



南開大學
Nankai University

计算机学院
并行程序设计期末报告

联想 Y7000P 2021 体系结构调研

姓名：卢麒萱
学号：2010519
专业：计算机科学与技术

2022 年 2 月 27 日

目录

1	引言	2
1.1	国内外并行硬件体系结构发展现状	2
1.2	国内外并行硬件体系结构发展趋势	2
1.3	选题目的及意义	3
2	体系结构	3
2.1	参数概览	3
2.2	CPU	3
2.2.1	Tiger Lake	3
2.2.2	i7-11800H 测评	4
2.2.3	12 代与 11 代酷睿对比	5
2.3	GPU	7
2.3.1	NVIDIA Ampere 架构	7
3	总结	9

1 引言

1.1 国内外并行硬件体系结构发展现状

并行计算机对于国家安全等重要领域有着举足轻重的作用。我国的并行计算研究始终紧跟国际步伐,尽管在方面开展的研究和应用较早,也拥有很多的并行计算资源,但成效相对美日等发达国家还存在较大的差距,在研究中形成了“理论-设计-实现-应用”的一体化并行计算研究体系。但并行计算的发展依旧是参差不齐的,目前并行应用的效率依然较低,无法高效的使用并行计算机的资源,导致大量资源闲置。并行程序编写难度较大,并行编程语言对于一般的编程人员还不够简单易行,也缺乏高效的并行编程环境工具,并行计算机本身的构建也存在着能耗过大、管理困难、可扩展性差等方面的问题。

当前,并行机发展基本状况可以大致归纳为:

- 并行软件的发展远远落后于并行计算体系结构的发展;
- 并行计算的应用远远落后于并行计算技术的发展;
- 大规模并行处理系统已不再是主要研究领域;
- 有高速网联成的各种类型、规模可伸缩计算机群,将进一步促使并行计算应用的发展;
- 计算系统的可扩展和可编程性已成为并行机未来发展的一对主要矛盾。

并行性包括时间并行性和空间并行性两部分。一般时间并行性的开发采用资源共享和时间重叠的方法,而空间并行性的提高则采用资源重复的方法。大规模并行结构尚有三大难题:节点负载均衡问题,Cache 一致性问题 and 通讯同步问题,均为全局优化问题。冯·诺依曼结构的一维顺序存储模型严重地制约了并行体系结构的发展,在此基础上进行并行性的挖掘只能有限地提高计算机性能。

1.2 国内外并行硬件体系结构发展趋势

在并行机硬件与用户需求之间有一巨大间隔,只能靠软件来填补。并行计算机的应用效率和可用性主要取决于并行软件。计算能力并不是衡量超级计算机的唯一重要标准,拿天河二号为例,它的运算能力世界第一,但是功能却远远落后于美国等国家的超级计算机。并行算法的效率与体系结构密切相关。将常用的并行算法经充分优化后做成并行算法库供用户调用,将目前广泛使用的库函数并行化变成标准的并行库函数,这是推广并行机必须要做的事,也是提高并行机实际性能的关键技术。

多核技术的出现与主流化,对于并行计算体系结构、并行算法、并行程序设计与并行应用的研究都分别产生了重要的影响,带来了新的挑战。一般说来,超级计算机系统可分为两大部分,一部分是包括硬件和系统软件的平台,另一部分是应用软件。对不具有可编程性的超级计算机,即使平台有较高的性价比,但却大大增加了移植原有应用软件的难度和开发新的应用软件的成本,从而导致整个系统的性价比大幅度下降。多核化趋势正在改变并行计算的面貌,人们开始不再仅仅关注并行计算机的高性能,也开始关注高效能。多核技术对并行计算机系统的算法和编程带来了很大的困难,程序代码迁移也是个问题,需要有一个清晰的迁移策略,来使其代码可以最大化地利用多核硬件资源,同时为之付出的迁移代价要尽可能的小。未来的并行计算机系统需要去适应多核系统的发展,多核技术促进了并行编程的普及以及并行程序设计的进步,降低了物理上共享存储的并行处理平台的门槛。

