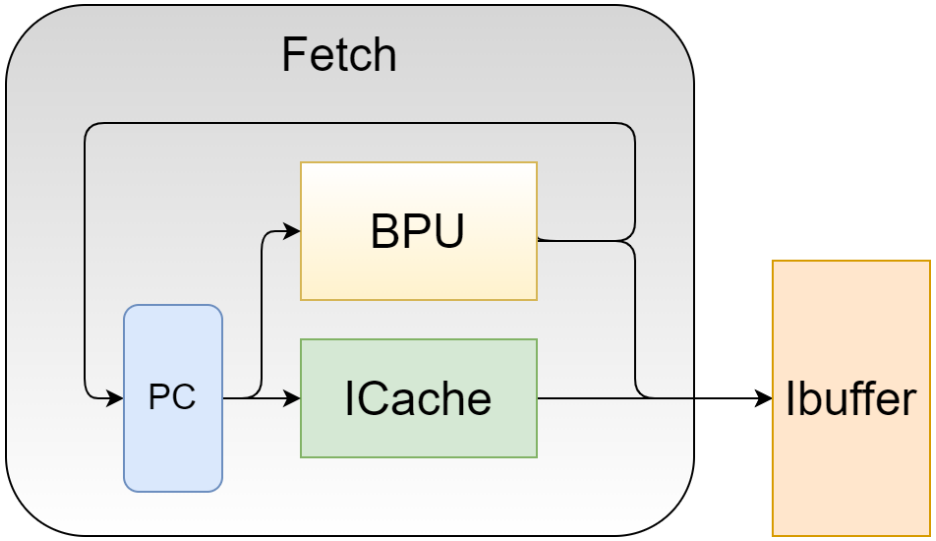
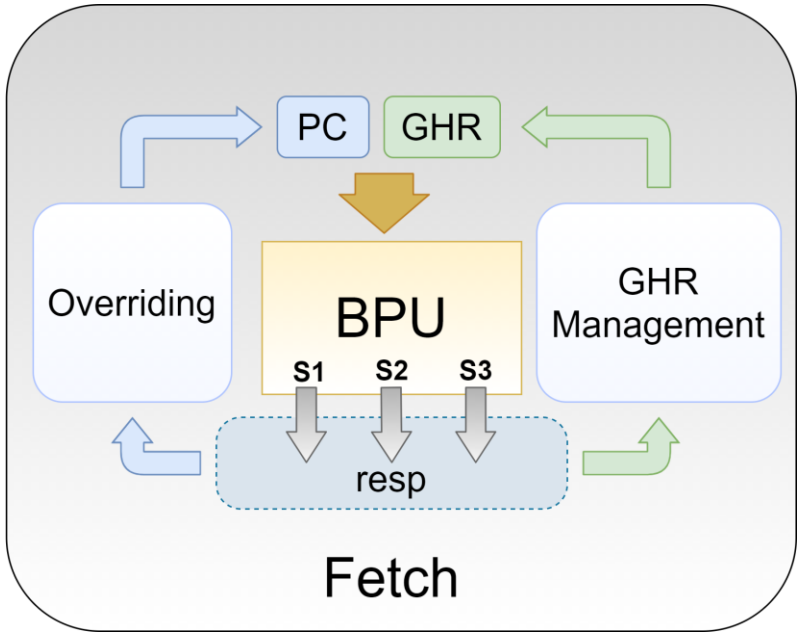
香山处理器分支预测部件设计实现

勾凌睿[1]　张林隽[1]　金越[1]　邹江瑞[2]
[1]中科院计算所
[2]深圳大学
2021年6月25日

分支预测部件（BPU）：保证指令供给

- 当前版本参考UC Berkeley Sonic BOOM(*)
- 和取指单元紧耦合，流水线同步
- 结合取指单元实现三级覆盖（overriding）预测、分支历史管理

