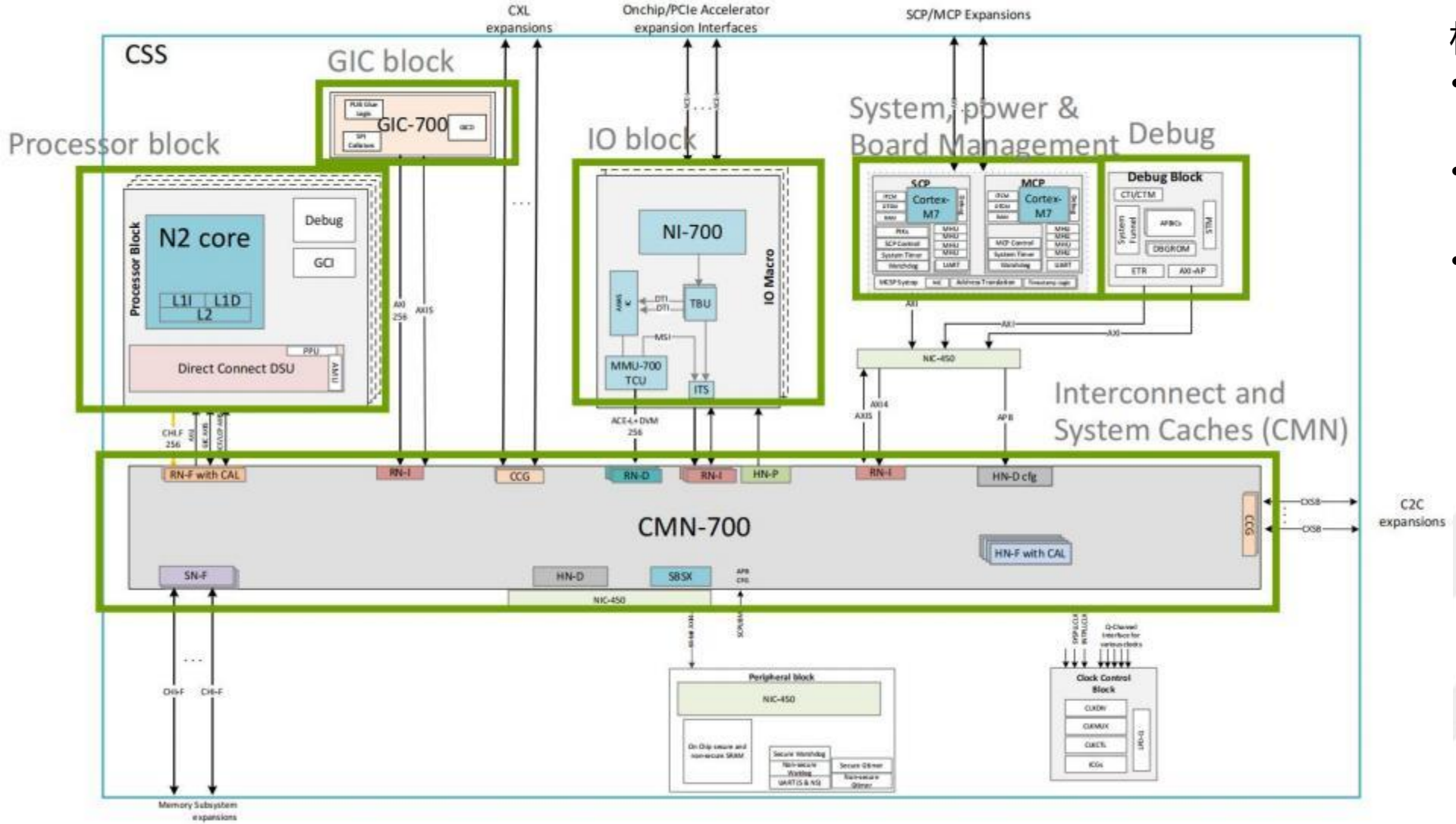
北京开源芯片研究院
BEIJING INSTITUTE OF OPEN SOURCE CHIP

开源高性能SOC IP技术路线图

2024年7月27日

目录

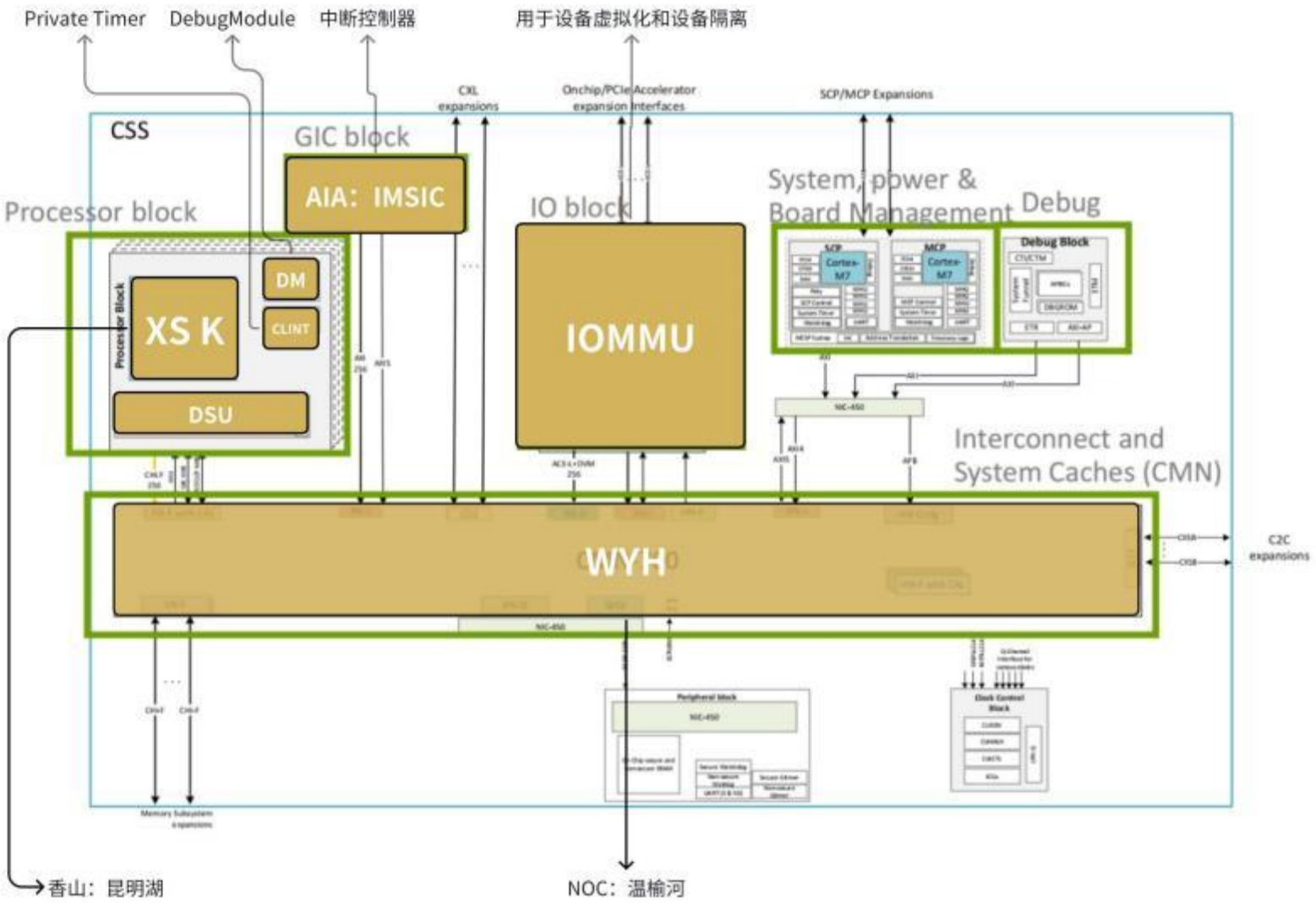
ARM服务器IP生态



相关上游规范/认证

- SBSA(Server Base System Architecture)
- SBBR(Server Base Boot Requirements)
- ARM Server Ready