1.3 选题目的及意义

联想拯救者系列，是当前国内市面上中高端定位主流游戏本系列。而 Y7000P 2021 款，也是畅销一时的大热机型，在同价位机型中性价比突出，深受民众青睐。本文意通过对该并行计算机体系结构进行调研与剖析，了解当前主流家用并行计算机架构的发展现状，同时与最新产品作对比，印证并行计算机未来的发展所趋。

2 体系结构

2.1 参数概览

处理器	Intel Core i7-11800H(2.3GHz/L3 24M)
核心/线程	八核心/十六线程
总线规格	DMI3 8GT/s
核心架构	Tiger Lake-H
处理器系列	第 11 代酷睿 i7
处理器主频	2.3GHz
最高频率	4.6GHz
三级缓存	L3 24M
制程工艺	10nm
功耗	80W（高耗电）

表 1: 处理器

显卡类型	独立显卡
显卡芯片	NVIDIA GeForce RTX 3060
显存容量	6GB
显存位宽	192bit
显存类型	GDDR6
显卡性能	支持光线追踪，支持 DirectX 12

表 2: 显示卡

2.2 CPU

酷睿 i7-11800H 作为英特尔的主流产品，用于大多数中高端笔记本电脑。搭载着英特尔推出的笔记本处理器 11 代酷睿，代号 Tiger Lake，使用 10nm+ 工艺，集成全新的 Willow Cove CPU 架构、Xe 架构的 Gen12 GPU 核显，并大大地强化了 AI 能力，最高加速频率可达 5GHz。11800H 配备 8 个核心和 16 个线程，24MB 的 L3 缓存，2.3GHz 的基本频率，功耗为 45W，睿频加速 2 个核心可达 4.6GHz，全核可达 4.2GHz。Xe 集成 GPU 有 32 个执行单元，频率速度高达 1450 MHz。

2.2.1 Tiger Lake

Tiger Lake 是英特尔第 11 代英特尔酷睿移动处理器的代号，它基于全新的 Willow Cove Core 微架构，采用英特尔第三代 10 纳米工艺节点 10SF（“10 纳米 SuperFin”）制造。Tiger Lake 取代了 Ice Lake 系列移动处理器，代表了英特尔流程-架构-优化模型中的优化步骤。

技术节点	英特尔 10 纳米 SuperFin (10SF) 工艺
架构	x86-64
微架构	Willow Cove
指令	x86-64
核心	2-8
GPU	基于 Intel Xe 的集成显卡

表 3: Tiger Lake 架构和规格

核心是英特尔最新的 Willow Cove 处理器内核，具有 1.25 MB 的专有 L2 缓存和每个内核高达 3 MB 的共享 L3 缓存，以及英特尔最新的 Xe-LP 显卡。这些都从 6 核和 12 线程开始，并为最昂贵的型号提供高达 5.0 GHz 的单核涡轮增压。

英特尔的顶级 Tiger Lake 'Core 11 th Gen' 处理器，即 Core i7-1185G7。这是具有超线程的四核处理器，总共提供八个线程。该处理器还具有全尺寸的新 Xe-LP 图形，具有 96 个运行单元，运行频率高达 1450 MHz。

核心线程	4 核 8 线程
12 W 时的基本频率	1200 兆赫
15 W 时的基本频率	1800 兆赫
28 W 时的基本频率	3000 兆赫
高达 50 W 的 1C Turbo	4800 兆赫
高达 50 W 的全核 Turbo	4300 兆赫
二级缓存	每个核心 1.25 MB
三级缓存	12 MB
集成显卡	Xe-LP 96 执行单元 1350 MHz Turbo
内存支持	32 GB LPDDR4X-4266 或 64 GB DDR4-3200

表 4: Intel Core i7-1185G7 'Tiger Lake'

在 12 W 时，英特尔列出了 1.2 GHz 的基本频率，而在 28 W 时，英特尔列出了 3.0 GHz 的基本频率。在 12 W 和 28 W 场景中，处理器可以在一个核心/两个线程上加速到 4.8 GHz。该系统专为散热或电源问题而设计，因此 CPU 在两种模式下均可提升至 4.8 GHz。不仅如此，对于任何 TDP 设置，在 Turbo 模式下的功耗都被限制为 55 W。

