



南開大學  
Nankai University

计算机学院  
并行体系结构调研报告

Apple Silicon M 系列芯片调研

姓名：卢海泉  
学号：2112649  
专业：计算机科学与技术

2024 年 3 月 17 日

# 目录

<b>1 基本要求：M3 系列芯片调研</b>	<b>2</b>
1.1 苹果 M3 系列芯片介绍及优势	2
1.2 关键指标	2
1.3 架构优势	3
1.3.1 统一内存架构 (UMA)	3
1.3.2 超大显存	4
1.3.3 超大带宽	4
1.3.4 超高效 ZRAM、SWAP	5
1.3.5 媒体引擎技术	5
<b>2 进阶要求：苹果 M 系芯片横向、纵向对比</b>	<b>5</b>
2.1 CPU 性能	5
2.2 GPU 性能纵向对比	6
2.3 GPU 性能横向对比	7
2.3.1 Geekbench 6	7
2.3.2 3DMARK	8
2.3.3 Blender 4.0.0	8
2.4 总结	9

# 1 基本要求：M3 系列芯片调研

## 1.1 苹果 M3 系列芯片介绍及优势

2023 年 10 月 31 日，苹果发布了全新 3 纳米工艺制成的 M3 芯片。Apple Silicon M 系列芯片从问世以来，就是近乎天花板级的台积电工艺制程、夸张的晶体管数量以及恐怖的单核心性能。更重要的是，Apple Silicon M 系列芯片的最大意义在于，它让主板的复杂设计，变得异常简洁。这是以前 MacBook 的主板设计，CPU、GPU、运行内存、I/O 芯片、安全芯片、雷雳控制器等，分散在主板各个位置。各种互相通信，各种带宽限制，各种高功耗发热。甚至是 CPU 和 GPU 都有着各自的运行内存，数据需要在两个独立内存之间频繁拷贝。而从 M1 开始，这些东西都做到了 All-in-one。使用集成大量功能的 M1 芯片以后的主板，不仅设计简单，故障率更低，而且做到了功耗极低，性能更强。当然，这些可不是单纯的技术分析。无论是搭载 Apple Silicon M 系列芯片的 MacBook 获得近乎翻倍的续航提升，与此同时还能带来极致的性能提升，还是离电性能等同于插电的表现和 60W 打人家 160W 的性能释放，都足以体现 Apple Silicon M 系列芯片在移动端的地位和价值了。从官方最新发布的数据来看，M3 相较于 M2 来说，CPU 提升 20%，GPU 提升 20%，NPU 提升 15%。而 M2 相较于 M1 来说，CPU 提升 18%，GPU 提升 35%，NPU 提升 40%。

除此之外，M3 还引入了动态缓存。不同于传统图形处理器，它具备动态缓存功能，因而可对硬件中本地内存的使用进行实时分配。在动态缓存功能的加持下，每项任务对内存的消耗精准符合所需。此项业界首创技术对开发者透明，为打造全新图形处理器架构提供了基石。它大幅提高了图形处理器的平均利用率，进而给要求更苛刻的专业级 app 及游戏的表现带来显著提升。同时带来首次登陆 Mac 的硬件加速光线追踪和网格着色等全新渲染功能，光线追踪技术能够模拟光线在场景中的表现，从而帮助 app 创造出栩栩如生、逼真的画面。网格着色功能实现图形处理能力和能效的双重提升，更可支持游戏和对图形处理要求高的 app 呈现视觉效果更复杂的场景。这一开创性的图形处理器架构为以上所有优化提升提供加持，同时延续了 Apple 芯片的高能效特色。

从上面的数据可以看出，苹果 M 系列芯片持续迭代的表现非常好。

## 1.2 关键指标

	M3	M3 Pro	M3 Max
晶体管数量	250 亿	370 亿	920 亿
制造工艺	台积电 N3B	台积电 N3B	台积电 N3B
CPU 性能核心	4 核，4.1GHz 主频	5/6 核，4.1GHz 主频	10/12 核，4.1GHz 主频
CPU 能效核心	4 核	6 核	4 核
CPU 单核心/多核心性能	3055/11690	3096/15231	3114/21061
GPU	8/10 核心，硬件追光	14/18 核心，硬件追光	30/40 核心，硬件追光
GPU OpenCL/Metal 跑分	2569/41002	50332/78547	92201/154234
NPU	16 核心,18TOPS	16 核心,18TOPS	16 核心,18TOPS
内存控制器	LPDDR5-6400,100GB/s	LPDDR5-6400,150GB/s	LPDDR5-6400,300GB/s
内存容量	8GB/16GB/24GB	18GB/36GB	48GB/64/128GB