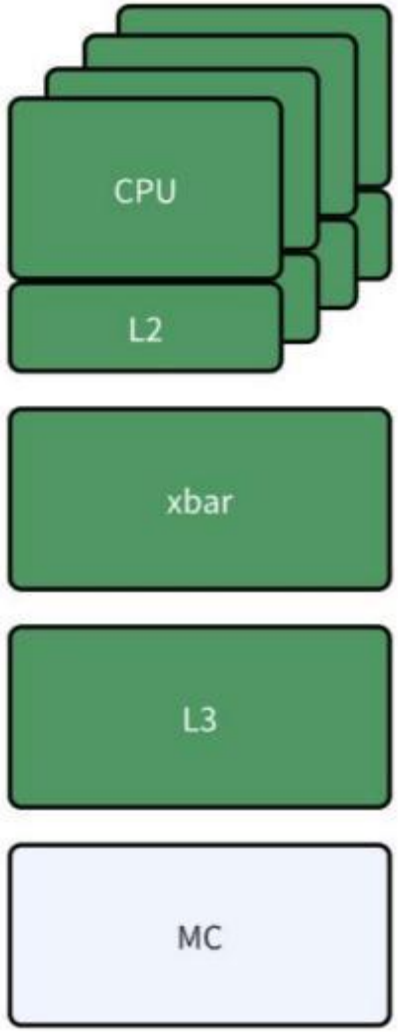
开芯院IP生态



相关上游规范

- RISC-V server platform specification
  - RVA profile
  - RISC-V server SOC specification
  - Boot and Runtime Services specification
  - RISC-V platform security model
- Certification(Certification Steering Commitee)

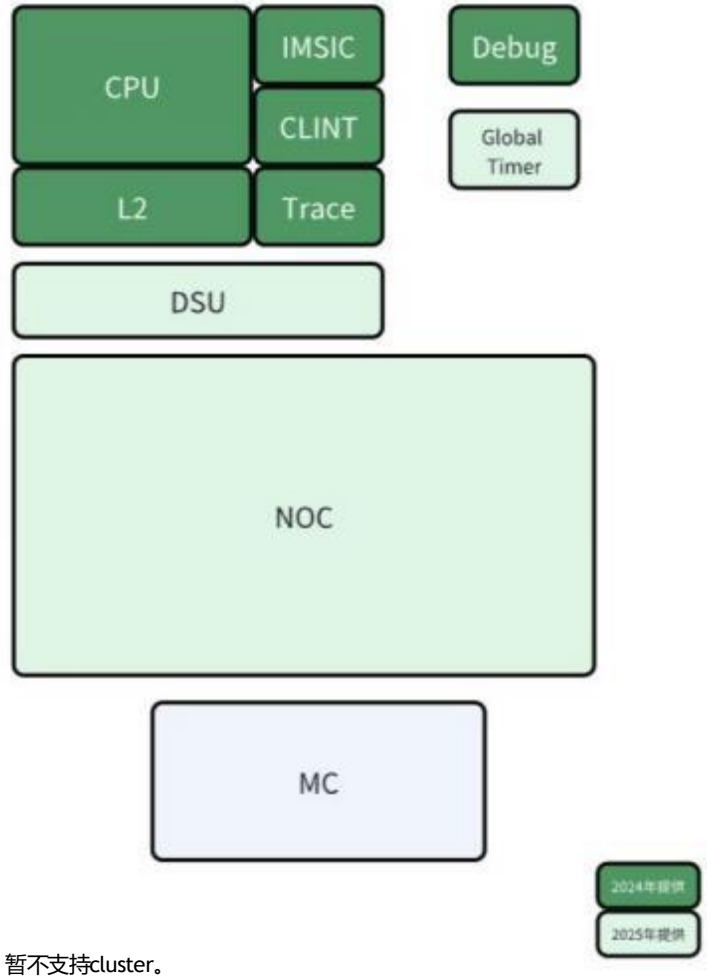
交付方式： 满足NOC拓扑的交付方式



已有交付方式：
CPU subsys交付

服务器场景，
分四部分交付

- CPU subsystem
- DebugModule
- Global Timer
- NOC



Q1： 交付形态有无其它要求?

- 暂不支持cluster。
- 2024年H2 多核验证支持4-8核
- NOC指WYH(温榆河)。竞合伙伴的商业IP不在交付范围之内，例如ARM CMN-600， CMN-700； 可以提供技术支持
- DSU处于研发状态， 2024Q3暂时无法提供。可以提供CHI， TL， AXI等接口的异步桥

服务器SOC IP技术路线图

- Bus protocol
- Memory
- 非标量运算
- 虚拟化
- 调测(debug and profiling)
- RAS
- 安全(security and crypo)
- 电源管理/低功耗
- 可扩展性

上半年已经完成的工作

模拟器
1. RVV 负载分析、 向量架构改进 ， RVV 版的 hmmer 初步获得性能提升
2. 多核 checkpoint 生成工具
3. 香山 GEM5 模拟器的 CHI 整合 ， 可运行多核 checkpoint 和多核 difftest
4. 帕拉丁运行单核 checkpoint 和 difftest ， 仿真速度达到 500+ KHz
5 .RTL 新后端 Chisel 编译速度优化 ， 从 35 min 降低到 3 min
6. 路预测调优 ， dcache 路预测准确率从 60+% 提高到 80+%

Frontend
1. ICache实现规格拆分 ， 缩减面积
2. FTQ重定向中ALU部分实现提前一拍读 ， 提高性能; FTQ去除折叠历史等冗余存储 ， 缩减面积
3. 补充部分clock gating ， 并优化clock gating efficiency ， 优化功耗表现
4. 实现动态关闭FTB ， 节省功耗
5. 修复ICache的X态传输bug; 修复fencei的功能bug; 修复BPU初始化bug

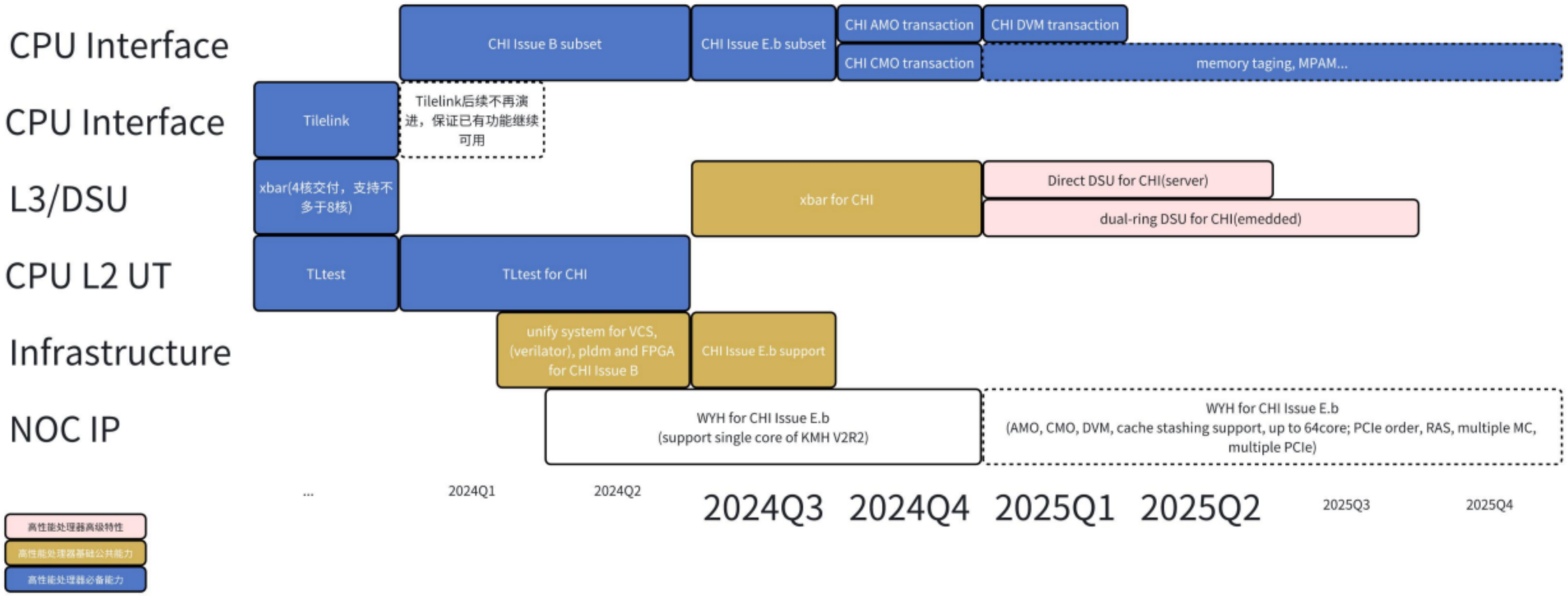
Backend
1. 支持H扩展 ； 实现V扩展运算和调度部分 ， SPEC06INT自动向量化性能接近标量
2. 实现发射后读寄存器堆、 ROB压缩等高性能CPU必备特性 ， 降低面积
3. 实现RV23A中约50%必选扩展 ， 提升对软件生态的适配性
4. 修复CSR里大量不符合规范的bug ， 提升稳定性和安全性
5. 乱序调度算法改进 ， 提升指令分派和发射效率