2.2.2 i7-11800H 测评

根据 ZMMOO 的性能测评，有以下的测评结果图参考。

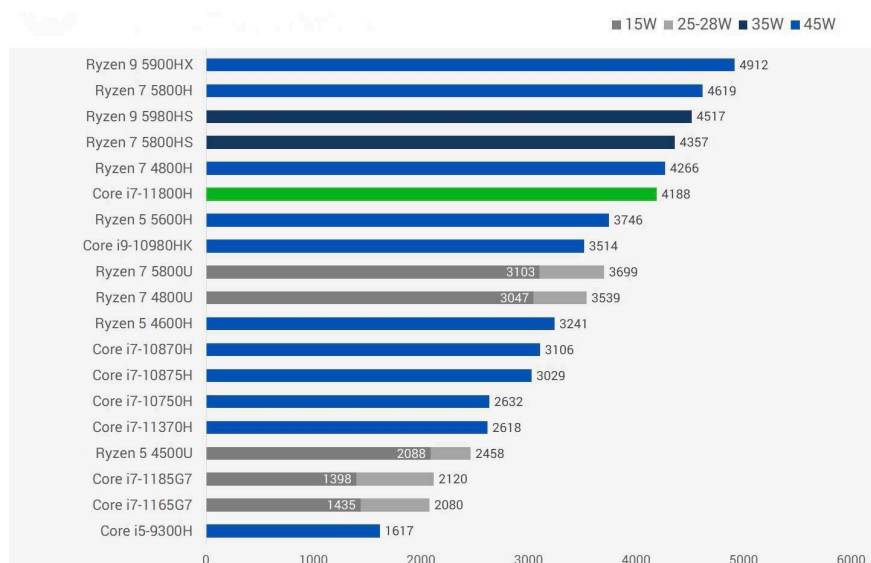


图 2.1: Cinebench R20 多线程性能

多线程方面酷睿 i7-11800H 比酷睿 i7-10870H 快了 35%，比 i7-10750H 领先近 60%。

尽管比上一代有巨大的进步，但英特尔所能实现的性能只能和 AMD 上一代的 Ryzen 7 4800H 媲美；11800H 仍然比 5800H 慢 9%。

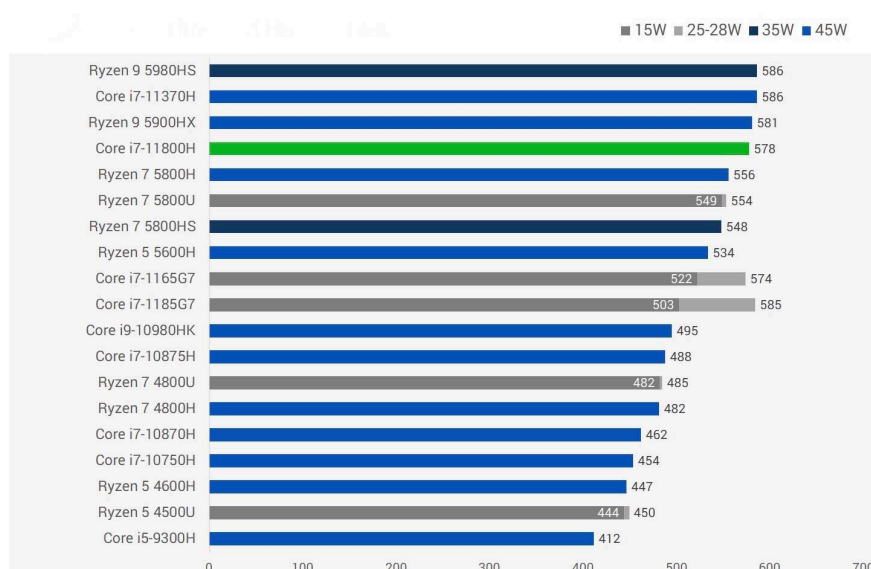


图 2.2: Cinebench R20 单线程性能

虽然英特尔在多线程方面无法与 AMD 匹敌，但正如表中所示，Tiger Lake 的单线程性能很强。即使频率只有 4.6GHz，11800H 也比 5800H 高出 4%，这使它的性能与 AMD 更高级别的 Ryzen 9 5900HX 相近。5900HX 的频率也达到了 4.6GHz。

