

ChatGPT开启AI发展新浪潮，算力紧缺和海量应用驱动AI硬件广阔空间

——AI行业跟踪报告之二



作者：光大证券电子通信行业首席分析师 刘凯，执业证书编号：S0930517100002

2023年2月15日



证券研究报告

- ❑ **ChatGPT 驱动AI浪潮趋势渐起，芯片等AI硬件构成核心底座**
 - 作为AI领域的最新成果，ChatGPT发展空间巨大。
 - 目前微软、谷歌等科技巨头坚定入局，展现出AI应用广阔的发展前景，AI行业有望开启新发展浪潮
- ❑ **云端算力硬件将成为AI新浪潮的关键基础设施**
 - GPU、CPU、FPGA、AI芯片、Chiplet、光模块、服务器等细分板块下游应用广泛，产品边际持续拓展。云、边、端需求拉动市场规模高速增长。
- ❑ **AI新浪潮未来将提升海量IoT设备的边缘算力需求**
 - NPU：在SoC与MCU中加入边缘算力，可广泛使用于各类AIoT设备。
 - RISC-V芯片架构：深度受益于未来AIoT趋势
- ❑ **投资建议：云端算力和边缘算力两个维度关注AI硬件投资机会**
 - 云端算力：建议关注（1）GPU：景嘉微、海光信息；（2）CPU：龙芯中科、中国长城、澜起科技、；（3）AI芯片：寒武纪、澜起科技；（4）FPGA：复旦微电、安路科技、紫光国微等；（5）光模块：天孚通信、德科立、新易盛、光迅科技、博创科技、源杰科技、中际旭创等；（6）CHIPLET：兴森科技、长川科技、方邦股份、华正新材、通富微电、长电科技等。
 - 边缘算力：建议关注（1）SoC：晶晨股份、富瀚微、瑞芯微等；（2）AIoT：乐鑫科技、恒玄科技等；（3）RISC-V：中科蓝讯、乐鑫科技、全志科技等。

ChatGPT 驱动AI浪潮趋势渐起，芯片等AI硬件构成核心底座

云端算力硬件将成为AI新浪潮的关键基础设施

AI新浪潮未来将提升海量IoT设备的边缘算力需求

投资建议：云端算力和边缘算力两个维度关注AI硬件投资机会

风险分析

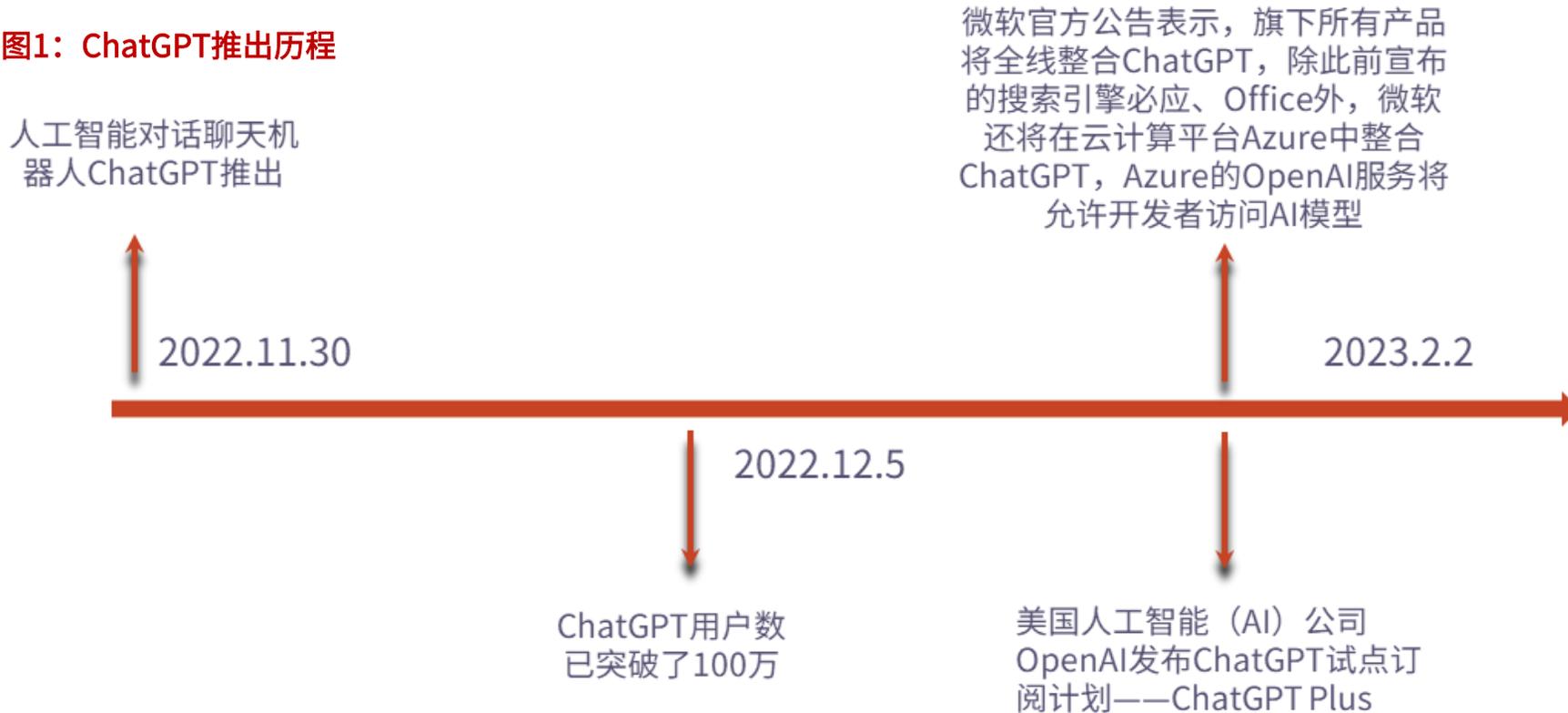
一、ChatGPT 驱动AI浪潮趋势渐起，芯片等AI硬件构成核心底座

- ❑ 1、ChatGPT应用空间广阔
- ❑ 2、AIoT终端浪潮趋势渐起
- ❑ 3、数据和算力是人工智能的基础，芯片等硬件构成AI核心底座

1.1、ChatGPT应用空间广阔

- ChatGPT是由人工智能实验室OpenAI研发的通用聊天机器人，使用了Transformer神经网络架构。GPT-3.5架构是一种用于处理序列数据的模型，于2022年11月30日上线。ChatGPT的网页应用允许用户免费使用，不限量向公众开放，用户与ChatGPT之间的对话互动包括了普通聊天、信息咨询、撰写诗词作文、修改代码等。
- 作为AI领域的最新成果，ChatGPT也拥有足够的想象力和发展空间。目前微软、谷歌等科技巨头坚定入局，展现出AI应用广阔的发展前景，AI行业有望开启新发展浪潮。

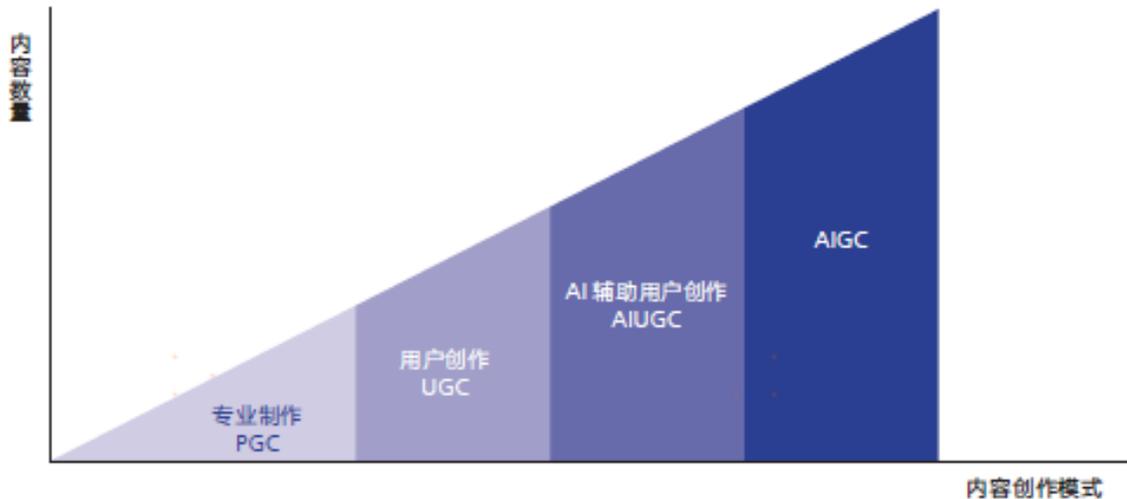
图1：ChatGPT推出历程



1.1、ChatGPT应用空间广阔

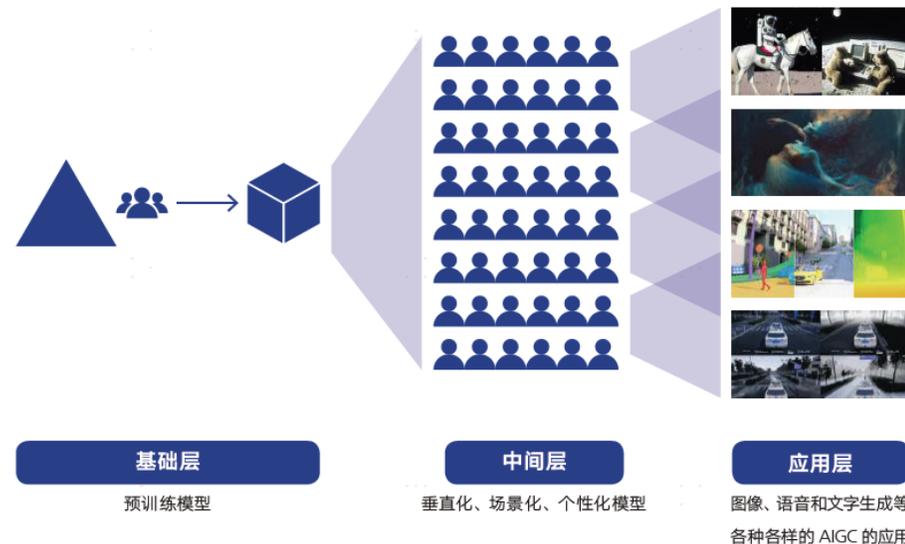
- 以ChatGPT为代表的AIGC备受关注。全球各大科技企业都在积极拥抱AIGC，不断推出相关技术、平台和应用。
- AIGC产业生态加速形成，走向模型即服务(MaaS) 的未来。