Memblock
1. 完成 L2 BOP 的虚地址预取 ， SPECfp 性能提升 4.28%
2. 完成乱序向量访存调度改造 ， 实现 Unit-stride 元素合并 ， 向量化 hmmer 分数相比于标量 hmmer预计提升约 14.3%
3. 完成 DCache Evict on Refill 的性能优化 ， SPECfp 性能提升约 1%
4. 优化 MemBlock 门控 ， 静态门控覆盖率由 75% 提升至 95% 以上

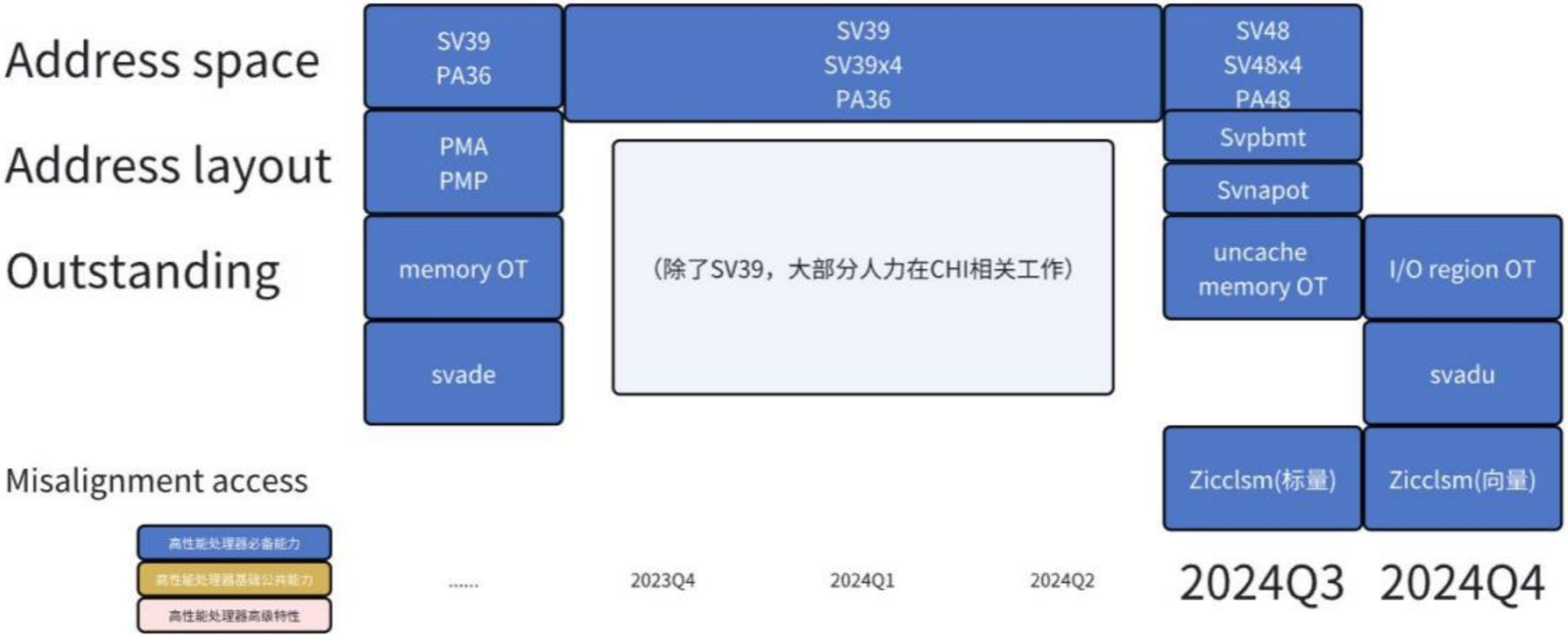
L2Cache
1. 完成原生CHI接口改造
2. 替换算法更新为 DRRIP
3. 添加关键字优先设计 （ L2 优先向 L1 发送【触发 load miss 地址】所在的数据 beat)
4. 采用更加准确的提前唤醒信号计算算法 （ L2 在返回数据的前 3 拍向 L1 发送 Hint 信号）
5. 添加了基于虚地址的 BOP 预取器 ； Temporal 预取所需的 meta 迁移至 L2 Data SRAM 存储 ， 且 meta 和 data 进行统一管理动态分配

2024Q3-2025Q2整体计划

IP support for bus protocol



Memory


Address space
Address layout
Outstanding
Misalignment access
SV39
PA36
PMA
PMP
memory OT
svade
SV39
SV39x4
PA36
（除了SV39，大部分人力在CHI相关工作）
SV48
SV48x4
PA48
Svpbmt
Svnapot
uncache
memory OT
I/O region OT
svadu
Zicclsm(标量)
Zicclsm(向量)
高性能处理器必备能力
高性能处理器高级特性
.....
2023Q4
2024Q1
2024Q2
2024Q3
2024Q4

Memory： 优先级和状态解释

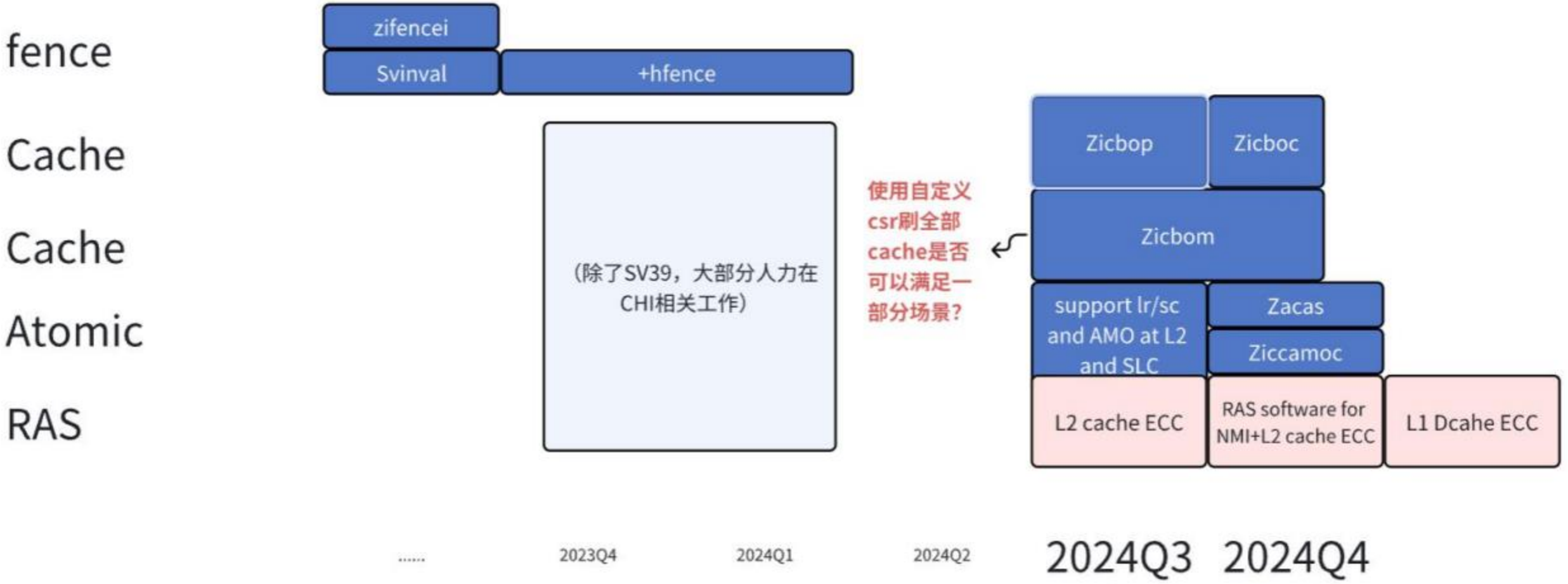
- 有两个特性的目的是满足服务器内存划分需求
  - SV48/SV48x4/PA48： 满足服务器对内存大小的需求；
  - Svpmbt： 支持memory（包块PCIe out-bound memory），I/O（包括PCIe in-bound memory）等服务器memory划分需求；

- 非对齐访问。
  - 根据水人的说法因为应当包括所有load/store（待补充链接），所以RVA23去掉了非对齐需要整数和向量的说法。
  - 昆明湖的想法是先把生态最需要的标量非对齐完成（930目标）。后续再做向量的非对齐（1230目标）。请会员反馈这个优先级是否可以，以及对服务器软件生态的影响。

