

终端产品对体系结构的需求和挑战

史公正 2018/08/08



终端CPU的生存环境介绍

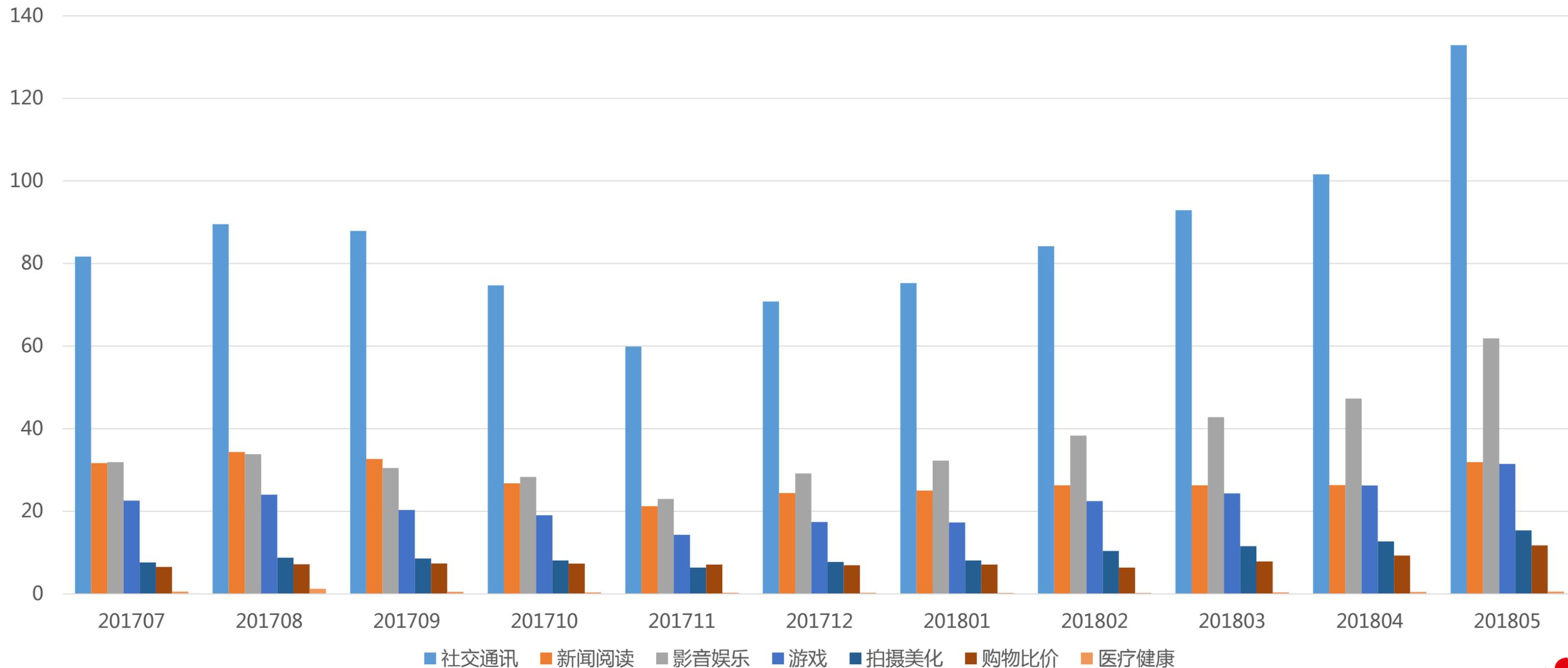
ARM处理器架构在安卓应用中的表现

终端TOP 应用所需求的处理器特点

对于未来国产自研CPU体系结构的期盼

终端产品对于CPU系统的主要挑战场景

各种终端应用使用日益增长

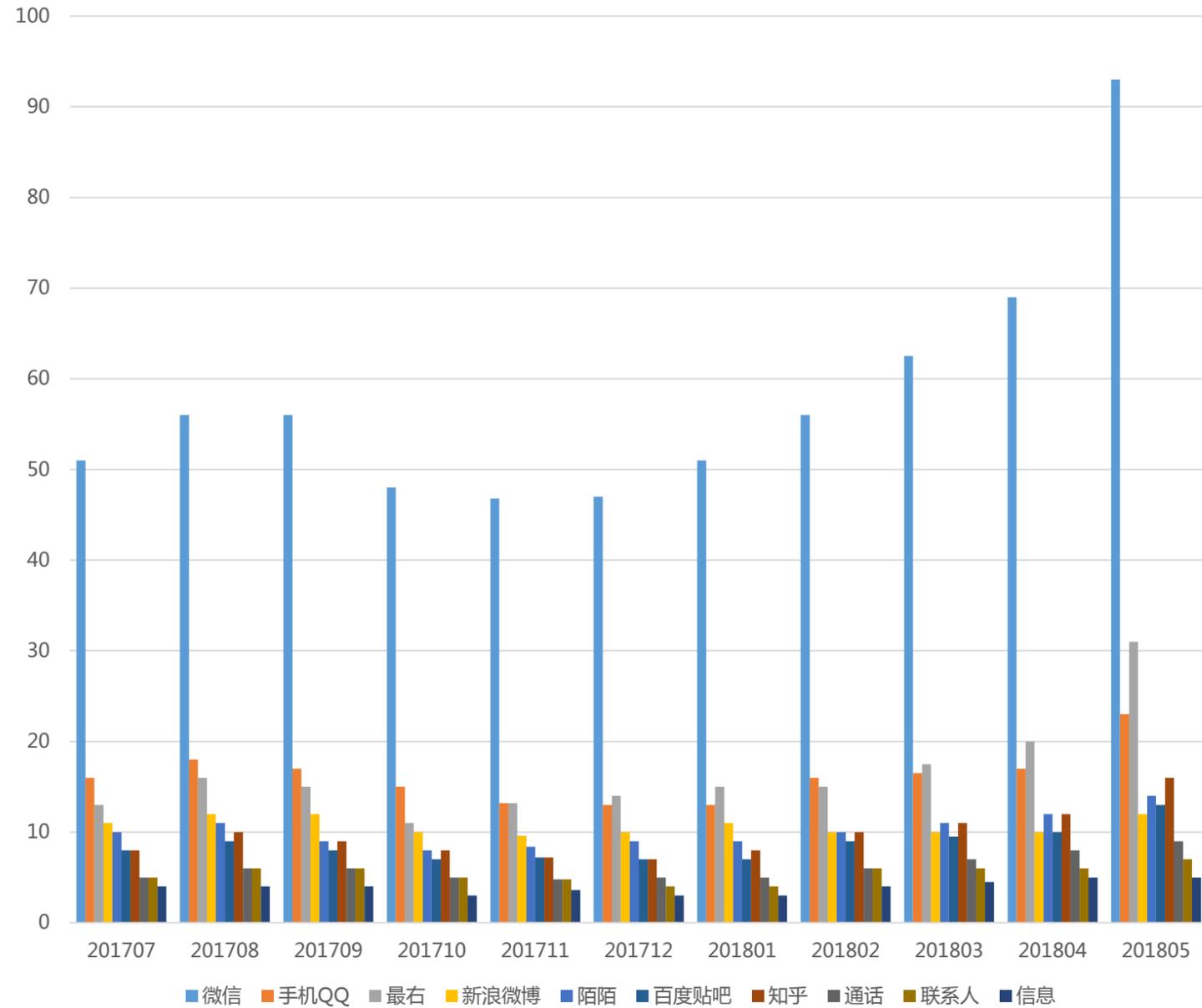


*数据来自华为终端用户体验分析

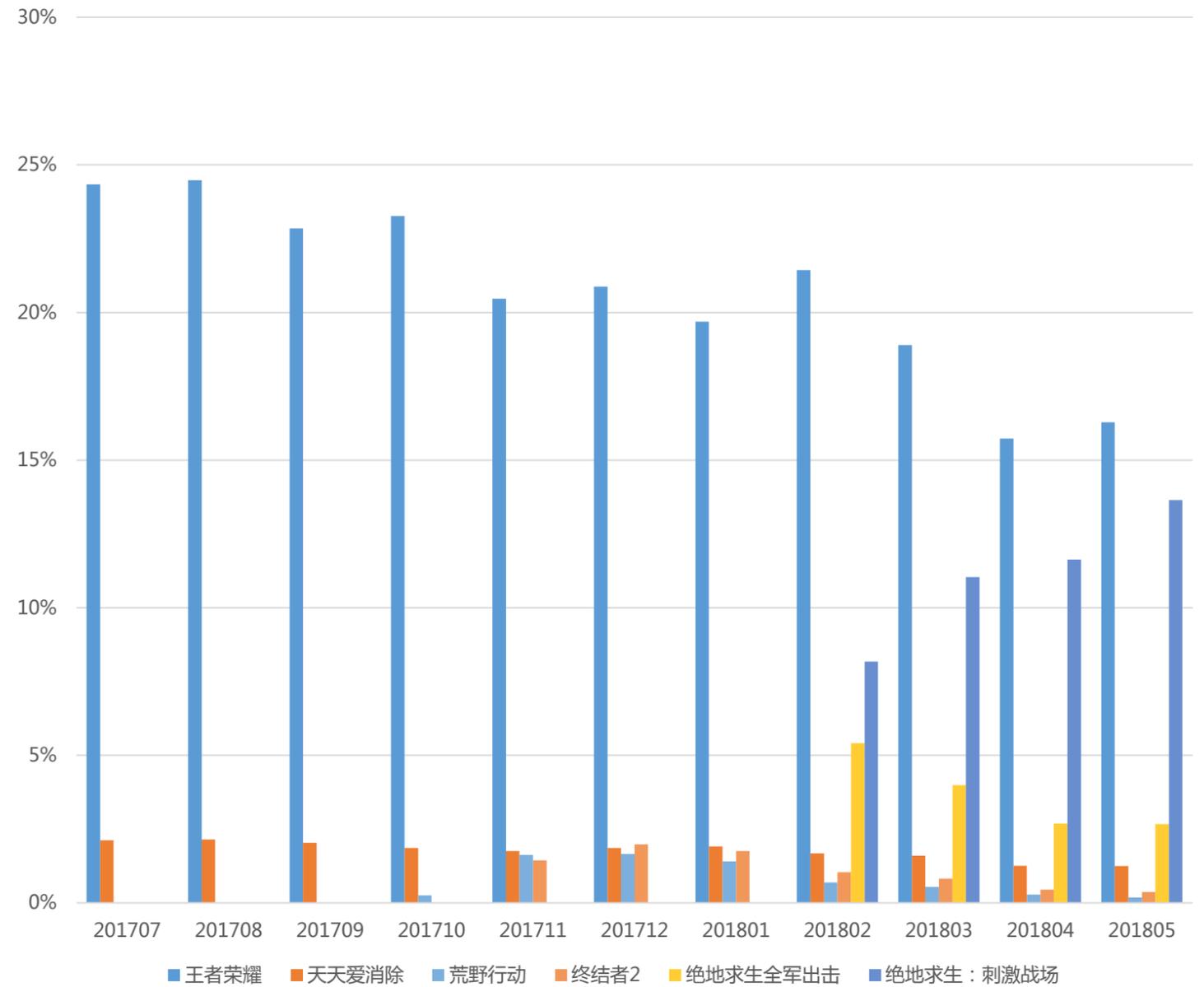
典型轻型、重型CPU负载类应用使用分布

MAKE IT POSSIBLE

轻载社交应用使用比例



重载游戏应用使用比例



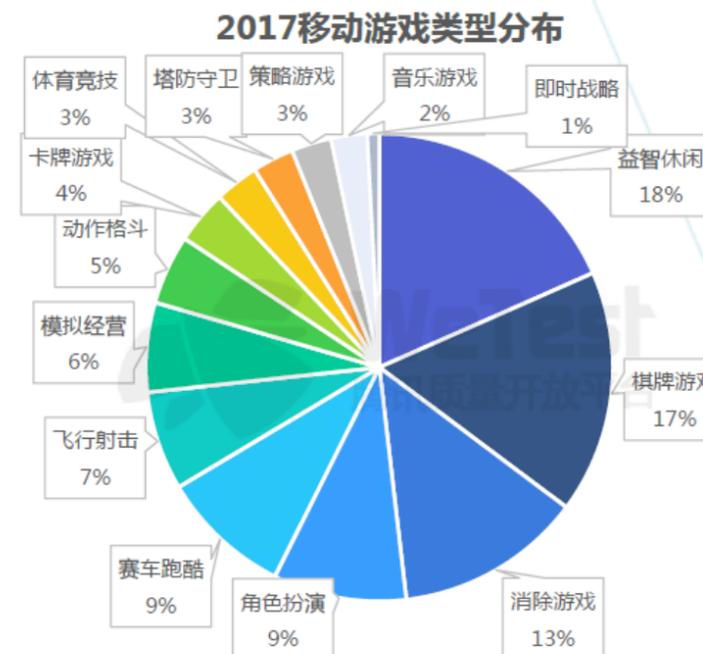
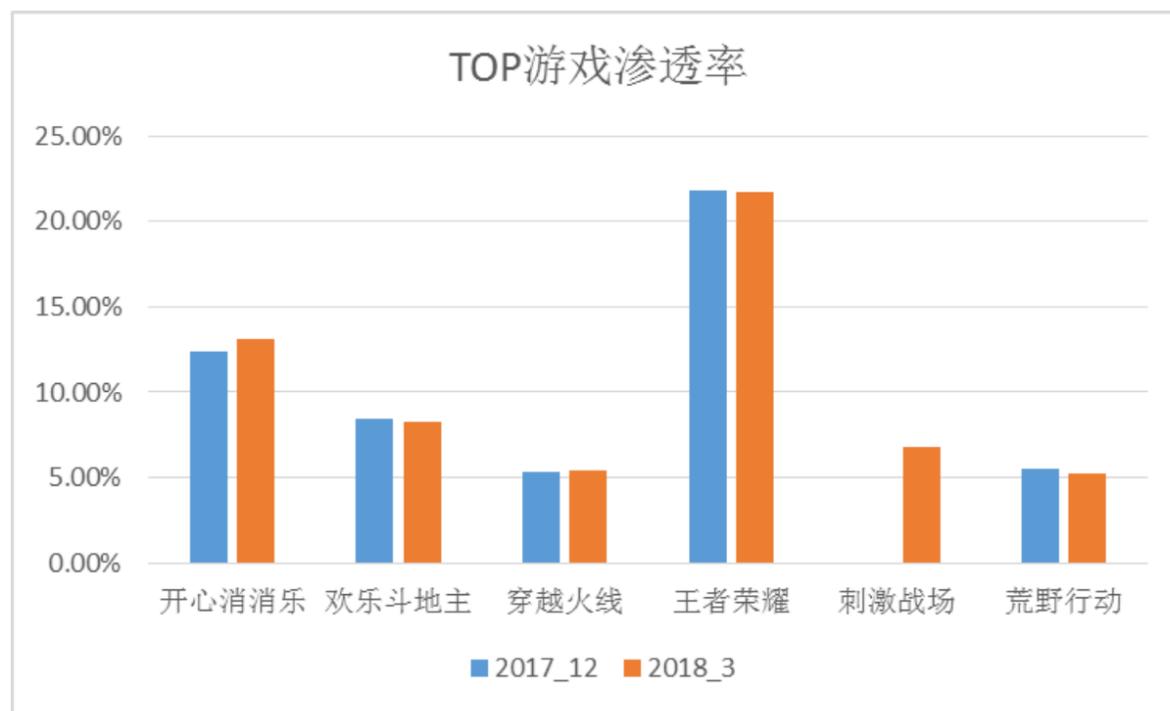
*数据来自华为终端用户体验分析