图2：内容创作模式的四个发展阶段



资料来源：腾讯研究院

图3：AIGC产业生态体系的三层构架

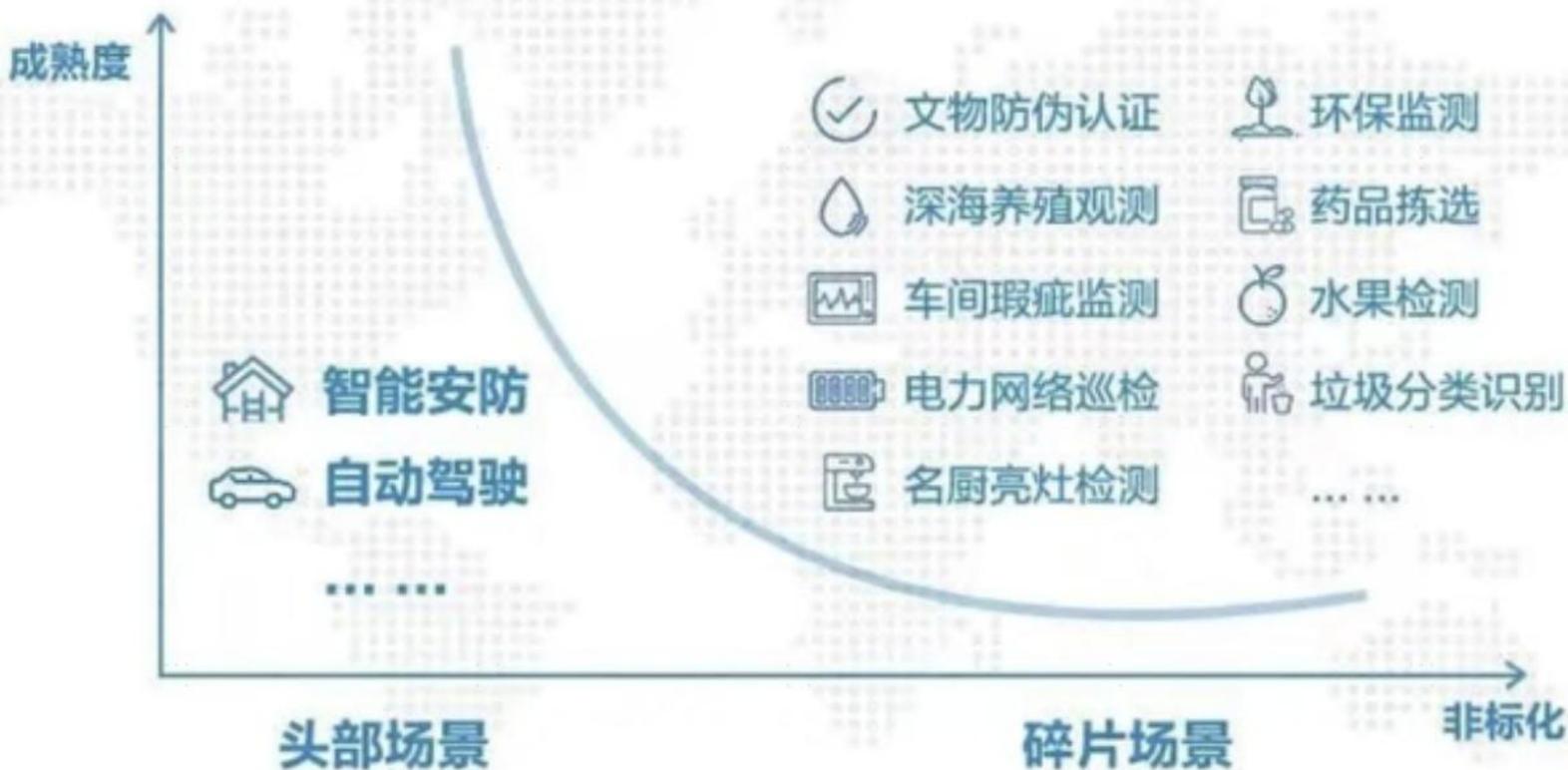


资料来源：腾讯研究院

1.2、AIoT终端浪潮趋势渐起

- AIoT时代拥有海量IoT终端。“智能”将是物联网时代最核心的生产力，AI技术将渗透到云、边、端和应用的各个层面，与IoT设备进行深度融合。在物联网时代，用户对于智能的需求呈快速增长，深入各个领域。

图4：AIoT时代拥有海量IoT终端

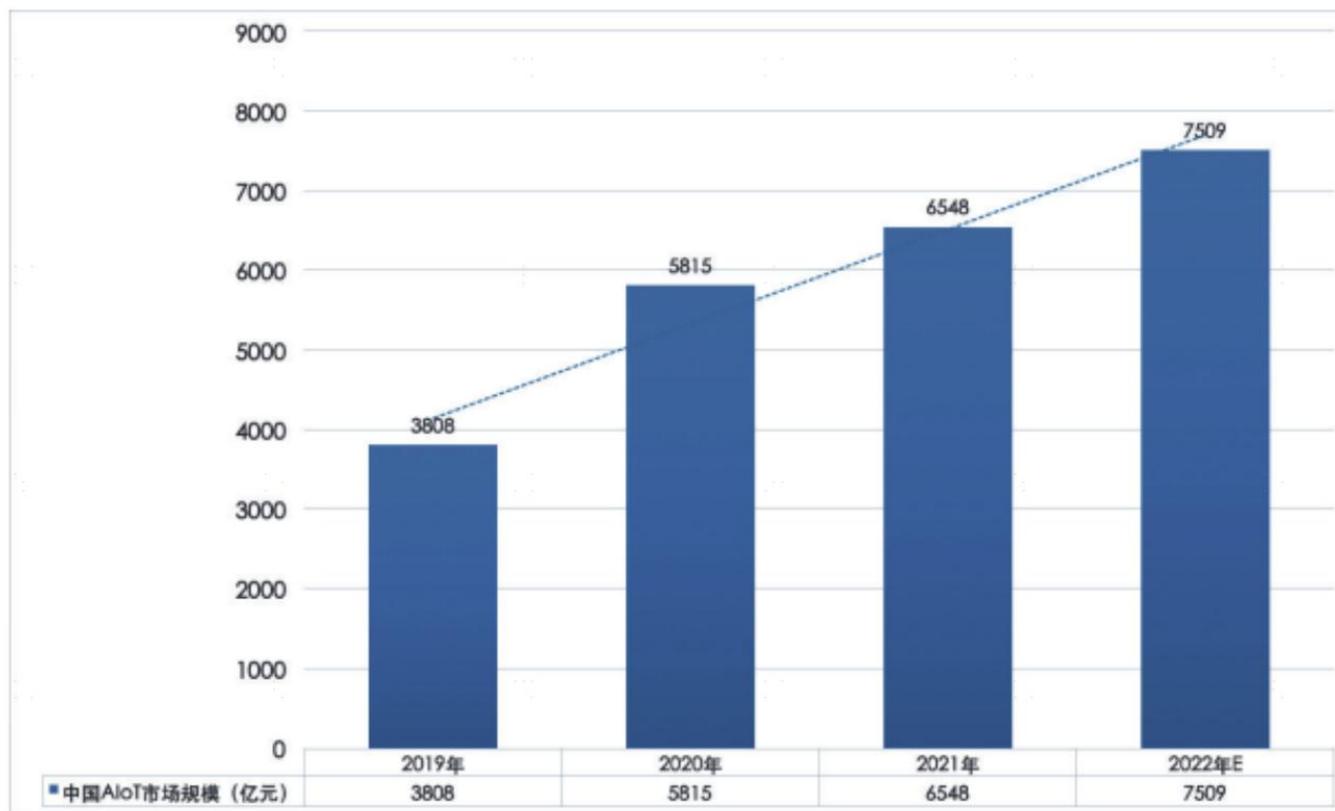


资料来源：猎云网

1.2、AIoT终端浪潮趋势渐起

- 根据IDC数据：2019年中国AIoT产业总产值为3808亿元，2021年将达到6548亿元。AIoT市场规模的快速增长，得益于以5G为代表的新技术的规模化商用和AIoT应用在消费及公共事业等领域的大规模落地。