2.2.3 12 代与 11 代酷睿对比

第 12 代英特尔® 酷睿™ 产品家族包括 60 款处理器，为来自合作伙伴的 500 多种机型设计提供动力。根据 2021 年英特尔架构日介绍，首次采用 Intel 7 制程工艺的全新高性能混合架构，带来了从

9 瓦到 125 瓦的可扩展性能，为从超轻薄笔记本电脑到发烧友台式机的每一个 PC 细分市场以及边缘计算设备提供卓越的计算性能。

英特尔 12 代酷睿处理器有很多创新的地方，最显著的变化就是采用了 Hybrid 混合架构：

- 性能核：Performance Core，简称 P-Core，采用 Golden Cove 架构；
- 能效核：Efficient Core，简称 E-Core，采用 Gracemont 微架构。

E-Core 负责负载较轻的任务，处理及时又省功耗，P-Core 负责高负载的任务性能更强，速度更快，如果能够合理的分配和调度这些处理器核心，就可以高效快速的处理任务。从结构上来说，异构核之间通过缓存的一致性可以解决相互间的同步问题，英特尔在 12 代酷睿处理器上重新设计了缓存架构：

- 每个 P-Core 都有单独的二级缓存；
- 每簇中的 E-Core 共享二级缓存；
- P-Core、E-Core 以及显卡之间智能共享 30MB 的三级缓存。

具体到英特尔 12 代酷睿处理器核之间的调度顺序，优先级为：P-Core—>E-Core—>HT 超线程。一个高性能任务进来，会首先进入到 P-Core 中，如果 P-Core 资源队列排满的话就安排到 E-Core，当有更高优先级的任务比如说浮点运算进来的话会优先安排到 P-Core，P-Core 中内排挤出的任务进入 E-Core，加入还需要更高性能，此时就需要打开超线程。这样会将处理器的性能发挥到最佳。

12 代酷睿处理器采用了 Intel 7 制程工艺 (10nm Enhanced SuperFin) 制造以及全新的微架构，因此在此 IPC (Instruction Per Clock) 上有了大幅的提升。相比 11 代，12 代酷睿处理器的 P-Core 在 IPC 上提升大约 28%，E-Core 提升 1%。在多核性能上，酷睿 i9-11900K 在实际功耗 250W 的负载中，酷睿 i9-12900K 实际功耗 241W 的情况下性能提升 50%，也就是说每瓦的性能提升了约 50%。更夸张的是酷睿 i9-12900K 只需要 65W 的功耗就能有酷睿 i9-11900K 的性能，只用约 1/4 的功耗就能达到相同的性能。不过这里需要注意的是只是酷睿 i9-11900K 的规格为 8 核 16 线程，而酷睿 i9-12900K 的规格为 16 核 24 线程，但是在性能上的提升是有目共睹的，这种提升与混合架构的设计也有关系。