手游越来越成为手机的重载场景

什么是市场渗透率

对于有形的商品，用户渗透率指的是在被调查的对象（总样本）中，一个品牌（或者品类、或者子品牌）的产品，使用（拥有）者的比例。也可以直接理解为用户渗透率或者消费者占有率，是一个品牌在市场中位置的总和，它是多年形成的结果。

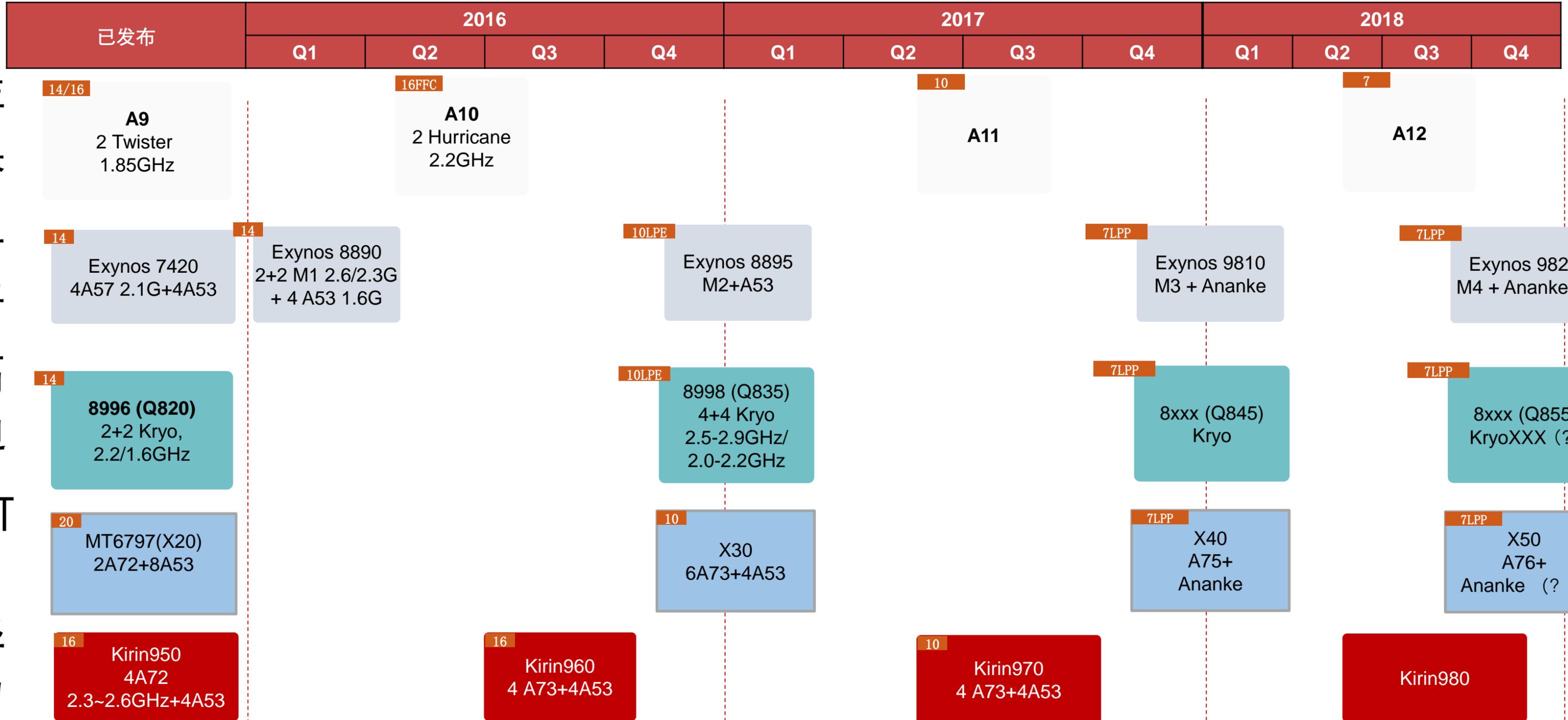


- 1. 截至2017年12月最后一周，中国手游app市场渗透率达76.1%，用户规模为7.76亿
- 2. 每名手游用户平均安装3.35个手游类app
- 3. 欢乐斗地主生命周期为55.2天，开心消消乐为54.1天，王者荣耀为50.7天
- 4. 截至2017年12月最后一周，王者荣耀渗透率为21.78%，12月月均DAU为6398万
- 5. 王者荣耀2017年12月日新增用户均值为180万，荒野行动日新增用户均值为166万
- 6. 2017年新增手游用户中，54.4%为26-35岁用户，女性占比43.2%