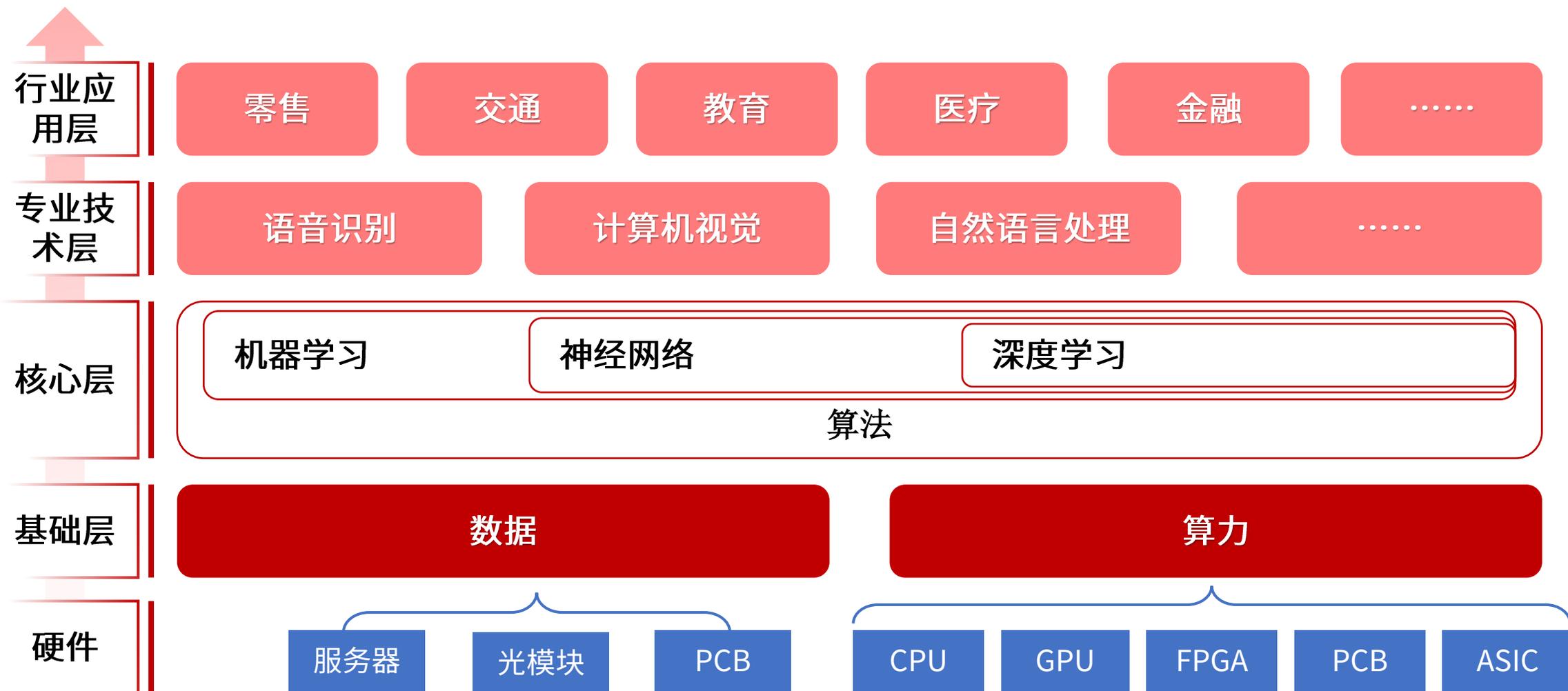
图5：2019-2022E中国AIoT市场规模



资料来源：IDC预测，a&s Research

1.3、数据和算力是人工智能的基础，芯片等硬件构成AI核心底座

➤ 人工智能的基础层是数据和算力，数据由服务器和光模块存储和运输；算力由CPU、GPU、FPGA、ASIC等芯片支撑





ChatGPT 驱动AI浪潮趋势渐起，芯片等AI硬件构成核心底座

云端算力硬件将成为AI新浪潮的关键基础设施

AI新浪潮未来将提升海量IoT设备的边缘算力需求

投资建议：云端算力和边缘算力两个维度关注AI硬件投资机会

风险分析

二、云端算力硬件将成为AI新浪潮的关键基础设施

- ❑ 1、GPU分为传统GPU与GPGPU， GPGPU占据人工智能 90%以上份额
- ❑ 2、CPU与AI融合将成大势所趋
- ❑ 3、FPGA与ASIC技术融合，在人工智能领域应用广泛
- ❑ 4、云、边、端需求拉动AI芯片市场规模高速增长
- ❑ 5、从AI Chip到AI Chiplet
- ❑ 6、数据流量与AI算力增长推动光模块需求
- ❑ 7、算力需求带动服务器 PCB 市场持续扩容

2.1、GPU分为传统GPU与GPGPU， GPGPU占据人工智能 90%以上份额

- 随着 GPU 在并行计算方面性能优势的逐步显现以及并行计算应用范围的逐步拓展，GPU 逐渐分化成两条分支：
- 一条是传统意义的 GPU，延续专门用于图形图像处理用途，内置了视频编解码加速引擎、2D 加速引擎、3D 加速引擎、图像渲染等专用运算模块。
- 另一分支是 GPGPU，作为运算协处理器，并针对不同应用领域的需求，增加了专用向量、张量、矩阵运算指令，提升了浮点运算的精度和性能，以满足不同计算场景的需要。

图6：GPGPU主要应用领域

类型	应用领域	算力类型	技术特点
商业计算和大数据处理	<ol style="list-style-type: none"> 1、CAE 仿真 2、物理化学 3、石油勘探 4、生命科学 5、气象环境 	<ol style="list-style-type: none"> 1、双精度浮点 2、单精度浮点 3、32 位整型 	<ol style="list-style-type: none"> 1、对芯片计算能力及运算精度要求高 2、科学运算指令集丰富 3、片上集成缓存容量大 4、内存带宽需求高 5、I/O 带宽高 6、支持多片一致性互连 7、可靠性高，RAS 功能丰富
人工智能	<ol style="list-style-type: none"> 1、模型训练 2、应用推理 	<ol style="list-style-type: none"> 1、混合精度浮点 2、半精度浮点 3、16 位整型 4、8 位整型 	<ol style="list-style-type: none"> 1、对计算性能要求高，精度需求相对低 2、能效比要求高 3、运算指令集丰富 4、内存带宽要求大 5、I/O 带宽高 6、支持多片互连 7、可靠性高、RAS 功能丰富 8、开放的生态环境

资料来源：海光信息招股说明书

2.1、GPU分为传统GPU与GPGPU， GPGPU占据人工智能 90%以上份额

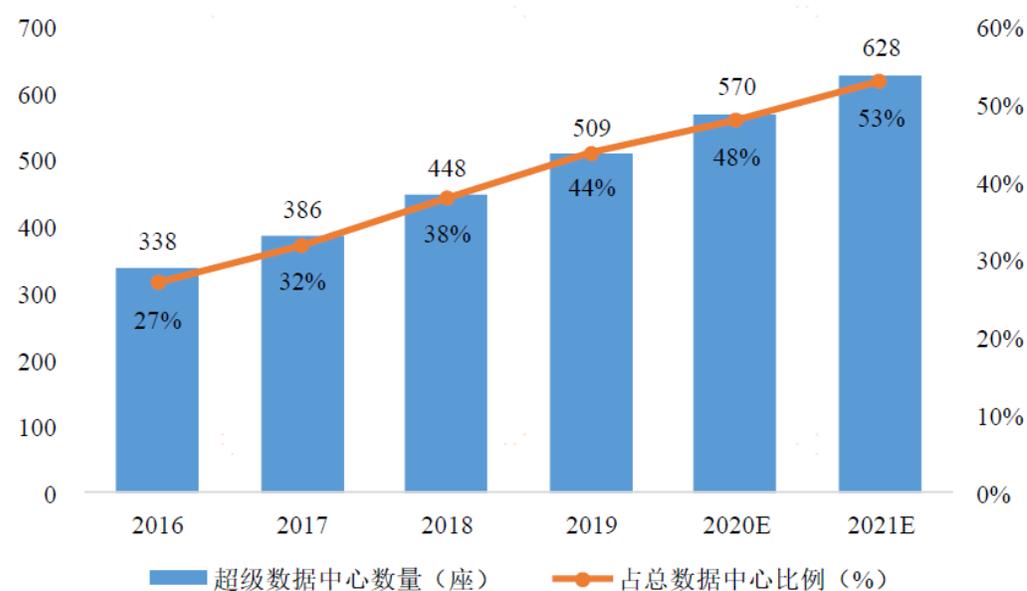
- 大规模张量运算、矩阵运算是人工智能在计算层面的突出需求，高并行度的深度学习算法在视觉、语音和自然语言处理等领域上的广泛应用使得计算能力需求呈现快速增长。
- 根据Cisco 的预计，2021 年全球数据中心负载任务量将超过 2016 年的两倍，从 2016年的不到250 万个负载任务量增长到 2021 年的近 570 万个负载任务量。
- 随着云计算的不断发展，全球范围内云数据中心、超级数据中心的建设速度不断加快，Cisco 预计到 2021 年，计算能力更强的超级数据中心将达到 628 座，占数据中心总量的53%。

图7：2016-2021年数据中心负载任务量变化



资料来源：海光信息招股说明书

图8：2016-2021年超级数据中心数量变化



资料来源：海光信息招股说明书

2.1、GPU分为传统GPU与GPGPU， GPGPU占据人工智能 90%以上份额

- 在人工智能领域，使用 GPGPU 在云端运行模型训练算法，可以缩短海量训练数据的训练时长，减少能源消耗，从而进一步降低人工智能的应用成本。
- GPGPU 能够提供完善的软件生态系统，便于各种已有应用程序的移植和新算法的开发，因此全球人工智能相关处理器解决方案仍然是以 GPGPU 为主。
- GPGPU 是人工智能领域最主要的协处理器解决方案，占据人工智能 90%以上的市场份额，在智能工厂、无人驾驶、智慧城市等领域具有广泛的市场空间。
- 根据前瞻产业研究院的数据，未来几年内，中国人工智能芯片市场规模将保持年均40%至50%的增长速度，到2024年，市场规模将达到785亿元。而随着人工智能相关技术的进步，应用场景将更加多元化，GPGPU通用性好和软件生态系统完善的优势会进一步展现出来，成为该领域的主流解决方案。GPGPU在我国人工智能芯片领域也将占据较大比例的市场份额。
- GPU 在商业领域的具体应用体现在：1.商业计算和大数据处理；2.人工智能处理器；3.智算中心。

2.2、CPU与AI融合将成大势所趋

- CPU 是计算机的运算和控制核心，是信息处理、程序运行的最终执行单元，是计算机的核心组成部件。CPU 的本质是超大规模集成电路，用于解释计算机指令和处理计算机软件中的数据，并负责控制、调配计算机的所有软硬件资源。
- CPU 可以应用在服务器、工作站、个人计算机（台式机、笔记本电脑）、移动终端和嵌入式设备等不同设备上，根据应用领域的不同，其架构、功能、性能、可靠性、能效比等技术指标也存在一定差异。

图9：CPU应用场景及技术特点

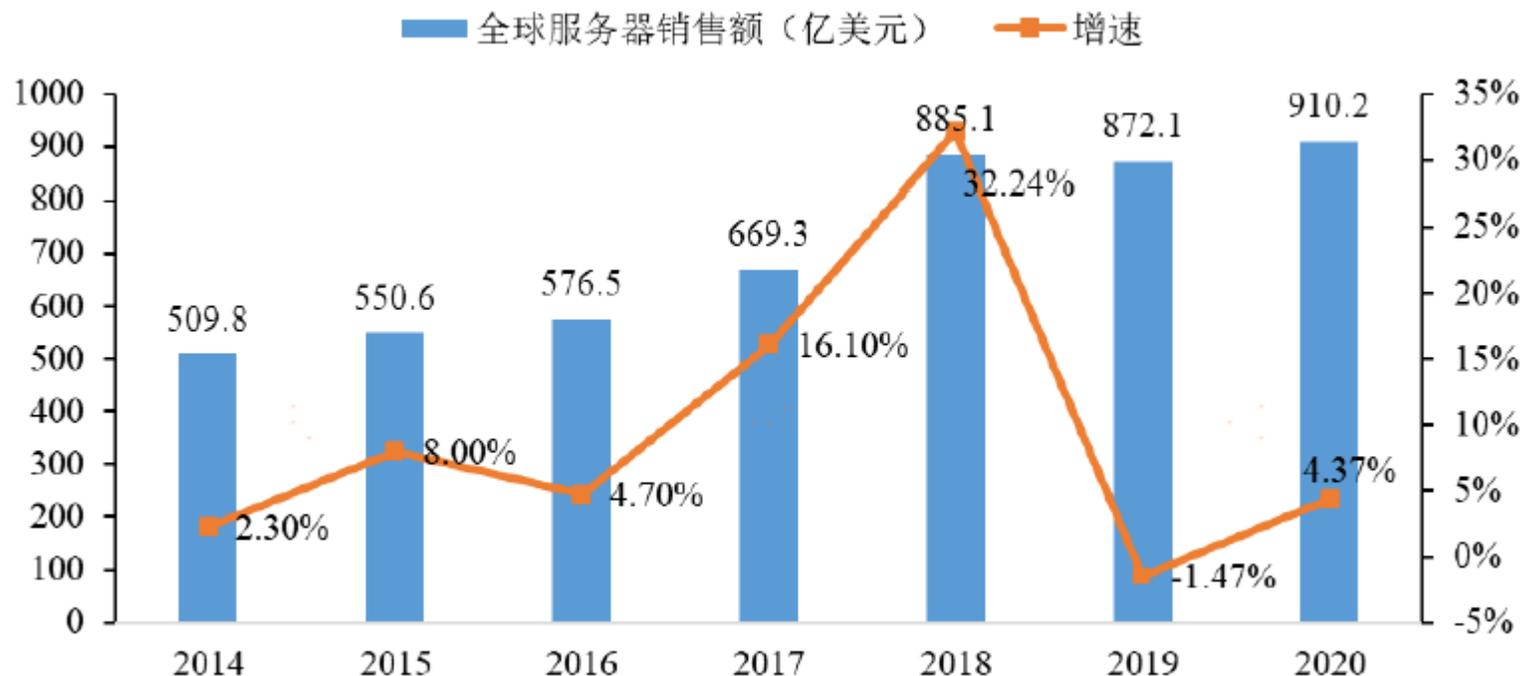
类型	主要性能指标	应用场景	技术特点
服务器	<ol style="list-style-type: none">1、单颗处理器核心数一般在 8 核~64 核，20 核以上居多2、支持多路互连，两路、四路、八路等3、可靠性、稳定性要求高，常年无故障运行4、高端内存，支持ECC 等可靠性要求5、功耗比较高，一般 100W 以上	<ol style="list-style-type: none">1、行业关键应用（电信、金融、教育、互联网等）2、政府国计民生关键应用（税务、电力、公安、社保等）	<ol style="list-style-type: none">1、微结构复杂、先进，制造工艺先进，核心数多，单核及多核性能皆优异2、指令集功能齐全3、片上集成缓存容量大4、内存通道数多5、I/O 带宽高6、支持多处理器一致性互连7、可靠性高，RAS 功能丰富8、TDP 功耗较高
计算机	<ol style="list-style-type: none">1、单颗处理器核心数一般在 10 核以下，4 核、8 核居多2、主要是单路形式3、可靠性、稳定性要求低4、低成本内存，可靠性要求相对较低，内存容量要求低5、功耗一般在100W 以下	<ol style="list-style-type: none">1、台式机2、笔记本电脑	<ol style="list-style-type: none">1、微结构复杂、先进，制造工艺先进2、性能与功耗较平衡3、指令集功能较齐全4、I/O 接口功能齐全5、内存通道数为1~2 个