- I/O region outstanding： Q3不支持会影响寄存器配置性能，但可以理论估计收益，由于人力原因放到Q4完成。

- Svdau： 是服务器标配，服务器场景，不支持会由于更多的page fault影响性能。 目前的想法access和dirty次数计算page fault实际开销，从而计算出没有page fault时的理想性能。所以属于重要不紧急的特性。

Memory(cont)


fence
Cache
Cache
Atomic
RAS
zifencei
Svinval
+hfence
(除了SV39，大部分人力在CHI相关工作)
使用自定义csr刷全部cache是否可以满足一部分场景?
Zicbop
Zicboc
Zicbom
support lr/sc and AMO at L2 and SLC
Zacas
Ziccamoc
L2 cache ECC
RAS software for NMI+L2 cache ECC
L1 Dcahe ECC
.....
2023Q4
2024Q1
2024Q2
2024Q3
2024Q4

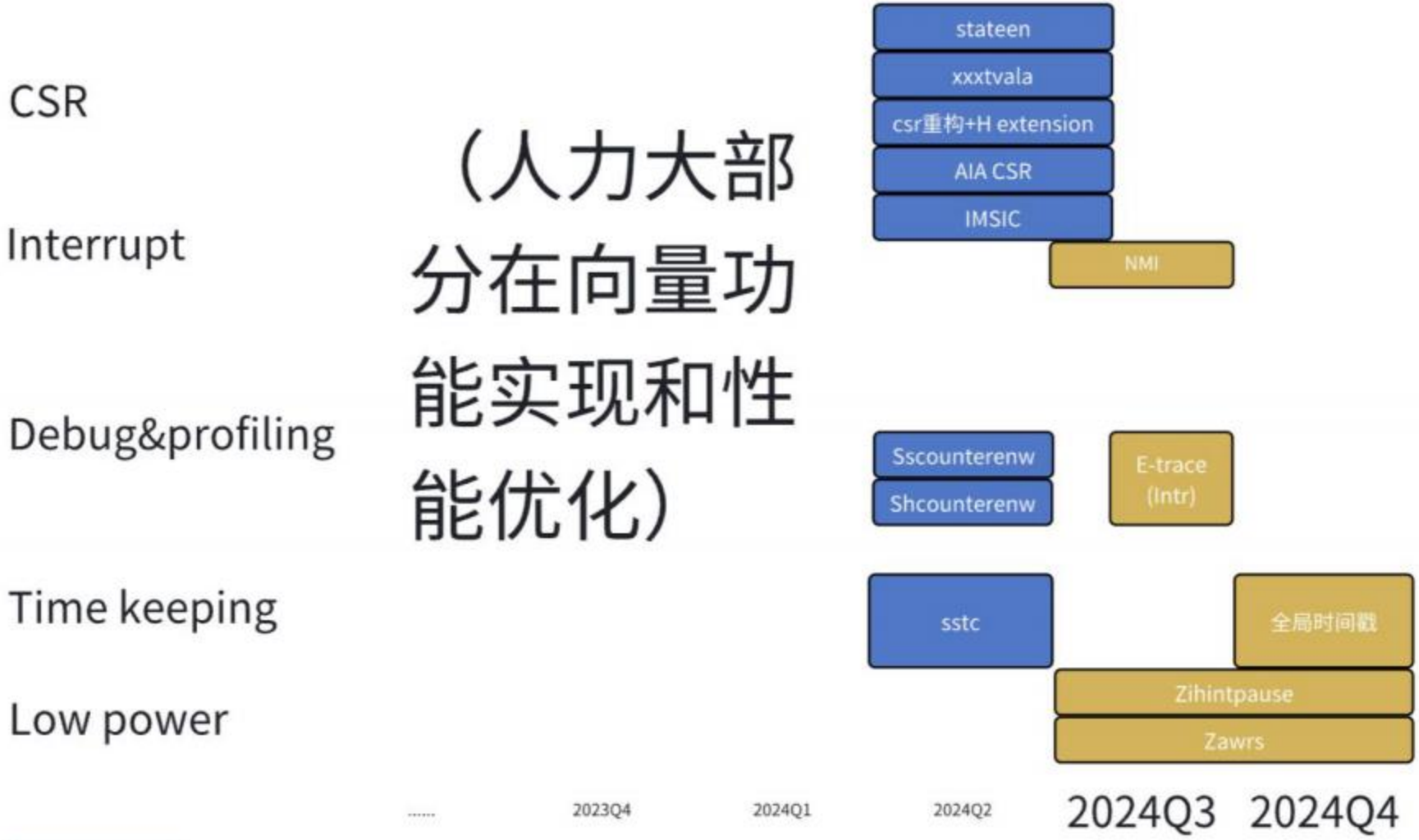
Memory： 优先级和状态解释（cont)

- zicbom
  - 完整实现在Q3来不及，主要是受限于访存的人力。（memory页的特性，大部分需要访存组(memblock组)投入，访存组除了RVA23特性，还有PPA，向量优化等任务）。
  - 对于虚拟机关机，单核关断两个场景来说，直接通过单独的csr flush/invalidates所有cache是否可以满足需要。

- Zifencei和Svinval均已实现。验证Q3完成。

- 原子操作：服务器多核场景，如果不支持所有层次cache的cas指令，当核数较多时，原子操作引起的cache争抢会非常严重，从而影响多核可扩展性。当前的计划是Q3先实现lrsc和原子操作的L2和SLC的支持。Q4实现cas功能，同时优化8核以上的性能（和今年同步准备的多核性能分析工具配合）

privilege level support



privilege level support： 优先级和计划解释

- 昆明湖认为Q3已经实现了服务器所需要的所有特权指令。 请会员补充其它重要特性。
- 对于64核芯片， 全局同步的时间戳是必备的。 当前人力不足， 考虑在Q4实现支持timer和trace的全局时间戳。