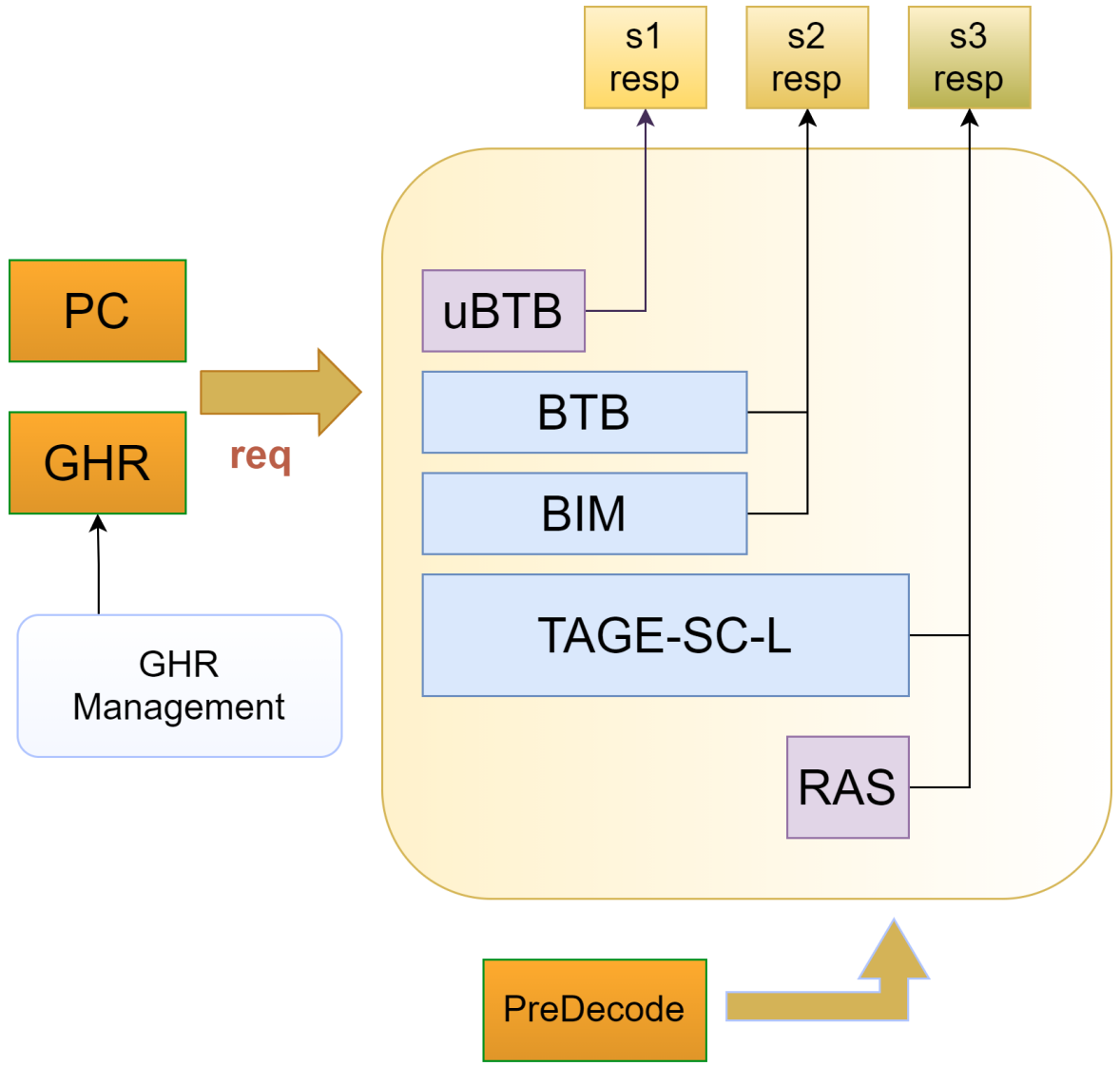




*: Zhao J, Korpan B, Gonzalez A, et al. Sonicboom: The 3rd generation berkeley out-of-order machine[C]//Fourth Workshop on Computer Architecture Research with RISC-V. 2020.

分支预测部件

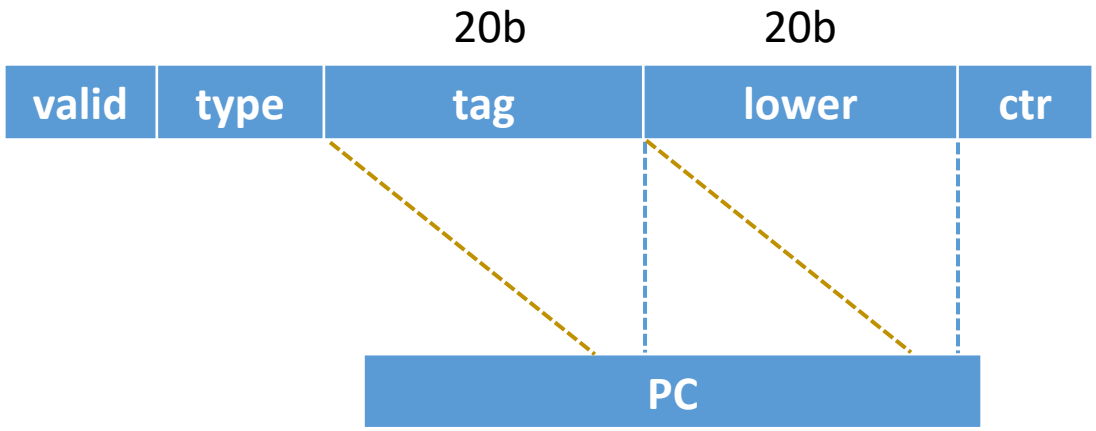
- 第一级：uBTB
- 第二级：BTB+BIM
- 第三级：
  - TAGE-SC-L
  - RAS
  - 预译码