表 1: M3 系列芯片的关键性能指标

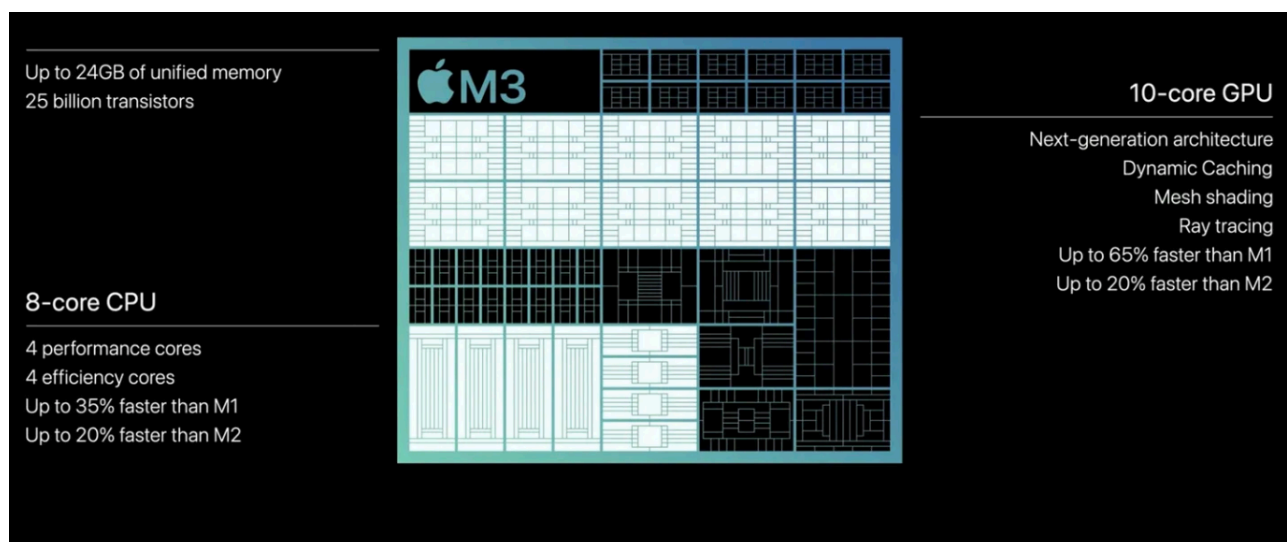


图 1.1: M3 芯片架构图

### 1.3 架构优势

#### 1.3.1 统一内存架构 (UMA)

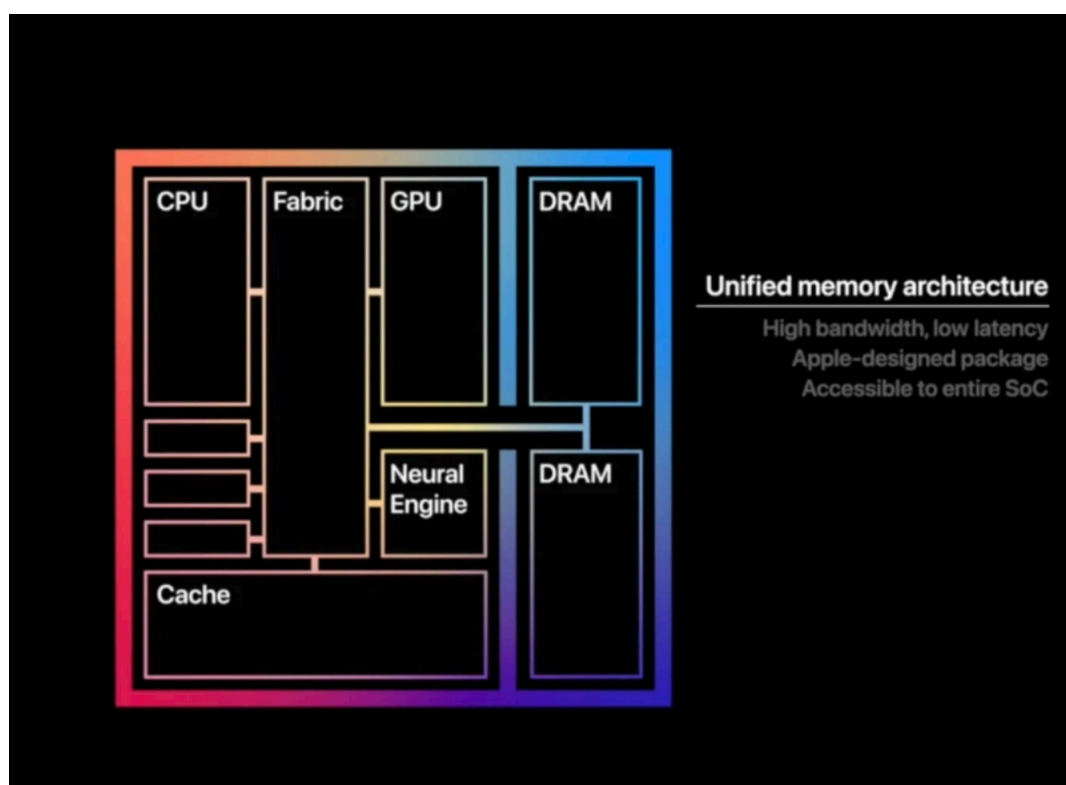


图 1.2: 统一内存架构 (UMA)

M1 问世以后，很多人看到的芯片就不再是单纯的一块硅片，而是有点诡异的全新封装。同时也把 UMA 这个词带到了用户面前，也就是统一内存。对于如今 CPU 超快的处理速度来说，哪怕是内存条距离 CPU、GPU 的距离，都足以增加访问内存时的开销。Apple Silicon M 系列芯片的这个封装，

无疑是把这个开销都降低到极致。简单来说，苹果这个 UMA 架构指的就是内存和 SoC 上的各个专用处理器通过 Fabric 实现互联，达到各个专用处理器都可以快速访问内存，甚至是访问同一块内存而设计的全新架构。CPU、GPU、NPU、视频引擎等处理器的内存不再是物理层面独立的内存块，在这个架构下不同专用处理器可以访问同一块内存区域。CPU 和 GPU 有着一大块共享内存，这部分数据 CPU 和 GPU 随时都可以按需直接使用。当然，GPU 也依然会有部分自己独占的内存。在这套架构下，CPU0 处理完的数据交给 GPU 处理，GPU 处理完交给 CPU1 处理，整个过程不需要任何内存数据拷贝。而处理大量数据的情况下，这个过程就可以非常顺滑，CPU0 可以一直在做逻辑处理，GPU 一直可以及时拿到数据，处理完的数据总是可以无缝衔接给下一个处理器。而以上这些，可不仅仅是集成显卡就能做到的。实际上，集成显卡往往也是有着自己独立的显存的，内存管理模式和独显没有什么区别。而只有配合 UMA 统一内存架构，才能实现以上这种高效的处理哇！

