

策略对话电子：AI 服务器需求牵引

——“AI 奇点时刻”系列（三）

报告摘要：

- **引言：**我们自 23.3.2 持续七篇报告推荐数字经济机遇，判断 AI 的“奇点时刻”已经到来，将贯穿 23 年投资机遇。广发电子组自 23.3.22 连续十一篇报告全面解读“AI 的 iPhone 时刻”，ChatGPT 带来 AIGC 奇点，驱动 AI 服务器、高性能计算（HPC）算力需求提升。本篇我们联合广发电子组，从三个视角判断本轮“AI+”带来的产业机遇，以问答形式全面解析 AI 主线中的电子行业投资机会。
- **策略：“AI+”奇点突破，如何判断产业机遇？**
 - （1）历史对比视角：类比 70 年代信息技术革命、99 年美股科网泡沫、13-15 年创业板“移动互联网+”，“技术+场景”革命催化；
 - （2）借鉴 13 年“移动互联网+”行情，我们认为本轮“AI+”行情也将分三阶段走：上游算力→中游软件服务→下游应用。当前行情仍在第一阶段，我们重点首推“上游算力”（AI 芯片/服务器交换机/光模块光芯片/液冷温控/互联网）。
- **对话电子：六问六答，“AI 的 iPhone 时刻”哪些环节受益？**

Q1: ChatGPT 产业链上下游电子行业的分布情况怎么样？看好哪些细分领域？（AI 服务器及相关电子产品/终端应用侧电子厂商）。

Q2: ChatGPT 拉动 AI 服务器需求提升，国内哪些领域和厂商直接受益？（服务器硬件/芯片/终端设备）。

Q3: 微软宣布 ChatGPT 支持第三方插件接入，支撑以后更多的类 ChatGPT 模型算力肯定需要芯片，具体有哪些 AI 芯片更关键更值得关注？（高性能 GPGPU）。

Q4: 英伟达 GTC 大会公布了不少新的产品，结合他们新品的发布情况，在硬件端和软件端有哪些投资机会值得关注？（高带宽显存突破内存带宽瓶颈/内存接口芯片量价齐升/高速率光模块加速渗透/PCB 产品量价齐升/多相电源产品单机价值增量显著）。

Q5: AIGC 产业链从成本的角度来看，还有哪些成本要素值得我们跟踪？（GPU 性能不断提升，有望优化 GPU 硬件成本）。

Q6: 怎么看待消费电子应用层上的商业化，建议关注哪些落地场景？（视觉领域，低成本、高效率地满足 VR/AR 用户海量内容需求/听觉领域，创造智能音箱交互新范式/驱动泛安防产业智能化升级）。
- **风险提示：**宏观经济下行压力超预期；全球疫情仍存在反复的风险；全球经济修复可能不及预期；全球/中国通胀高位流动性可能边际收紧；中美贸易/金融领域的关系仍有不确定性。AIGC 发展不及预期；AI 服务器出货量不及预期；国产厂商技术和产品进展不及预期。

分析师：戴康



SAC 执证号：S0260517120004

SFC CE No. BOA313



021-38003560



daikang@gf.com.cn

分析师：许兴军



SAC 执证号：S0260514050002



021-38003661



xuxingjun@gf.com.cn

分析师：耿正



SAC 执证号：S0260520090002



021-38003660



gengzheng@gf.com.cn

分析师：倪康



SAC 执证号：S0260519070001



021-38003561



nigeng@gf.com.cn

请注意，许兴军、耿正、倪康并非香港证券及期货事务监察委员会的注册持牌人，不可在香港从事受监管活动。

目录索引

一、策略：“AI+”奇点突破，如何判断产业机遇？	4
（一）本轮 “AI+”判断：类比 13 年 “移动互联网+”，首推上游算力	4
（二）AIGC 产业链：电子行业涉及硬件、软件、应用	5
（三）风险提示	6
二、对话电子：六问六答，“AI 的 IPHONE 时刻”哪些环节受益？	7
（一）Q1：CHATGPT 产业链上下游电子行业的分布情况怎么样？看好哪些细分领域？	7
（二）Q2：CHATGPT 拉动 AI 服务器需求提升，国内哪些领域和厂商直接受益？ ..	8
（三）Q3：微软宣布 CHATGPT 支持第三方插件接入，支撑以后更多的类 CHATGPT 模型算力肯定需要芯片，具体有哪些 AI 芯片更关键更值得关注？	8
（四）Q4：英伟达 GTC 大会公布了不少新的产品，结合他们新品的发布情况，在硬件端和软件端有哪些投资机会值得关注？	12
（五）Q5：AIGC 产业链从成本的角度来看，还有哪些成本要素值得我们跟踪？ ..	16
（六）Q6：怎么看待消费电子应用层上的商业化，建议关注哪些落地场景？	17
（七）风险提示	18

图表索引

图 1: 美科网泡沫酝酿期的美债利率	5
图 2: 13-15 年创业板-沪深 300 业绩向上	5
图 3: AIGC 产业链图谱全览	5
图 4: DGX A100 系统硬件组成示意图	7
图 5: 更多应用场景软件均有布局/引入 LLM 模型	8
图 6: AIGC 搭建 3D 景观	8
图 7: 不同类型服务器硬件占比	9
图 8: 中国 GPGPU 芯片板卡市场规模	9
图 9: 训练侧算力需求对 AI 服务器出货量的拉动	11
图 10: 推理侧算力需求对 AI 服务器出货量的拉动	11
图 11: HBM 结构示意图	12
图 12: 全球光模块市场规模及预测	13
图 13: 光模块结构示意图 (SFP+封装)	13
图 14: 源杰科技主要产品的开发及演变情况	14
图 15: 澜起科技 PCIe Retimer 产品在服务器中的应用场景	14
图 16: 按产品类型划分的全球 PCB 市场	15
图 17: 服务器用 PCB 市场规模以及增速预测	15
图 18: Nvidia V100 采用 16 相供电解决方案	16
图 19: NVIDIA H100 相对 A100 的训练性能提升	17
图 20: NVIDIA H100 相对 A100 的性能提升	17
图 21: VR/AR 市场规模	17
图 22: 苹果 MR 头显四大亮点	17
图 23: 阿里发布“鸟鸟分鸟”	18
图 24: 家庭智能设备市场规模	18
图 25: 2021-2024 年内计划部署人工智能场景企业应用率	18
图 26: 海康威视 AI 一站式服务	18
表 1: 不同精度数据类型应用场景	9
表 2: NVIDIA A100/H100 GPU 参数	10
表 3: 海光信息 DCU 产品与国际厂商高端 GPU 产品对比	11

一、策略：“AI+”奇点突破，如何判断产业机遇？

我们自23.3.2持续七篇报告推荐数字经济机遇，这是23年最值得重视的或将贯穿全年的投资机遇！报告详细梳理数字经济“三大预期差→三大趋势→三大投资线索”、数据要素&数字基建两大基石，数字经济“位置感”，以及由AI引发第四次场景革命。我们于4.3发布《与风共舞：从经典科技牛看AI浪潮》提出，AI的奇点时刻已经到来，我们当前站在新一轮产业浪潮的拐点，并于4.7发布《“AI+”堪比13年的“移动互联网+”》，进一步深度挖掘“AI+”与13年“移动互联网+”的异同，从策略视角自上而下理解本轮“AI+”的行情。