资料来源：海光信息招股说明书

2.2、CPU与AI融合将成大势所趋

- 过去十多年，全球服务器市场总体保持了稳健的增长。根据 IDC 数据，2020 年，受全球互联网行业资本投入收缩和“新冠疫情”的影响，全球服务器出货量为1212.9万台，销售额 910.2 亿美元，同比分别增长 3.26%和 4.37%，增速低于前期平均水平。
- 随着云计算的不断发展，全球范围内云数据中心、超级数据中心的建设速度不断加快，Cisco 预计到 2021 年，计算能力更强的超级数据中心将达到 628 座，占数据中心总量的53%。

图10：2014-2020年全球服务器销售额



资料来源：海光信息招股说明书

2.2、CPU与AI融合将成大势所趋

- CPU 技术发展趋势：1，单个处理器核心性能持续提升；2，处理器设计复杂度提高，核心数逐步增加，I/O 性能持续提升；3，微体系结构持续优化。
- CPU 在商业领域的具体应用：1，“企业上云”需求强烈，云计算可助力企业完成数字化转型；2，5G 网络的快速铺开，加速“万物互联”时代提前到来。
- 国产 CPU 市场空间广阔：随着国产 CPU 性能的不断提高和软件生态的不断完善，基于国产 CPU 的信息产品已经得到批量应用。对信息安全、供应链安全要求相对较高的领域，均是国产 CPU 的优势市场，伴随着未来信息化的加速，桌面、服务器 CPU 的需求量有望大幅增加。
- 英特尔 Sapphire Rapids 第四代至强可扩展处理器深度支持 AI 运算。英特尔在这代产品中增添全新内置 AI 加速器——英特尔高级矩阵扩展（AMX）技术。英特尔透露该 GPU 的基础算力平均提升值为 53%，而在 AMX 的助推下，其在 PyTorch 上的 AI 实时推理速度，可提升至上一代产品（FP32）的 5.7-10 倍，训练性能提升最高也能提升到上一代产品的 10 倍。

图11：龙芯1H芯片示意图



资料来源：龙芯中科官网

图12：龙芯2K1000LA芯片示意图



资料来源：龙芯中科官网

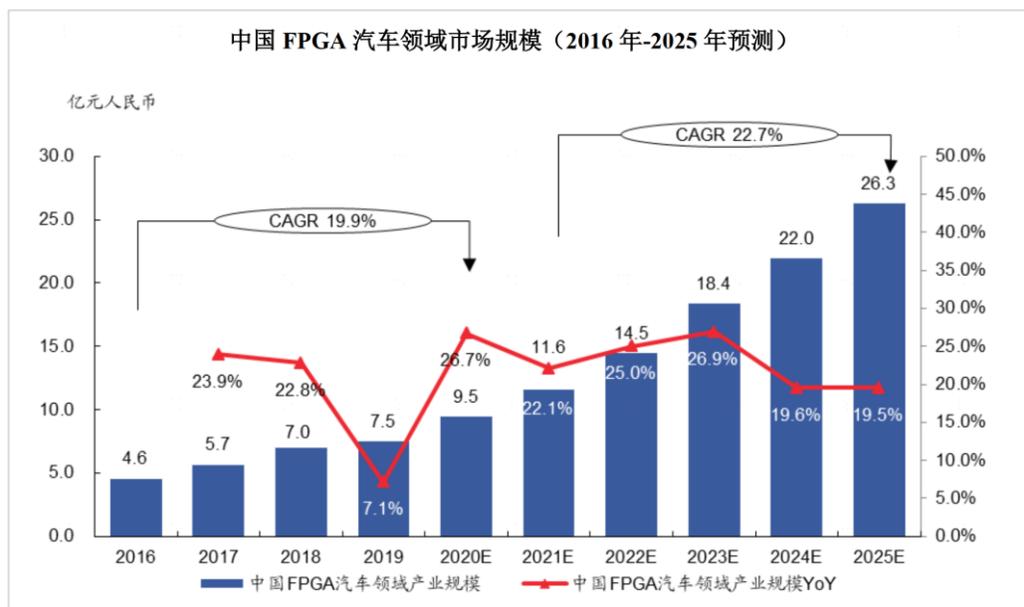
2.3、FPGA与ASIC技术融合，在人工智能领域应用广泛

- 现场可编程门阵列（Field-Programmable Gate Array, FPGA），是基于通用逻辑电路阵列的集成电路芯片，和ASIC芯片不同，其最大的特点是芯片的具体功能在制造完成以后由用户配置决定，用户可通过配套的FPGA专用EDA软件实现具体功能，首先由专用EDA软件接受用硬件语言描述的用户电路，其次编译生成二进制位流数据，最后将位流下载到芯片中实现用户所需的功能。
- FPGA芯片由可编程的逻辑单元（LogicCell, LC）、输入输出单元（InputOutputBlock, IO）和开关连线阵列（SwitchBox, SB）三个部分构成。逻辑单元通过数据查找表（look-uptable, LUT）中存放的二进制数据来实现不同的电路功能。LUT的本质是一种静态随机存取存储器（StaticRandomAccessMemory, SRAM），其大小是由输入端的信号数量决定的，常用的查找表电路是四输入查找表（4-inputLUT, LUT4）、五输入查找表（5-inputLUT, LUT5）和六输入查找表（6-inputLUT, LUT6）。

2.3、FPGA与ASIC技术融合，在人工智能领域应用广泛

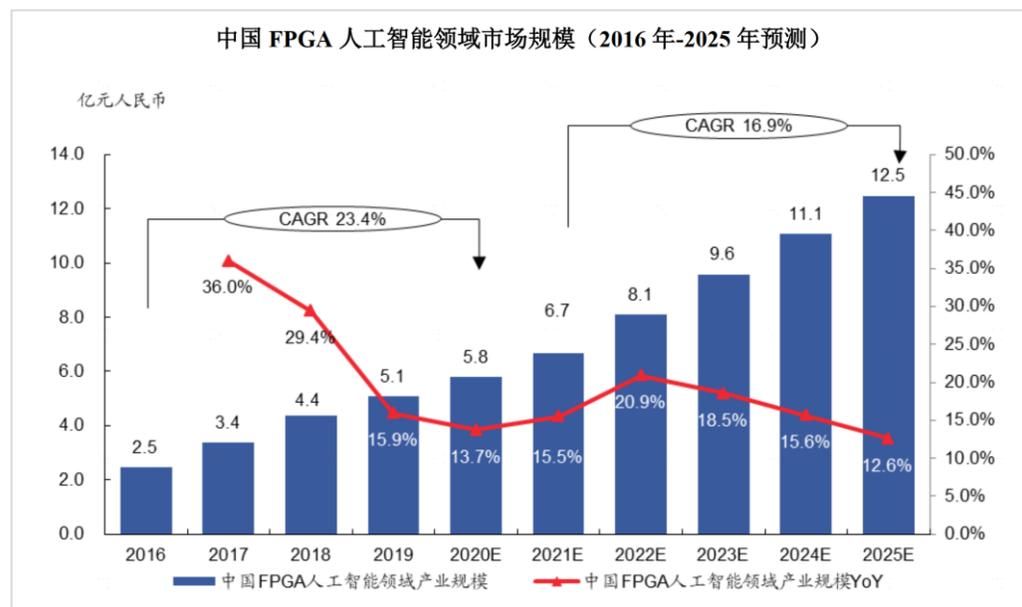
- FPGA在汽车电子中的应用：在系统接口及控制领域，可用于控制和驱动电动汽车电机控制系统，连接驾驶系统、仪表盘、雷达、超声波传感器等各种车载设备，实现激光雷达、毫米波雷达等信号处理和控制在；在视频桥接和融合领域，可用于实现多个图像传感器的信号桥接、3D环视视频融合、倒车辅助视频、辅助驾驶视频等功能；在辅助驾驶和自动驾驶领域，可用于实现机器视觉与目标检测等各种功能。
- 在云侧与端侧的不同任务中，FPGA芯片均已与GPU及ASIC等芯片一起成为人工智能处理芯片的重要选择之一，Frost&Sullivan数据显示人工智能领域FPGA芯片2020年中国销售额达到5.8亿元，占中国FPGA芯片市场份额的3.9%，2021年至2025年年均复合增长率将达到16.9%。