终端处理器升级更快-每年一代

MAKE IT
POSSIBLE

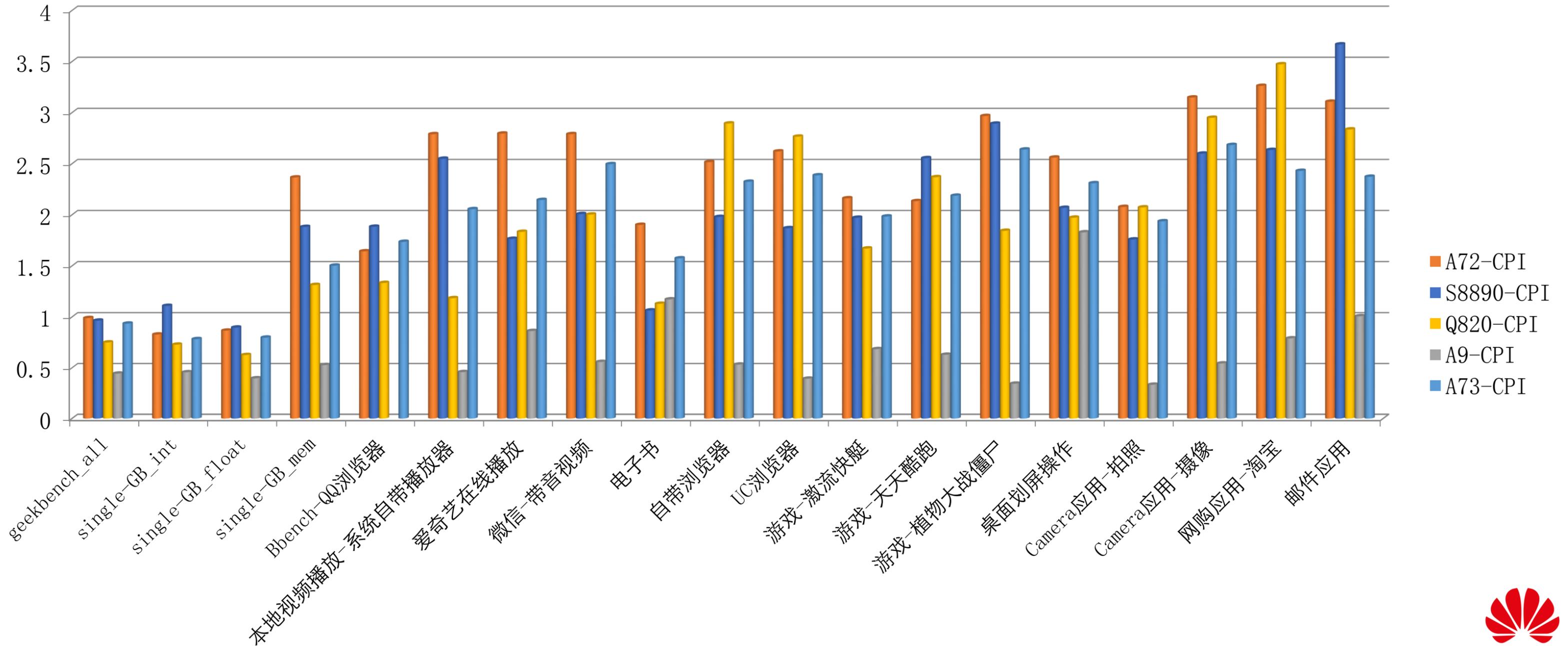
苹果
三星
高通
MTK
华为



传统的BENCHMARK无法反应各类应用的特点

MAKE IT POSSIBLE

各个场景的CPI (Cycle Per Instruction) 对比



注：数值越低表示CPU性能越高

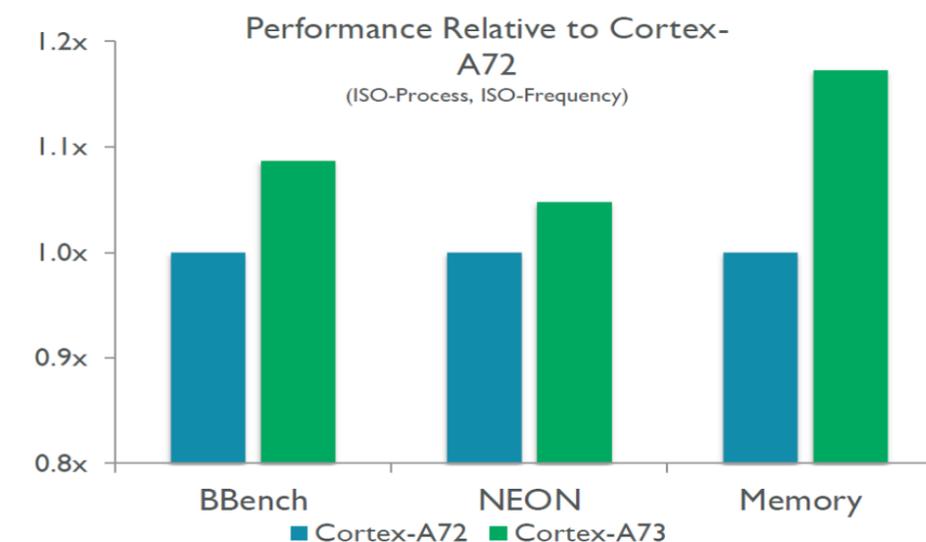


性能、功耗、热是衡量终端CPU关键用户体验的三维度

性能

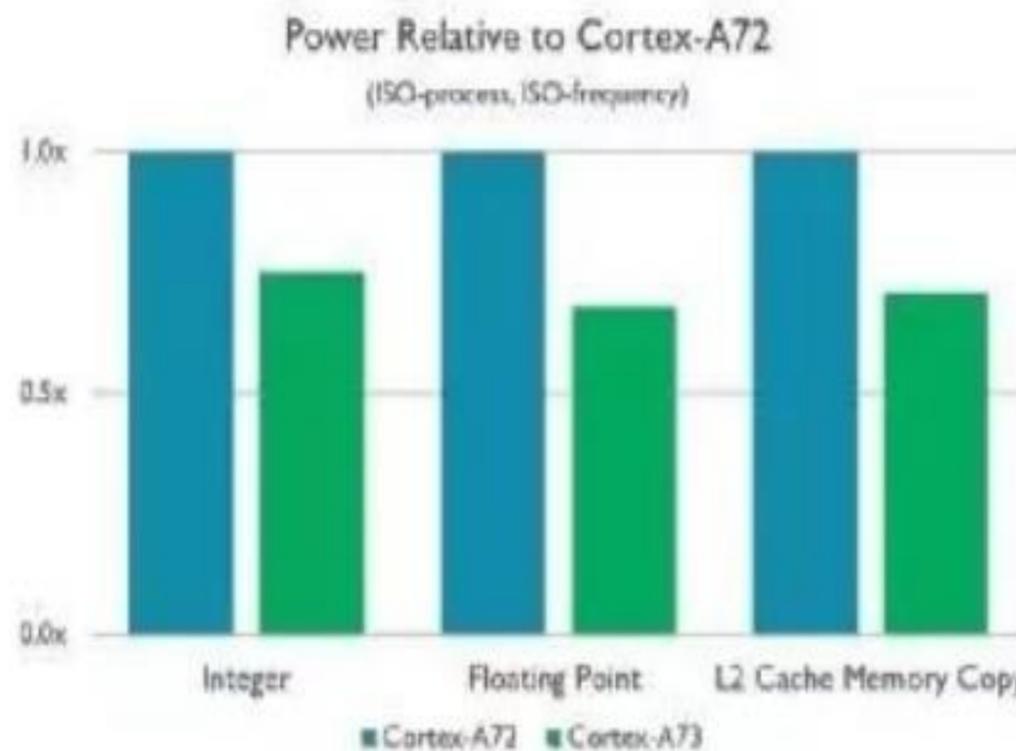
瞬时性能：应用启动性能、跟手性、触控反应速度

平均性能：帧率流畅度瞬时触控速度



功耗

单位性能、单位主频的功耗要越低越好



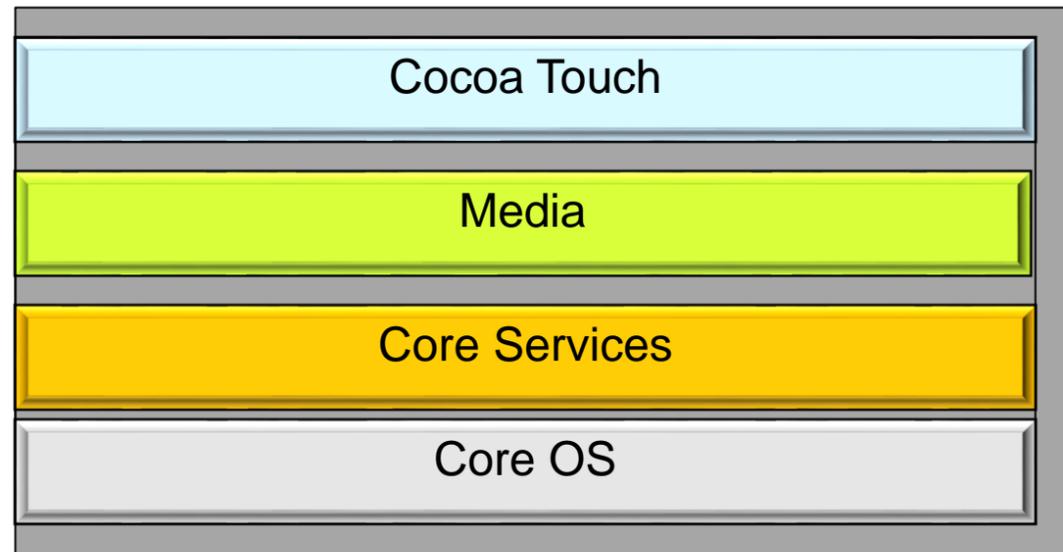
热

持续性能下的功耗不可超标，超标的性能不可用

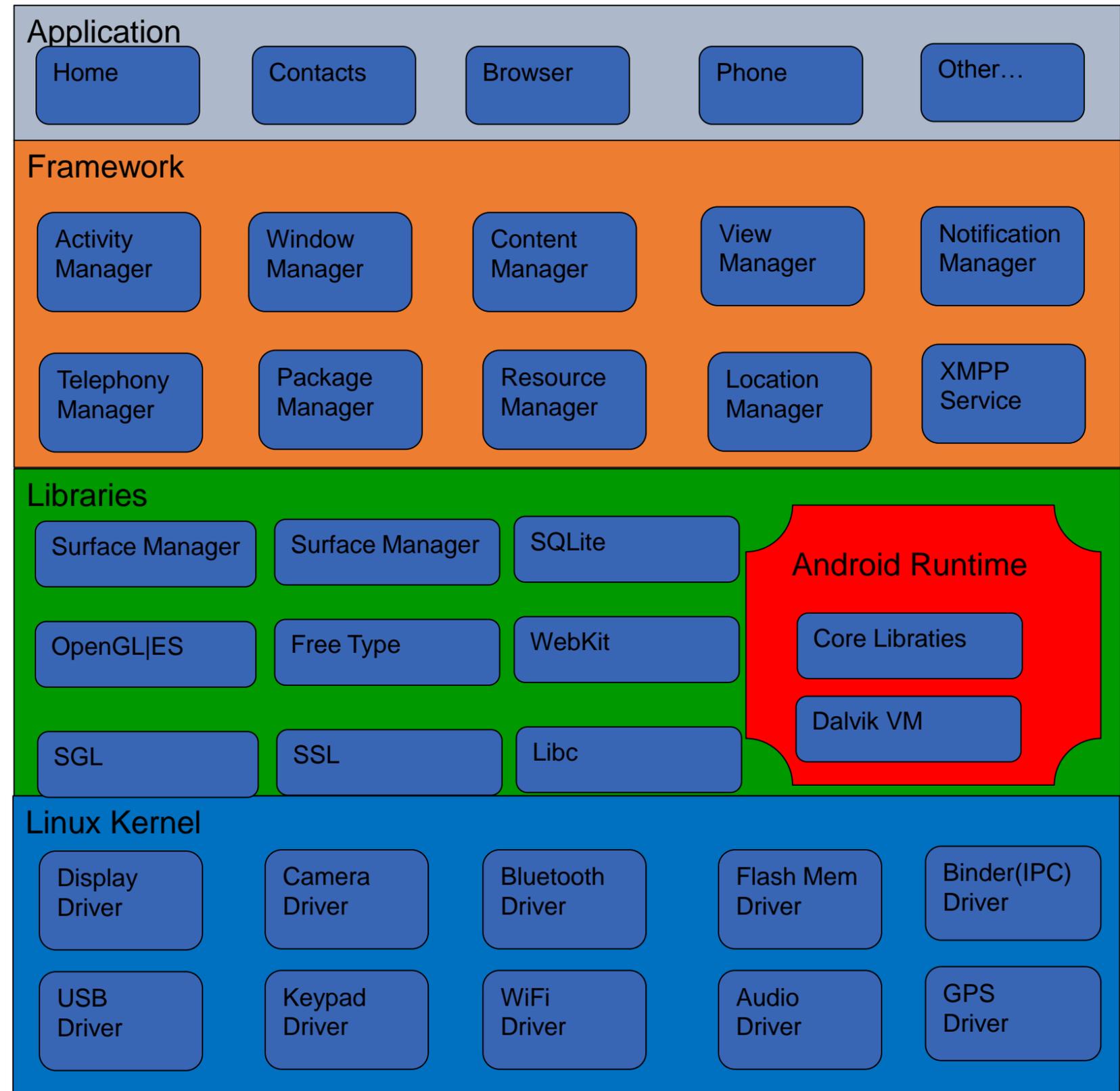
安卓先天性特性-应用平台的开放性带来的挑战

MAKE IT
POSSIBLE

IOS架构



Android架构

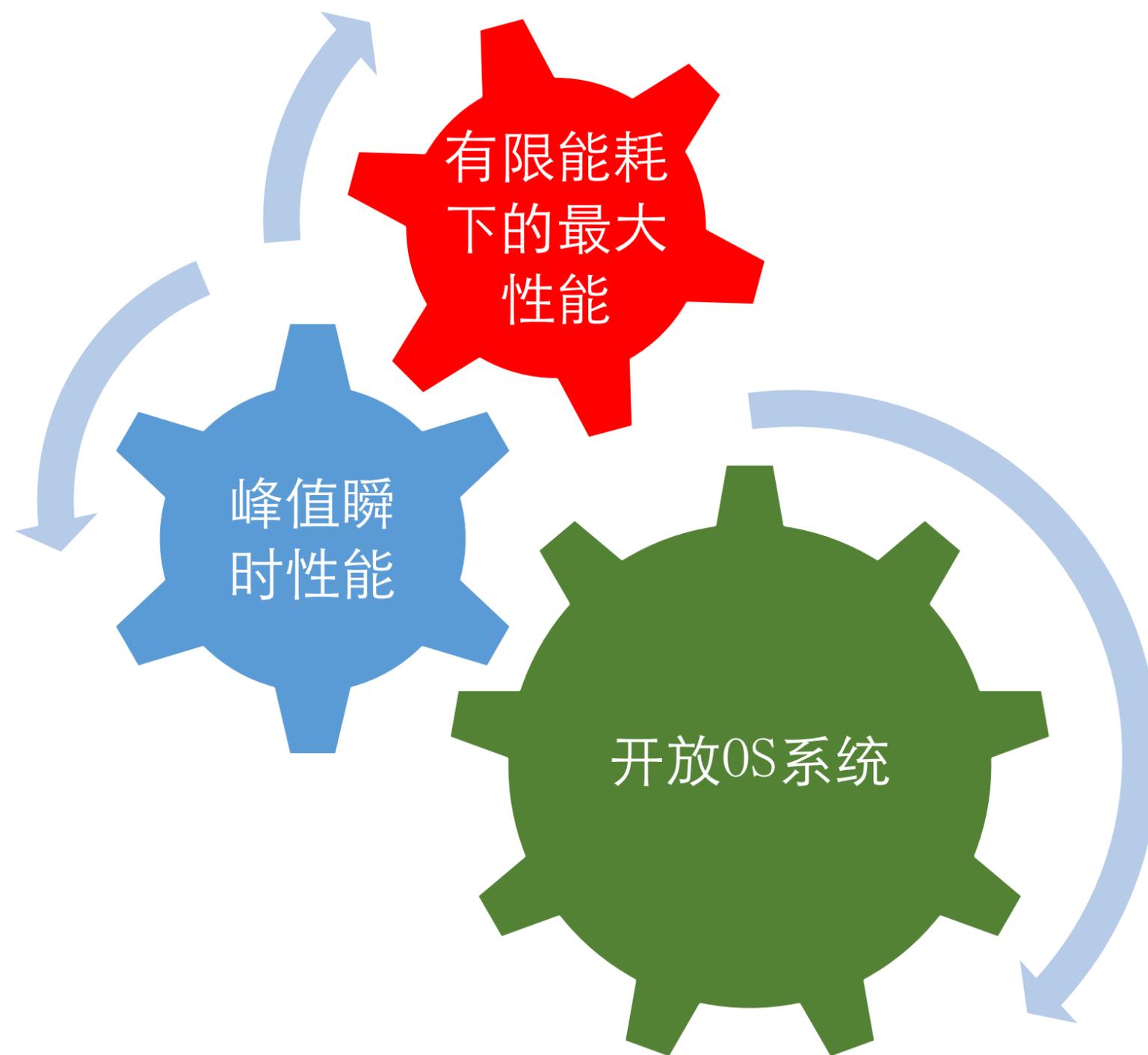


- ◆ IOS: 架构传统但高效, 使用面向对象C, 无需虚拟机
- ◆ Android分层架构, 基于虚拟机技术, 通用性强, 但效率偏弱
- ◆ 开放式的内存管理
- ◆ 自由的后台管理
- ◆ 基于Java应用编程, 跨平台、通用性强, 但应用本身代码效率相比IOS的Objective-C偏弱
- ◆ 触控优先级在原始系统架构中不明显



终端产品挑战CPU体系结构的关键点

MAKE IT
POSSIBLE



性能分析方法-影响终端CPU性能的几个维度

MAKE IT
POSSIBLE

软件编程效率

软件编译器效率

系统OS调度效率

存储子系统效率 (CPI)