广发电子组自23.3.22连续十一篇报告全面解读“AI的iPhone时刻”，英伟达GTC大会揭开“AI的iPhone时代”序幕，ChatGPT引爆AIGC奇点，驱动AI服务器、高性能计算（HPC）算力需求提升，光芯片步入黄金时代，并于4.3发布《如何初步理解GPU算力》中进行模型训练&推理算力需求测算，关注算力需求下，算力芯片/连接芯片/存储芯片/服务器硬件的投资机遇，同时在4.9《AI服务器需求风起，PCB升级浪潮已至》中关注AI服务器高算力需求爆发，推动PCB单机价值量提升。

本篇报告我们联合广发电子组，从三个视角判断本轮“AI+”带来的产业机遇，并全面解析AI主线中的电子行业投资机会。

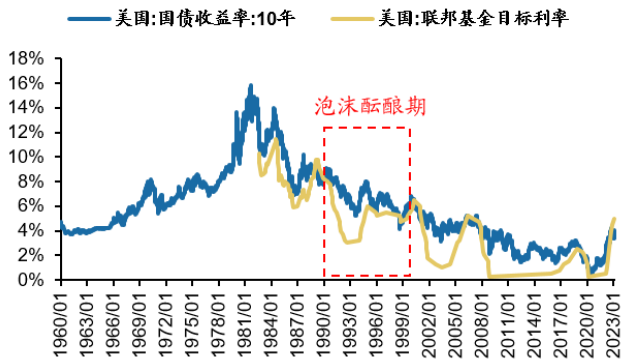
（一）本轮“AI+”判断：类比13年“移动互联网+”，首推上游算力

类比70年代信息技术革命，AI技术驱动全要素生产率提升，本质上是“技术革命+场景革命”的催化，将会呈现“认识技术→升级技术（降低成本或实现规模化）→找到商业化模式→全要素生产率提升”的路径。

类比99年科网泡沫和13-15年创业板“移动互联网+”，则可以发现本轮AI行情与之的相似处和不同点相似处包括：（1）“技术+场景”的孵化模式类似，但99年“互联网+”和15年“移动互联网+”在C端落地速度快，而AI则目前以B端为主、C端尚在孵化；（2）业绩确定性不同，99年“互联网+”典型龙头公司的业绩处于高位、13-15年创业板业绩提速，而当前AI业绩尚未能显现、仅出现算力需求爆发驱动相应环节；（3）当前没有极度宽松的货币政策、源源不断的增量资金，99年“互联网+”和15年“移动互联网+”流动性十分宽松、利率中枢持续下移，并且增量资金相对较多。

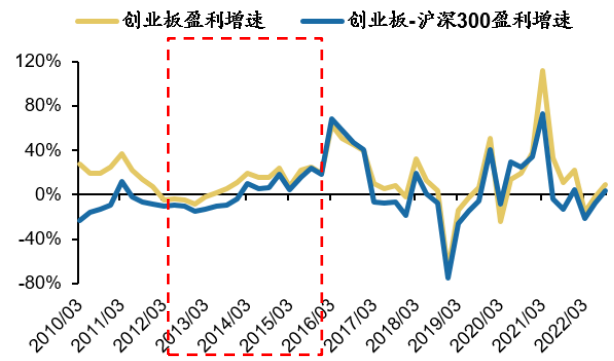
借鉴13年“移动互联网+”行情，我们认为本轮“AI+”行情也将分三阶段走：上游算力→中游软件服务→下游应用。当前行情仍在第一阶段，我们重点首推“上游算力”；同时投资者需要关注向第二阶段“中游软件服务”和第三阶段“下游应用”的切换信号。

图 1: 美科网泡沫酝酿期的美债利率



数据来源: Wind, 广发证券发展研究中心

图 2: 13-15 年创业板-沪深 300 业绩向上



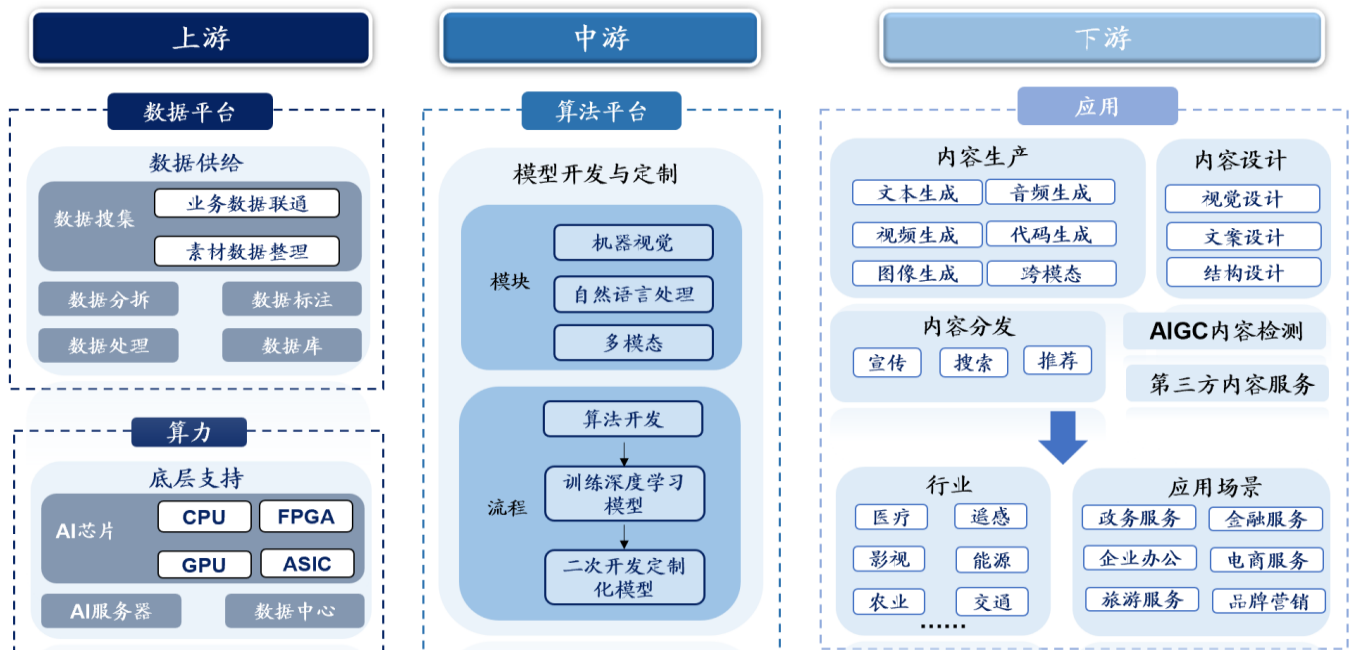
数据来源: Wind, 广发证券发展研究中心

(二) AIGC 产业链: 电子行业涉及硬件、软件、应用

目前AIGC产业链可分为上游(数据)、中游(算法)、下游(应用), 上游主要包括数据平台——提供数据和算力层——提供底层支持; 中游为算法层, 在原始数据的基础上进行模型的训练与二次开发; 下游则为应用层, 包括内容的生产与分发两个方面, 涉及营销、推荐等多个应用场景, 未来有望在更多行业铺开。