分支预测部件：S1

■第一级：uBTB（寄存器）

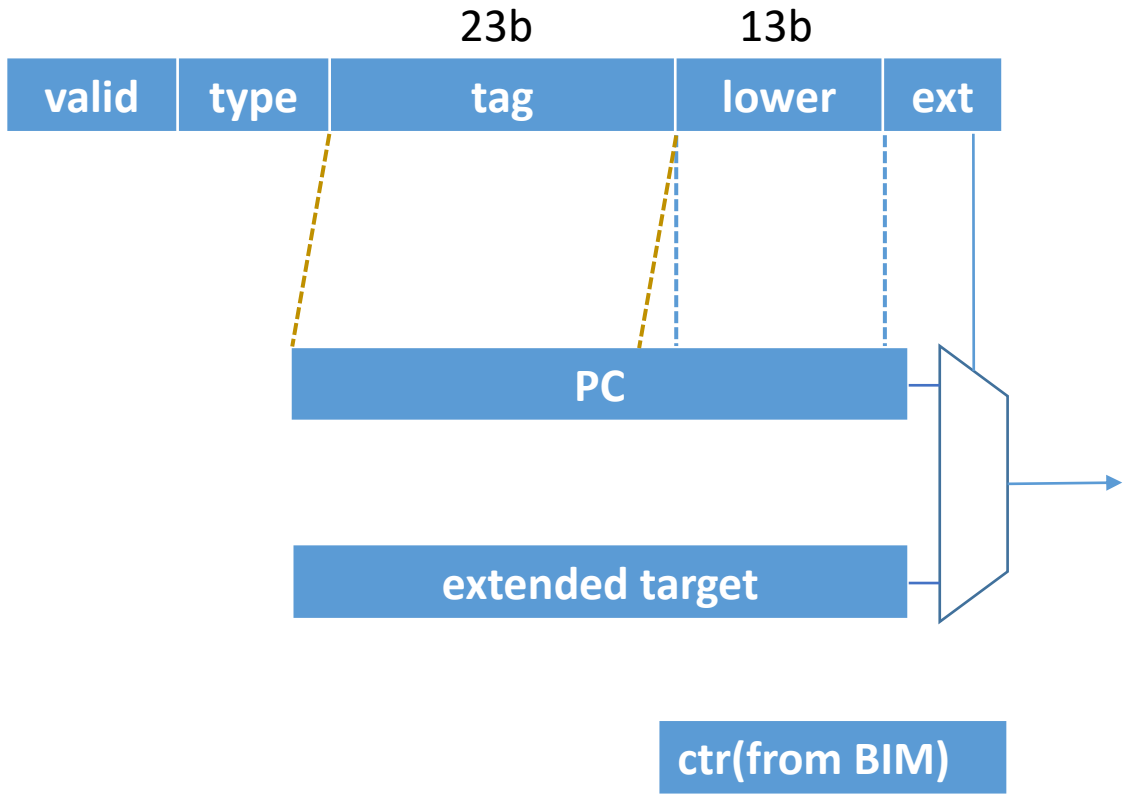
- 全相联
- 拼位计算目标地址
- 两位饱和计数器



分支预测部件：S2

■第二级：BTB+BIM（SRAM）

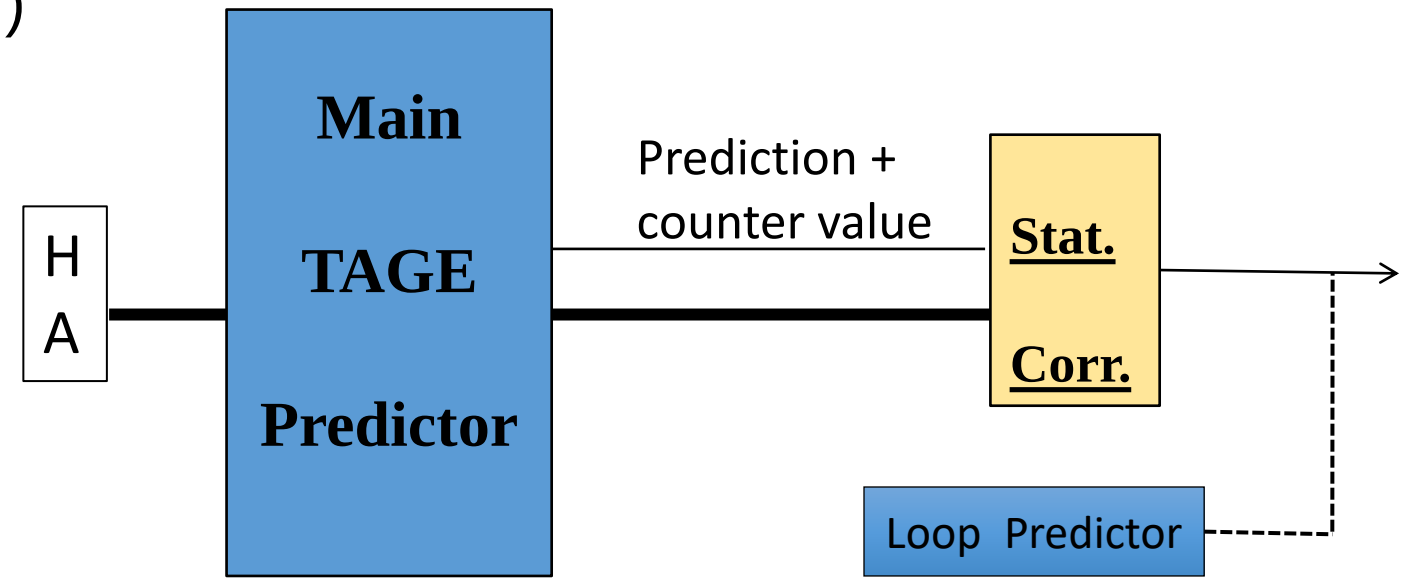
- 直接索引
- 两路组相联
- 拼位计算目标地址
- 更多的两位饱和计数器（给TAGE复用）



分支预测部件：S3

■第三级：

- TAGE-SC-L（SRAM+寄存器）
  - 64位历史
  - 6个历史表
  - 6张SC表
- RAS
- 预译码



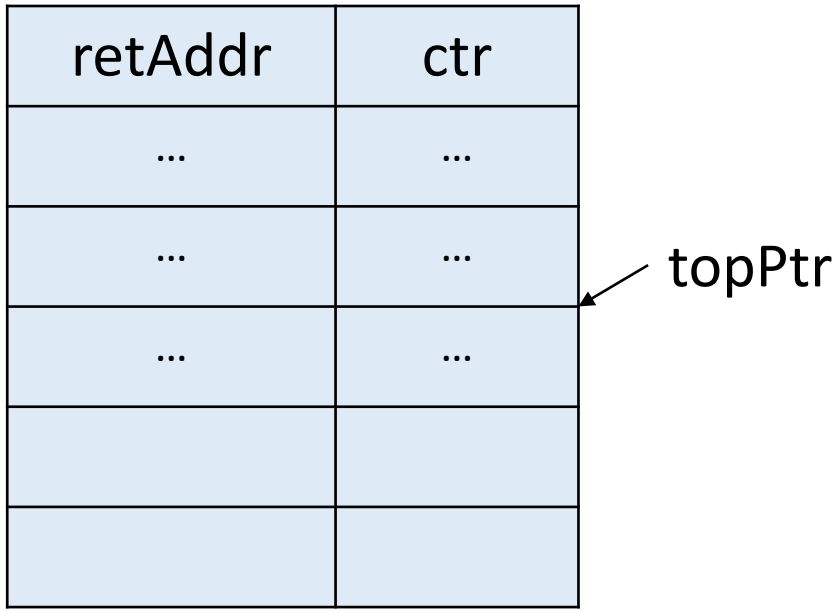
TAGE-SC-L示意图 （*）

*: https://www.jilp.org/jwac-2/program/cbp3_03_seznec.pptx

分支预测部件：S3

■第三级：

- TAGE-SC-L
- RAS（寄存器）
  - 16项
  - 带有计数器
- 预译码（指令码）
  - 转移指令类型信息
  - 计算br、jal目标地址