### 1.3.2 超大显存

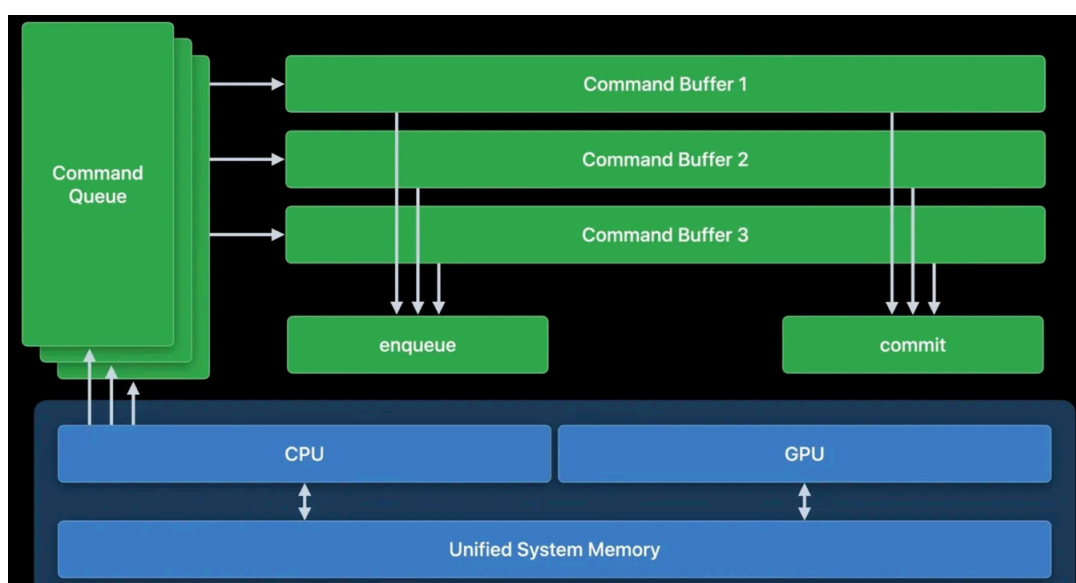


图 1.3: 超大显存

UMA 架构下，GPU 就是直接拿系统内存当显存用的。对于 Apple Silicon M 系列芯片而言，DRAM 可以做到拿出 2/3 以上当成显存使用。比如 32GB 版本的 MacBook，可以拿出 21GB 当成显存，64GB 则可以拿出 48GB 当显存。如今 M3 Max 支持 128GB 的大内存，显存预估能够达到 85GB 以上。而 M2 Ultra 已经支持到 192GB 内存，未来 M3 Ultra 那就是 256GB。简单对比一下显存容量，个人显卡天花板的 RTX 4090，也仅仅给到 24GB 显卡内存。哪怕是 RTX 6000 Ada，显存也就给到 48GB，这等于是 M1 Max 就能给到的容量。这些不是要说 Apple Silicon M 系列芯片在显卡领域超越英伟达了，而是想说，UMA 架构确实给个人用户带来了前所未有的大显存，尤其是碰上这个 LLM 流行的时代，超大显存的价值更是不言而喻了。

### 1.3.3 超大带宽

M1 在使用 LPDDR4X-4266、128bit 位宽的内存的情况下，提供了  $4266 \times 128 / 8 / 1000 \text{ GB/s} = 68.256 \text{ GB/s}$  的内存带宽。而 M2/M3 则换成了 LPDDR5-6400、128bit 位宽内存，内存带宽则达到了 102.4GB/s。其实高通最新发布的，用来对标 M2 Max 的 X Elite，借助高频的 LPDDR5x-8533，也才把内存带宽做到 136GB/s，而 M2 Max 的内存带宽已经是 400GB/s。

### 1.3.4 超高效 ZRAM、SWAP

苹果的 Mac 至今依然是 8GB 起步，这也是很多人吐槽的点。之所以会有这种想法，很多时候后正是因为依然没有真正理解 UMA 内存的优势，以及 macOS 下内存的管理和使用模式。考虑到 ZRAM 和 SWAP 可以额外提供不止 10GB 的“虚拟内存”，其实整体效果是可以高于 16GB 的。合理利用 SSD 来提升系统加载程序和数据的能力，这就是最基础、最自然的设计理念。