图13：2016-2025年中国FPGA汽车领域市场规模（亿元）



资料来源：Frost&Sullivan、光大证券研究所整理

图14：2016-2025年中国人工智能领域市场规模（亿元）



资料来源：Frost&Sullivan、光大证券研究所整理

2.3、FPGA与ASIC技术融合，在人工智能领域应用广泛

- **在人工智能领域：** AI大规模发展目前仍存在许多芯片层面的问题，AI的基础硬件需要满足低延迟、高性能、高可靠性、高性价比等要求，这些刚好是FPGA能够充分满足的。FPGA芯片已与GPU及ASIC等芯片一起成为人工智能处理芯片的重要选择之一。在AI市场，FPGA芯片可以作为加速卡加速人工智能算法的硬件级运算速度，高密度FPGA面向云侧并行运算需求，中低密度面向端侧推断任务转移。
- **云端：** FPGA流水线并行结构，可以实现更高的并发处理。FPGA可针对数据包步骤数量搭建同等数量流水线，数据包经多个流水线处理后可即时输出，无需像GPU一样数据单元一致输出，因此计算密集型任务（如图像处理、机器学习等）中，拥有流水线并行和数据并行的高密度FPGA效率更高。
- **端侧：** 当推断决策任务转至智能终端，中低密度FPGA迎来用武之地。随着智能终端对实时响应和多样化应用的需求，越来越多的推断任务被转移到端侧来完成。为实现推断任务的转移，通常对训练后的人工智能算法模型进行压缩，进而应用到推理环节，FPGA芯片可实现快速推断决策的特点也使其可广泛应用于该领域。
- **FPGA发展至今与ASIC技术融合、向系统级发展成为新的趋势**，软硬件相结合，兼具性能和灵活性，诞生SoC FPGA和eFPGA两种路径。与独立FPGA器件相比，eFPGA是一种相对较新的技术，其可被集成到ASIC，或提供FPGA晶粒进行多芯片封装（MCM）而集成到其中。eFPGA的典型应用包括汽车驾驶员辅助系统、计算性存储加速器、人工智能/机器学习和5G基础设施等。这些应用都使用了集成了eFPGA IP的定制ASIC器件，以提供工作负载和算法灵活性。

2.4、云、边、端需求拉动AI芯片市场规模高速增长

- 人工智能算法与应用必须以计算机硬件作为物理载体方能运转，其效果、效率与核心计算芯片的计算能力密切相关。
- 当前以深度学习为代表的人工智能技术对于底层芯片计算能力的需求一直在飞速增长，其增速已经大幅超过了摩尔定律的速度。
- 人工智能运算常常具有大运算量、高并发度、访存频繁的特点，且不同子领域（如视觉、语音与自然语言处理）所涉及的运算模式具有高度多样性，对于芯片的微架构、指令集、制造工艺甚至配套系统软件都提出了巨大的挑战。



2.4、云、边、端需求拉动AI芯片市场规模高速增长

图15：传统芯片与智能芯片比较

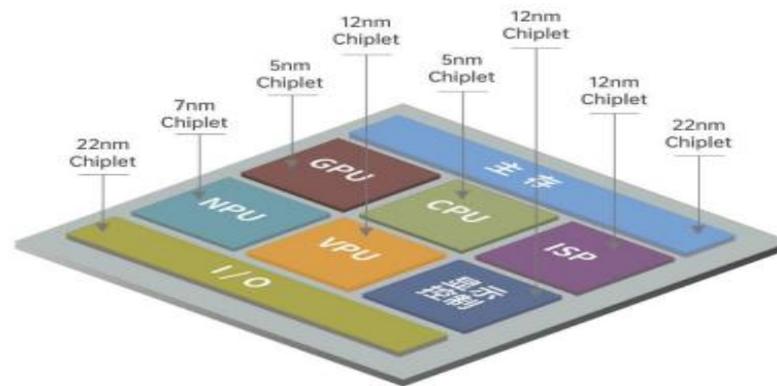
芯片类型		技术原理	技术发展情况与技术特点	技术优势与局限性	未来发展、演化或融合的趋势	在云边端训练与推理场景的市场渗透率
传统芯片	CPU	(1) 基本原理为：通过灵活的控制单元、细粒度的运算单元、多层次的缓存、多发射流水线，实现对于通用计算任务灵活和高效的支持 (2) 具体对于智能训练和推理应用，通过 CPU 的基本指令组合出训练或推理需要的运算操作，从而实现对智能算法的支持	技术成熟，通用性最强可执行各种类型的计算机应用程序，非常适合传统的控制密集型计算任务	人工智能应用开发生态成熟，但性能已无法满足人工智能快速增长的计算能力需求	CPU的演化趋势为集成更多更高的外部接口，长期看仍将主要应用于通用计算	广泛应用于个人电脑、移动终端、传统服务器等领域，在人工智能芯片市场渗透率相对较低
	GPU	(1) 基本原理为：通过简化控制单元并集成大规模的并行运算单元，实现对图形渲染等并行任务的良好支持 (2) 具体对于智能训练和推理应用，通过 GPU 的向量等指令组合出训练或推理需要的运算操作，从而实现对智能算法的支持	技术成熟，通用性较好，擅长数据级并行处理，为图形处理、科学计算等传统任务提供了良好的硬件支持	峰值运算性能高，但整体能耗较高；在云端具备成熟的应用开发生态，但在终端生态尚不成熟	GPU 的演化趋势为持续保持其在图形渲染和科学计算领域的技术优势，加强对人工智能领域的支持	在人工智能领域，GPU 多用于服务器与数据中心，是目前渗透率最高且最主流的芯片类型，在终端应用较少
	DSP	(1) 基本原理：通过面向数字信号处理的专用架构和专用指令集，实现对数字信号处理任务（如快速傅里叶变换）的高效支持 (2) DSP 一般仅用于智能终端的推理任务。在支撑推理任务时，DSP 利用其数字信号处理指令来实现智能推理任务中的运算操作，例如通过信号卷积指令组合来实现对于深度学习模型中的卷积层的支持	技术成熟，最初为数字信号处理任务设计，早期多用于传统的通信和音视频信号处理	在人工智能领域主要用于处理图像、语音等任务，编程开发门槛高	随着5G等通信技术的发展，DSP将持续保持在通信领域的广泛应用，在人工智能领域的应用前景尚不明朗	在云端应用较少，但在手机等终端设备中有一定生态基础
	FPGA	(1) 基本原理：通过集成大量的可重构逻辑单元阵列，可支持硬件架构的重构，从而灵活支持不同的人工智能模型 (2) FPGA 一般仅用于智能推理，其技术原理通常是将对应的智能算法通过硬件编程的方式实现在 FPGA 上，从而支持推理任务	技术成熟，包含充裕的可重构逻辑单元阵列，硬件编程开发和调试门槛较高	可通过硬件重构方式灵活实现适合于人工智能应用的架构，但成本和能效与主流技术路径仍有差距	FPGA 将与CPU、DSP 等传统类型处理器融合到同一颗 SoC 芯片中，实现多样化功能	在芯片原型验证与仿真中有着广泛应用，在人工智能芯片市场渗透率相对较低
智能芯片	通用型	(1) 基本原理：通过对各类智能应用和算法的计算和访存特点进行抽取和抽象，定义出一套适用于智能算法且相对灵活的指令集和处理器架构，从而广泛支持多样化的人工智能算法和应用 (2) 智能芯片的指令通常与人工智能算法中的关键运算操作相匹配 (3) 在具体的训练和推理应用中，对于关键运算操作，智能芯片指令可直接支持，从而实现高效的训练和推理	相关技术持续发展中，全新指令集完备高效，可覆盖各类智能算法所需的基本运算操作	性能、功耗比较传统芯片优势明显，可适应各种场景和规模的人工智能计算需求	云端智能芯片将集成更高计算能力、更高速的外围接口及更先进的集成电路工艺；边缘及终端智能芯片将集成多样化的模块，沿SoC技术路径继续深度发展	在云端、边缘端和消费类电子终端都开始出现广泛应用，渗透率将逐渐提升
	专用型 (ASIC)	基本原理：针对面向特定的、具体的、相对单一的人工智能应用专门设计的芯片，具体实现方法为在架构层面对特定智能算法作硬化支持，多用于推理任务	相关技术持续发展中，在架构层面对特定智能算法作硬化支持，指令集简单或指令完全固化	成本相对较低，软件栈相对简单，设计和生产周期短但通用性较差	专用型智能芯片逐渐融入各行业专用SoC芯片（如智能音箱芯片）中	常用于在低功耗、成本敏感的终端上支撑特定的智能应用，在云端、边缘端等场景渗透率相对较低

资料来源：寒武纪招股书、光大证券研究所整理

2.5、从AI Chip到AI Chiplet

- **Chiplet可以大幅提高大型芯片的良率。**近年来，随着高性能计算、AI等方面的巨大运算需求，集成更多功能单元和更大的片上存储使得芯片不仅晶体管数量暴增，芯片面积也急剧增大。芯片良率与芯片面积有关，随着芯片面积的增大而下降。Chiplet可将单一die面积做小以确保良率，并用高级封装技术把不同的芯粒集成在一起。
- **Chiplet有利于降低设计的复杂度和设计成本。**利用小芯片（具有相对低的面积开销）的低工艺和高良率可以获得有效降低成本开销。此外，研发成本也逐渐占据芯片成本的重要组成部分，通过采用已知合格裸片进行组合，可以有效缩短芯片的研发周期及节省研发投入。同时Chiplet芯片通常集成应用较为广泛和成熟的芯片裸片，可以有效降低Chiplet芯片的研制风险，从而减少重新流片及封装的次数，有效节省成本。
- **Chiplet有望降低芯片制造的成本。**SoC中具有不同计算单元，以及SRAM、I/O接口、模拟或数模混合元件等。除了逻辑计算单元以外，其他元件并不依赖先进制程也通常能够发挥很好的性能。所以，将SoC进行Chiplet化之后，不同的芯粒可以根据需要来选择合适的工艺制程分开制造，然后再通过先进封装技术进行组装，不需要全部都采用先进的制程在一块晶圆上进行一体化制造，这样可以极大的降低芯片的制造成本。