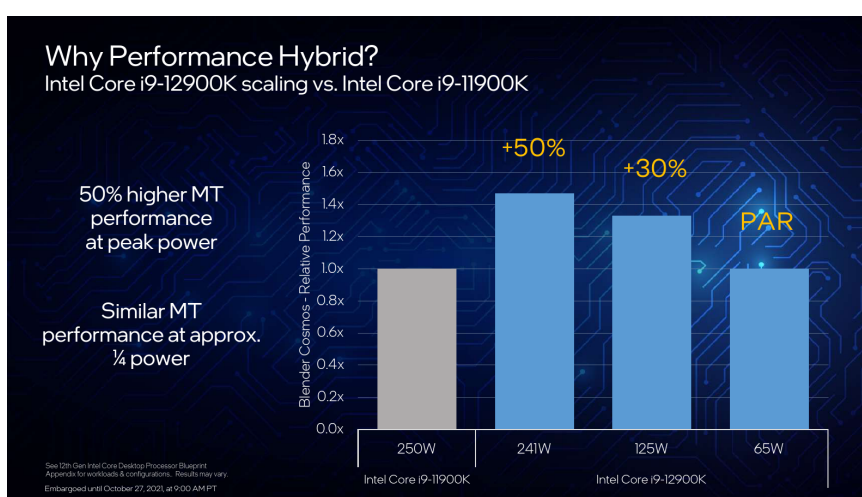


图 2.3: Intel Core i9-12900K vs. Intel Core i9-11900K

2.3 GPU

RTX 3060 采用了 GA106-300-A1 核心，核心面积 276mm²，CUDA 数量为 3584 个，生产工艺是三星为 NVIDIA 定制的 8nm 工艺。RTX 3060 拥有 1.32-1.78GHz 频率，配备了 12GB 192bit 位宽的 GDDR6 显存。这款显卡的官方功耗为 170W，配合 i9-10900K 需要 550W 电源。其浮点性能为 13 TFLOPS，张量性能可达 101 TFLOPS，光追性能则是 25 RT-TFLOPS。仅从规格上看，它的传统光栅化性能是 GeForce GTX 1060 的 2 倍，而光追性能则相比 GeForce GTX 1060 提升高达 10 倍。

型号	RTX 3060	RTX 3070	RTX 2060 SUPER
核心代号	GA106-300	GA104-300	TU106-410
核心面积 (mm ²)	276	392.5	445
制程 (nm)	8	8	12
GPCs	3	6	3
SMs/CUs	28	46	34
CUDA Cores/SP	3584	5888	2176
Tensor Cores	112 (第三代)	184 (第三代)	272 (第二代)
RT Cores	28 (第二代)	46 (第二代)	34 (第一代)
纹理单元	112	184	136
光栅单元	48	96	64
基础频率 (MHz)	1320	1500	1470
Boost 频率 (MHz)	1882	1725	1650
单精度浮点性能 (TFLOPS)	12.7	20.3	7/7.2
显存	12GB GDDR6	8GB GDDR6	8GB GDDR6
显存数据速率 (Gbps)	15	14	14
显存位宽 (bit)	192	256	256
显存带宽 (GB/s)	360	448	448
视频接口	HDMI 2.1×2 DisplayPort 1.4×3	HDMI 2.1×1 DisplayPort 1.4×3	HDMI 2.0b×1 DisplayPort 1.4×2 USB-C×1 DVI-D×1
TDP/TGP (W)	170	220	175
推荐电源 (W)	550	650	550
供电接口	8 Pin	12 Pin	8 Pin

表 5: 热门显卡规格参数对比

2.3.1 NVIDIA Ampere 架构

安培微架构 (Ampere) 是 NVIDIA 于 2020 年 5 月发布的一个 GPU 架构。用以取代图灵微架构 (Turing microarchitecture)。命名为“安培”以向法国物理学家安德烈-马里·安培 (André-Marie Ampère) 致敬。Ampere 架构拥有晶体管达 540 亿，是三星 8nm 级芯片。是世界上晶体管最多的芯片，直到后来被苹果 M1 Max 击败。

Ampere 包含六项关键性的突破创新：

- **CUDA® 核心：**与上一代相比，NVIDIA Ampere 架构的 CUDA 核心可将单精度浮点 (FP32) 运算处理速度提升一倍，并将能效提升 2 倍。这显著改善了 3D 模型开发等图形工作流程的性能，另外还为计算机辅助工程 (CAE) 的桌面模拟等工作负载提供了强大算力。
- **第二代 RT Cores：**第二代 RT Core 的计算吞吐量是上一代的 2 倍，并能同时运行光线追踪和着色或降噪功能，从而大幅加快工作负载的运行速度，例如电影内容的逼真渲染和产品设计的虚拟

原型创建。这项技术还可加速渲染具有光线追踪效果的动态模糊画面，从而更快获得视觉准确性更高的结果。

- 第三代 Tensor Cores：新的 Tensor Float 32 (TF32) 精度提供的训练吞吐量达到上一代的 5 倍，而且无需更改代码即可加速 AI 和数据科学模型的训练。从硬件上支持结构化稀疏使推理吞吐量提升一倍。Tensor Core 还通过 DLSS、AI 降噪等功能将 AI 引入到图形处理中，并增强了特定应用程序的编辑功能。

另外，为了提高 Tensor Cores 训练 AI 时的效率，NVIDIA 新创了一种名为 TF32 的数据类型，它拥有 FP32 的范围和 FP16 的精度，对于调用 Tensor Core 的操作，它会自动启用 TF32 进行处理。而没有调用 Tensor Cores 的操作将仍然走 FP32 的数据路径，Tensor Cores 会自动读取 FP32 数据，在内部减精度进行运算，在最终输出的时候会将数据还原成 IEEE 标准。