目前上游算力(包含AI服务器)和中游大模型需求率先爆发, 下游应用需求尚未被挖掘, 处于雏形概念阶段。电子行业主要涉及上游硬件(包含AI服务器中的GPGPU/ASIC/CPU、服务器芯片、服务器硬件等)、下游软件和应用(如AIGC应用相关的VR硬件、泛安防产业链等), 其中AI服务器受益ChatGPT带来的大量高性能计算算力需求, 当前处于需求增长阶段。

图3: AIGC产业链图谱全览



数据来源: 广发证券发展研究中心

（三）风险提示

宏观经济下行压力超预期，国内“稳增长”政策落地效果仍有不及预期风险。

全球疫情仍存在反复的风险，尤其是奥密克戎疫情带来更大的不确定性。

疫情变异/反复可能导致全球经济修复不及预期并影响中国出口韧性。

全球/中国通胀高位流动性可能边际收紧。

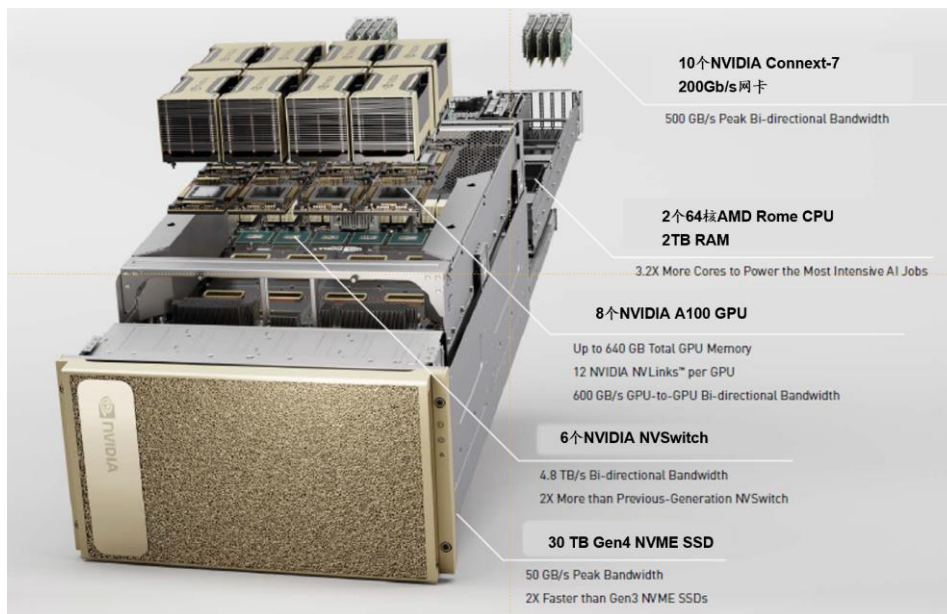
中美贸易/金融领域的关系仍有不确定性。

二、对话电子：六问六答，“AI的iPhone时刻”哪些环节受益？

（一）Q1: ChatGPT产业链上下游电子行业的分布情况怎么样？看好哪些细分领域？

电子行业在ChatGPT产业链中主要分布位于AI服务器及相关电子产品。AI服务器是以ChatGPT为代表的大模型进行训练/推理的算力载体和基础。相比于通用型服务器，AI服务器价值量最大的提升来自于配备4/8颗GPGPU，GPGPU在AI服务器硬件成本中占比可达70%。数据量和算力的提升也明显拉升了AI服务器中存储芯片、高速互联芯片、光模块/光芯片、多相电源、PCB基板等电子产品的量与价，给相关电子厂商带来了机会。

图 4: DGX A100系统硬件组成示意图



数据来源：NVIDIA 官网，广发证券发展研究中心

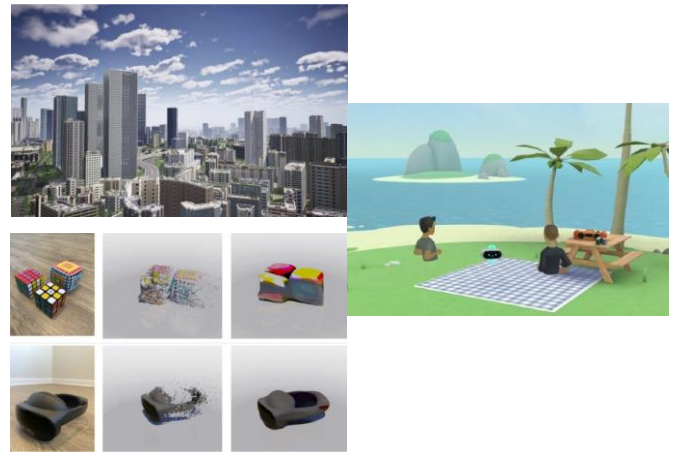
终端应用侧的电子厂商亦有望受益于ChatGPT产业链的发展。在终端应用侧，AIGC将有望为XR、智能音箱、安防等终端设备打开新成长空间。一方面，AIGC作为生产力工具能够低成本、高效率地满足用户海量的内容需求，在内容端为XR产业持续赋能；另一方面，AIGC的应用提升了用户与智能音箱等终端设备的交互体验，打开了新的使用场景。在安防领域，ChatGPT和AIGC技术的应用，可以帮助安防产品实现更高效的数据处理和管理，大幅提升安防系统的反应速度和处理能力，实现一定程度降本增效，推动智慧安防乃至智能物联网行业发展。

图 5: 更多应用场景软件均有布局/引入LLM模型



数据来源：各公司官网整理，广发证券发展研究中心

图 6: AIGC搭建3D景观



数据来源：维深 Wellsenn XR, Meta, 广发证券发展研究中心

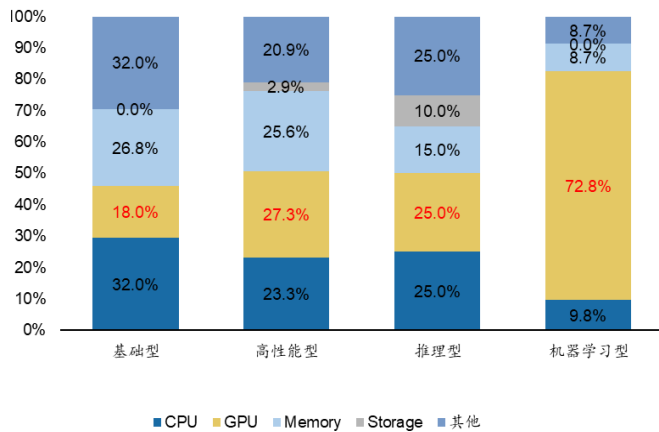
（二）Q2: ChatGPT 拉动 AI 服务器需求提升，国内哪些领域和厂商直接受益？

在服务器硬件厂商领域，主要受益的国内厂商有PCB厂商胜宏科技、沪电股份，ODM厂商工业富联等。在AI服务器产业链领域，主要受益的细分领域为算力芯片、存储芯片和连接芯片。主要受益的厂商有算力芯片厂商海光信息、寒武纪、芯原股份、龙芯中科等；连接芯片厂商源杰科技、澜起科技、聚辰股份、裕太微等；存储芯片相关厂商深科技、北京君正、兆易创新、东芯股份、国芯科技、雅克科技等；电源芯片：杰华特等。在AIGC应用层的终端设备领域，主要受益的厂商有大华股份、海康威视、韦尔股份、思特威、恒玄科技、晶晨股份、全志科技、瑞芯微等。