图16: Chiplet示意图

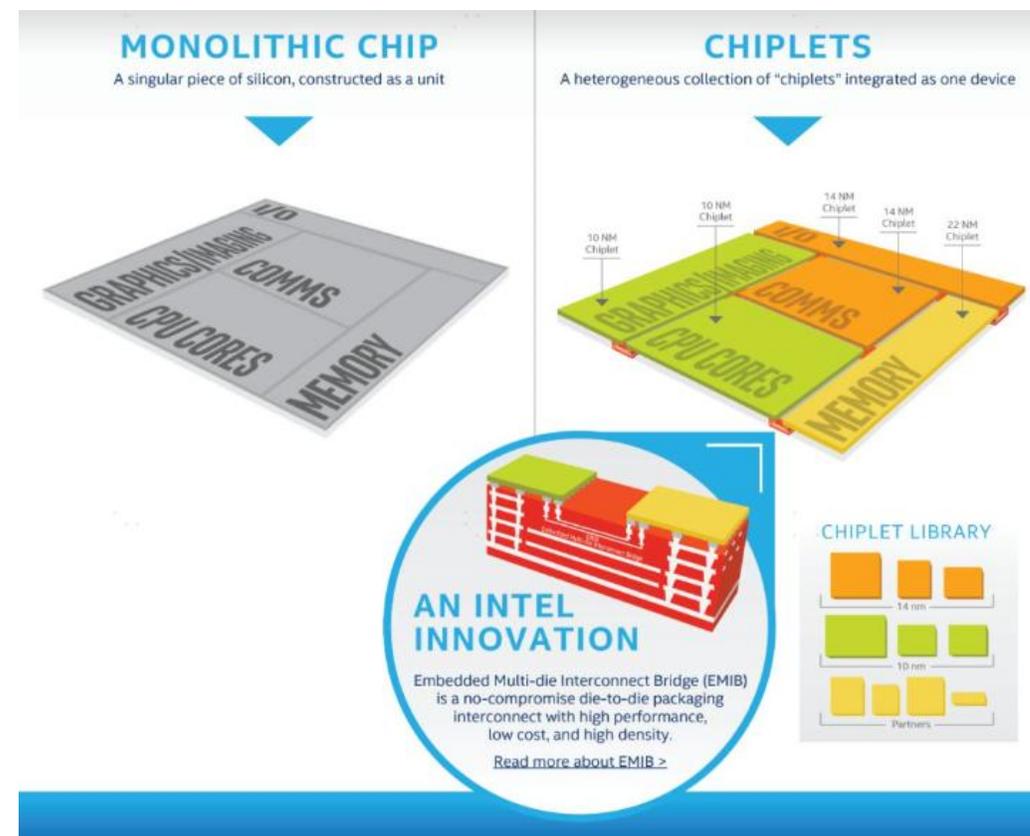


资料来源：芯原股份2021年报、光大证券研究所整理

AI Chiplet优势

- Chiplet模式一个系统里可集成多个工艺节点的硅片（如右图）：
 - 支持快速开发，根据需要选择不同的工艺节点，实现降本
 - 开发商可以以Chiplet IP的形式提供产品，和其他不同工艺的功能模块集成在一起，无需受限于Foundry工艺的进展
- 架构设计的灵活性：
 - 结合工艺灵活性，可以在架构设计中有更合理的功能/工艺权衡，有利于AI SoC或AIoT芯片更好适应场景需求
 - 系统架构的设计，特别是功能模块间的互联有更多优化空间
- 商业模式的灵活性：
 - 对于聚焦AI加速芯片厂商，chiplet可以提供一种新的产品形式，增加潜在市场，或拉长一代产品的生命周期
 - 对于垂直领域的做集成AI加速功能的SoC厂商，集成合适AI chiplet而不是IP可以大大节约项目开发时间

图17: Chiplets和Chip的不同



资料来源：壁仞科技研究院，Intel，光大证券研究所整理

2.6、数据流量与AI算力增长推动光模块需求

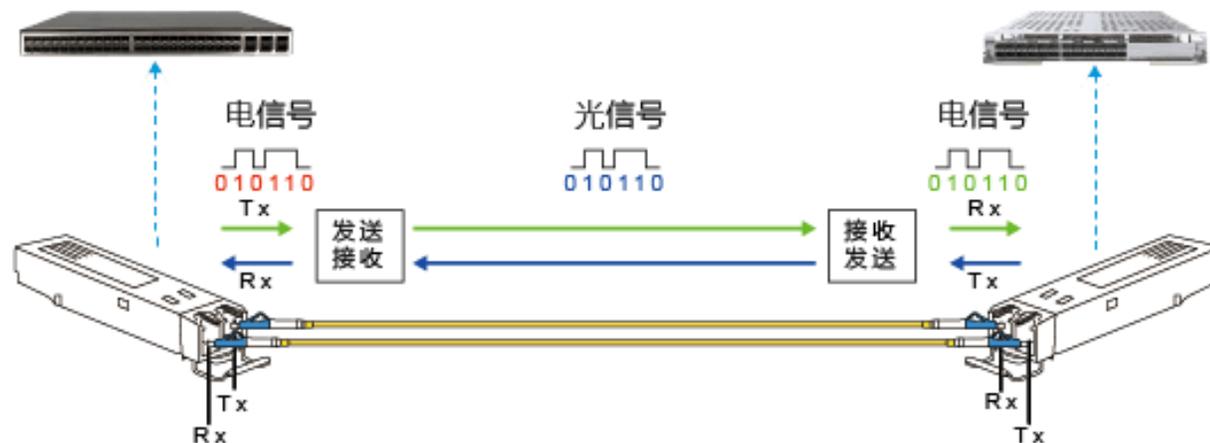
- 光模块是光通信系统的核心器件之一，由各种无源器件以及光电芯片组合封装。光模块构成了数据中心互连、5G承载网络 and 全光接入网络的基础单元，主要完成光电/电光转换功能。近年来随着速率的逐渐提升，其在系统设备中的成本占比不断攀升，已成为各应用领域高带宽、广覆盖、低成本和低能耗的关键要素。

图18：光模块与交换机的配合使用



资料来源：睿海光电官网

图19：光模块进行光电转换

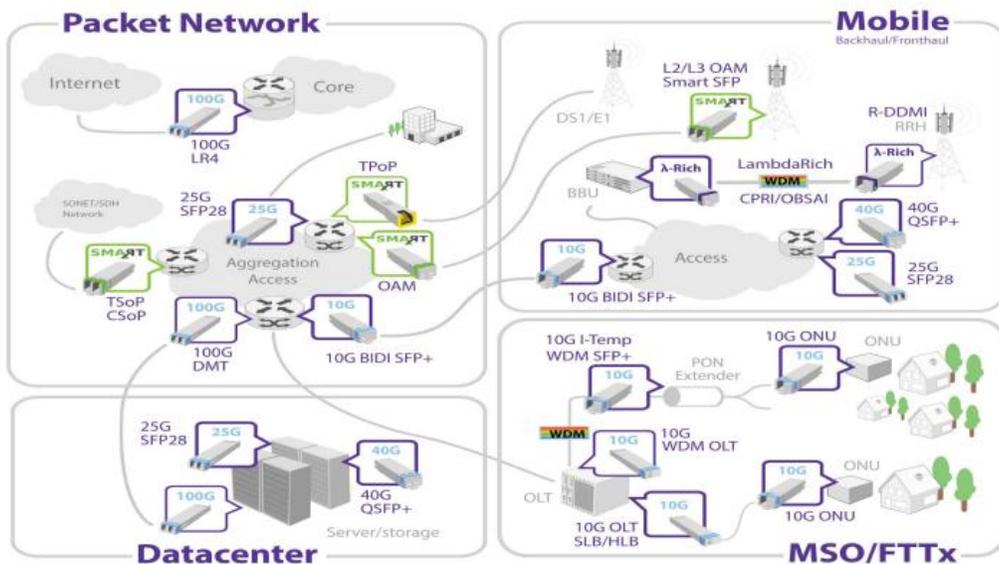


资料来源：CSDN

2.6、数据流量与AI算力增长推动光模块需求

- 随着5G、云计算、大数据等技术与应用的快速发展，全球数据流量持续增长，光通信与光网络需求得到了快速提升。光模块是光通信产业链中的关键器件，将充分受益于产业新趋势。

图20：光模块主要应用场景



资料来源：OE Solutions

2.7、算力需求带动服务器 PCB 市场持续扩容

- 根据中国信息通信研究院数据，2021 年，全球数字经济 GDP 占比已达到 46%，我国数字经济 GDP 占比达到 40%，预计 2025 年，我国数字经济 GDP 占比将达到 62%。数字经济产业的发展带来全球数据流量的迅速增长。根据 Cisco Global Index 数据显示，2016 年全球数据中心流量规模为 6.8ZB，到2021 年规模增长至 20.6ZB，复合增长率为 25%；根据 IDC 数据，预计 2025 年全球数据量达 175ZB。
- 随着全球数据流量的指数级增长以及全球信息化建设速度加快，服务器作为最重要的算力基础设施，全球范围内的出货量与市场规模得以大幅增长。根据Gartner 数据，2021 年全球服务器总体市场规模达到 6634 亿元，预计 2026 年达到 10600 亿元，复合增长率达 9.83%。
- PCB 是服务器的重要组成部分，是承载服务器运行的关键材料。服务器出货量的大幅增长也使得服务器 PCB 市场规模扩容，成为 PCB 市场中复合增长率最快的下游细分市场。根据 Prismark 数据，2021 年全球服务器领域 PCB 市场规模为 78.04 亿美元，预计 2026 年达到 132.94 亿美元，复合增长率为 11.2%。

ChatGPT 驱动AI浪潮趋势渐起，芯片等AI硬件构成核心底座

云端算力硬件将成为AI新浪潮的关键基础设施

AI新浪潮未来将提升海量IoT设备的边缘算力需求

投资建议：云端算力和边缘算力两个维度关注AI硬件投资机会

风险分析

三、AI新浪潮未来将提升海量IoT设备的边缘算力需求

- 1、NPU将成为人工智能的重要趋势
- 2、RISC-V芯片架构：深度受益于未来AIoT趋势

3.1、NPU将成为人工智能的重要趋势

- NPU IP是指专用于加速神经网络运算、机器视觉和机器学习等人工智能应用的数字IP。以芯原股份与ARM公司为代表的IP公司可以为SoC与MCU公司提供NPU IP，在SoC与MCU中加入边缘算力。
- 以芯原股份的神经网络处理器技术为例，其包括自主可控的卷积神经网络加速、可编程的浮点运算加速、指令集和可编程的浮点运算专用编译器、优化器等工具设计，支持国际标准OpenVX1.2和OpenCL1.2EP/FP；支持最大32位浮点精度数据处理和张量处理的硬件加速；支持0.5TOPs到6TOPs性能的单卷积运算核的可扩展架构设计，多卷积运算核扩展后的NPUIP运算能力可达10TOPs。

图21：NPU的核心技术与应用



资料来源：芯原股份招股说明书

3.1、NPU将成为人工智能的重要趋势

- 当前的边缘侧NPU AI算力可以覆盖从汽车电子到手机再到可穿戴设备的各类设备。
- 从应用的归类来看，机器视觉和语音识别是NPU在边缘侧的两大重要方向。

图22：芯原股份NPU产品系列以及应用方向



资料来源：芯原股份招股说明书

3.1、NPU将成为人工智能的重要趋势

- NPU在SoC中的典型应用为机器视觉。以瑞芯微的新一代机器视觉方案RV1106及RV1103为例，两颗芯片在NPU、ISP、视频编码、音频处理等性能均有显著升级，集成度与性价比较高，可以在低待机功耗的同时提供优秀的边缘AI算力。瑞芯微高性能的RV1126具备四核ARM Cortex-A7和RISC-V MCU，2.0Tops NPU，活体检测率可以高达98.48%。
- 晶晨股份A311D在采用高性能A73内核的同时，搭载了5TOPS的高性能NPU，可以广泛使用于各类中高端AIOT设备。