RAS

Statistical Corrector效果评估

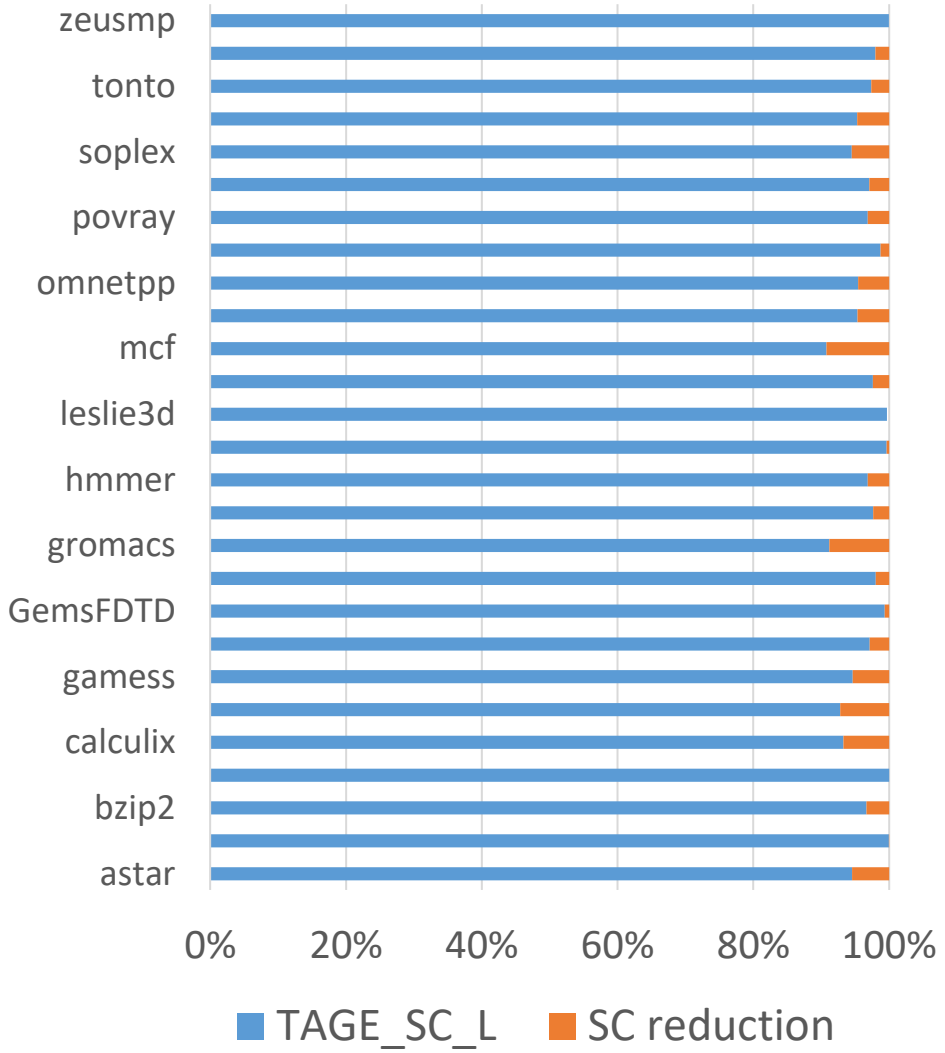
- 基准：100%

  TAGE-L（无SC）的MPKI

- 橙色：SC减少的MPKI

- 蓝色：TAGE-SC-L的MPKI

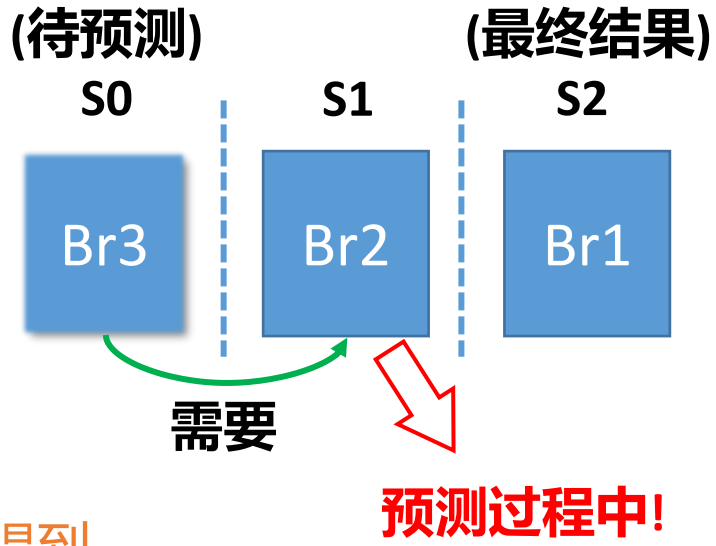
- **平均MPKI降低~3%**



注：
- MPKI代表每千条指令误预测次数
- 测试片段通过SimPoint生成

分支历史管理机制——背景

- 分支历史推测更新的理想模型
  - 出发点：用预测结果代替真实执行结果更新分支历史
  - 理想：用最新、最准确的预测结果推测更新
- 多级预测机制下实现推测更新的困境
  - 模拟器一般不会考虑的**实现**问题
  - 某一条分支指令的预测过程中：
    - 产生的多个预测结果可能得到不同的分支历史
  - 分支历史只在预测开始的时间点使用
    - 开始预测时，上一条分支指令的最终准确预测结果尚未得到



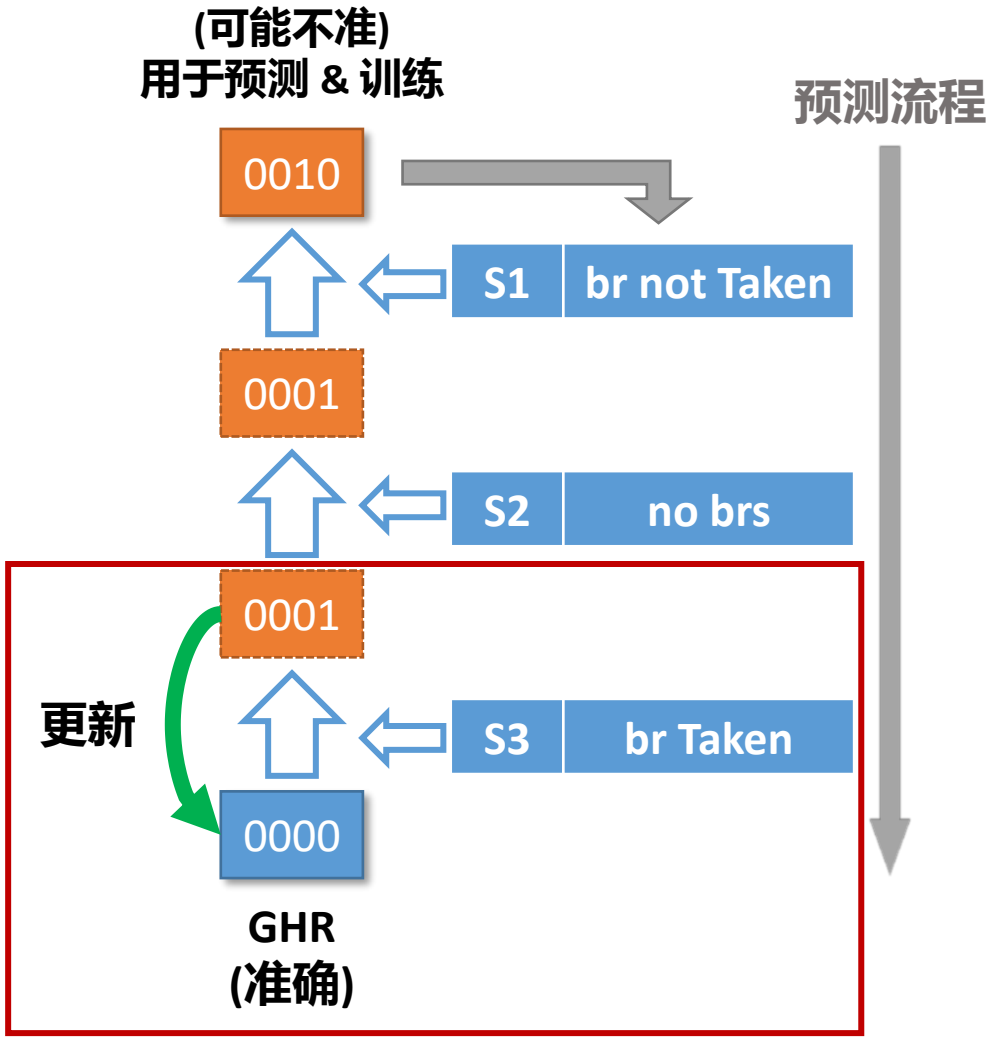
分支历史管理机制——BOOM [*]

- 做法：覆盖重定向
  - 当多级预测产生不同的历史结果时，用后级覆盖并冲刷前级流水线
- 此方法优缺点：
  - 保证了每次预测使用的历史准确
  - 产生很多**取指空泡**

*: Zhao J, Korpan B, Gonzalez A, et al. Sonicboom: The 3rd generation berkeley out-of-order machine[C]//Fourth Workshop on Computer Architecture Research with RISC-V. 2020.