（三）Q3: 微软宣布 ChatGPT 支持第三方插件接入，支撑以后更多的类 ChatGPT 模型算力肯定需要芯片，具体有哪些 AI 芯片更关键更值得关注？

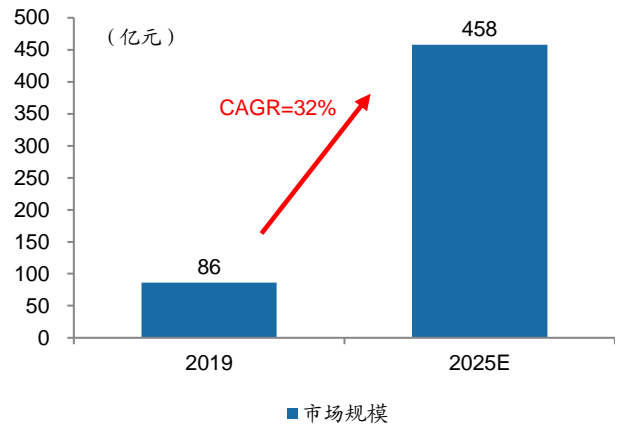
高性能GPGPU为AI服务器算力核心，是支撑AIGC在用户端广泛应用的算力基础。GPGPU是一种利用图形处理器进行高性能计算的通用图形处理器。作为运算协处理器，GPGPU去掉或减弱了GPU的图形显示部分能力，将其余部分全部投入与图形处理无关的通用计算，并成为AI加速卡的核心，在AI服务器硬件成本中占比可达70%。在AI领域，使用GPGPU在云端运行模型训练算法可以显著缩短训练时间，减少能源消耗，从而降低成本。与此同时，GPGPU提供了完善的软件生态系统，便于各种应用程序的移植和新算法开发。因此，GPGPU是人工智能领域最主要的协处理器解决方案，占据90%以上的市场份额，在智能工厂、无人驾驶、智慧城市等领域广泛应用。

图 7: 2021年不同类型服务器硬件占比



数据来源: IDC, 广发证券发展研究中心

图 8: 中国GPGPU芯片板卡市场规模



数据来源: Verified Market Research, 广发证券发展研究中心

理解GPU的算力性能参数是理解AI芯片的基础和前提。 GPU计算时常用的数据类型有浮点数: FP64、FP32、FP16、TF32 (Nvidia提出)、BF16 (Google提出); 整点数: INT8, INT16, INT32等。他们的核心区别在于数据精度。对于AI训练、AI推理、Advanced HPC等不同使用场景, 所需求的数据类型也有所不同, 根据英伟达官网的表述, AI训练为缩短训练时间, 主要使用FP8、TF32和FP16; AI推理为在低延迟下实现高吞吐量, 主要使用TF32、BF16、FP16、FP8和INT8; HPC (高性能计算) 为实现在所需的高准确性下进行科学计算的功能, 主要使用FP64。

表 1: 不同精度数据类型应用场景

	AI 训练	AI 推理	HPC
应用需求	AI 网络和数据集规模持续扩大, 计算量庞大, 使用较高进度对这些数据进行集中计算需要耗费较长的时间, 降低了训练的效率。使用较低精度的数学计算可大幅提升性能, 缩短训练所需时间	提供出色的性能、加速不同神经网络的通用性, 可靠地部署推理, 在低延迟下实现高吞吐量	要求严谨, 对性能的追求永无止境, 如理解复杂分子结构以支持药物发现, 通过模拟物理效果来寻找潜在的能源, 以及通过模拟大气数据来更好地预测极端天气状况并为之做好准备
常用数据类型	FP8、TF32 和 FP16	TF32、BF16、FP16、FP8 和 INT8	FP64

数据来源: 英伟达, 广发证券发展研究中心

GPU峰值算力的测算公式为: 峰值计算能力= GPU Core的运行频率*GPU SM 数量*单个SM对应的特定数据类型的指令吞吐量*2。 以NVIDIA GA100为例, A100 中总共有108个SM, 每个SM有64个FP32 CUDA核, 64个INT32 CUDA核, 32个FP64 CUDA核, 以及4个第三代Tensor Core, 其中Tensor Core专注于矩阵运算, 在模型训练与推理的过程中, Tensor Core将是主要的计算内核。根据上述公式, 可以计算

得到A100 FP16 (Tensor Core加速) 峰值算力为312T FLOPS, FP32 (Cuda Core) 峰值算力=19.5T FLOPS, 与英伟达官方披露的性能参数一致。由此可见, GPU峰值算力与主频和硬件配置数量等参数相关。

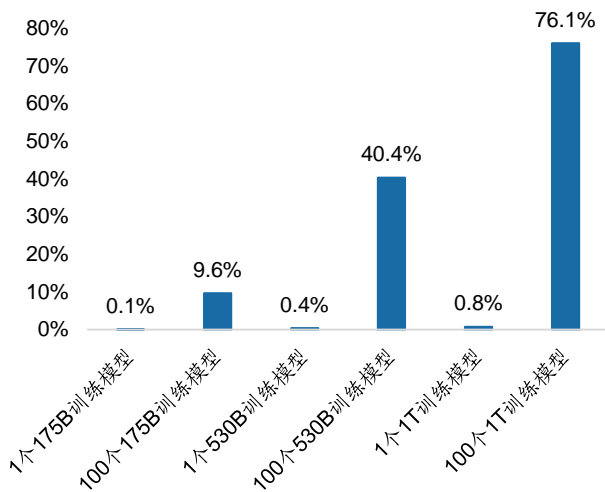
表 2: NVIDIA A100/H100 GPU参数

	A100 SXM4	H100 SXM5	H100 PCIe
SM 数	108	132	114
运行频率-FP16 Tensor Core	1.41GHz	1.83Ghz	1.62GHz
运行频率-FP32 Cuda Core	1.41GHz	1.98Ghz	1.755GHz
单个 SM FP32 Cuda Core 数	64	128	128
单个 SM FP64 Cuda Core 数	32	64	64
单个 SM INT32 Cuda Core 数	64	64	64
FP16 Tensor Core 峰值算力	312 TFLOPS	989.4 TFLOPS	756 TFLOPS
FP32 Cuda Core 峰值算力	19.5 TFLOPS	66.9 TFLOPS	51.2 TFLOPS
INT8 Cuda Core 峰值算力	19.5 TOPS	33.5 TOPS	25.6 TOPS