图23: 瑞芯微RV1126



资料来源：瑞芯微官网

图24: 晶晨股份A311D



CPU 四核 ARM Cortex-A73 +双核 ARM Cortex-A53

GPU ARM Mali-G52 MP4 (6EE)
OpenGL ES 3.2, Vulkan 1.1和OpenCL 2.0

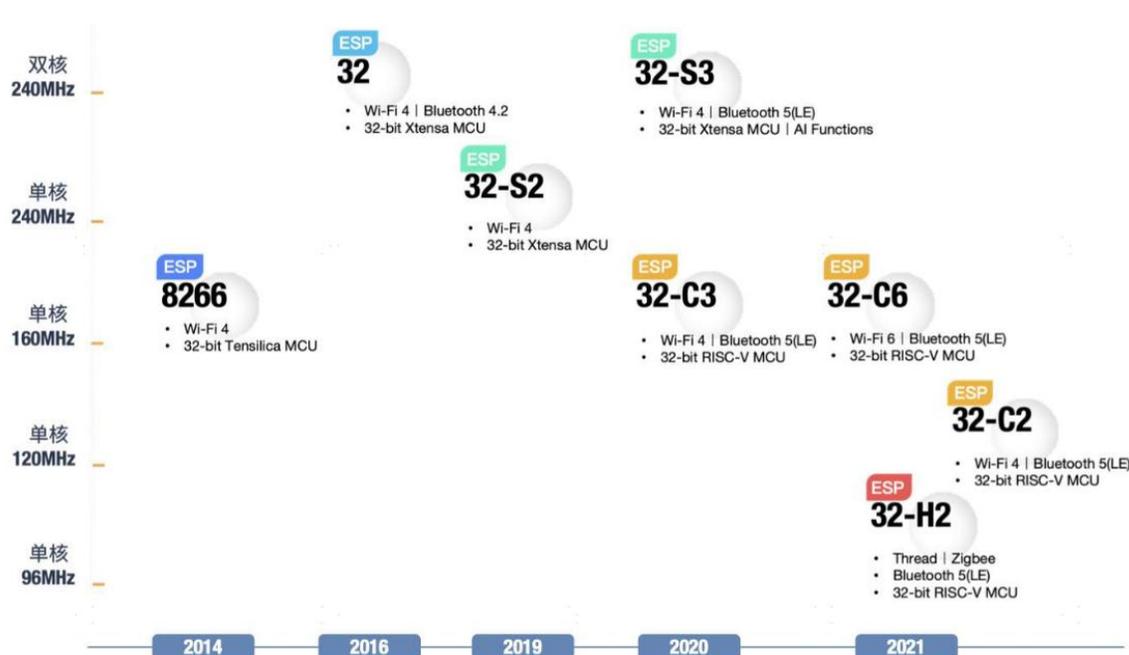
NPU 5 TOPS

资料来源：晶晨股份官网

3.1、NPU将成为人工智能的重要趋势

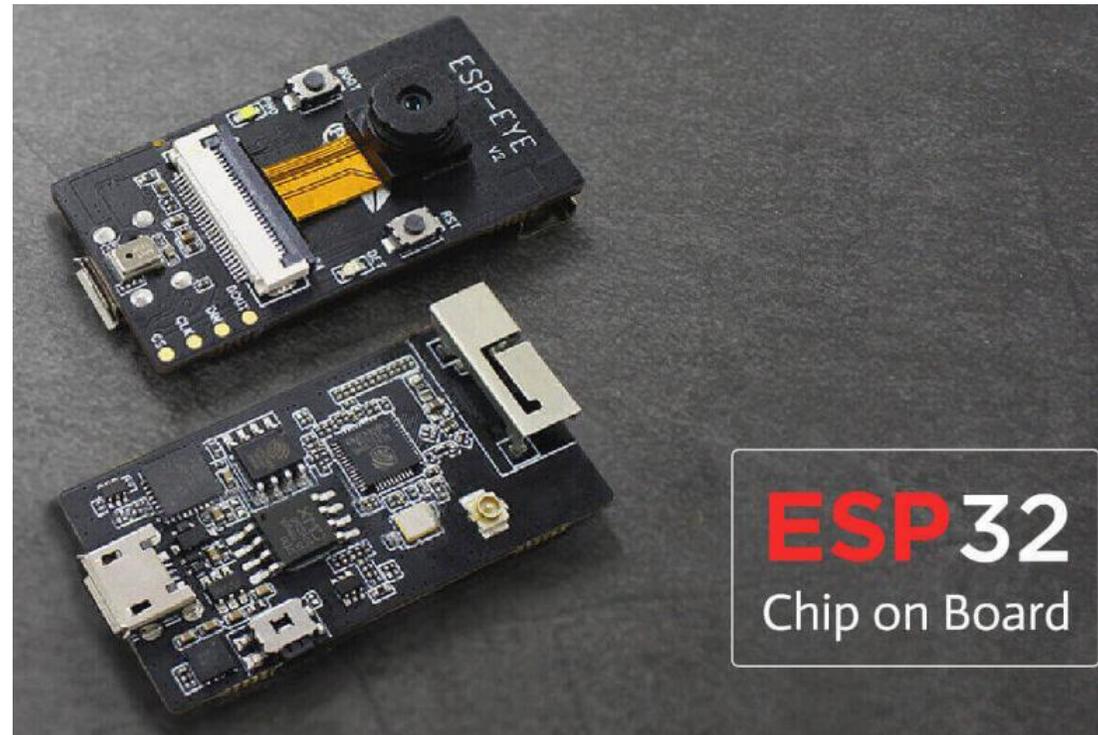
- 乐鑫科技在其高端产品ESP32-S3 MCU中增加了用于加速神经网络计算和信号处理等工作的向量指令 (vector instructions)。AI 开发者们通过 ESP-DSP 和 ESP-NN 库使用这些向量指令，可以实现高性能的图像识别、语音唤醒和识别等应用。ESP-WHO 和 ESP-Skainet 也将支持此功能。
- ESP32-S3-EYE 是一款搭载 ESP32-S3 的小型人工智能开发板。它集成了一个 200 万像素的摄像头、一个 LCD 显示屏和一个麦克风，适用于图像识别和音频处理等应用。

图25：乐鑫科技产品线



资料来源：乐鑫科技年报

图26：ESP 32 EYE开发板



资料来源：乐鑫科技官网

3.2、RISC-V芯片架构：深度受益于未来AIoT趋势

- RISC-V是一个基于精简指令集（RISC）原则设计的开源指令集架构，秉承简单有效的设计哲学，具备开放、简洁、模块化、可扩展的技术优势。
- RISC-V指令集可以自由地用于任何目的，允许任何人任何企业设计、制造、销售RISC-V芯片和软件。RISC-V能满足从微控制器到超级计算机等各种尺寸的处理器，支持从FPGA到ASIC等各种实现，能高效地实现各种微结构，支持大量的定制与加速功能，能和现有软件栈与编程语言很好的适配。RISC-V技术标准的维护和推广由总部位于瑞士的RISC-V国际基金会持续负责，以保证RISC-V的开放和中立，技术供应稳定安全。

图27：RISC-V特点



3.2、RISC-V芯片架构：深度受益于未来AIoT趋势

- ARM架构和RISC-V架构都源自1980年代的精简指令计算机RISC，区别在于RISC-V推崇的大道至简的技术风格和彻底开放的技术模式。
- ARM是一种封闭的指令集架构，应用ARM架构的厂商只能根据自身需求调整产品频率和功耗，不得改变原有设计。经过几十年的发展演变，CPU架构变得极为复杂和冗繁，ARM架构文档长达数千页，指令数目复杂、版本众多，且彼此之间既不兼容，也不支持模块化，并且存在着高昂的专利和架构授权问题。
- 相反，RISC-V在设计之初就定位为完全开源的架构，架构文档只有二百多页，基本指令数目仅40多条，同时一套指令集支持所有架构，模块化使得用户可以根据需求自由定制不同的指令子集。

图28：RISC-V特点

项目	X86 架构	ARM 架构	RISC-V 指令集架构
开发厂商	英特尔	ARM Holdings PLC	加州大学伯克利分校开发，目前由 RISC-V 基金会负责保护和推广
发布时间	1978 年	1985 年 (ARM1 Sample)	2010 年
指令集架构类型	CISC (复杂指令集架构)	RISC (精简指令集架构)	RISC (精简指令集架构)
架构文档篇幅	数千页	数千页	不足 300 页
指令集数目	指令集数目繁多	指令集数目繁多	基本指令集 40 多条，合计低于 300 条
模块化设计	不支持	不支持	支持
可扩展性	不支持	不支持	支持
版本兼容性	不同模块不兼容	不同的版本架构不兼容	兼容性较高
技术自主	否	否	是
技术可控	否	是	是
授权费收取情况	目前仅英特尔、AMD、VIA (台湾威盛) 三家拥有授权，不对外授权	一次性授权费、与产品销量挂钩的版税提成 Royalty、技术咨询费用	免费开源
主要应用领域	个人计算机、服务器、超级计算机	移动智能终端、物联网	物联网

3.2、RISC-V芯片架构：深度受益于未来AIoT趋势

- 计算“边缘化”趋势将更多AI和计算能力赋予边缘设备，为SoC设计公司提供更多机会的同时也提出了更高的PPA要求（PPA是Performance（性能）、Power（功耗）、Area（尺寸）三者的缩写）。
- 此外，计算架构“开放”激发开源硬件创新，RISC-V掀起了开源硬件和开放芯片设计的热潮，现已得到全球很多大中企业、科研机构 and 初创公司的支持，围绕RISC-V成长起来的生态和社区也发展迅猛，从基础RISC-V ISA、内核IP到开发环境和软件工具，都在推动RISC-V生态的进一步扩大。

图29：AI浪潮下RISC-V芯片应用领域



ChatGPT 驱动AI浪潮趋势渐起，芯片等AI硬件构成核心底座

云端算力硬件将成为AI新浪潮的关键基础设施

AI新浪潮未来将提升海量IoT设备的边缘算力需求

投资建议：云端算力和边缘算力两个维度关注AI硬件投资机会

风险分析

4、投资建议

- 云端算力：建议关注（1）GPU：景嘉微、海光信息；（2）CPU：龙芯中科、中国长城、澜起科技、海光信息；（3）AI芯片：寒武纪、澜起科技；（4）FPGA：复旦微电、安路科技、紫光国微等；（5）光模块：天孚通信、德科立、新易盛、光迅科技、博创科技、源杰科技、中际旭创等；（6）CHIPLET：兴森科技、长川科技、方邦股份、华正新材、通富微电、长电科技等。
- 边缘算力：建议关注（1）SoC：晶晨股份、富瀚微、瑞芯微等；（2）AIoT：乐鑫科技、恒玄科技等；（3）RISC-V：中科蓝讯、乐鑫科技、全志科技等。