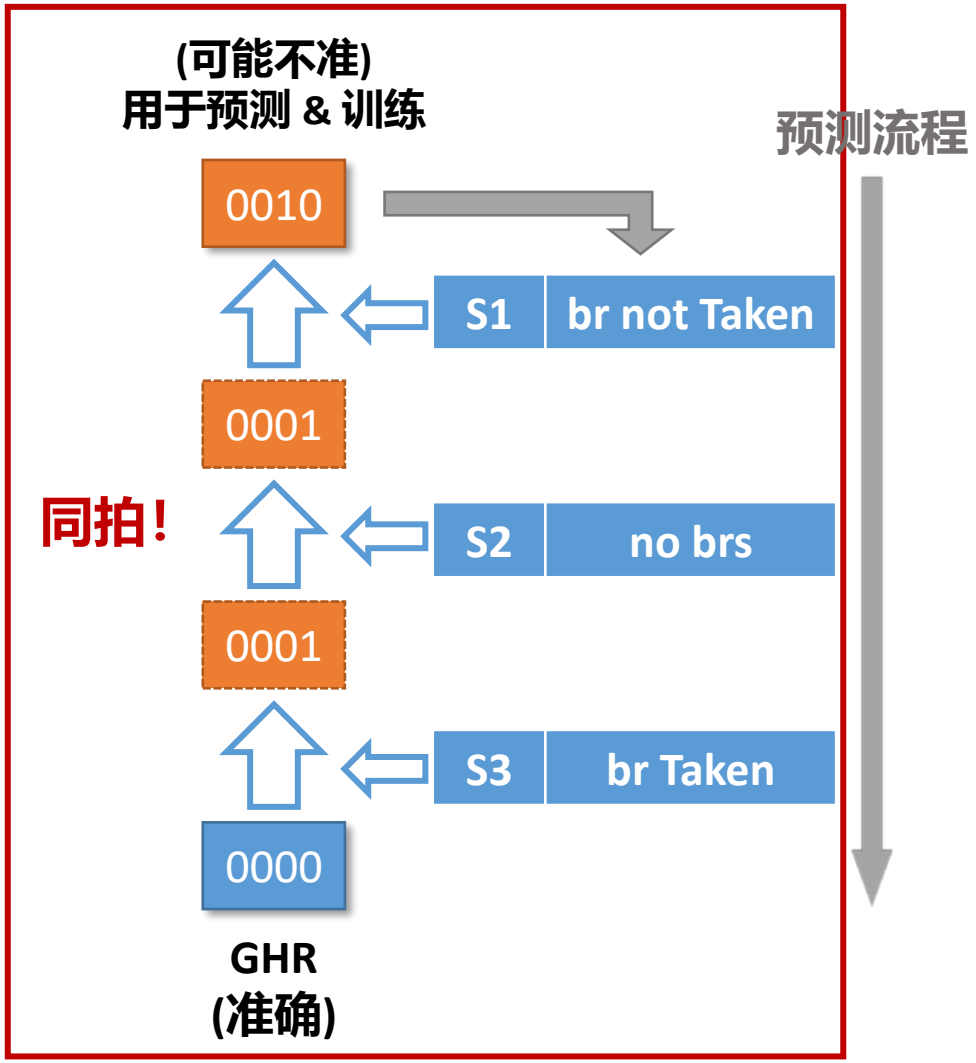
分支历史管理机制——香山处理器

- 目的：解决取指空泡问题
  - **准确性**：分支历史寄存器（GHR）
    - 在 BPU 末端流水级，由最终预测结果推测更新
    - 正确路径在**该流水级**见到的历史永远是准确的
  - **及时性**：最新的分支历史（back to back）
    - 以 GHR 为基准，决定大部分位
    - 按流水级情况，**当拍动态**决定最新的分支历史位



分支历史管理机制——香山处理器

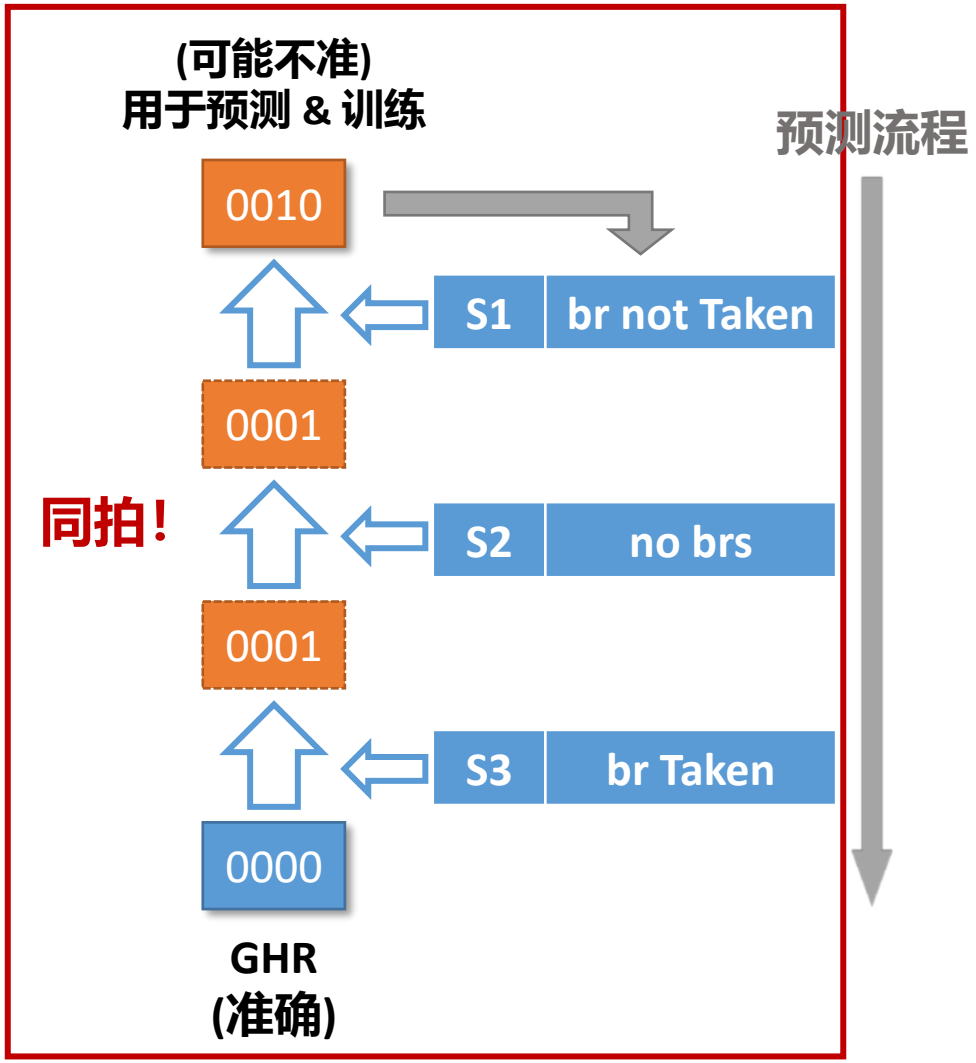
- 目的：解决取指空泡问题
  - **准确性**：分支历史寄存器（GHR）
    - 在 BPU 末端流水级，由最终预测结果推测更新
    - 正确路径在**该流水级**见到的历史永远是准确的
  - **及时性**：最新的分支历史（back to back）
    - 以 GHR 为基准，决定大部分位
    - 按流水级情况，**当拍动态**决定最新的分支历史位