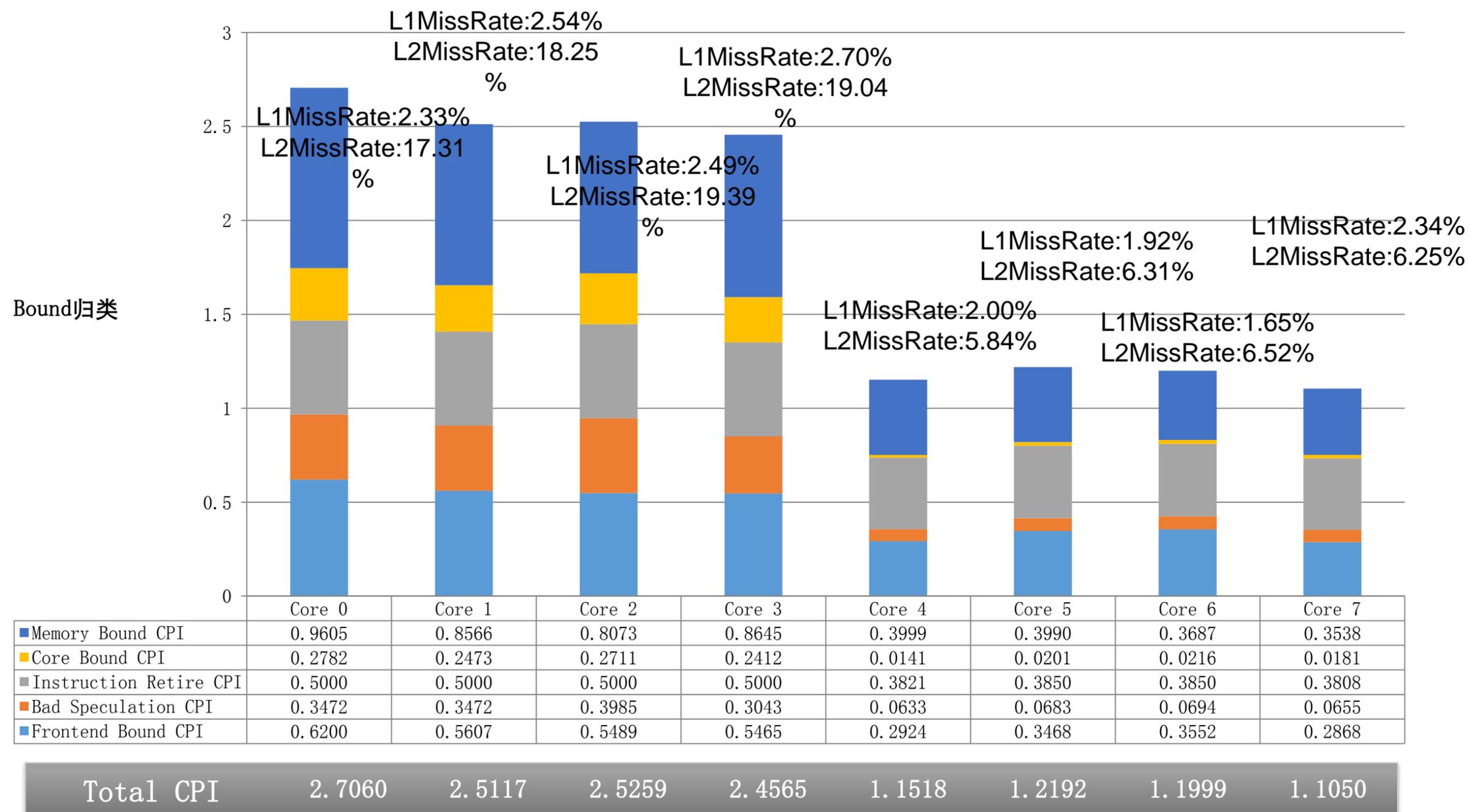
微架构性能 (IPC, mis-predicat
penalty, reorder buffer depth...)

◆软件编程/编译器/OS调度方面的效率，主要与软件本身设计相关，而与芯片本身相关度较弱，所以后面的分析主要集中在微架构性能IPC、以及存储子系统效率，只对前面的三个方面通过测试，估算其性能影响，然后得出硬件方面的性能影响比例

◆对于CPI和微架构的影响主要通过跑分BENCH MARK工具测试以及典型用户场景等多类场景分析，得出相关性能和能效比数据



A72/A53架构的Bbench效率表现



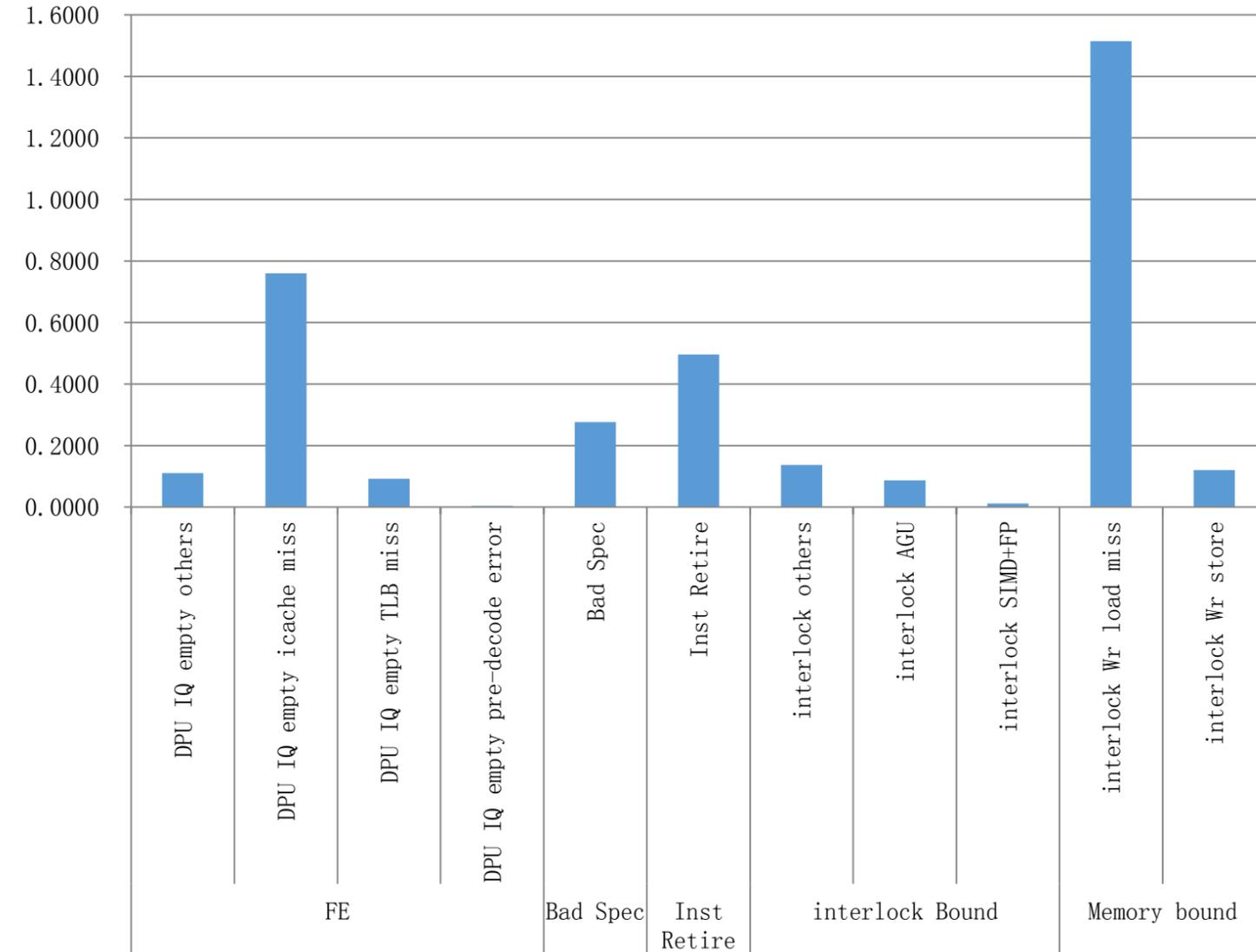
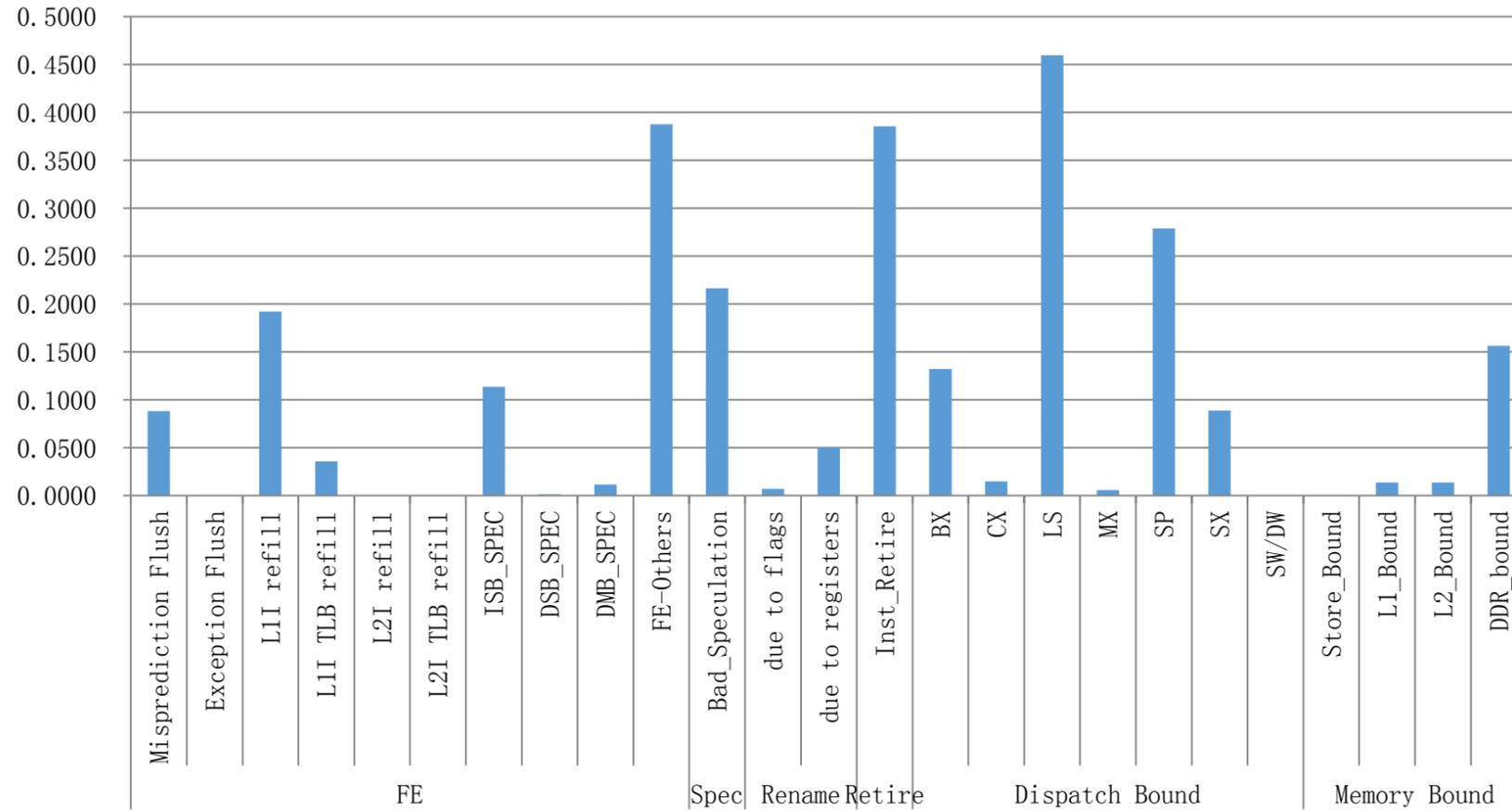
◆ A53在Bbench中负荷较轻，可以不关注

◆ 即使在BBENCH这种BENCHMARK场景，L2 MISS率仍然较高，这也是造成前后端BOUND的主要原因

ARM Big-Little架构在微信场景下的效率表现

微信即时通信-A53 CPI 分解

微信即时通信-A72 CPI 分解



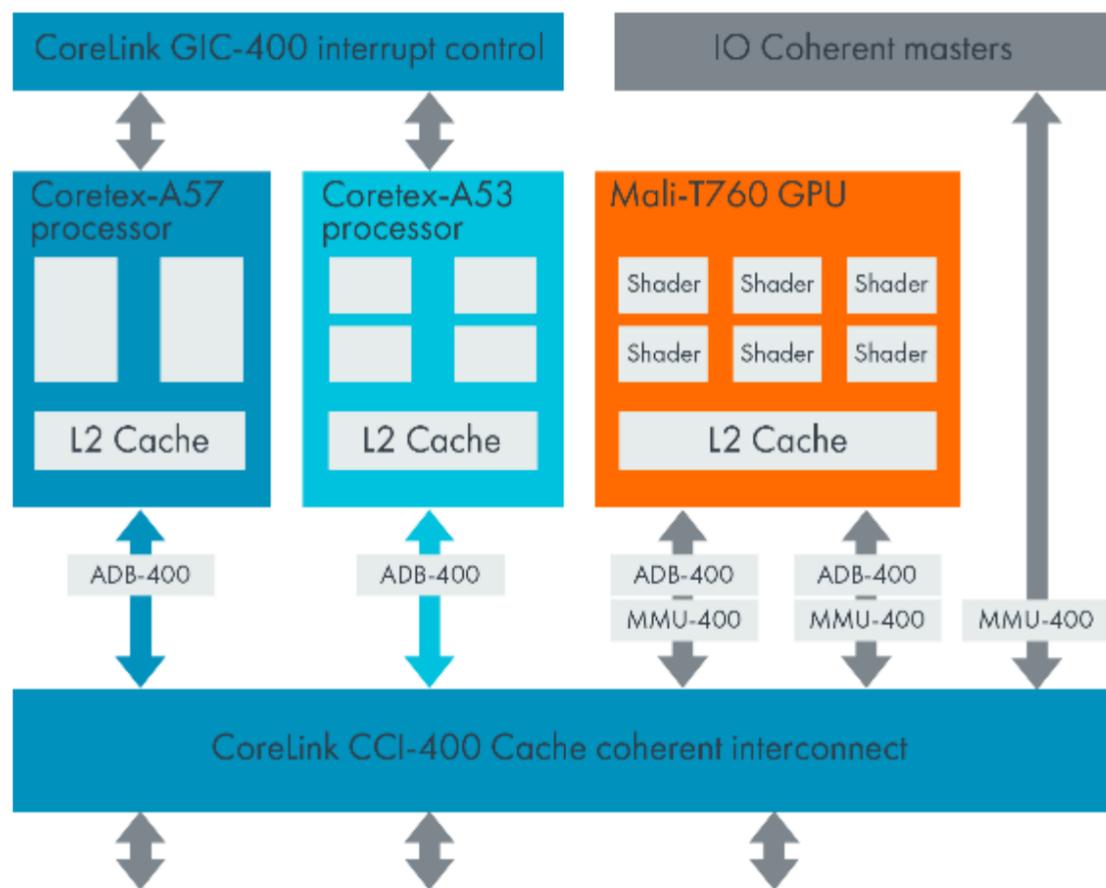
A72主要性能瓶颈: Load/Store 数据单元吞吐率、L2 MISS单元、FE-Others、Instruction Retire、SP执行单元和Bad SPEC