### 1.3.5 媒体引擎技术

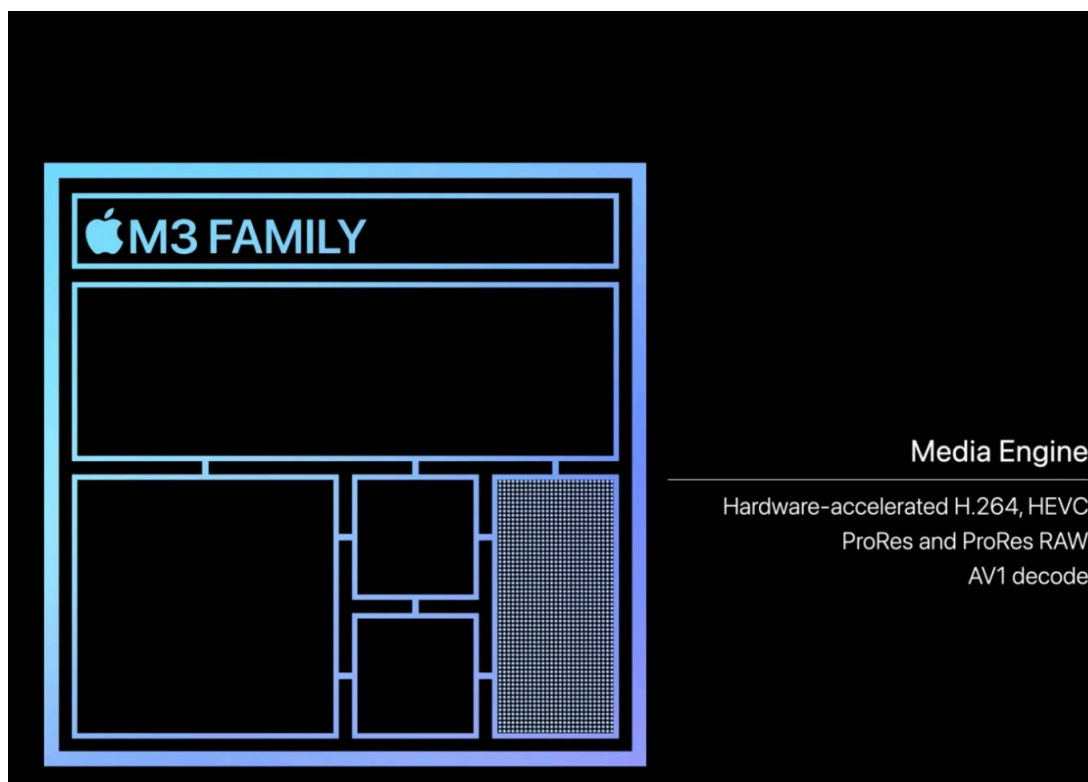


图 1.4: 媒体引擎技术

M1 和 M2 对于 H.264 和 HEVC 加解码都有着硬件加速，但是 M3 和 M2 的视频引擎对于 ProRes 和 ProRes RAW 同样有硬件加速。M3 更是加入了全新的 AV1 解码能力，可以让设备拥有更好的流媒体能力。M3 和 M2 集成了更高带宽的视频解码器，可以支持 8K H.264 和 HEVC 视频。而 ProRes 视频引擎更是可以支持同时播放多条 4K 或 8K 视频流。

## 2 进阶要求：苹果 M 系芯片横向、纵向对比

### 2.1 CPU 性能

M3、M2 和 M1 同样都是 4 个性能核心和 4 个能效核心的架构，不过 M1 采用的是和 A14 相同的 Firestorm 性能核心和 Icestorm 能效核心，而 M2 采用的则是和 A15 一致的 Avalanche 性能核心和 Blizzard 能效核心。M3 采用的是和 A17 Pro 一致的核心架构，不过目前苹果没有透露任何关于核心架构的细节，所以目前还不知道 A17 Pro 和 M3 架构的具体细节。一定程度上讲，A 系芯片的设计很大程度上可以体现 M 系芯片未来的走势。M1 性能核心有着 192KB 指令缓存和 128KB 数据缓存，四

	Single-Core Score	Multi-Core Score
M1	2352	8394
M2	2633	9748
M3	3056	11691

表 2: Geekbench 6 基准测试跑分

个性能核心共享 12MB 的 L2 缓存。M1 的能效核心则有着 128KB 指令缓存和 64KB 数据缓存，四个能效核心共享 4MB 的 L2 缓存。M1 还有着 16MB 的 SLC，也就是片上缓存。以上 M1 的大部分数据，M2、M3 基本保持一致。明显不同的是，M2 把四个性能核心的 L2 缓存提升到了 16MB。不过 M3 比较神秘，很多数据苹果并未对外公布，逆向数据也不多。M1 的主频当年只做到了 3.2GHz，Geekbench 6 测试基准下，单核跑分 2352 分，多核 8394 分。M2 把主频提升到了 3.5GHz，Geekbench 6 测试基准下，单核跑分 2633 分，多核 9748 分。M3 的提升则非常大，主频直接上到 4.1GHz，Geekbench 6 测试基准下，单核跑分 3056 分，多核 11691 分。

M1 当年发布的时候，在移动端轻薄本领域，带来了非常大的震撼。对标同期竞品，10W 功耗输出竞品 2 倍性能，而达到竞品峰值性能，功耗仅仅是竞品的 1/4 功耗。在 M1 之前，每年 CPU 的能效提升都微乎其微，基本都是在以能耗提升换性能提升，而 M1 一下子带来了 3 倍能效提升，可谓是降维打击。而后续的 M2、M3 可都是在 M1 这种巨幅提升的基础上进一步提升。M2 做了全新的核心设计，晶体管提升到 200 亿颗，面积更大，提供了更大的二级缓存，也把主频从 M1 的 3.2GHz 提升到了 3.5GHz。对标 M1，CPU 提升 18% 左右。对标当时最新的 10 核心 PC 笔记本芯片，M2 在 15W 左右的峰值性能差不多是同功耗下竞品的 2 倍。对标当时最新的 12 核心的 PC 笔记本芯片，M2 峰值性能弱了 10% 左右，但是功耗却仅仅只有竞品的 1/4。而 M3 的 CPU 性能和能效表现也依然很亮眼，简单来说，性能稳步提升，而能效依然行业领先。M3 的 CPU 实现了只用一半功耗就能达到 M1 的峰值性能，只用了短短三年时间。而在绝对性能上，M3 CPU 的高性能核心相较于 M2 提升 15%，相较于 M1 提升 30%。四个高能效核心，相较于 M1 提升 50%，相较于 M2 提升 30%。