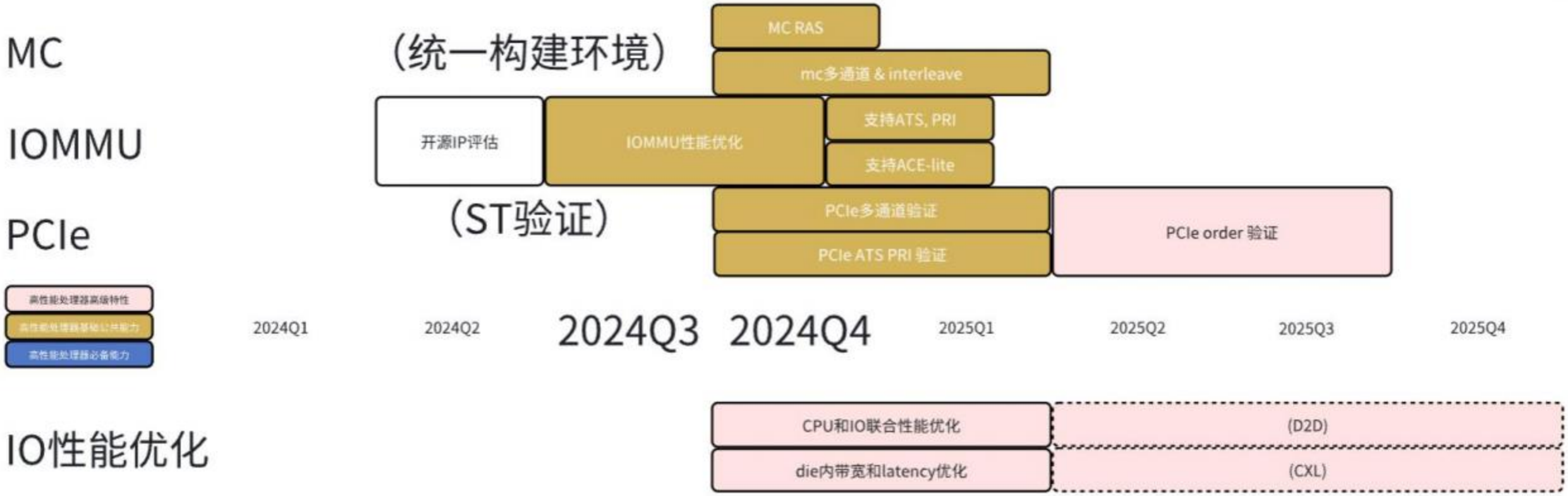
- 昆明湖暂时没有考虑跨die时间戳同步问题。

- Zihintpause和Zawrs涉及CPU低功耗架构， 当前昆明湖尚未支持。 计划Q3都按fencei实现。Q4实现更准确的功能。

非标量运算

- 浮点：实现基本指令集；
- 向量：实现基本指令集；
- ~~BFloat16；~~
- ~~张量；~~
- ~~AME。~~

MC&IO子系统

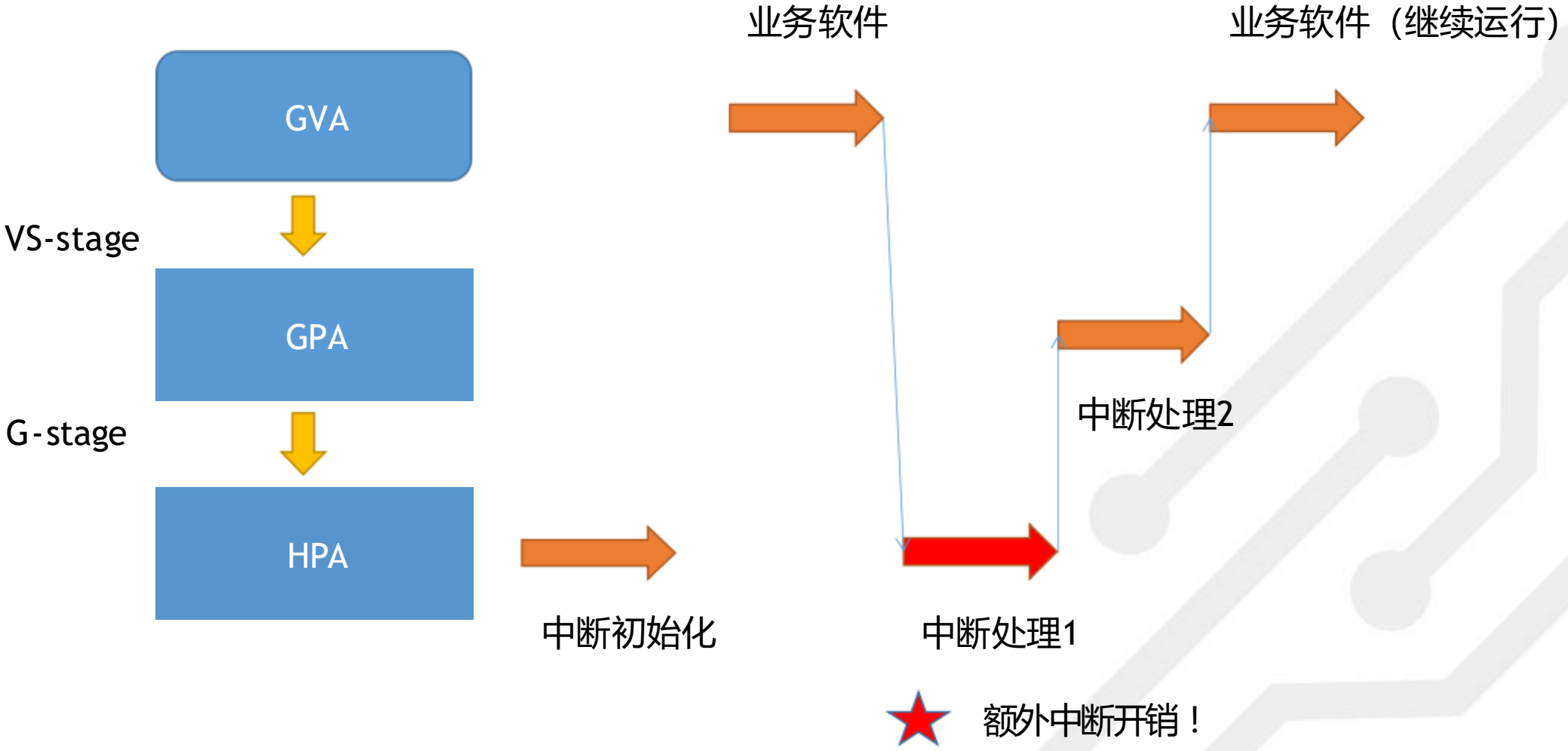

MC
IOMMU
PCIe
IO性能优化
（统一构建环境）
（ST验证）
MC RAS
mc多通道 & interleave
开源IP评估
IOMMU性能优化
支持ATS, PRI
支持ACE-lite
PCIe多通道验证
PCIe ATS PRI 验证
PCIe order 验证
CPU和IO联合性能优化
die内带宽和latency优化
(D2D)
(CXL)
2024Q1
2024Q2
2024Q3
2024Q4
2025Q1
2025Q2
2025Q3
2025Q4

虚拟化

- CPU虚拟化
- 内存虚拟化
- 中断虚拟化
- 设备虚拟化
  (PCIe虚拟化)