分支历史管理机制——香山处理器

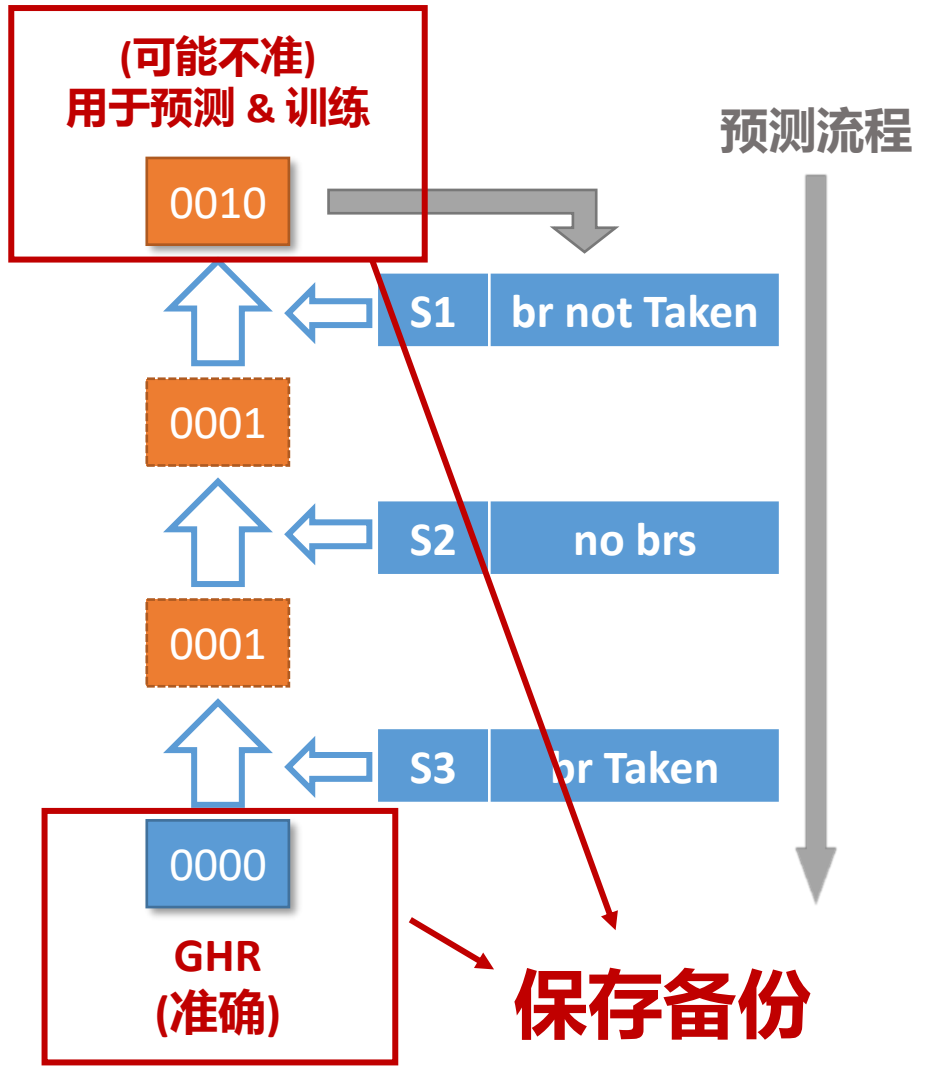
- 目的：解决取指空泡问题
  - **准确性**：分支历史寄存器（GHR）
    - 在 BPU 末端流水级，由最终预测结果推测更新
    - 正确路径在**该流水级**见到的历史永远是准确的
  - **及时性**：最新的分支历史（back to back）
    - 以 GHR 为基准，决定大部分位
    - 按流水级情况，**当拍动态**决定最新的分支历史位

无空泡！



分支历史管理机制——香山处理器

- 目的：解决取指空泡问题
  - **准确性**：分支历史寄存器（GHR）
    - 在 BPU 末端流水级，由最终预测结果推测更新
    - 正确路径在**该流水级**见到的历史永远是准确的
  - **及时性**：最新的分支历史（back to back）
    - 以 GHR 为基准，决定大部分位
    - 按流水级情况，**当拍动态**决定最新的分支历史位
  - **存储两份备份**分别用于恢复和训练：
    - **恢复**：全部预测结束时见到的 GHR → 保证准确性
    - **训练**：预测开始时见到的最新历史 → 和预测**对应！**