A53主要性能瓶颈: Memory Bound、Icache miss、Instruction Retire和Bad SPEC



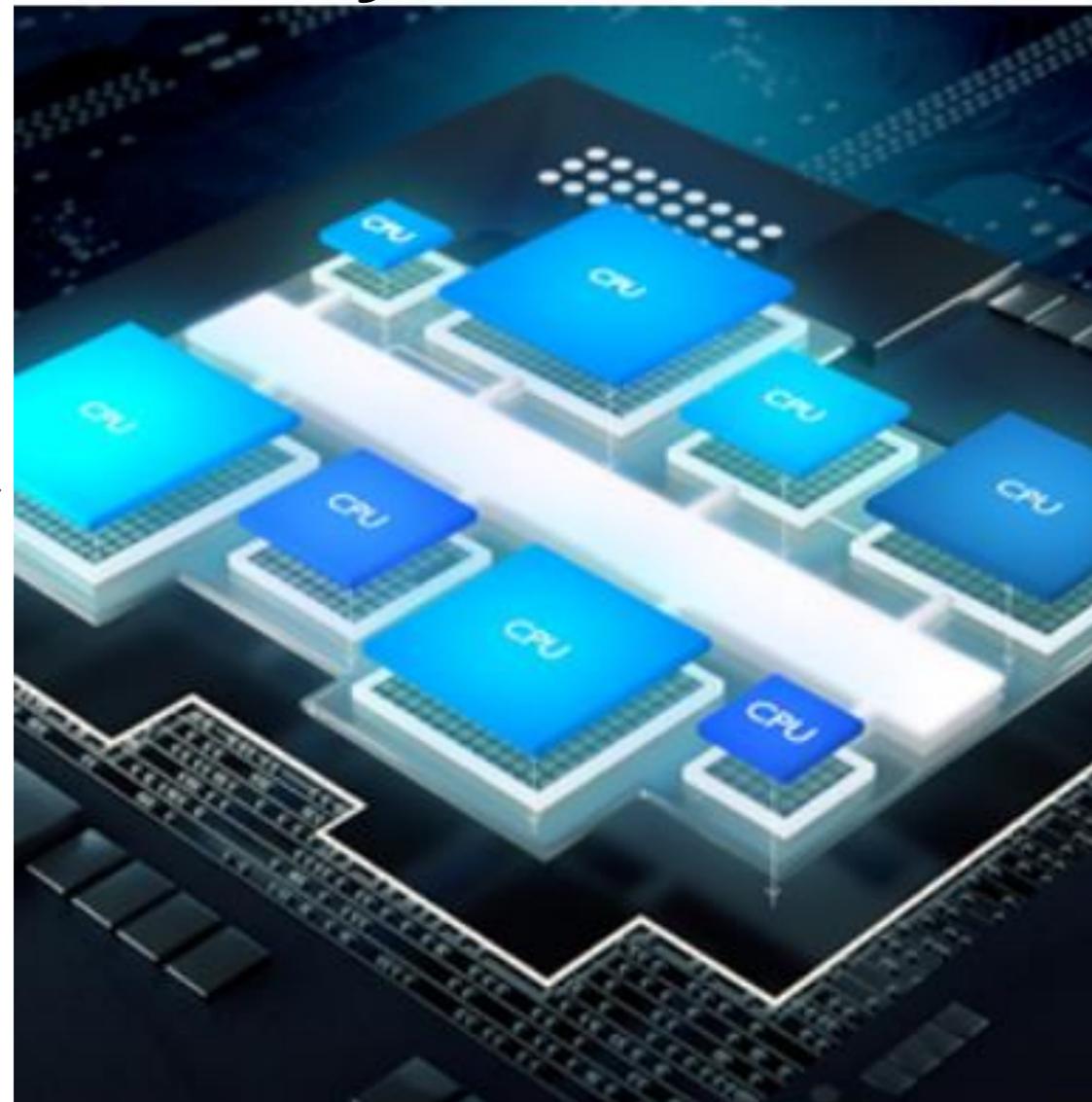
从Big-Little到Dynamic架构支持多核异构

Big-Little



双Cluster/高性能低功耗两类核异构
不支持单核独立DVFS

DynamIQ



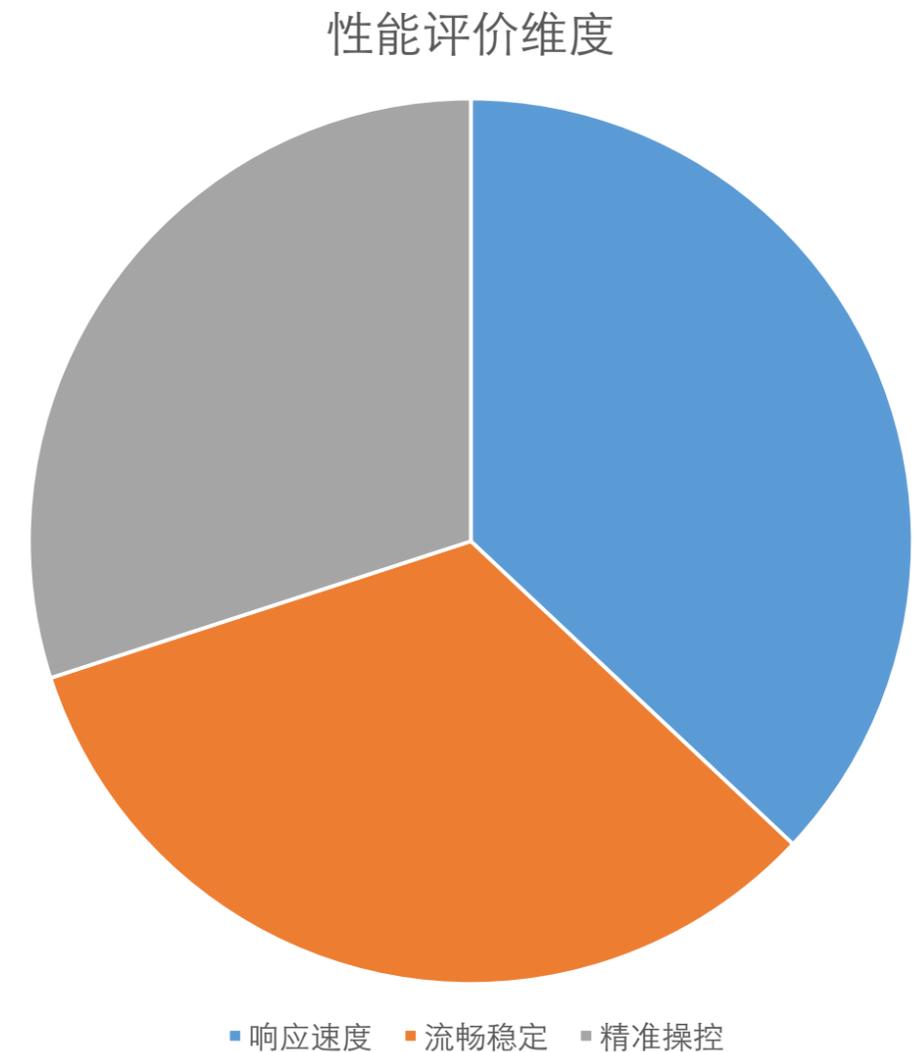
单Cluster/最大八核异构
支持单核独立DVFS

终端TOP应用所要求的处理器性能特点

MAKE IT
POSSIBLE

瞬时性能无极限-触控感知

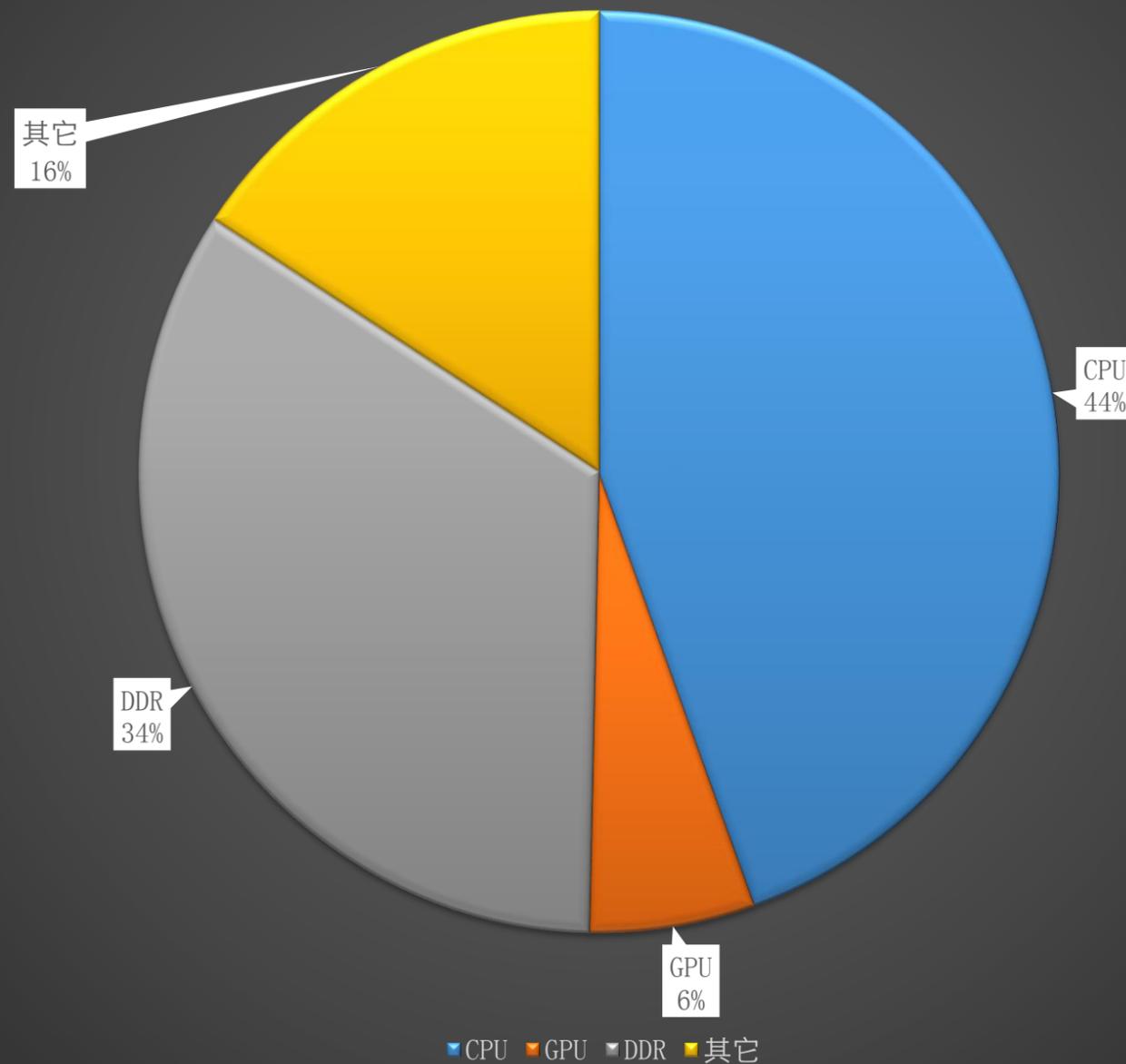
- 响应速度是手机用户对性能的最关键体验
- 100ms左右的运算和处理要最快，如网页浏览等
- 响应处理往往单线程，单线程系统性能必须瞬时最大化
- 流畅稳定需要一定功耗约束下的性能最大化