图30：行业重点上市公司盈利预测与估值

公司代码	公司名称	市值	归母净利润（亿元）				PE (X)				PS(X)			
		(亿元)	21	22E	23E	24E	21	22E	23E	24E	21	22E	23E	24E
603160.SH	汇顶科技	260	8.6	3.2	6.9	10.2	30	82	37	25	5	6	5	4
300474.SZ	景嘉微	385	2.0	2.9	4.2	5.8	132	133	91	67	35	34	22	16
688041.SH	海光信息	1,221	3.3	9.5	15.6	23.2	373	128	78	53	53	24	16	11
688047.SH	龙芯中科	473	2.4	1.7	3.1	4.8	200	274	152	99	39	45	26	18
000066.SZ	中国长城	428	6.0	3.5	6.1	10.5	72	121	70	41	2	2	2	1
688008.SH	澜起科技	714	8.3	13.4	19.0	27.4	86	53	38	26	28	18	12	9
688256.SH	寒武纪-U	351	-8.2	-11.2	-7.1	-3.8	N/A	N/A	N/A	N/A	49	41	29	21
688385.SH	复旦微电	458	5.1	10.9	14.9	18.8	89	42	31	24	18	13	10	8
688107.SH	安路科技-U	286	-0.3	0.4	0.9	1.9	N/A	652	330	150	42	26	17	12
002049.SZ	紫光国微	1,024	19.5	29.2	40.4	54.7	52	35	25	19	19	14	10	8
300394.SZ	天孚通信	165	3.1	5.1	6.4	-	54	32	26	-	16	10	8	-
688205.SH	德科立	59	1.3	1.1	1.6	2.2	46	53	36	27	8	8	6	5

资料来源：Wind，光大证券研究所整理；注：归母净利润预测为Wind一致预期，股价时间为2023年2月13日

4、投资建议

图31：行业重点上市公司盈利预测与估值（续）

公司代码	公司名称	市值	归母净利润（亿元）				PE (X)				PS(X)			
		(亿元)	21	22E	23E	24E	21	22E	23E	24E	21	22E	23E	24E
300502.SZ	新易盛	171	6.6	9.4	10.3	12.4	26	18	17	14	6	5	4	3
002281.SZ	光迅科技	163	5.7	6.4	7.4	8.4	29	25	22	20	3	2	2	2
300548.SZ	博创科技	59	1.6	2.0	2.7	3.5	37	29	22	17	5	4	3	2
688498.SH	源杰科技	93	1.0	1.1	1.5	2.1	97	86	61	44	40	32	24	18
300308.SZ	中际旭创	288	8.8	12.0	14.6	17.4	33	24	20	17	4	3	2	2
002436.SZ	兴森科技	202	6.2	6.3	8.1	10.4	32	32	25	19	4	4	3	2
300604.SZ	长川科技	280	2.2	5.4	8.5	11.4	128	52	33	25	18	10	7	5
688020.SH	方邦股份	57	0.4	-0.7	0.4	2.8	163	N/A	129	21	20	18	10	4
603186.SH	华正新材	43	2.4	0.5	1.2	2.4	18	80	36	18	1	1	1	1
002156.SZ	通富微电	320	9.6	9.6	13.3	16.9	33	33	24	19	2	2	1	1
600584.SH	长电科技	502	29.6	32.5	36.2	42.7	17	15	14	12	2	1	1	1
688099.SH	晶晨股份	334	8.1	9.3	12.6	17.2	41	36	27	19	7	6	4	3
300613.SZ	富瀚微	155	3.6	4.9	6.4	8.2	43	32	24	19	9	6	5	4
603893.SH	瑞芯微	370	6.0	4.7	7.6	11.1	62	78	49	33	14	15	10	7
688018.SH	乐鑫科技	98	2.0	1.6	2.5	3.5	49	61	40	28	7	7	5	4
688608.SH	恒玄科技	165	4.1	2.0	3.8	5.6	40	82	43	29	9	10	7	5
688332.SH	中科蓝讯	78	2.3	2.3	2.8	3.3	34	34	28	24	7	6	5	4
688018.SH	乐鑫科技	98	2.0	1.6	2.5	3.5	49	61	40	28	7	7	5	4
300458.SZ	全志科技	157	4.9	2.3	3.4	4.3	32	69	46	36	8	10	8	7
301099.SZ	雅创电子	61	0.9	1.7	2.6	3.6	66	37	23	17	4	3	2	2
688381.SH	帝奥微	119	1.7	2.1	3.0	4.4	72	57	39	27	24	21	13	10
688061.SH	灿瑞科技	73	1.3	2.0	3.0	4.2	58	36	24	17	14	9	7	5

资料来源：Wind，光大证券研究所整理；注：归母净利润预测为Wind一致预期，股价时间为2023年2月13日

AI浪潮趋势渐起，ChatGPT应用空间广阔

云端算力硬件将成为AI新浪潮的关键基础设施

AI新浪潮未来将提升海量IoT设备的边缘算力需求

投资建议：云端算力和边缘算力两个维度关注AI硬件投资机会

风险分析

市场竞争风险

随着下游产业的发展和市场需求的不断增加，AI各个细分领域的市场竞争愈发激烈。在通用处理器领域，Intel、AMD的CPU产品在全球市场中占据绝对优势地位；在协处理器领域，英伟达、AMD的GPGPU产品占据绝对优势地位。总体上国内各厂商的市场份额与国际龙头企业相比差距较。

行业景气度持续低迷

国际政治冲突、新冠疫情、极端气候等因素影响下，全球经济复苏承压，造成下游需求疲软，从而会影响行业景气度。若行业景气度持续低迷，则会对行业内公司业绩产生负面影响。

研发不及预期风险

本行业重视人才的经验积累，产品研发周期较长，需要厂商持续高研发投入，若新品研发不及预期，可能影响后续相关公司成长增速。

衷心 感谢

光大证券研究所



通信电子研究团队

刘凯

📄 执业证书编号: S0930517100002
☎ 电话: 021-52523849
✉ 邮件: kailiu@ebscn.com

杨德珩

📄 执业证书编号: S0930522110003
☎ 电话: 021-52523805
✉ 邮件: yangdh@ebscn.com

蔡微未

📄 执业证书编号: S0930522040001
☎ 电话: 021-52523818
✉ 邮件: caiweiwei@ebscn.com

林仕霄

📄 执业证书编号: S0930522090003
☎ 电话: 021-52523818
✉ 邮件: linshixiao@ebscn.com

孙啸

📄 联系人
☎ 电话: 021-52523587
✉ 邮件: sunxiao@ebscn.com

石崎良

📄 执业证书编号: S0930518070005
☎ 电话: 021-52523856
✉ 邮件: shiql@ebscn.com

于文龙

📄 执业证书编号: S0930522100002
☎ 电话: 021-52523587
✉ 邮件: yuwenlong@ebscn.com

何昊

📄 执业证书编号: S0930518070005
☎ 电话: 021-52523869
✉ 邮件: hehao1@ebscn.com

朱宇澍

📄 执业证书编号: S0930522050001
☎ 电话: 021-52523805
✉ 邮件: zhuyushu@ebscn.com

王之含

📄 联系人
☎ 电话: 021-52523818
✉ 邮件: wangzhihan@ebscn.com

分析师声明

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，并对本报告的内容和观点负责。负责准备以及撰写本报告的所有研究人员在此保证，本研究报告中任何关于发行商或证券所发表的观点均如实反映研究人员的个人观点。研究人员获取报酬的评判因素包括研究的质量和准确性、客户反馈、竞争性因素以及光大证券股份有限公司的整体收益。所有研究人员保证他们报酬的任何一部分不曾与，不与，也将不会与本报告中具体的推荐意见或观点有直接或间接的联系。

行业及公司评级体系

买入—未来6-12个月的投资收益率领先市场基准指数15%以上；
增持—未来6-12个月的投资收益率领先市场基准指数5%至15%；
中性—未来6-12个月的投资收益率与市场基准指数的变动幅度相差-5%至5%；
减持—未来6-12个月的投资收益率落后市场基准指数5%至15%；
卖出—未来6-12个月的投资收益率落后市场基准指数15%以上；
无评级—因无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使无法给出明确的投资评级。

基准指数说明：A股主板基准为沪深300指数；中小盘基准为中小板指；创业板基准为创业板指；新三板基准为新三板指数；港股基准指数为恒生指数。

特别声明

光大证券股份有限公司（以下简称“本公司”）创建于1996年，系由中国光大（集团）总公司投资控股的全国性综合类股份制证券公司，是中国证监会批准的首批三家创新试点公司之一。根据中国证监会核发的经营证券期货业务许可，本公司的经营范围包括证券投资咨询业务。

本公司经营范围：证券经纪；证券投资咨询；与证券交易、证券投资活动有关的财务顾问；证券承销与保荐；证券自营；为期货公司提供中间介绍业务；证券投资基金代销；融资融券业务；中国证监会批准的其他业务。此外，本公司还通过全资或控股子公司开展资产管理、直接投资、期货、基金管理以及香港证券业务。

本报告由光大证券股份有限公司研究所（以下简称“光大证券研究所”）编写，以合法获得的我们相信为可靠、准确、完整的信息为基础，但不保证我们所获得的原始信息以及报告所载信息之准确性和完整性。光大证券研究所可能将不时补充、修订或更新有关信息，但不保证及时发布该等更新。

本报告中的资料、意见、预测均反映报告初次发布时光大证券研究所的判断，可能需随时进行调整且不予通知。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。客户应自主作出投资决策并自行承担投资风险。本报告中的信息或所表述的意见并未考虑到个别投资者的具体投资目的、财务状况以及特定需求。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。对依据或者使用本报告所造成的一切后果，本公司及作者均不承担任何法律责任。

不同时期，本公司可能会撰写并发布与本报告所载信息、建议及预测不一致的报告。本公司的销售人员、交易人员和其他专业人员可能会向客户提供与本报告中观点不同的口头或书面评论或交易策略。本公司的资产管理子公司、自营部门以及其他投资业务板块可能会独立做出与本报告的意见或建议不相一致的投资决策。本公司提醒投资者注意并理解投资证券及投资产品存在的风险，在做出投资决策前，建议投资者务必向专业人士咨询并谨慎抉择。

在法律允许的情况下，本公司及其附属机构可能持有报告中提及的公司所发行证券的头寸并进行交易，也可能为这些公司提供或正在争取提供投资银行、财务顾问或金融产品等相关服务。投资者应当充分考虑本公司及本公司附属机构就报告内容可能存在的利益冲突，勿将本报告作为投资决策的唯一信赖依据。

本报告根据中华人民共和国法律在中华人民共和国境内分发，仅向特定客户传送。本报告的版权仅归本公司所有，未经书面许可，任何机构和个人不得以任何形式、任何目的进行翻版、复制、转载、刊登、发表、篡改或引用。如因侵权行为给本公司造成任何直接或间接的损失，本公司保留追究一切法律责任的权利。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。