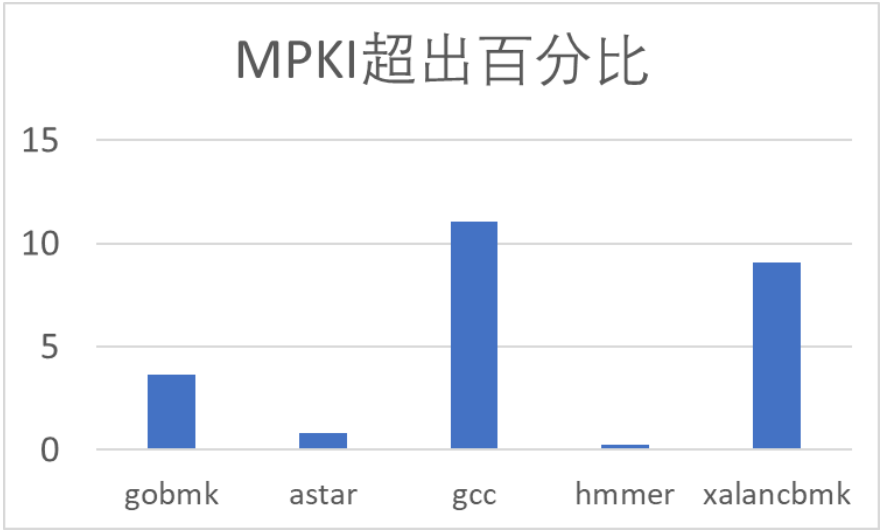


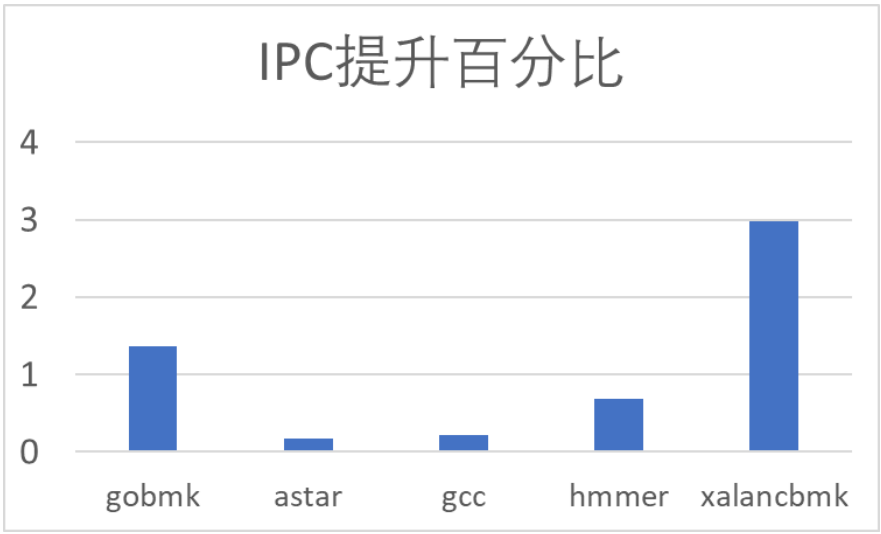
分支历史管理机制——对比

- 相比覆盖重定向方法的优缺点：
  - **无额外取指空泡**
  - 可能降低预测准确率
  - 存储开销相对较大（+3Kb SRAM）

香山处理器分支历史管理机制评估结果

- 我们的机制 **v.s.**覆盖重定向方法：



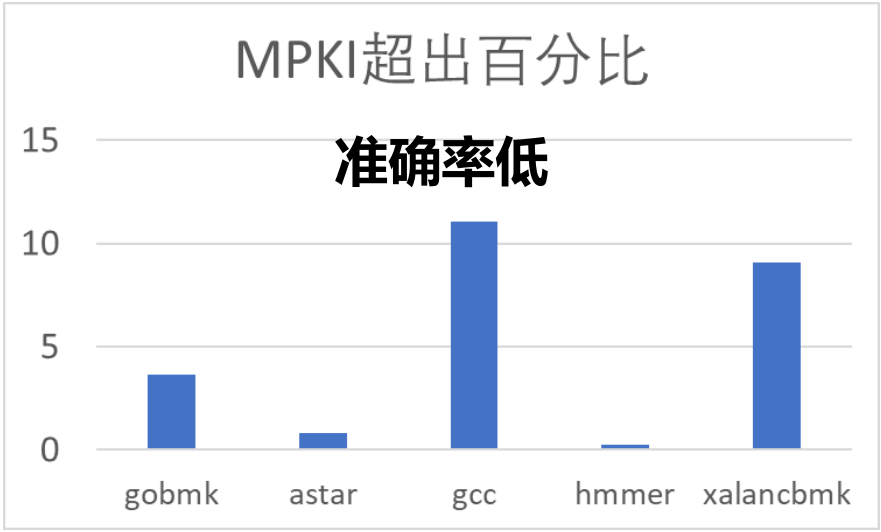


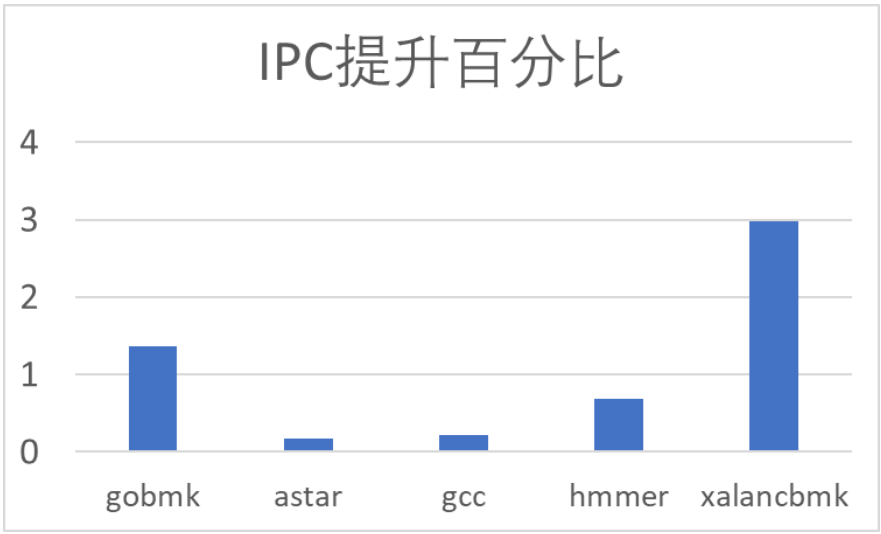
注：测试片段通过SimPoint生成

香山处理器分支历史管理机制评估结果

- 我们的机制 **v.s.**覆盖重定向方法：
  - 增加误预测

注：测试片段通过SimPoint生成

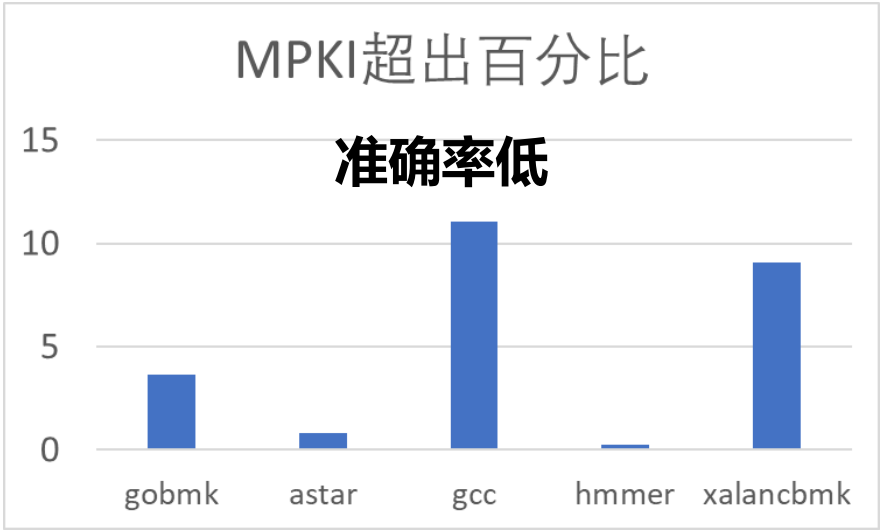


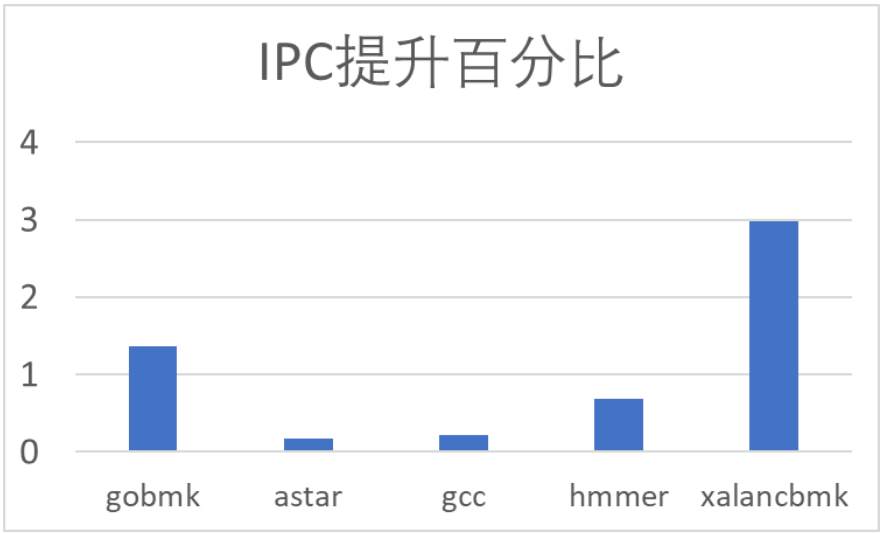


香山处理器分支历史管理机制评估结果

- 我们的机制 **v.s.**覆盖重定向方法：
  - 增加误预测

但是…



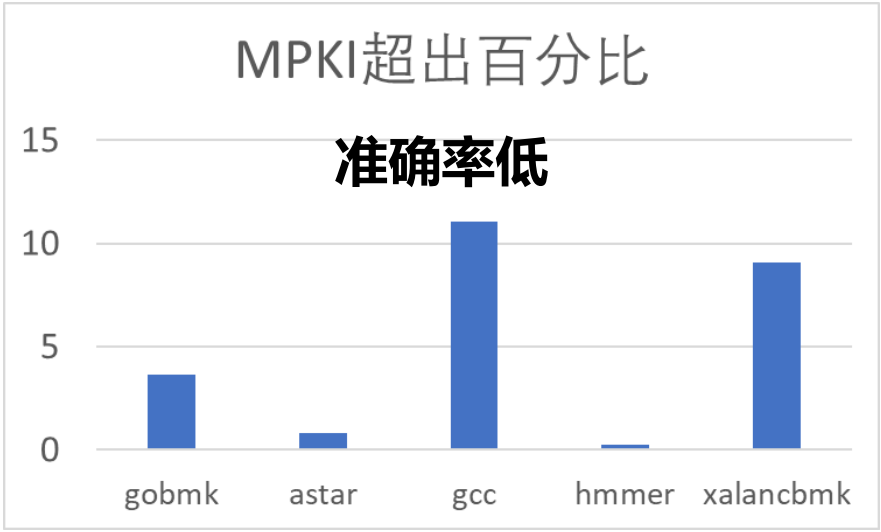


注：测试片段通过SimPoint生成

香山处理器分支历史管理机制评估结果

- 我们的机制 **v.s.**覆盖重定向方法：
  - 增加误预测
  - 但**提升总体IPC**

但是…



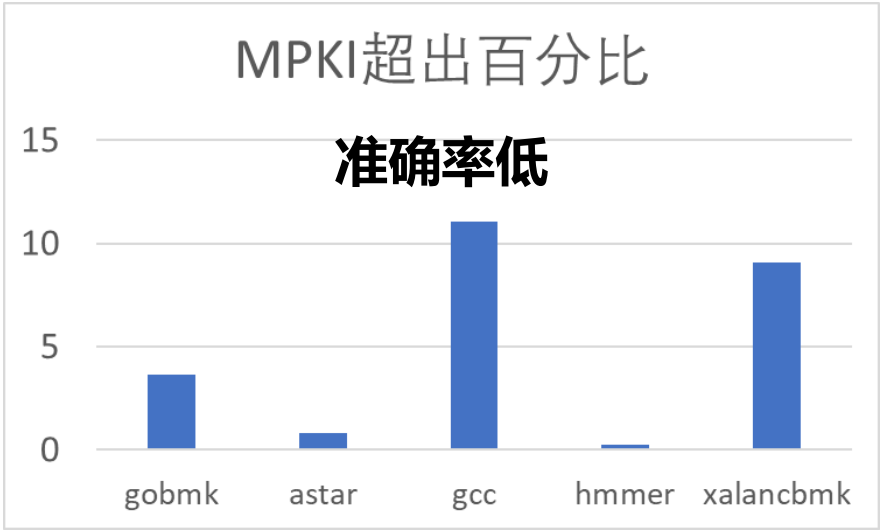
注：测试片段通过SimPoint生成

香山处理器分支历史管理机制评估结果

- 我们的机制 **v.s.**覆盖重定向方法：
  - 增加误预测
  - 但**提升总体IPC**

- **误预测并不是唯一评价指标**

注：测试片段通过SimPoint生成



下一版架构的可能方向

- 和取指单元解耦[1]
