AMD MI100 的 SE 与 CU 禁用统计

统计了手头 62 张 MI100 后, 我得到了它们的分布.

---

debugfs 是 Linux 内核提供的调试文件系统, 把驱动的内部状态和调试操作做成文件接口. AMDGPU 的节点通常位于 `/sys/kernel/debug/dri/N/`, 需要 root 权限访问. 其中 `name` 用于确认 GPU, `amdgpu_gca_config` 提供计算资源规模等硬件配置, `amdgpu_regs` 则允许读取寄存器, 也就是硬件保存状态和配置的存储单元.

可直接 cat 的状态信息包括 `name` 和 `amdgpu_pm_info`.

此外, debugfs 还能提供温度, 功率, 固件和任务队列等信息..

MI100 有 8 个 SE (Shader Engine, 着色器引擎), 编号 0-7. 每个 SE 有 1 个 SH (Shader Array, 着色器阵列), 包含 16 个 CU (Compute Unit, 执行计算的单元), 编号 0-15. 每卡共有 128 个 CU 位置, 本次样本每个 SE 禁用 1 个, 实际启用 120 个.

通过 `amdgpu_regs` 按 SE 读取 `CC_GC_SHADER_ARRAY_CONFIG` (底层禁用配置) 和 `GC_USER_SHADER_ARRAY_CONFIG` (用户额外禁用配置), 就能得到禁用位图. 按驱动逻辑, 将两个值按位或, 再取高 16 位: 从最低位开始编号, bit i 为 1 表示 CUi 被禁用. `0x` 表示十六进制, 例如 `0x00200001` 的高 16 位是 `0x0020`, 即 bit 5 为 1, 表示禁用 CU5.

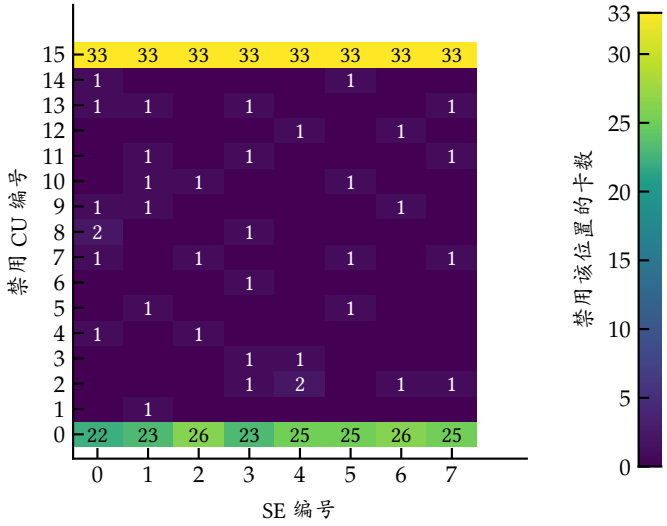


图 1　62 张 MI100 的 CU 禁用位置统计.