数据来源: 英伟达, 广发证券发展研究中心

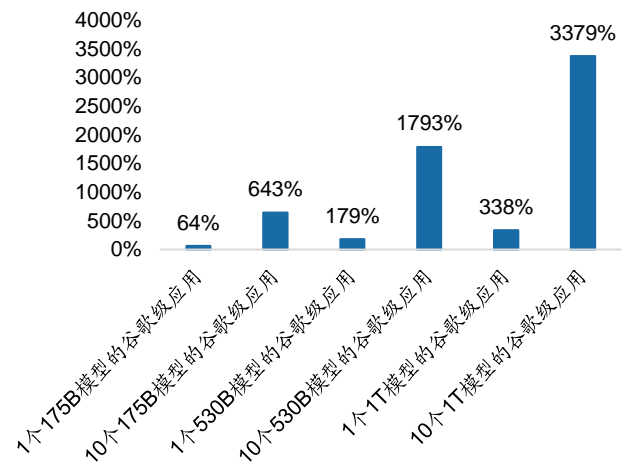
AIGC时代AI芯片的增量市场需求可以通过AI大模型训练/推理的算力需求进行测算。我们参考NVIDIA发表论文中对不同参数GPT模型算力需求的计算方法及NVIDIA A100在模型训练过程的参数, 对以GPT-3.5为代表的大模型的训练/推理算力需求进行测算。根据测算: 在训练侧, 训练一个GPT-3.5 175B模型的NVIDIA A100需求为1080个、AI服务器需求为135台; 训练一个万亿参数量AI大模型对A100、AI服务器的需求分别为8521个、1065台。在推理侧, 一个谷歌级应用使用GPT-3.5 175B进行推理, 对NVIDIA A100需求为72万个、对AI服务器需求为9万台; 一个谷歌级应用使用万亿参数量大模型进行推理, 对NVIDIA A100需求为378万个、对AI服务器需求为47.3万台。同时以2022年全球AI服务器出货量14万台为基数, 测算了大模型算力需求对全球AI服务器市场的拉动。

图 9：训练侧算力需求对AI服务器出货量的拉动



数据来源：英伟达，Trend Force，广发证券发展研究中心

图 10：推理侧算力需求对AI服务器出货量的拉动



数据来源：英伟达，Trend Force，广发证券发展研究中心

海光信息是国产微处理器领航者，AIGC时代迎新机遇。海光信息主要产品包括高端通用处理器（CPU）和高端协处理器（DCU）。与NVIDIA公司高端GPU产品（型号为A100）及AMD公司高端GPU产品（型号为MI100）进行对比，公司DCU产品深算一号在生产工艺、内核频率等性能参数上达到国际上同类型高端产品水平。公司产品自带片上中国原生安全基因，兼顾性能和安全需求，可以在数据处理过程中为用户提供更高效的安全保障，满足中国境内最为严苛的安全计算要求。公司发展有望充分受益于信创领域国产替代大趋势与AIGC大时代的来临。

表3：海光信息DCU产品与国际厂商高端GPU产品对比

项目	海光	NVIDIA	AMD
品牌	深算一号	Ampere 100	MI100
生产工艺	7nm FinFET	7nm FinFET	7nm FinFET
核心数量	4096 (64 CUs)	2560 CUDA processors 640 Tensor processors	120CUs
内核频率	Up to 1.5GHz (FP64) Up to 1.7Ghz (FP32)	Up to 1.53Ghz	Up to 1.5GHz (FP64) Up to 1.7Ghz (FP32)
显存容量	32GB HBM2	80GB HBM2e	32GB HBM2
显存位宽	4096 bit	5120 bit	4096bit
显存频率	2.0 GHz	3.2 GHz	2.4 GHz
显存带宽	1024 GB/s	2039 GB/s	1228 GB/s
TDP	350 W	400 W	300W
CPU to GPU 互联	PCIe Gen4 x 16	PCIe Gen4 x 16	PCIe GEN4 x 16
GPU to GPU 互联	xGMI x 2, Up to 184 GB/s	NVLink up to 600 GB/s	Infinity Fabric x 3, up to 276 GB/s

数据来源：公司招股书，广发证券发展研究中心

（四）Q4：英伟达 GTC 大会公布了不少新的产品，结合他们新品的发布情况，在硬件端和软件端有哪些投资机会值得关注？

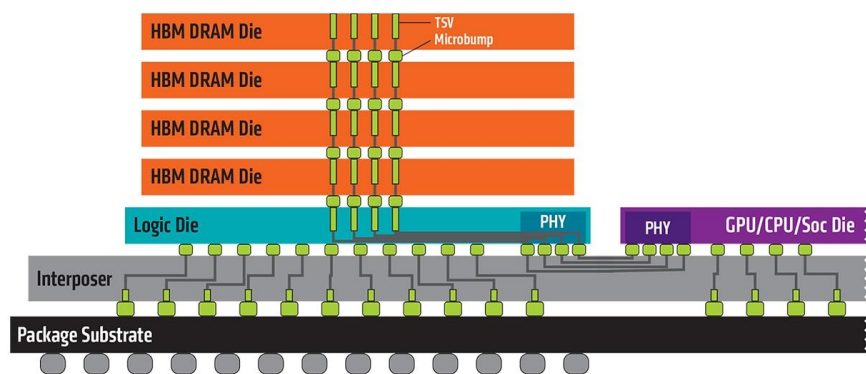
英伟达在GTC大会针对训练和推理发布多款新产品，包括全新GPU推理平台（L4，L40，H100 NVL，Grace Hopper芯片），4种配置分别用于加速AI视频、图像生成、大型语言模型部署和推荐系统。其中适用于训练阶段的DGX H100，其拥有8个H100 GPU模组，在FP8精度下可提供32PetaFLOPS的算力，并提供完整的英伟达AI软件堆栈，助力简化AI开发。芯片的算力提升是AI硬件产品发展的主线规律，建议持续关注本土算力芯片厂商在产品研发及产品批量出货应用方面的进展。

AI芯片算力的提升对系统的存储容量与读取速度、芯片/服务器间高速连接、供电管理等性能的高可靠性与稳定性提出了更高的要求，相应电子产品的量/价在AI服务器中也有望迎来量价齐升。

1. 存储

高带宽显存（HBM）突破内存带宽瓶颈，提升数据交换速率。 HBM 是一款新型的 CPU/GPU 内存芯片，通过将多个DDR芯片堆叠在一起，实现大容量，高位宽的DDR组合阵列。这种设计能够大大提升信息交换的速率。这些堆叠的芯片通过称为“中介层 (Interposer)”的超快速互联方式连接至 CPU 或 GPU。将HBM的堆栈插入到中介层中，放置于 CPU 或 GPU 旁边，然后将组装后的模块连接至电路板。此外，HBM在功耗和面积上也有优越的表现，HBM重新调整了内存的功耗效率，使每瓦带宽比 GDDR5 高出 3 倍还多；相比于GDDR5，HBM能够节省94%的表面积。

图 11：HBM结构示意图



数据来源：AMD，广发证券发展研究中心

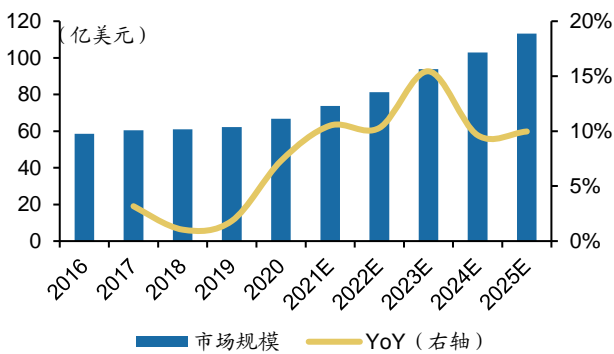
DDR5时代内存接口芯片量价齐升，内存接口芯片是服务器内存模组的核心逻辑器件，以提升内存数据访问的速度稳定性，满足服务器CPU对内存模组日益增长的高性能及大容量需求。目前行业正从DDR4升级至DDR5。据JEDEC定义，DDR5时代，RCD和DB需支撑的传输速率从3200MT/s提升至DDR5第一子代的4800MT/s，且子代还在继续升级中。此外，DDR5世代还需配置1颗SPD、1颗PMIC

