



对比不同论文的负载描述方案 (鲍成)

- 研读论文了解其负载定义：Dietrich & Chakraborty、Co-Cap、CRAVE、zTT、Lin et al.
- 设想负载描述目标：用实时硬件信号，将任务映射到一组“原型负载”上；可选择离散分类，也可得到连续的混合系数向量，供调频/建模直接使用
- 具体方案：
 - 负载特征： f_{cpu} , f_{gpu} , f_{ddr} ; $util_{cpu_compute}$ / $util_{cpu_stall}$ / $util_{gpu}$; MB(GB/s), R/W比
 - 预处理：按设备做 z-score 标准化，滑窗聚合
 - 离线阶段：采集各典型任务（计算密集/访存密集/混合/IO 等）特征，固化为“原型库”
 - 在线阶段：对当前窗口特征向量 x_t 计算各原型相似度（马氏或余弦或MSE），归一化得到混合系数

内存频点、带宽与性能能耗的关系 (鲍成 & 程诗淇)

$$\text{Bandwidth} = f_{clock} \times (\text{transfers per clock}) \times \frac{\text{bus width (bits)}}{8} \quad (\text{注：这里是峰值带宽})$$

- CPU 频点主要决定算力上限，内存频点主要决定数据等待时延
- 读写比例影响带宽-时延曲线形状
- 频点提升若不匹配带宽，将导致性能反向退化

CPU、GPU频点对于给定应用的影响 (任德上)

以在jetson上使用deepstream为例，设置了三种情形并分别统计调节CPU和GPU频率对视频fps的影响情况：

- 只播放时，CPU频率是影响帧率的主要瓶颈
- 只推理时，CPU频率是影响端到端延迟的主要瓶颈，GPU频率是影响推理时间的主要因素，CPU频率对推理时间也有一定影响
- 播放+推理时，CPU频率和GPU频率都会影响帧率，GPU的影响更大

调试工具调研及人为干预带宽方法 (OpenHarmony OS) (蔡东辰)

- 调研了一些调试工具：Hiprofiler、HiPerf、HitraceMeter、HiTraceChain、Graphics Profiler、Energy，了解其工具类型以及核心功能
- 人为干预带宽：Intel的RDT(资源调度技术)提供了MBA(内存带宽分配)，可以限制特定进程的可用带宽（适用于intel cpu且系统必须有RDT 驱动与 resctrl 文件系统支持，例如linux系统）