## 2.2 GPU 性能纵向对比

	M1	M2	M3
机器学习图像优化放大性能提速	1	14.9 倍	17.7 倍
Final Cut Pro 中渲染性能提速	1	5.9 倍	7.4 倍
Affinity Photo 2 中图像处理性能提速	1	5.4 倍	5.9 倍
Excel 中电子表格性能提速	1	2.9 倍	3.5 倍

表 3: Geekbench 6 基准测试跑分

Apple Silicon M1 在 10W 下的 GPU 性能输出是同期最新笔记本集显性能的 2 倍。而达到同期最新笔记本集显峰值性能的功耗仅仅只有 1/3。

Apple Silicon M2 在 M1 同功耗下，可以有着 25% 的性能提升。而 M2 最大功耗下，性能则可以比 M1 高出 35%。M1 GPU 的最大功耗被控制在 10W 左右，而 M2 已经到 13W 左右了。根据 Geekbench Metal 的实际跑分来看，M2 可以跑到 30000 分，而 M1 差不多是 21000 分。对标 10 核心的 i7-1255U，M2 GPU 在最大功耗下的性能可以达到 2.3 倍。同样是对标 10 核心的 i7-1255U，M2 GPU 达到其峰值性能，能耗仅仅只有其 1/5。



Apple Silicon M3 跑出 M1 的性能仅仅需要一半的功耗，而峰值性能可以比 M1 高出 65%。可见集成显卡的进化速度相当之快，尤其是在不强烈依赖堆核心规模、拉高功耗的情况下。对标 12 核心的 i7 1360P，集成 iRIS Xe 显卡，M3 的 GPU 只用到 1/5 的功耗就可以达到 i7 1360P 的峰值性能。关键是，作为笔记本的处理器，这功耗直接拉高到 30W+。M3 的 GPU 真正强大之处不在于这种纯粹 GPU 性能的提升，而是 M3 终于补足了最后一块短板，那就是硬件光线追踪能力。在此基础上，M3 还加入了动态缓存以及硬件加速的网格着色功能。硬件光追可以很好的模拟真实世界，阴影和反光更加真实，无论是游戏还是专业 3D 渲染软件都能受益于此。渲染速度方面，相较于 M2 最快提速 1.8 倍。硬件加速的网格着色对于几何模型处理的功能和效率大幅提升，可以打造视觉效果更复杂的场景。而 M3 GPU 还新引入了 Dynamic Caching，也就是动态缓存。传统 GPU 还在根据最大任务需求提前预留内存呢！M3 已经可以根据实际需求动态调整局部内存使用了，GPU 平均利用率大幅提升。这个技术本身对开发者透明，也就是说，专业级的 app 可以更加充分发挥 GPU 的性能了。

## 2.3 GPU 性能横向对比

需要注意的是，比较显卡性能其实会受到很多因素的限制，比如测试基准，不同的测试会有不同的性能表现，测试设备本身的设计，同款显卡轻薄本一般就跑不过游戏本，低频内存就跑不过高频内存。甚至是那些跨平台的测试基准，也很难绝对反应显卡的真实水平。显卡性能的释放其实受限于软硬件平台的各种约束，甚至也受限于软件本身的优化。比如同样是 M2 芯片，导出同样的素材，在机身散热处于劣势的 iPad Pro 上，却可以比 MacBook M2 还要快。同一颗芯片搭载在不同平台上，不同的内存配置，不同的散热条件，也一样会有一定的性能差异。以下给出几个经典的测试集上定性的跨平台的性能比较。