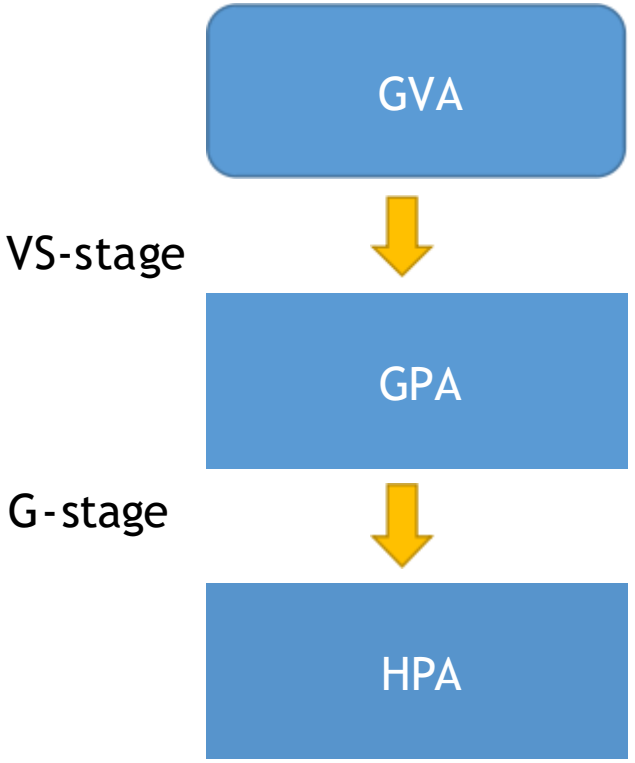
中断虚拟化——为什么

- CPU虚拟化
- 内存虚拟化
- **中断虚拟化**
- I/O虚拟化
- PCIe虚拟化



中断虚拟化——怎么办： sstc

- CPU虚拟化
- 内存虚拟化
- **中断虚拟化**
- I/O虚拟化
- PCIe虚拟化



需要处理的三类中断

- Tick -> RVA23 sstc
- IPI -> AIA IMSIC
- I/O -> AIA IMSIC + AIAAPLIC

RVA23 sstc

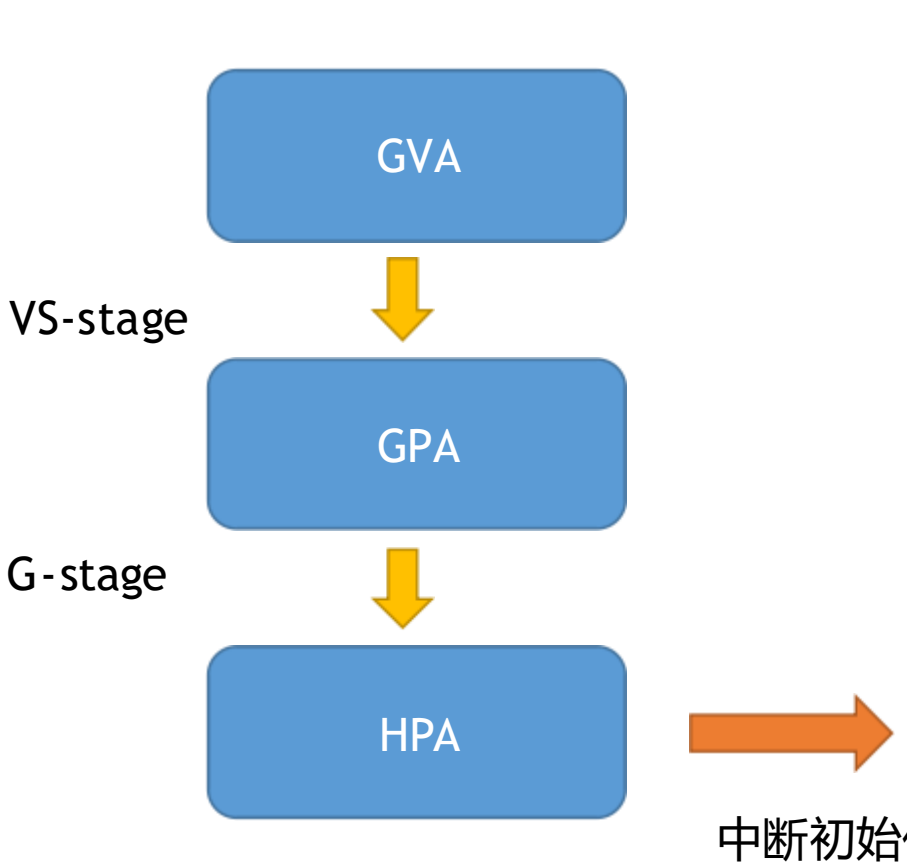
- 支持在s mode的tick
- 支持在vs mode的tick

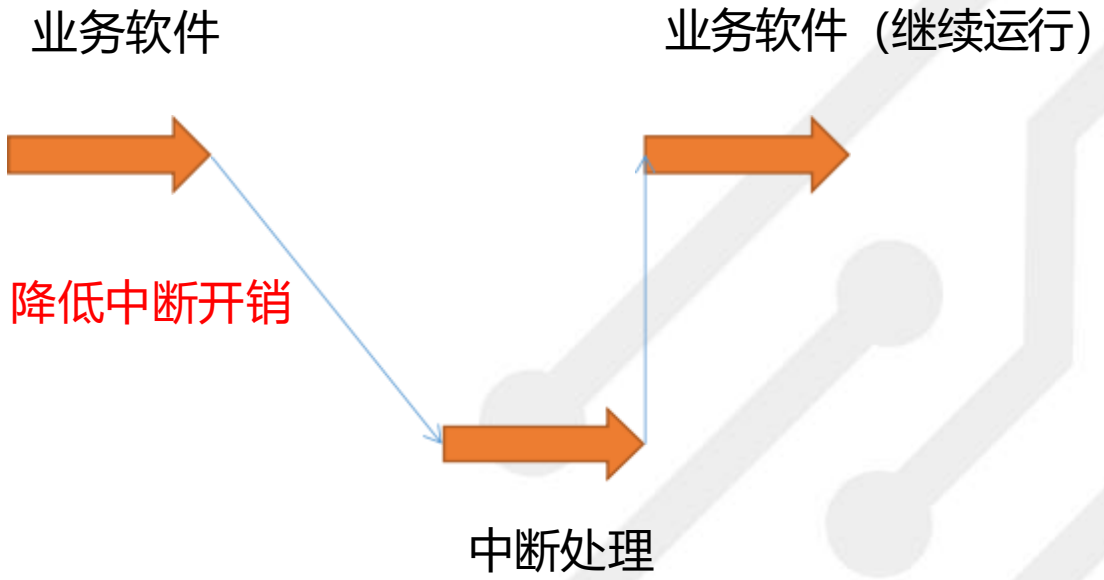
中断虚拟化——怎么办： AIA

- CPU虚拟化
- 内存虚拟化
- **中断虚拟化**
- I/O虚拟化
- PCIe虚拟化
- 机密虚拟机

需要处理的三类中断

- Tick -> RVA23 sstc
- IPI -> AIA IMSIC
- I/O -> AIA IMSIC + AIAAPLIC

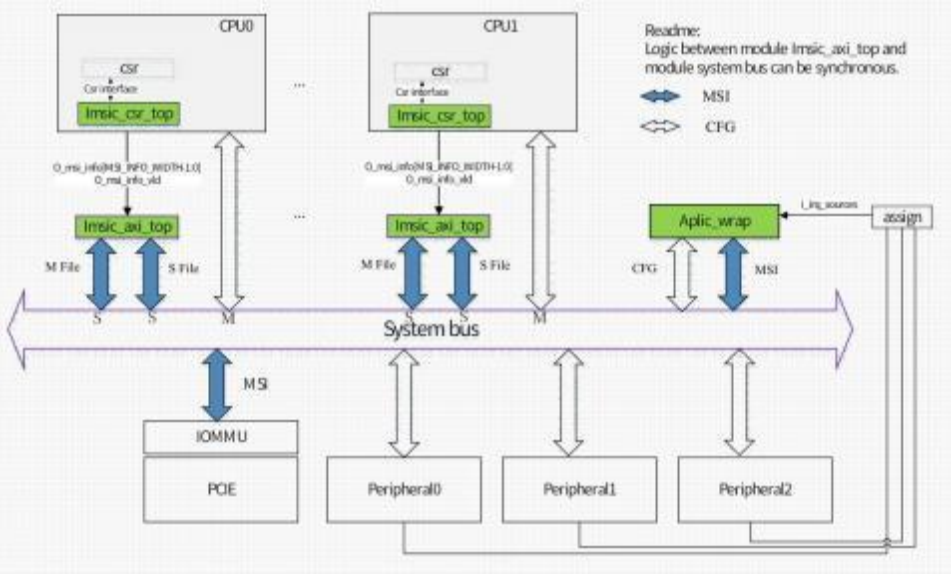