和2颗TS等配套芯片，行业迎来量价齐升。澜起科技是目前全球唯二可提供DDR5全套芯片的龙头厂商（其中，SPD与聚辰股份合作提供）。

2. 高速互联

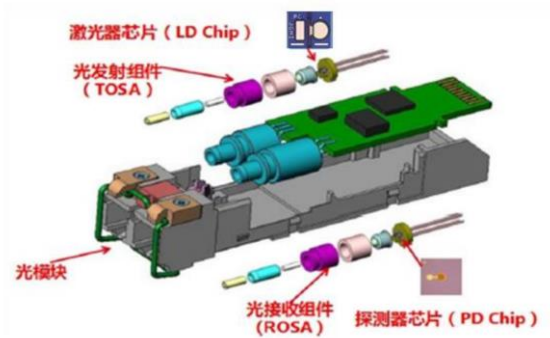
AIGC应用驱动400G/800G高速率光模块在AI服务器中加速渗透。光模块是数据中心内部互连和数据中心相互连接的核心部件。随着AIGC应用所需数据流量持续提升，交换机互联速率逐步由100G向400G/800G升级。根据LightCounting的数据，2020年全球光模块市场规模为66.7亿美元。预测2025年全球光模块市场将达到113亿美元，为2020年的1.7倍，

图 12: 全球光模块市场规模及预测



数据来源: LightCounting, 广发证券发展研究中心

图 13: 光模块结构示意图 (SFP+封装)

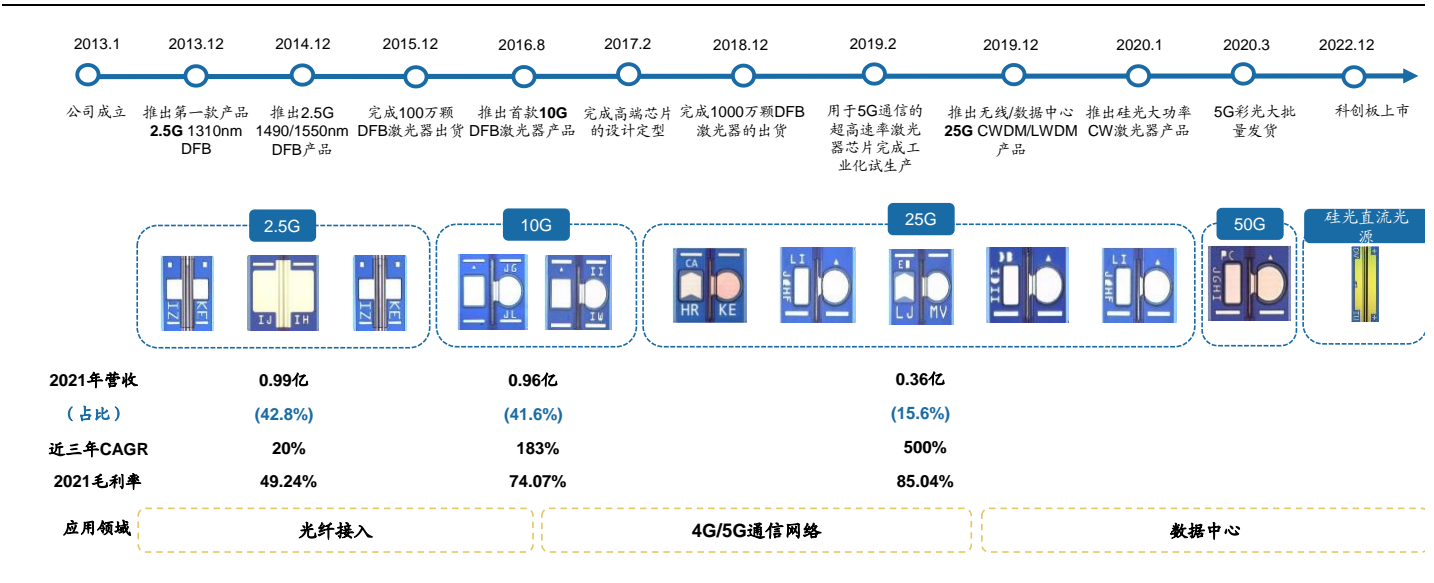


数据来源: 源杰科技招股书, 广发证券发展研究中心

光芯片作为光模块核心元件，迎来量价齐升。光芯片是光模块中的核心元件，在对高速传输需求不断提升背景下，更高价值量的25G及以上高速率光芯片市场增长迅速。根据Omdia预测，2019年至2025年，25G以上速率光模块所使用的光芯片占比逐渐扩大，市场空间将从13.56亿美元增长至43.40亿美元，CAGR为21.40%。封装方式上，CPO是解决超高算力过载并提高能效的核心解决方案之一，据Light Counting预计，按照端口数量统计，CPO的全球发货量将从2023年的5万件逐步增长到2027年的450万件。

源杰科技: 本土光通信芯片领军者。源杰科技聚焦于光芯片行业，主要产品包括2.5G、10G、25G及更高速率激光器芯片系列产品等，目前主要应用于光纤接入、4G/5G移动通信网络和数据中心等领域。公司应用于数据中心的25G DFB芯片已实现批量供货，2021年公司数据中心业务营业收入同比增长459%。更高阶产品方面，公司100G PAM4 EML激光器芯片处于送样阶段；公司CW大功率光源可用于CPO领域，产品与海外差距较小，已在多家客户送样测试。

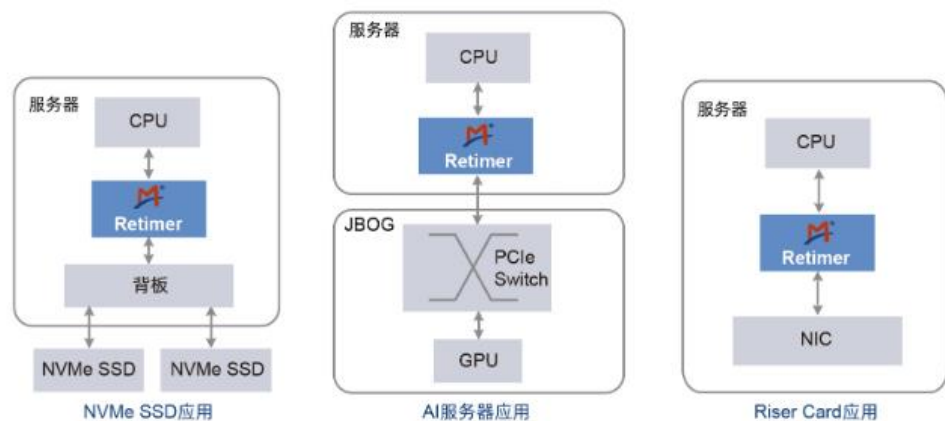
图 14: 源杰科技主要产品的开发及演变情况



数据来源: 源杰科技官网, 源杰科技招股说明书, 广发证券发展研究中心

PCIe5.0拉动PCIe Retimer需求, 澜起科技具备先发优势。 PCIe Retimer芯片是适用于 PCIe 高速数据传输协议的超高速时序整合芯片, 主要解决数据中心数据高速、远距离传输时, 信号时序不齐、损耗大、完整性差等问题, 还可灵活地切换 PCIe 或 CXL 模式, 多应用于NVMe SSD、AI 服务器、Riser卡等典型应用场景, AI服务器中由于增加8颗GPU, 他们通过PCIe与CPU互联, 因此AI服务器对PCIe Retimer会有比较明显的增量需求。澜起科技布局PCIe较早, 是全球可量产PCIe 4.0 Retimer的唯一中国公司, 公司PCIe 5.0 Retimer研发进展顺利。