### 2.3.1 Geekbench 6

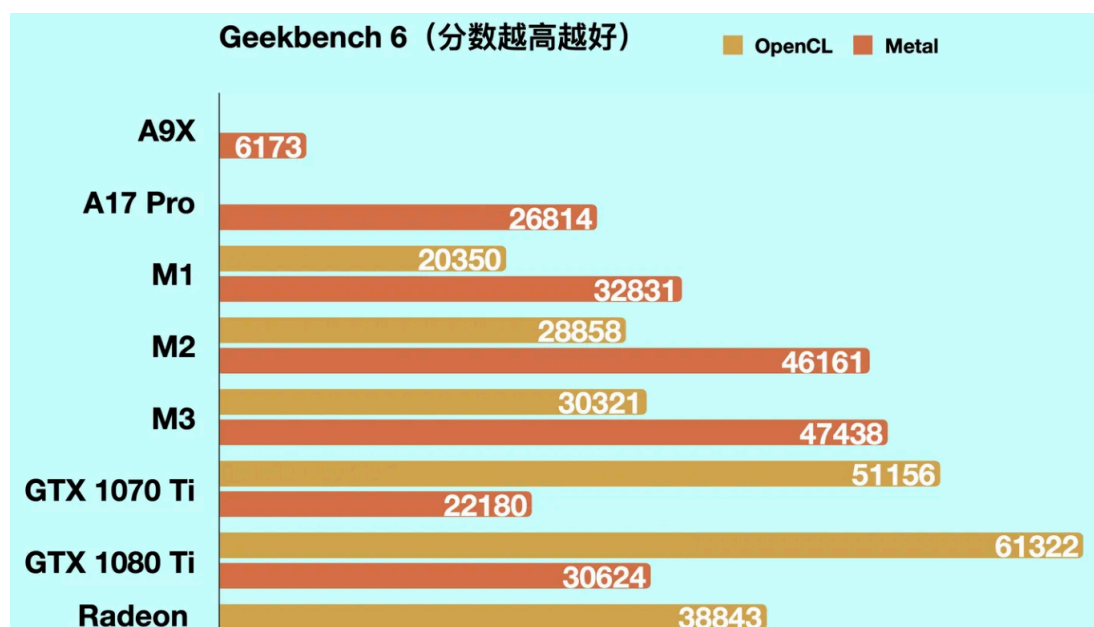


图 2.5: Geekbench 6 测试

A17 Pro 的 Metal 成绩已经接近 M1，更是 A9X 的 4 倍多。M2、M3 的 Metal 和 OpenCL 分数基本一致，而相较于英伟达和 AMD 的入门显卡，M 系列芯片在原生 Metal 上更有优势，而 OpenCL



就明显被拉开差距了。

### 2.3.2 3DMARK

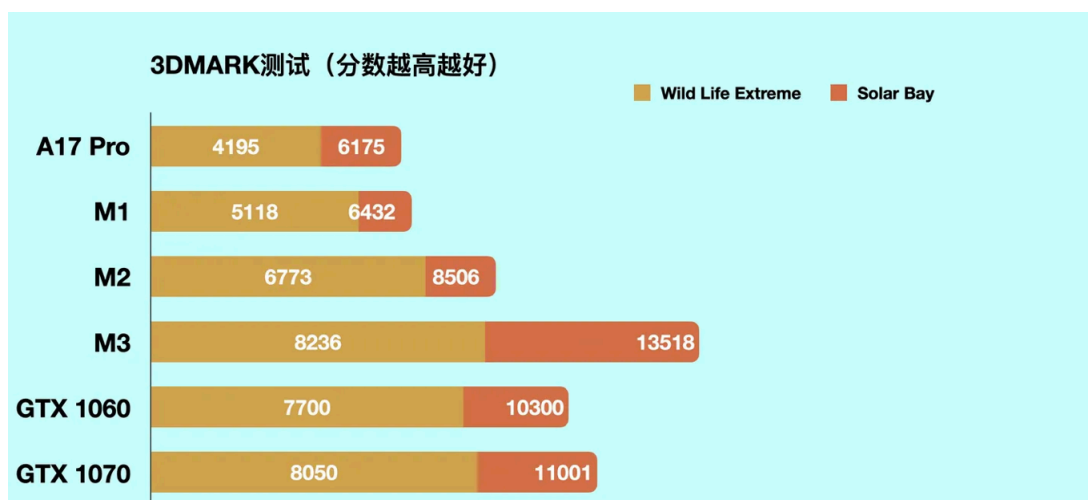


图 2.6: 3DMARK 性能测试

在 M1/M2/M3 之外，加入了 A17 Pro 和 GTX 1060/1070 作为对比。Wild Life Extreme 测试中，M3 最强，比 M2 强了 22%，比 M1 强了 61%。用到光追的 Solar Bay 测试中，M3 可谓是大放异彩，比 M2 直接强了 59%，比 M1 更是翻倍还多。其实同样有硬件光追加持的 A17 Pro，已经基本追上 M1 的 Solar Bay 成绩了。