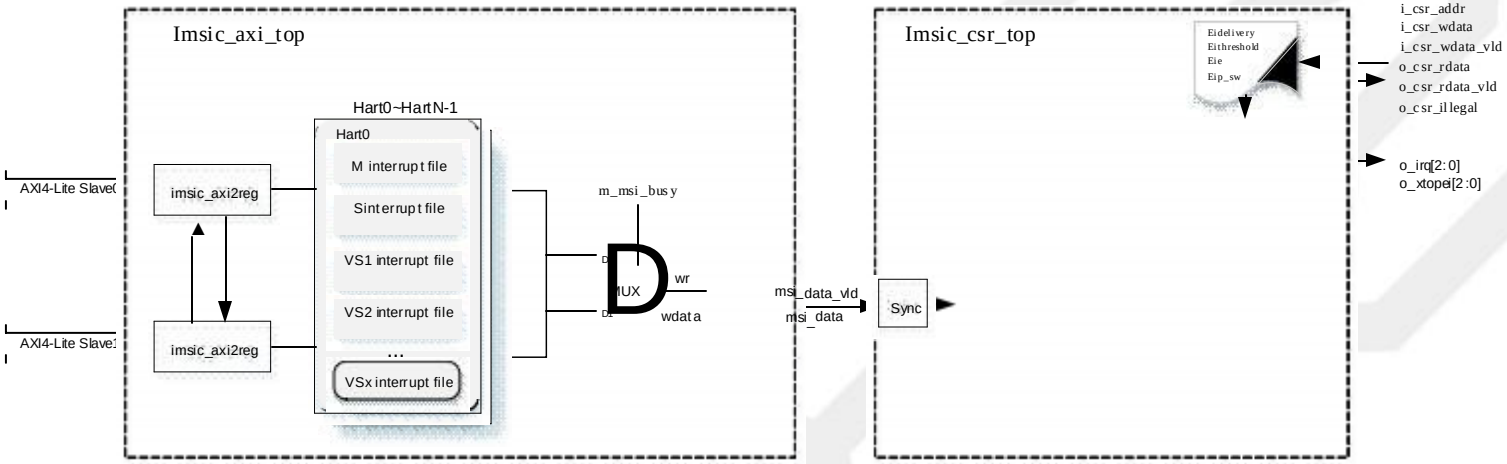


中断虚拟化——AIA实现

- CPU虚拟化
- 内存虚拟化
- **中断虚拟化**
- I/O虚拟化
- PCIe虚拟化
- 机密虚拟机

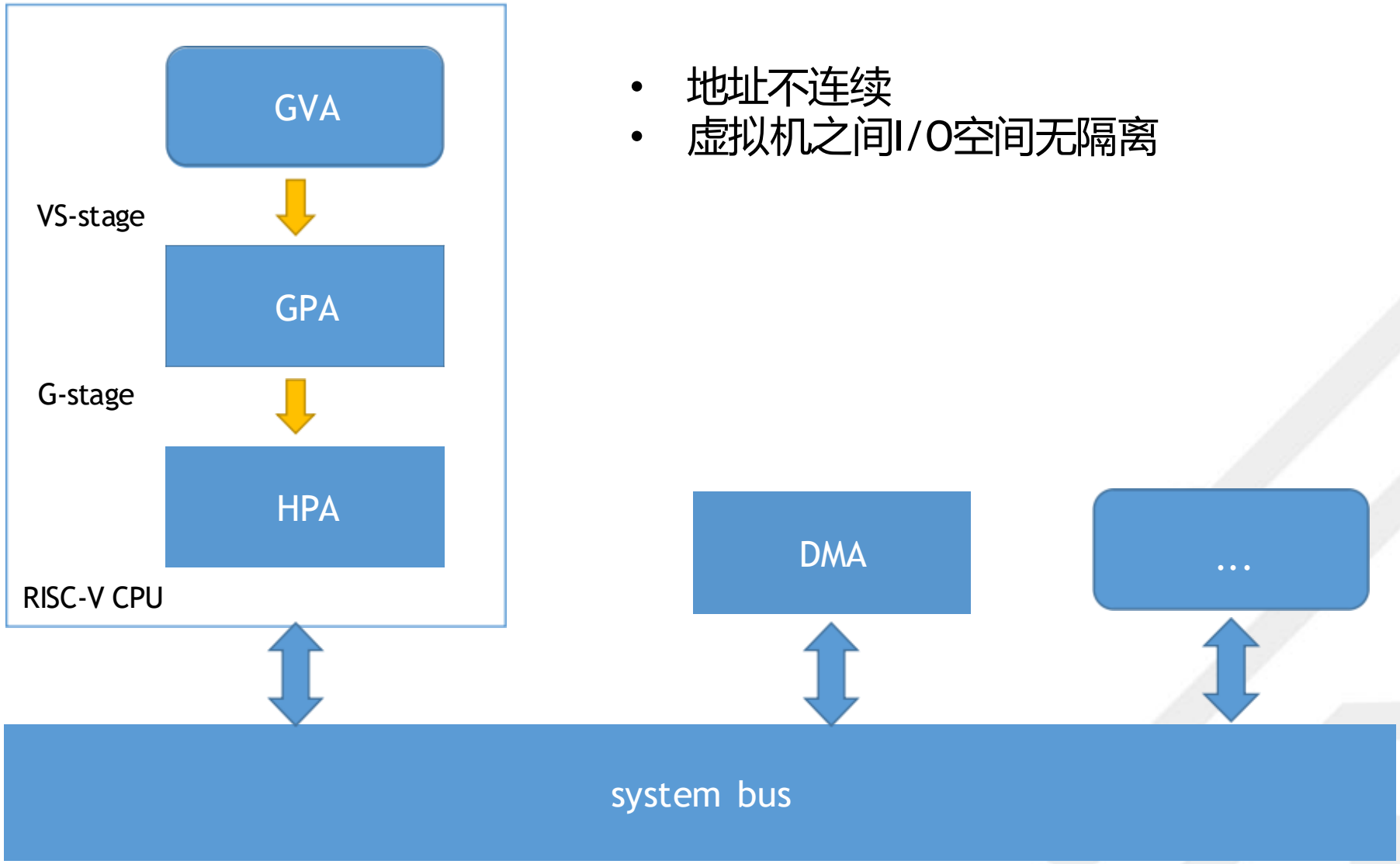
AIA: IMSIC+APLIC			
组件	交付点	交付目标	交付件
IMSIC	2024.4.25	imsic典型用例冒烟通过	1.imsic详细设计方案; 2.imsic RTL代码; 3.imsic用例以及验证结果说明
	2024.5.25	imsic完备用例验证通过	imsic_testplan
	2024.6.15	imsic lint及综合完成	imsic lint及综合check报告
aplic	2024.7.30	完成aplic详细设计方案及代码初版	1.aplic详细设计方案; 2.aplic RTL代码;
	2024.8.30	完成aplic UT验证	1.aplic testplan 2.aplic验证结果说明





I/O虚拟化——为什么

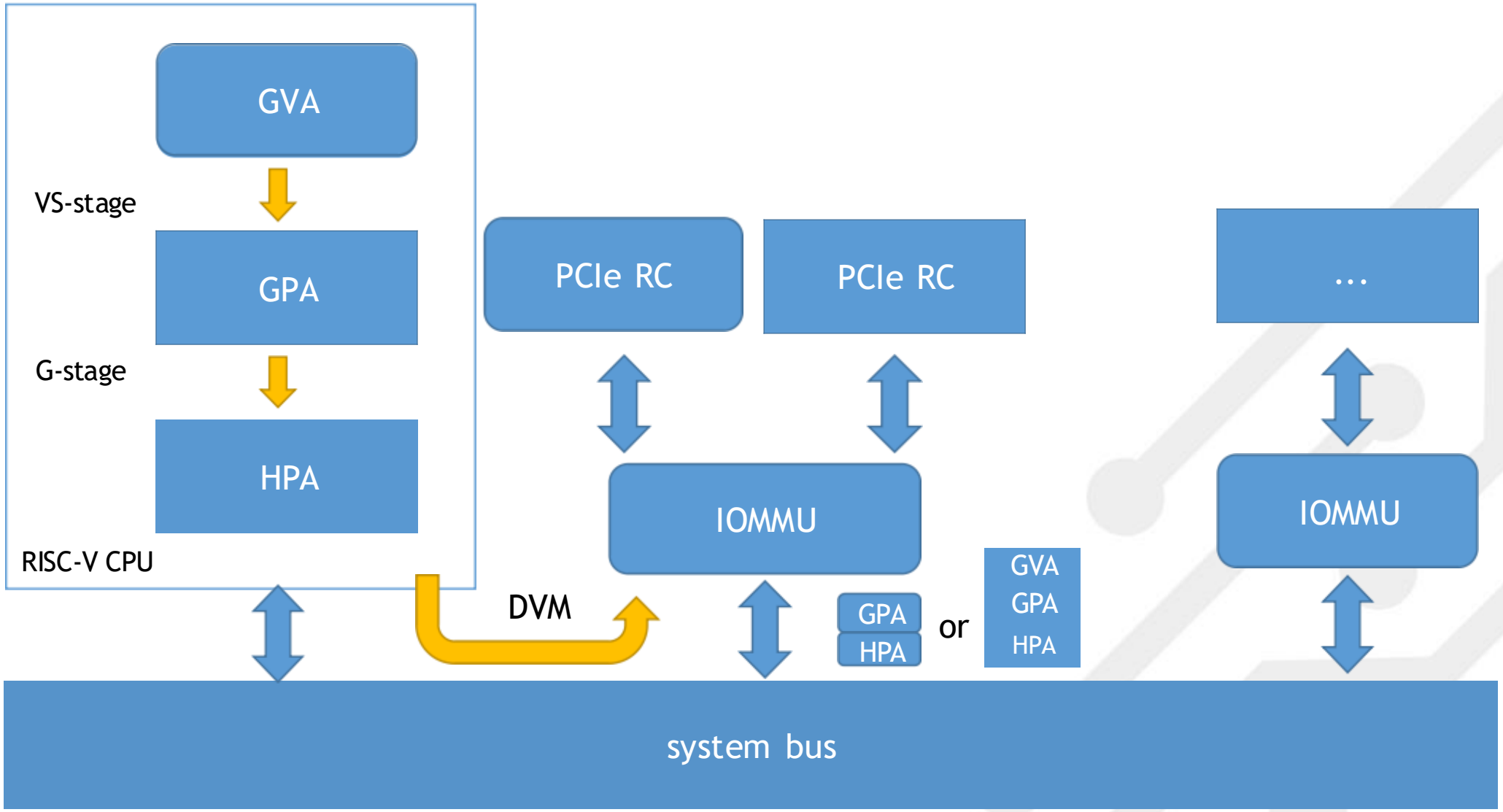
- CPU虚拟化
- 内存虚拟化
- **中断虚拟化**
- I/O虚拟化
- PCIe虚拟化



