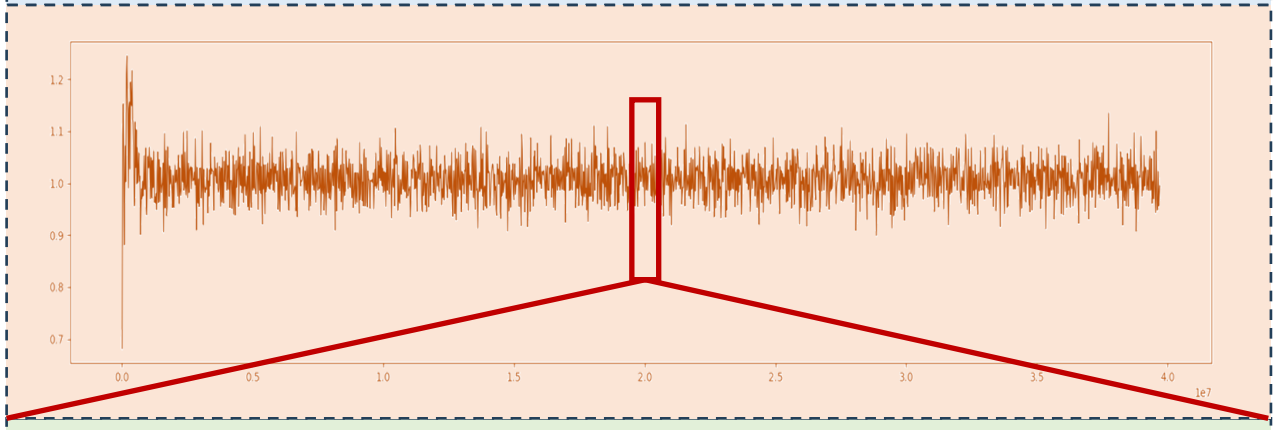
模拟器与 RTL 的性能对齐

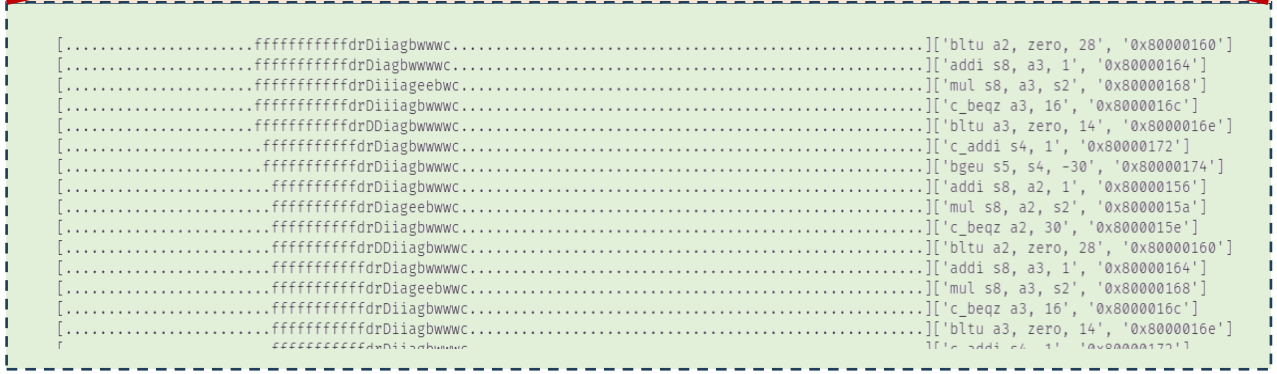
TopDown

IPCRolling

PerfCCT

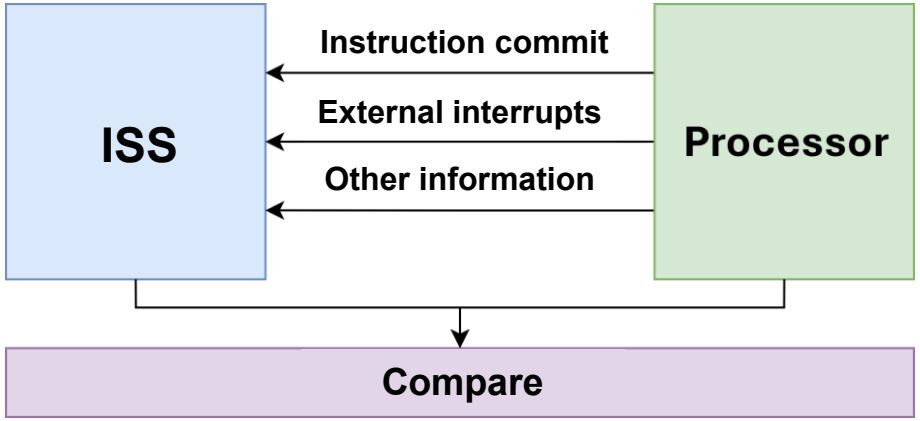






Debug工具 : Difftest

- **集成DIFFTEST 框架**
  - ISA 级行为校验
  - NEMU / Spike 作为参考模型（REF）
  - 实时对比模拟器与 REF 的寄存器状态
  - 快速定位错误指令



