

北京大学 - 智能硬件体系结构

2024 年秋季 作业 5

有一款 DRAM 内存系统，其结构如下：

- **Channel**: 2 个通道
- **DIMM**: 每个通道包含 1 个 DIMM
- **Rank**: 每个 DIMM 包含 2 个 Rank
- **Chip**: 每个 Rank 包含 8 个 Chip
- **Bank**: 每个 Chip 包含 16 个 Bank

其中，不同的通道和 Chip 之间可以并行访问；而不同 Rank 和 Bank 之间需要进行交错（interleaving）访问。假设该 DRAM 内存系统总共提供 96GB/s 的带宽，并且每一个需要进行交错访问的层级的数据准备时间都可以被刚好重叠（overlap）掉。

1. 对于单次 64 位的访问（位地址范围[0:63]），请绘制各层级的数据访问情况图示。(30 分)
2. 请分析每一个层级中每个单元的带宽。(30 分)
3. 对于一个内存受限（memory-bound）类型的问题，假设：
 - CPU 处理数据的速度为 960GB/s。
 - 采用存算一体架构后，计算单元被放置在每个 Bank 旁边，DRAM 中计算单元整体处理数据的速度为 192GB/s。

请分析处理 1GB 数据量时，传统从 DRAM 加载到 CPU 运算与 Bank 级别存算的所需时间。
(40 分)