- 优化TAGE-SC-L（更长历史…）
- 在S2加入更强的分支方向预测器（GShare[2], Perceptron[3]…）
- 加入间接跳转地址预测部件
- …

[1] Reinman G, Austin T, Calder B. A scalable front-end architecture for fast instruction delivery[J]. ACM SIGARCH Computer Architecture News, 1999, 27(2): 234-245.
[2] McFarling S. Combining branch predictors[R]. Technical Report TN-36, Digital Western Research Laboratory, 1993.
[3] Jiménez D A, Lin C. Dynamic branch prediction with perceptrons[C]//Proceedings HPCA Seventh International Symposium on High-Performance Computer Architecture. IEEE, 2001: 197-206.

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

中国开放指令生态（RISC-V）联盟 China RISC-V Alliance

鹏城实验室 Peng Cheng Laboratory

感谢 北京微核芯科技有限公司 BEIJING VCORE TECHNOLOGY CO.,LTD. **提供产业经验、联合完成结构设计及物理设计**

招募香山处理器二期联合开发合作伙伴

北京微核芯科技有限公司 BEIJING VCORE TECHNOLOGY CO.,LTD.　ByteDance 字节跳动　ESWIN　优矽科技

欢迎更多伙伴加入！

联系人：李迪 13811881360

敬请批评指正！