图 15: 澜起科技PCIe Retimer产品在服务器中的应用场景

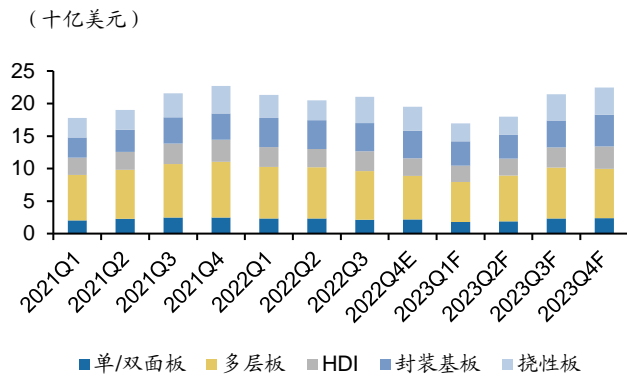


数据来源: 澜起科技官网, 广发证券发展研究中心

AI服务器中PCB产品量价齐升, 单机价值增量显著。 AI服务器PCB板价值量提升主要来自三方面: (1) PCB板面积增加。AI服务器中除了搭载CPU的主板外, 每颗GPU需要分别封装在GPU模块板, 并集成到一块主板上, 相比传统服务器仅使用一块主板, PCB面积大幅增加。(2) PCB板层数增加。AI服务器相对于传统

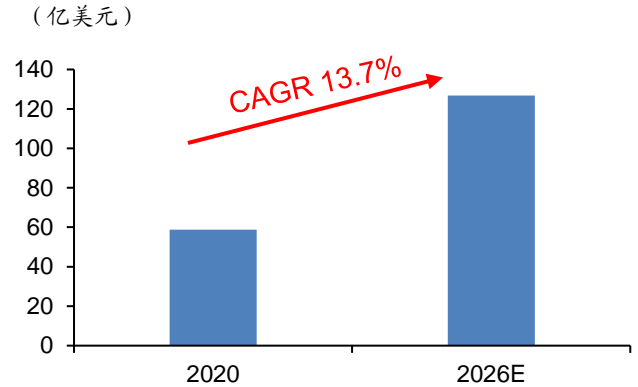
服务器具有高传输速率、高内存带宽、硬件架构复杂等特征，需要更复杂的走线，因而需要增加PCB层数以加强阻抗控制等性能。（3）PCB用CCL材料标准更高。AI服务器用PCB需要更高的传输速率、更高散热需求、更低损耗等特性，其核心材料CCL需要具备高速高频低损耗等特质，CCL材料等级需要提升，材料的配方以及制作工艺复杂度攀升，推动价值量提升。

图 16: 按产品类型划分的全球PCB市场



数据来源: Prisma, 广发证券发展研究中心

图 17: 服务器用PCB市场规模以及增速预测



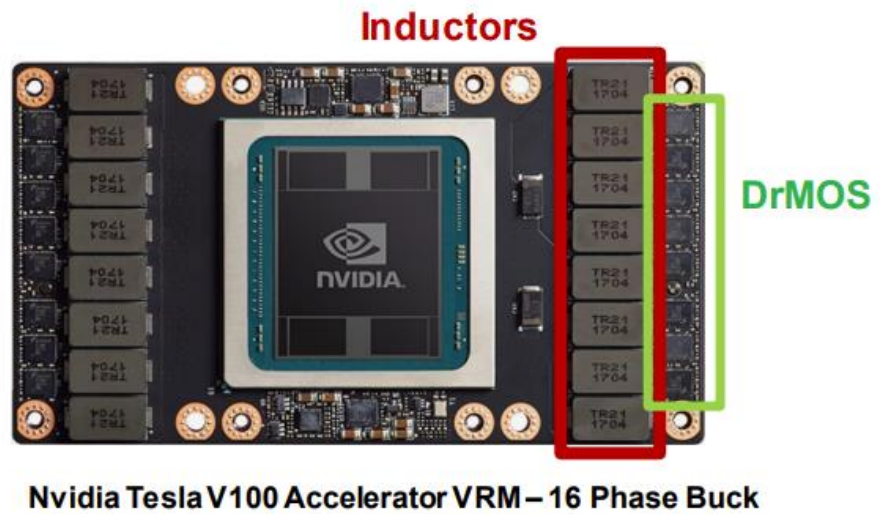
数据来源: 沪电股份 2021 年年报, 广发证券发展研究中心

本土PCB厂商积极布局AI服务器市场。沪电股份立足于既有的企业通讯市场板、汽车板等主导产品的技术领先优势，及时把握通信、汽车等高端产品需求，应用于AI加速、Graphics、GPU、OAM、FPGA等加速模块类的产品以及应用于UBB、BaseBoard的产品已批量出货。胜宏科技具备70层高精密线路板、20层五阶HDI线路板的研发制造能力，平台服务器主板小批量试产，服务器硬盘用高频主板试样中。

3. 其他

多相电源产品在AI服务器中单机价值增量显著。多相控制器+DrMOS组成的多相电源解决方案是GPU的主流供电形式。参考NVIDIA V100的多相电源配置，A100至少需要16相电源解决方案（1颗多相控制器+16颗大电流DrMOS的配置），则该AI服务器相较于普通服务器增加了8颗多相控制器、128颗大电流DrMOS需求，参考TI官网产品价格，则搭载8张A100的AI服务器单机新增多相电源产品价值量超过300美元。根据MPS数据，单台普通服务器CPU主板电源解决方案（包含多相控制器+DrMOS+POL+EFuse等产品）合计价值量约80美元；因此AI服务器单机多相电源产品价值量相较于普通服务器有数倍提升。

图 18: Nvidia V100采用16相供电解决方案



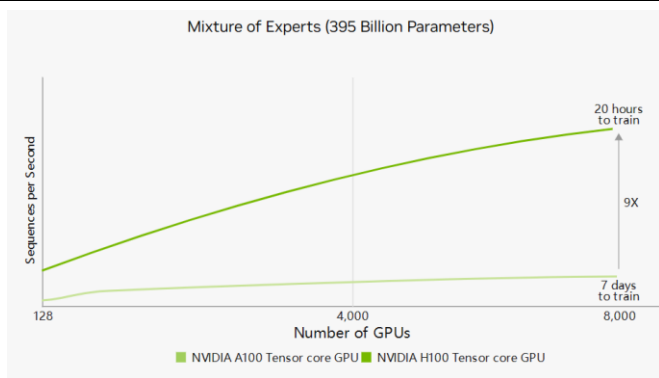
数据来源：Princeton University Power Electronics Research Group，广发证券发展研究中心