### 2.3.3 Blender 4.0.0

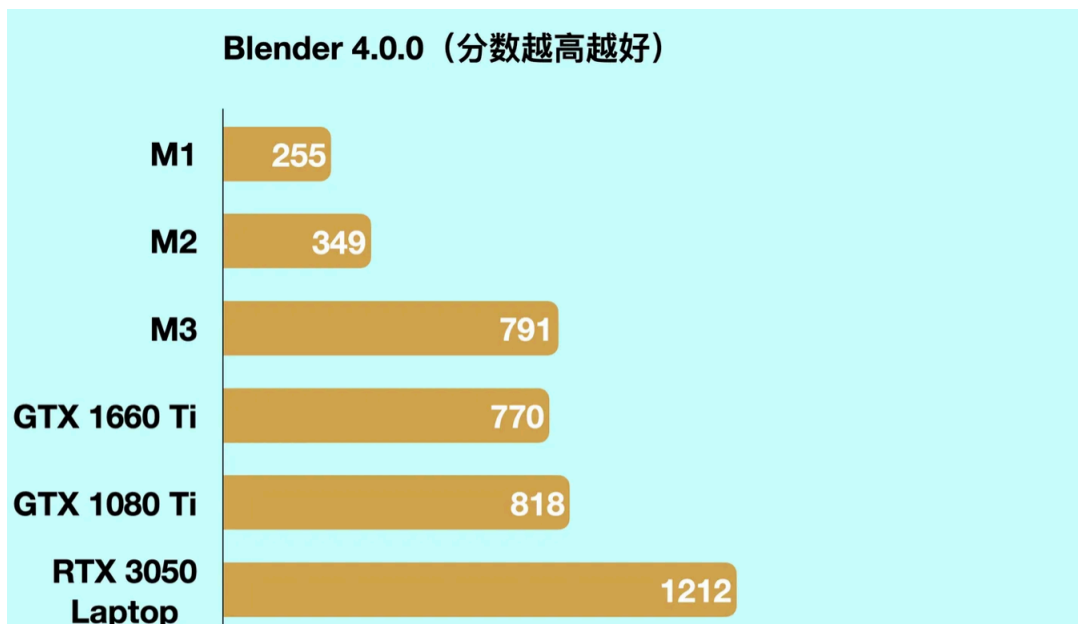


图 2.7: Blender 性能测试

在这个场景中，M3 有着巨幅提升，也应证了 GPU 性能很依赖特定场景。在 macOS 下，甚至仅仅是 app 层面的优化，都可以让 app 的处理速度大幅提升，而这其实就是软件生态的价值。

## 2.4 总结

苹果 M 系列芯片自 M1 到 M2 再到 M3 的发展趋势和特点主要体现在以下几个方面：

1. 性能提升：每一代的 M 系列芯片都会带来显著的性能提升。这包括 CPU、GPU、神经处理单元 (NPU) 等方面的提升。M1 芯片首次引入了苹果自研的基于 ARM 架构的 CPU 和 GPU 设计，以及强大的神经引擎，为苹果设备带来了新的性能水平。随后的 M2 和 M3 芯片在性能上进一步提升，为用户提供更快的运算速度和更流畅的体验。
2. 能效优化：除了提高性能外，苹果 M 系列芯片还致力于提升能效。这意味着在提升性能的同时，尽量减少功耗，延长电池续航时间，降低设备发热量。通过优化架构设计、制程工艺和功耗管理技术，每一代的 M 系列芯片都在能效方面取得了显著的进步。
3. 集成度提升：随着技术的进步，苹果 M 系列芯片在集成度上不断提升。更多的功能单元被集成到芯片中，包括 CPU、GPU、NPU、图像处理器等，同时集成了更多的外设控制器和加速器，从而减少了对外部芯片的依赖，提高了整体系统的一体化程度。
4. 改进的架构设计：每一代 M 系列芯片都会带来架构设计的改进，以进一步提高性能和能效。这可能包括对 CPU 核心、GPU 核心、NPU 架构等方面的改进和优化，以及对内存控制器、缓存系统、存储子系统关键组件的优化。
5. 面向特定应用的优化：随着 M 系列芯片的发展，苹果会越来越注重对特定应用场景的优化。这可能包括针对图形设计、视频编辑、机器学习等领域的优化，通过硬件与软件的深度结合，提供更好的用户体验和更高的性能表现。