平均能效无极限-连续娱乐

MAKE IT
POSSIBLE

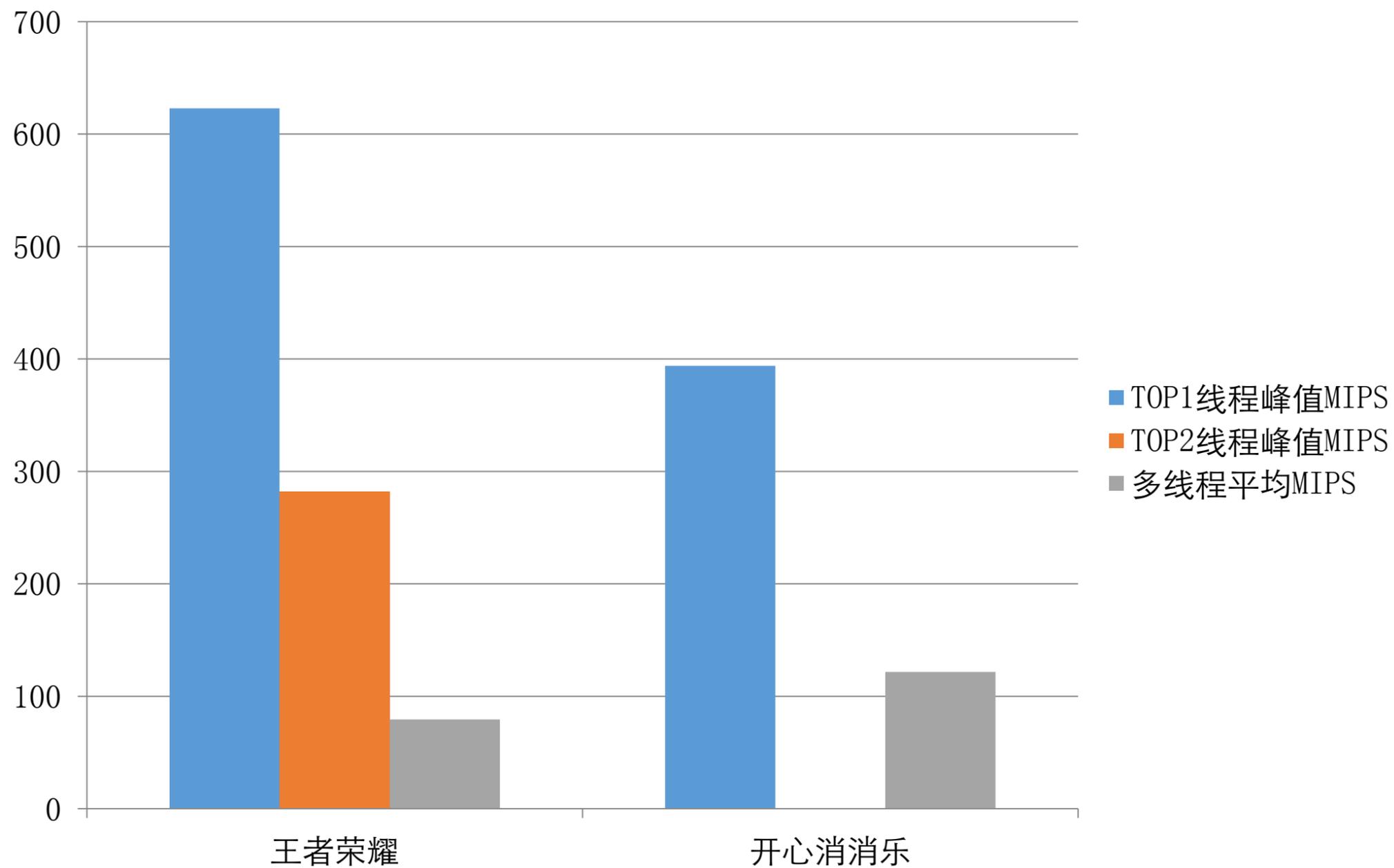
手机亮屏续航关键部件耗电分布



- CPU在整体芯片耗电中占了近一半
- 确保流畅性前提下,降低单位帧能量消耗是关键

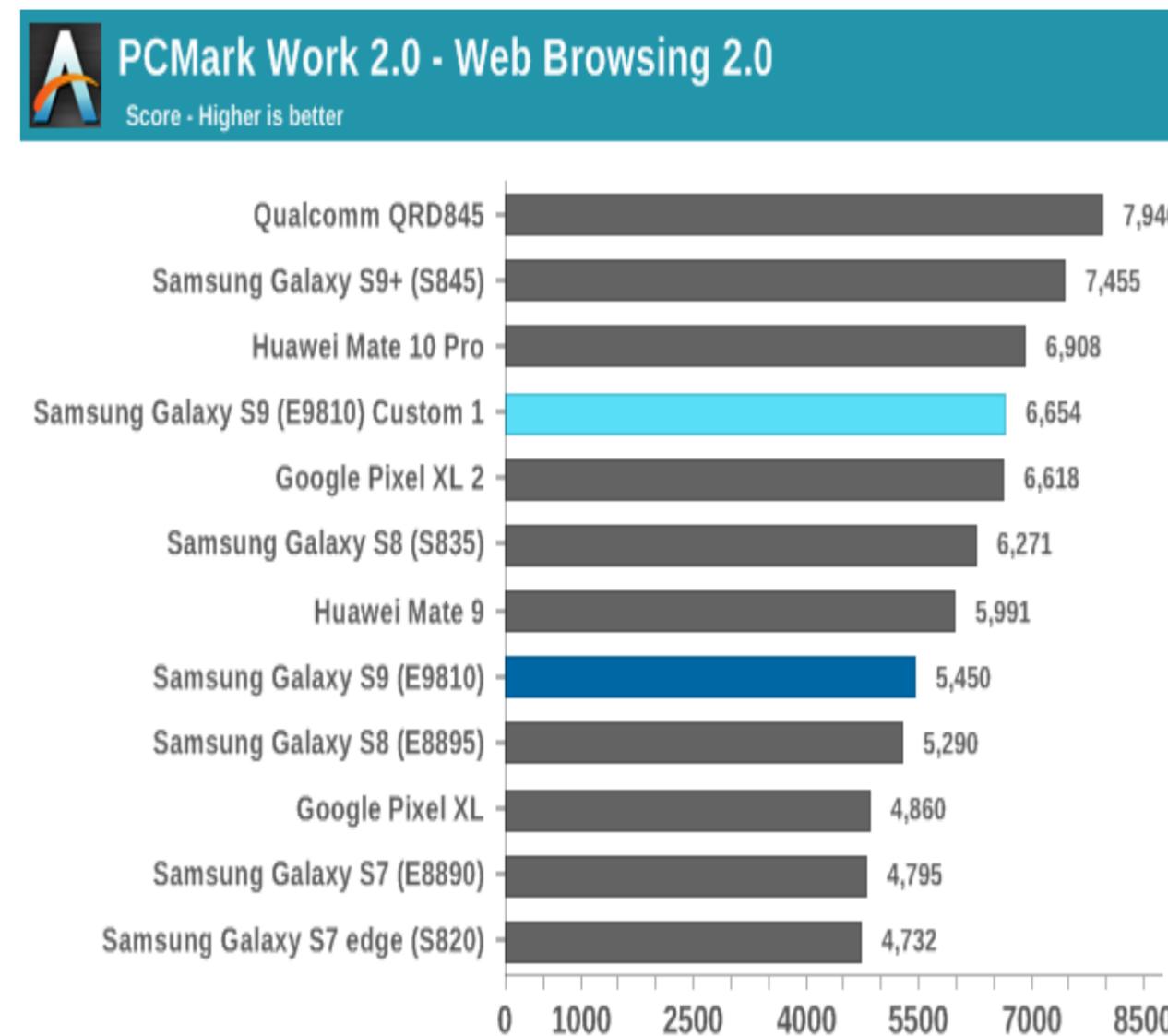
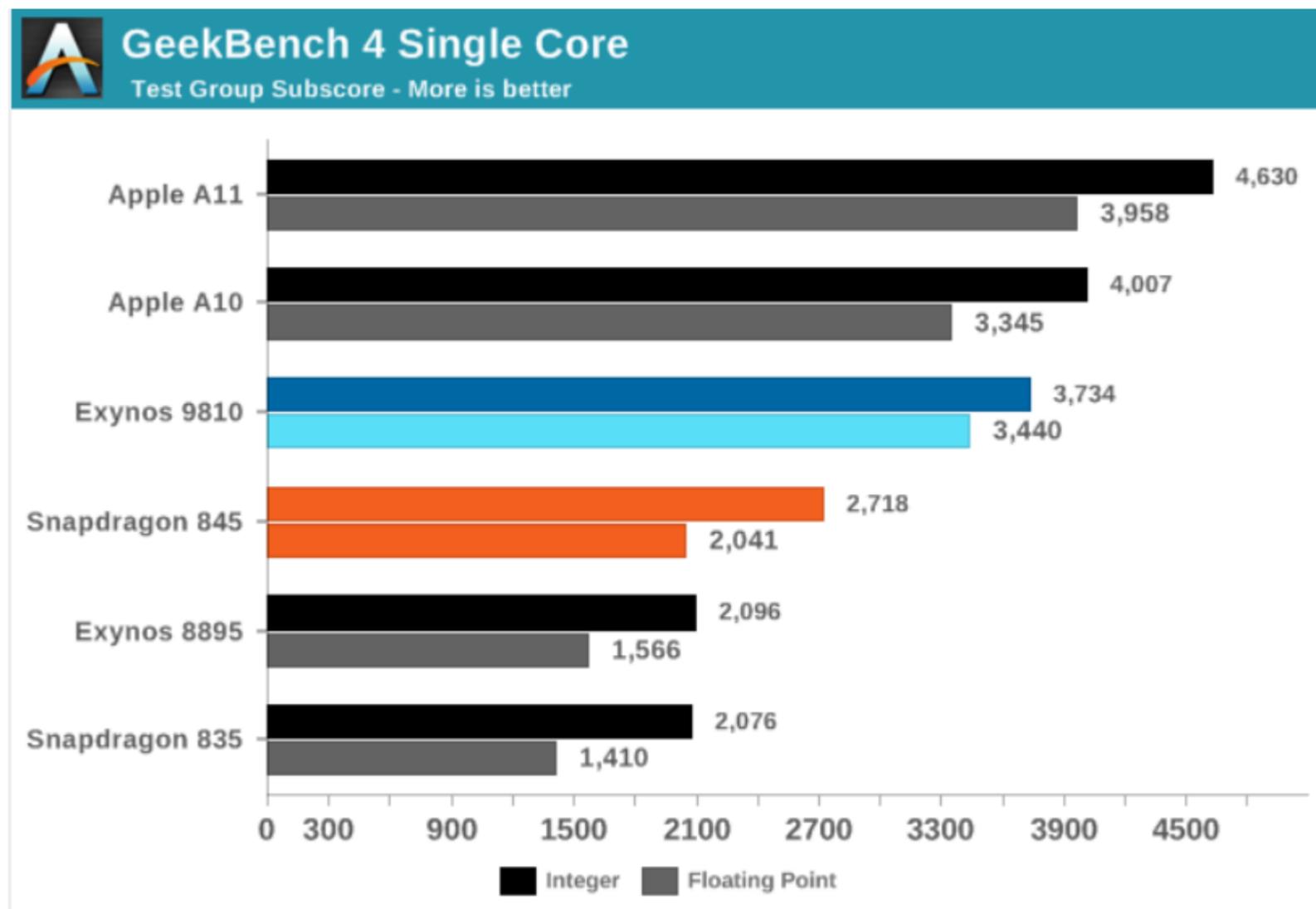