- 地址不连续
- 虚拟机之间I/O空间无隔离

I/O虚拟化——怎么办

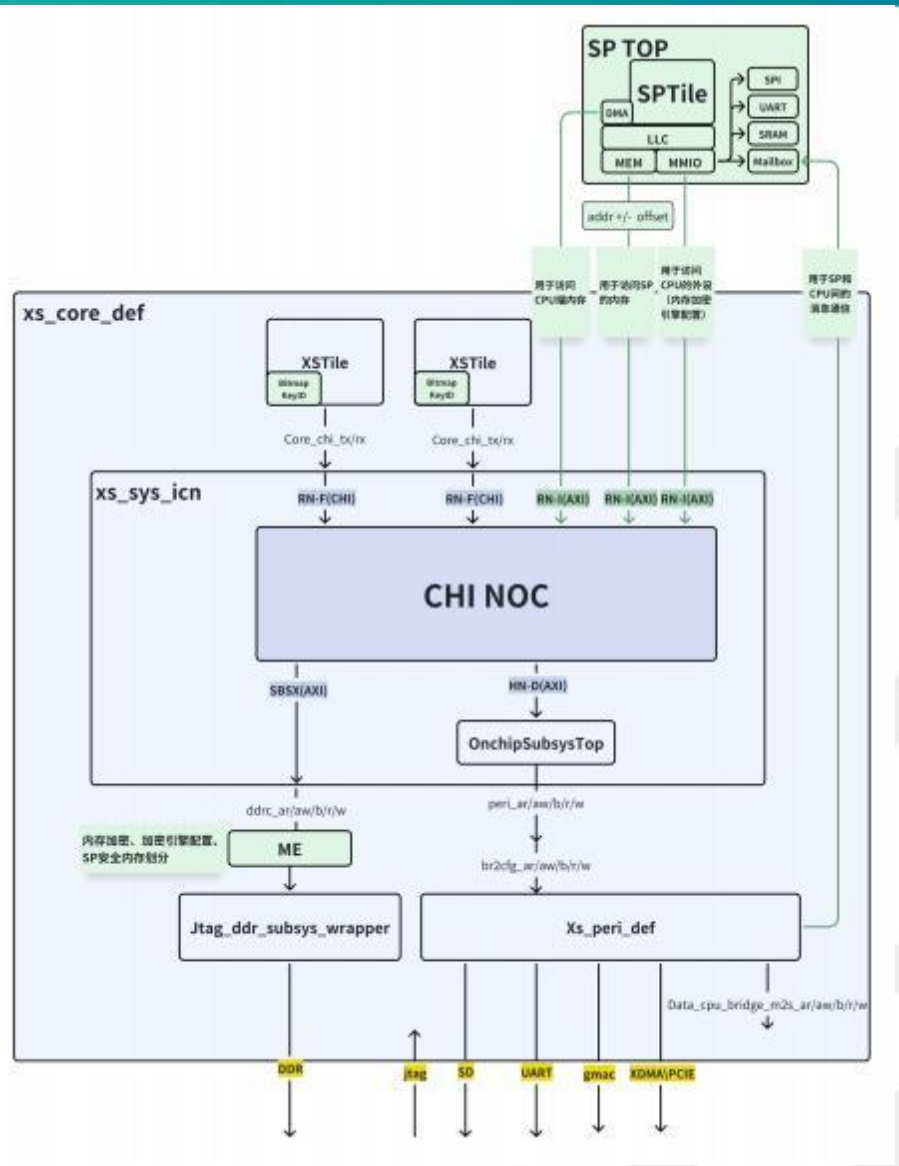
- CPU虚拟化
- 内存虚拟化
- **中断虚拟化**
- I/O虚拟化
- PCIe虚拟化



机密虚拟机

- CPU虚拟化
- 内存虚拟化
- 中断虚拟化
- I/O虚拟化
- PCIe虚拟化
- **机密虚拟机**

- 内存隔离
- ROT（可信启动）
- 机密虚拟机
- 指针鉴权
- 防御侧信道攻击
- 控制流验证



RAS(RISC-V server SOC/platform spec)

- CPU(cache ECC)
- SLC
- memory controller

电源管理（2024Q4-2025Q2）

- 低功耗指令集（见异常部分）
- CPU WFI之后的行为 ： WFI后是否需要正确处理snoop
- 对齐P channel能力 ： 是希望支持retention和powerdown ， 还是希望按P channel实现。
- 对齐Q channel能力 ： 是希望按照Q channel协议实现 ， 还是可以只实现功能 ， 使用自定义协议。

调测(debug & profiling)（930之前完成）

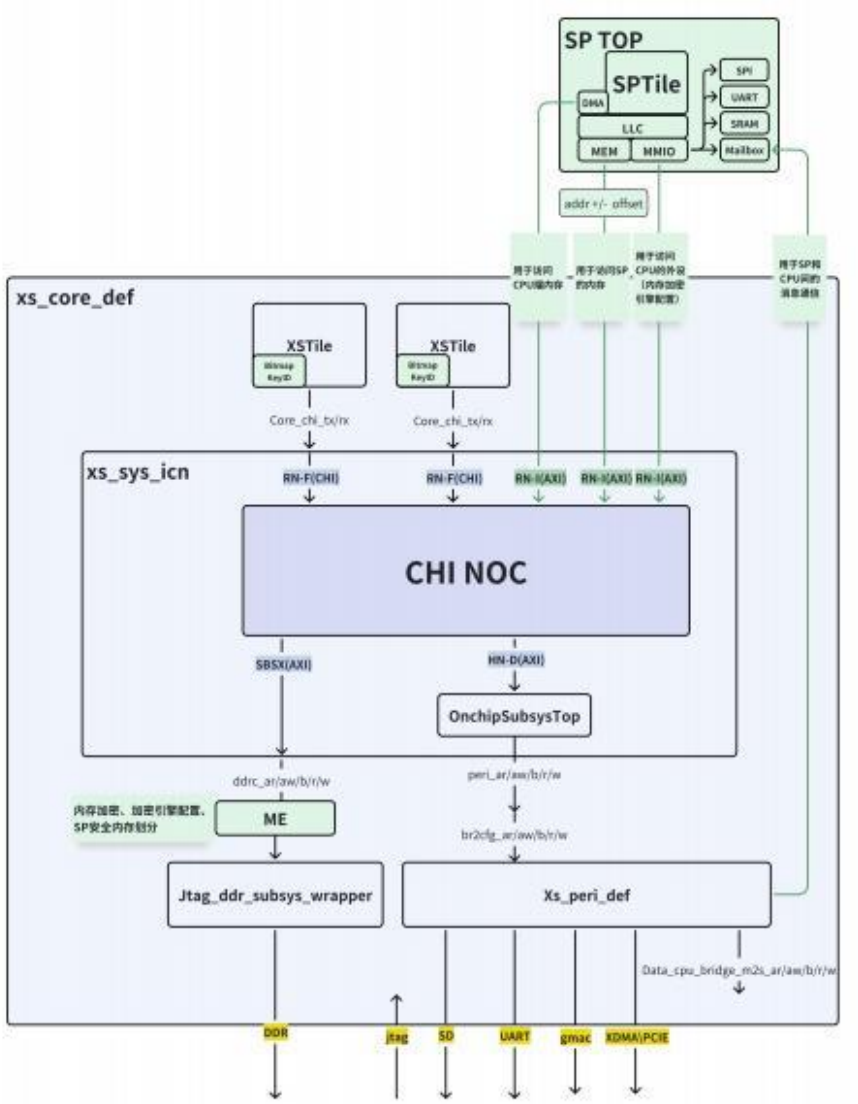
- HPM(Hardware Performance Monitor)： DONE
- debug module： 基于0.13版本
- trace(E-trace, intruction trace)
  - 基于2.0.3版本的E-trace
  - 与西门子IP集成，开芯院提供trace cpu interface接口（指令trace部分），外部提供西门子IP的集成方案。
  - 开发，集成和验证工作量待评估

可扩展性（2025-2026计划）

- 异构
- 多通道memory
- 多通道PCIe
- 跨die
- CXL

安全

- 内存隔离
- ROT（可信启动）
- 机密虚拟机
- 指针鉴权
- 防御侧信道攻击
- 控制流验证



Thanks

谢谢

版本记录

- 2024-1-5 v0.1 初稿
- 2024-1-5 v0.2 第一次讨论
  - 第一次讨论： 唐丹， 陈键， 李贤飞， 刘珊， 马久跃， 张健
  - 香山高性能芯片场景有两个： 服务器芯片和复杂终端SOC；
  - 按128核+计算场景设计服务器芯片， 网络， 存储和视频编解码场景的定制和剪裁由开芯院的商业公司负责；
  - 复杂终端SOC表示手机AP（不含baseband）和chromebook AP等芯片， 按8核设计。
  - 复杂终端SOC是采用ring还是mesh有争议， 争议点同时做两套方案是否有足够资源； mesh能否支持低功耗设计。 唐丹老师倾向于用ring；
  - （陈键）： CPU确定需要做CHI的接口， 具体时间点待定
  - （陈键）： AIA优先级高， 尽快做工作评估；
  - （陈键）： 对标x86， ARM的top-down实现香山HPM；
  - （陈键+李贤飞） P0： RAS继续分析；
  - （李贤飞） P1： IOMMU优先级高， 尽快做工作评估， 尤其是对CPU核是否有影响， 能不能买NI700解决问题；
  - （李贤飞） P2， PCIe ACPI验证
  - （李贤飞） P2， 搭建FPGA开源平台， 推广香山生态。
  - （李贤飞）发行版支持openEuler（ 已经可以运行最新版本23.09， 后续推进与社区合作）

版本记录（续）

- 2024-1-11： v0.3 添加香山当前状态框图， RISC-V标准演进框图
- 2024-7-17: 240717 和唐老师初步讨论
- 2024-7-23: 240723 和张林隽， 陈熙， 胡轩（部分） ，陈键（部分）讨论
- 2024-7-27： 与知合计算CPU设计总监刘畅交流后刷新
- 2024-8-9： 为了RISC-V中国峰会刷新： 题目： 香山服务器现状和IP技术路线图。

备用：香山对安全的支持