Basic architecture

```
while (1) {
    icnt = gem5_step();
    ref_step(icnt);
    r1s = gem5_getregs();
    r2s =  ref_getregs();
    if (r1s != r2s) { abort(); }
}
```

Online Comparison Mechanism

分析工具: DSE

- XS-GEM5 基于 CI 构建了一套便捷的**设计空间探索框架**:
  - 通过自然语言描述优化目标与约束条件，利用 AI 和已有 Skill 自动生成 DSE Spec
  - 通过 GitHub Actions 手动指定 Spec，启动自动化设计空间搜索
  - 搜索完成后自动汇总结果，便于后续分析

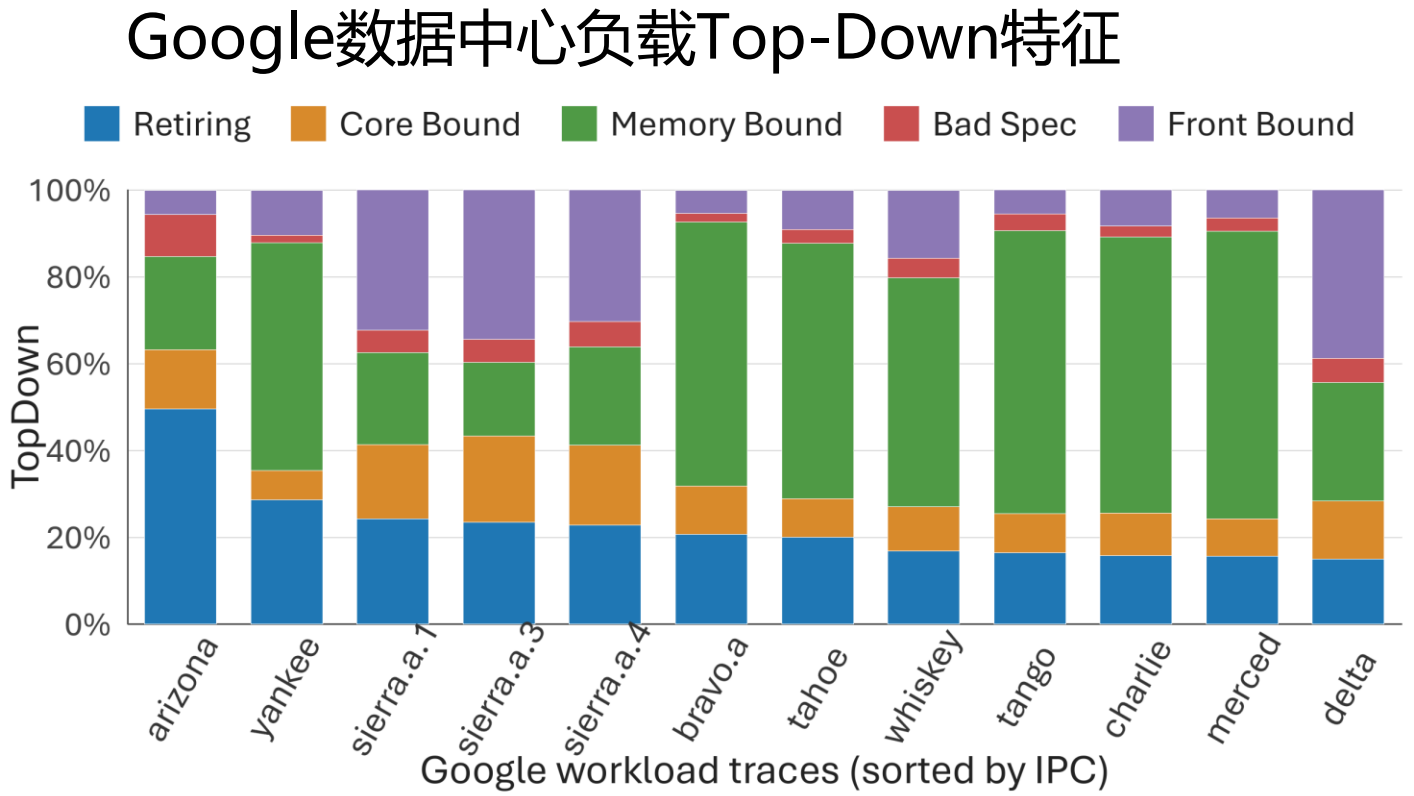




例. 有限资源开销下搜索性能最优的 TAGE 参数

支持Trace驱动

- 支持 Trace-driven Simulation
- 可以运行更多的复杂应用负载
- 面向真实业务场景进行评估



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