纯粹的多线程应用不好写-一个大个子撑高一间房



终端应用更看重CPU单位能量的性能而非CPU绝对性能

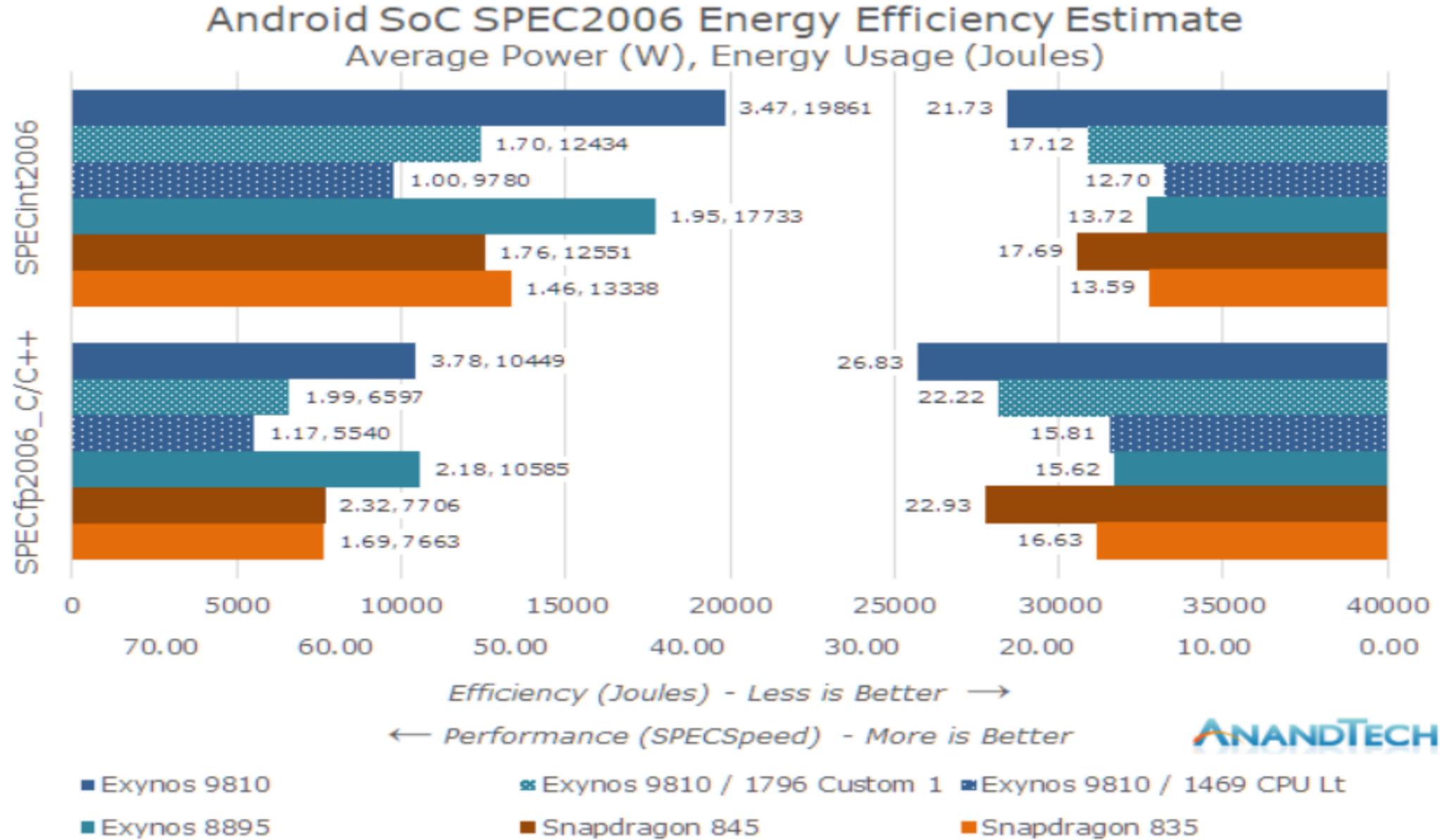
单纯的微架构提升不一定能够直接获得性能提升
六发射可以提升跑分性能，但对用户场景性倒不一定最优



9810安卓CPU跑分最高，但实际应用性能一般

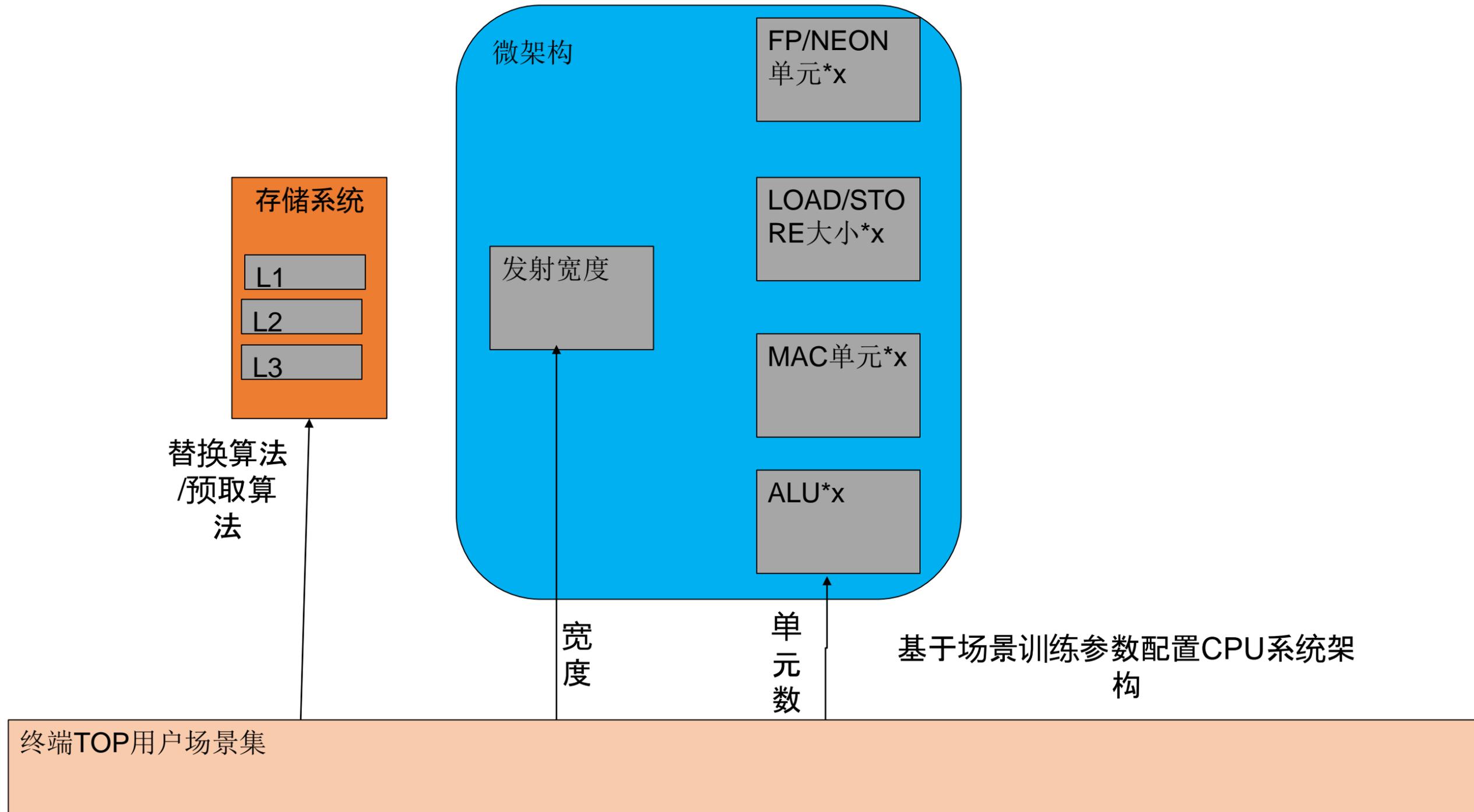


终端应用更看重CPU单位能量的性能而非CPU绝对性能



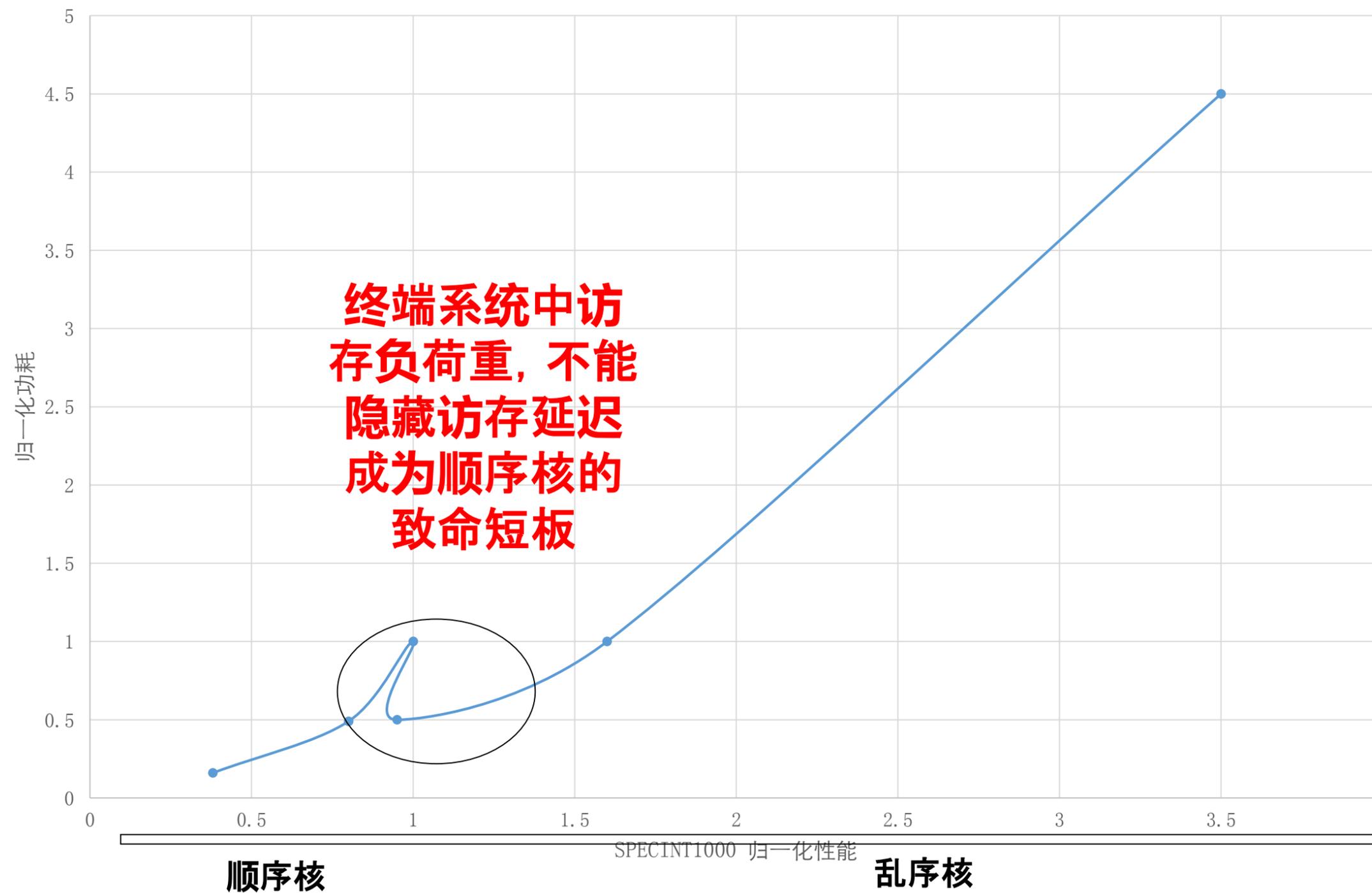
采用6发射的9810性能最高，但能效最差；而限制频率后能效接近Q845，但性能又丧失竞争力