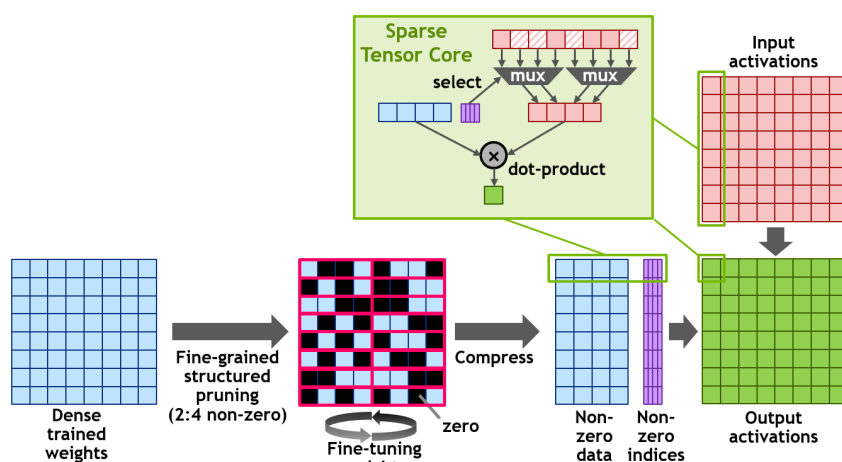


图 2.4: 计算流程演示

新版 Tensor Cores 还支持稀疏矩阵运算。稀疏矩阵指的是大部分元素为 0 的矩阵，对于这种矩阵，NVIDIA 使用了自己开发出来的稀疏计算方式，它支持 2:4 的结构化稀疏运算，需要参与计算的矩阵在每四个元素中有 2 个以上的 0 元素，它可以将 Tensor Cores 的计算吞吐量翻一倍。

- PCI Express 4.0：基于 NVIDIA Ampere 架构的 GPU 支持 PCI Express 4.0，该规范提供的带宽是 PCIe 第 3.0 代的 2 倍。这提高了从 CPU 内存传输数据的速度，可更好地执行 AI 和数据科学等数据密集型任务。更快的 PCIe 性能还能加速 GPU 直接显存访问 (DMA) 传输，从而能让支持视频的设备通过 GPUDirect® 更快速地传输视频数据，以及利用 GPUDirect Storage 加快输入/输出 (I/O) 速度。
- 第三代 NVLink：第三代 NVIDIA NVLink® 技术允许用户将 2 个 GPU 连接起来，以分享 GPU 性能和显存。借助高达 112 千兆字节/秒 (GB/s) 的双向带宽和高达 96 GB 的组合显存，专业人员可以应对大型的渲染、AI、虚拟现实和视觉计算工作负载。新的 NVLink 连接器还具有更低矮的外形，可在更多型号的机箱中实现 NVLink 功能。

在图形计算方面，考虑到目前复杂的图形计算任务，NVIDIA 采用的是 FP32+INT32 混组核心的设计，能够带来每晶体管性能的显著提升。

3 总结

- i7-11800H 在同系列处理器中属于中高产品，无明显短板，单线程性能强，多线程相对于 AMD 的 5800H 略差。
- 12 代酷睿处理器的表现对得起众多 DIY 玩家长久的期盼，混合架构的设计被引入到了 PC 领域并且有着非常优秀的表现，相信未来 PC 领域的处理器都将追随英特尔的脚步推出混合架构设计的处理器。从这一点来看英特尔无愧是 PC 领域的引领者，与此同时也让人期待英特尔的移动标压处理器。
- Ampere 架构的更新并不是革命性的，而是延续了 NVIDIA 这几年在架构设计上的一贯思路，微观上在 SM 单元中延续分精度计算，并加强 Tensor Cores 这个对深度学习计算非常有用的单元，宏观上面增大 GPU 的规模，不仅将整个 GPU 包含的 SM 单元数量扩大到 128 组这个数字，更是把整片 GPU 上面的缓存系统都放大了，尤其是 40MB 的二级缓存，让人印象深刻。

参考文献