可配置的CPU系统架构适配不同应用的最佳能效



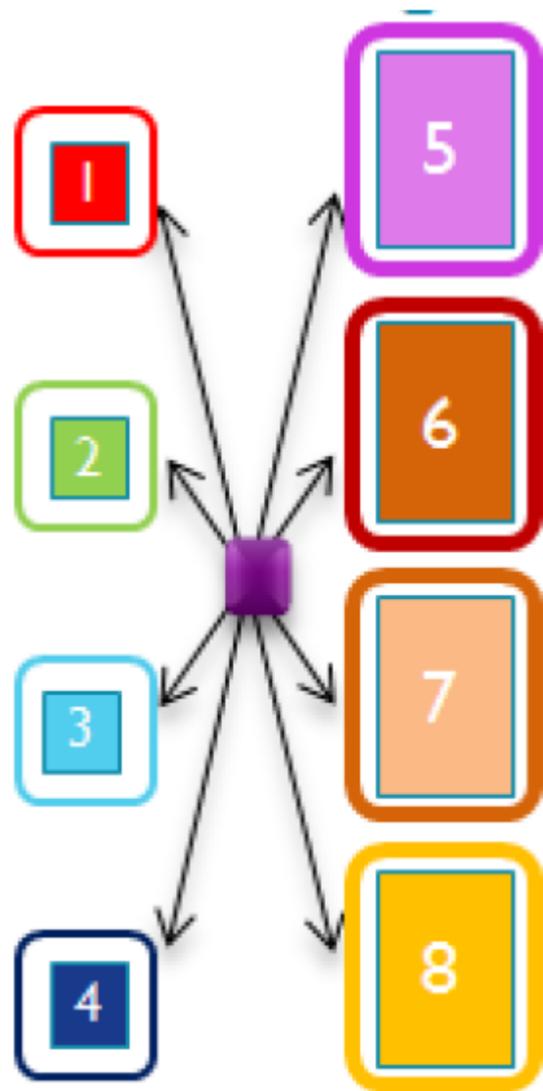
低功耗乱序核是否有可能？

归一化功耗（按照乱序核最高功耗）

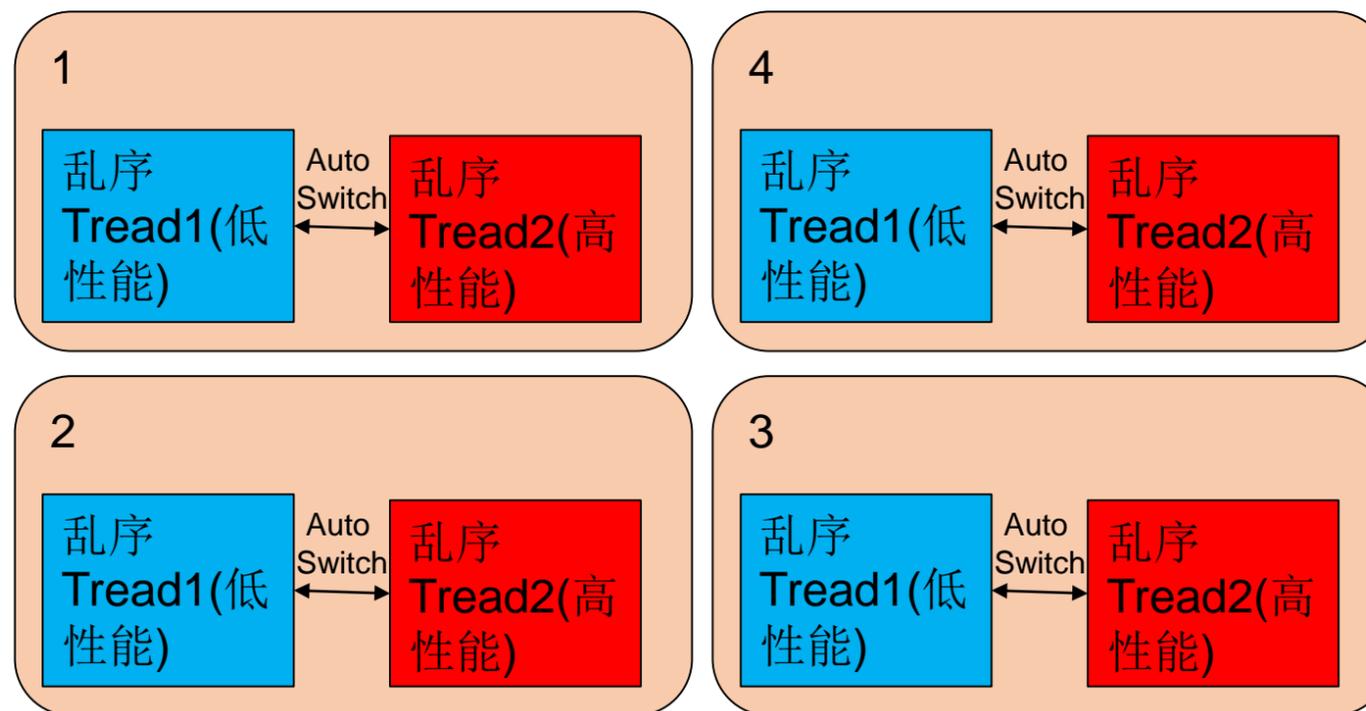


Auto-Switch是否有可能？-更简单的调度带来更好的能效

大小核架构中复杂的调度，一不小心就会因调度损失能效

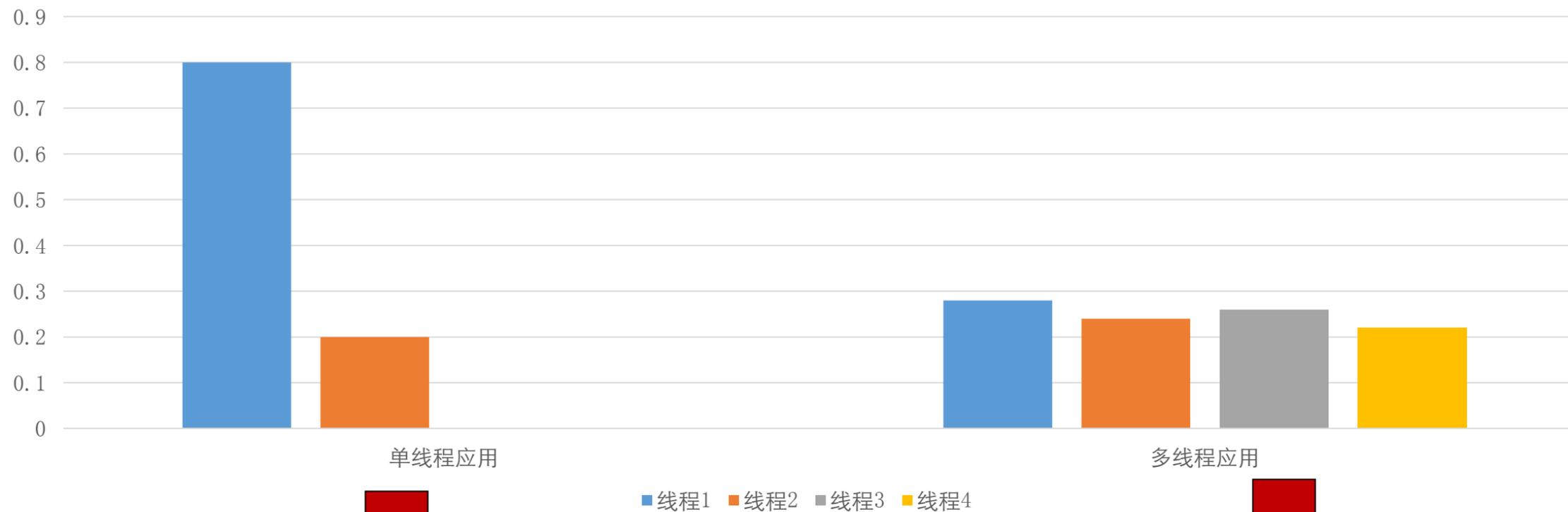


简单的调度

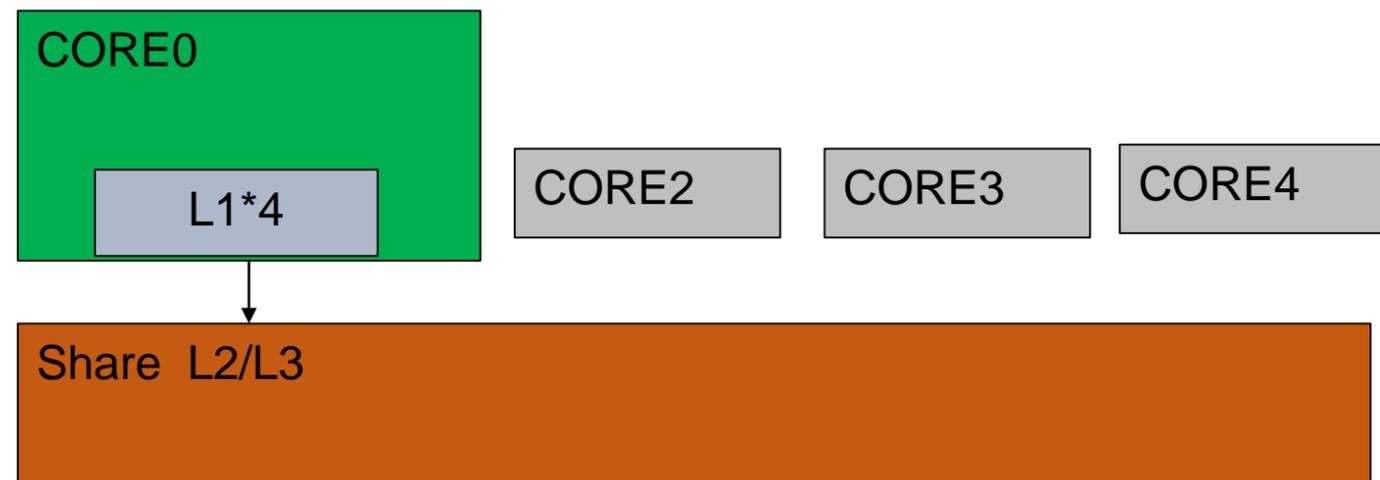


单核/多核资源动态配置

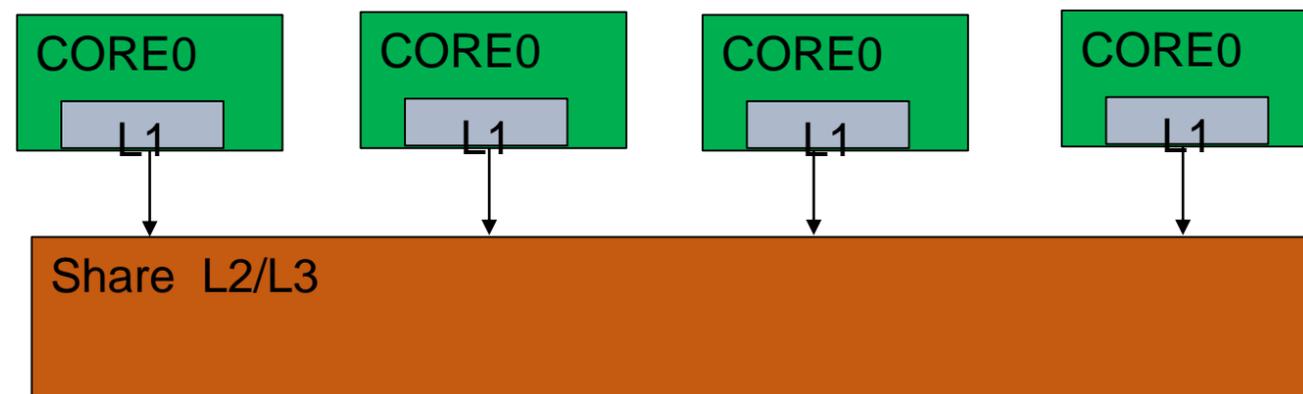
典型应用任务负荷示意图



单线程模式

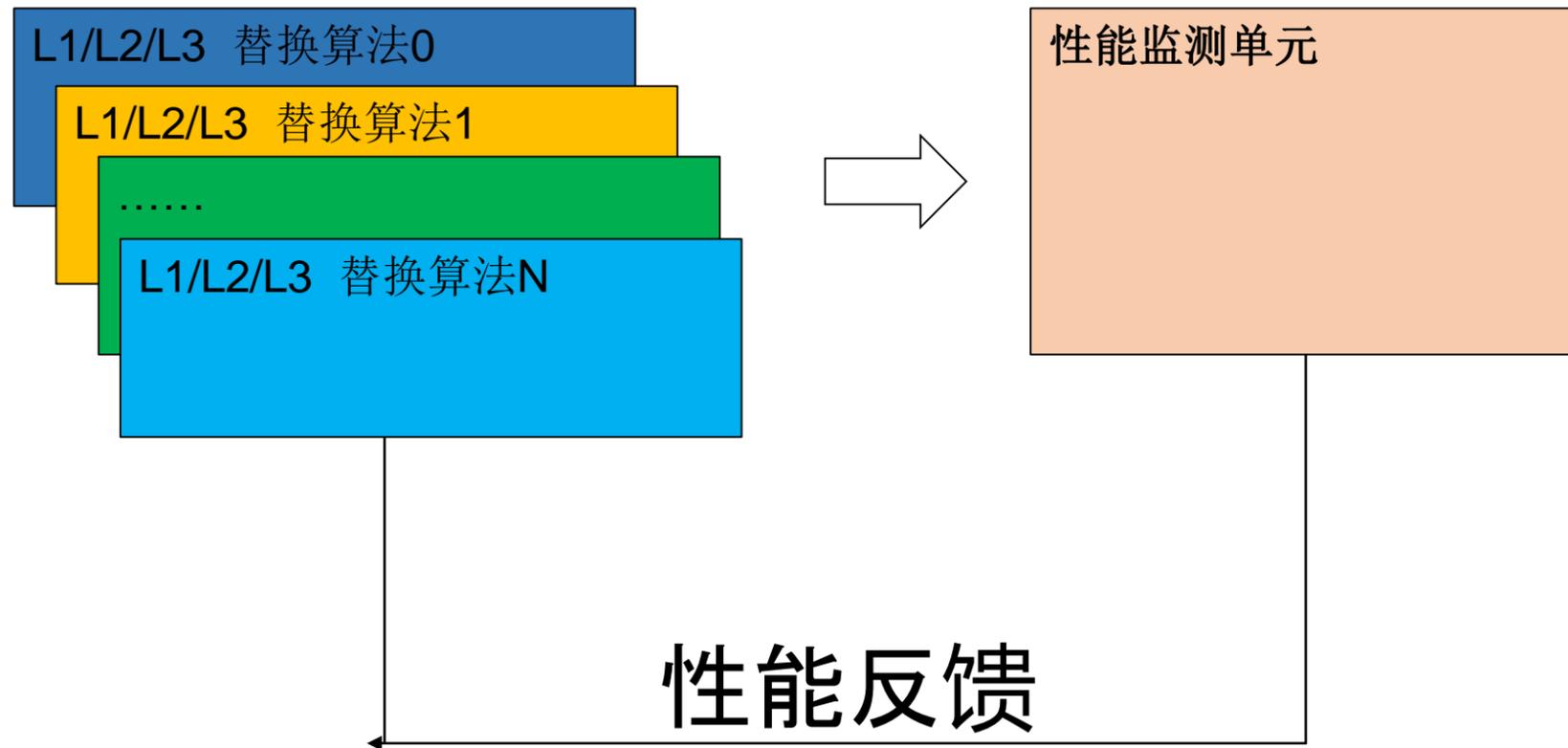


多线程模式



最强的自适应CACHE替换算法

MAKE IT
POSSIBLE



终端应用更宽泛，原始的服务器CPU存储系统策略不完全适合终端应用



其它建议

MAKE IT
POSSIBLE

- ✓ 从自研CPU核到自研CPU系统
- ✓ 提高处理器性能利用率
- ✓ 高性能和低功耗不应该成为矛盾
- ✓ BOOST性能技术动态可配置



华为CBG SOC 能效架构团队介绍

MAKE IT
POSSIBLE

职责: 为CBG产品的SOC能效竞争力负责, 持续提升用户性能和功耗的用户体验

合作愿景: 愿同各高校、科研院所展开强力合作, 包括CPU、GPU、DDR等各类性能和功耗优化技术方向, 一起为国产处理器能效达到国际TOP1水平而努力!



Thank you