杰华特：本土多相电源解决方案领军者。根据杰华特招股书、投资者活动关系记录，公司于2020年成功研发出基于公司自有工艺和技术的DrMOS产品，单芯片可支持60A输出电流，打破了欧美厂商的垄断；2021与2022年，公司相继研发车规DrMOS产品以及90A DrMOS产品，持续增强市场竞争力；目前公司DrMOS产品在服务器、笔电领域均实现批量出货。公司多相控制器产品目前处于送样阶段，未来会与DrMOS配套出货。

（五）Q5: AIGC 产业链从成本的角度来看，还有哪些成本要素值得我们跟踪？

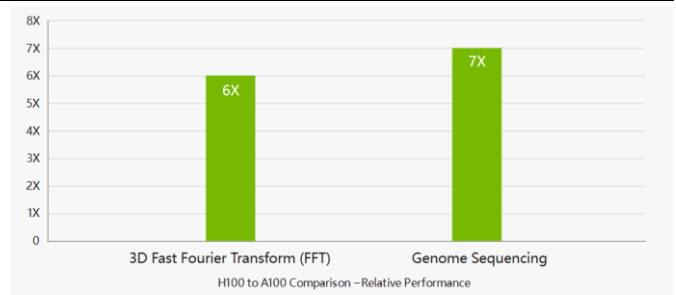
GPU性能不断提升，有望优化GPU硬件成本。NVIDIA最新一代H100产品配置了第四代Tensor Cores及FP8精度的Transformer engine。在执行训练任务时，相比于上一代配置MoE模型的A100计算集群，大规模H100计算集群在配置NVLink的情况下最高可将训练速度提升9倍；在执行推理任务时，第四代Tensor Cores提高了包括FP64、TF32、FP32、FP16、INT8和FP8在内的所有精度下的推理速度，在保持LLM精度的同时减少了内存使用并提高性能，最高可将推理速度提升30倍。因此随着技术进步，单个GPU在训练和推理任务上的效率提升，有望降低大模型高昂的训练和推理成本。

图 19: NVIDIA H100相对A100的训练性能提升



数据来源: NVIDIA 官网, 广发证券发展研究中心

图 20: NVIDIA H100相对A100的性能提升

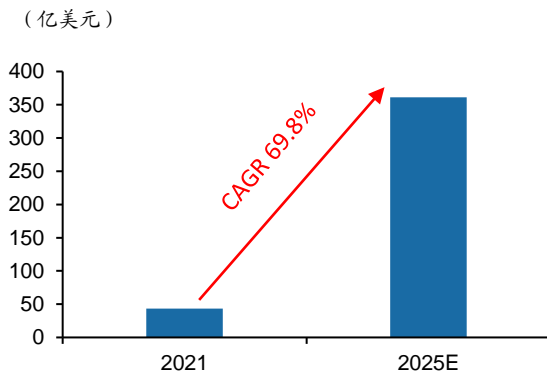


数据来源: NVIDIA 官网, 广发证券发展研究中心

(六) Q6: 怎么看待消费电子应用层上的商业化, 建议关注哪些落地场景?

从视觉领域来看, AIGC作为生产力工具赋能元宇宙, 逐渐实现元宇宙的可扩展性潜力, 低成本、高效率地满足VR/AR用户海量内容需求。根据Omdia数据, 2023年近眼显示面板出货量有望达到2406万台, 同比增长67.3%。随着索尼PSVR 2、苹果MR、Meta Quest3等主要产品的推出在即, 以及AIGC未来在内容端的持续赋能, XR产业有望迎来新一轮增长。

图 21: VR/AR市场规模



数据来源: IDC, 广发证券发展研究中心

图 22: 苹果MR头显四大亮点



数据来源: Metaverse 元宇宙, 广发证券发展研究中心

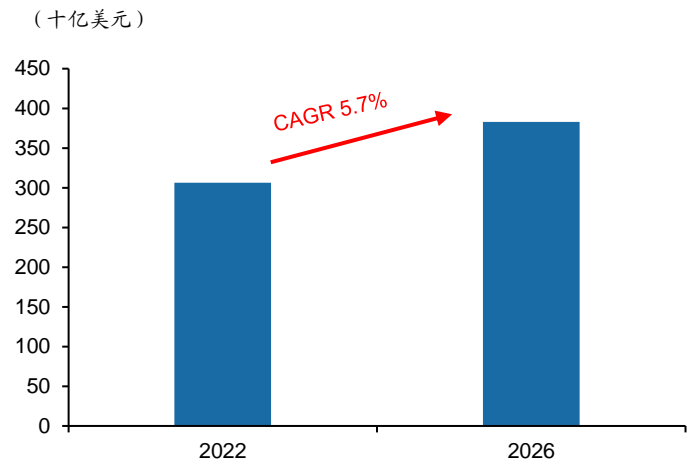
从听觉领域来看, AIGC有望创造智能音箱交互新范式。智能音箱是家庭场景交互中心与IoT控制中心, AIGC符合智能音箱用户个性化交互需求, 在家庭内部家庭教育、电器控制、起居辅助等方面进一步提升用户体验, 开启听觉盛宴。阿里发布了“鸟鸟分鸟”, 将AI大模型装进智能终端, 探索了智能终端与大模型的交互范式。根据IDC数据, 2022年包括智能音箱在内的家庭智能设备市场规模为306.3亿美元, 预计到2026年将达到382.9亿美元, CAGR为5.7%。随着AIGC热潮上升, 音频内容个性化需求进一步凸显, 打开终端设备销量未来市场空间。

图 23: 阿里发布“鸟鸟分鸟”



数据来源: 智东西, 广发证券发展研究中心

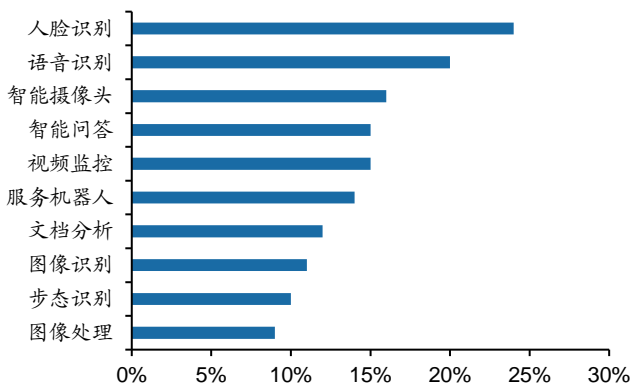
图 24: 家庭智能设备市场规模



数据来源: IDC, 广发证券发展研究中心

AIGC有望与安防紧密结合, 驱动泛安防产业智能化升级。 ChatGPT和AIGC技术的应用, 可以帮助安防产品实现更高效的数据处理和管理, 大幅提升安防系统的反应速度和处理能力, 实现一定程度降本增效, 推动智慧安防乃至智能物联网行业发展。具体而言, ChatGPT可以帮助识别并分类人、车辆、动物等不同对象, 帮助安全系统快速识别异常事件等。同时, ChatGPT的应用还将促进安防与其他相关领域合作, 以谷歌和微软等厂商致力于利用神经网络进行机器人控制为例, 将视觉数据以及大型语言模型结合起来, 可推动安防机器人的发展。

图 25: 2021-2024年内计划部署人工智能场景企业应用率



数据来源: IDC, 广发证券发展研究中心

图 26: 海康威视AI一站式服务



数据来源: 海康威视 AI 开发平台, 广发证券发展研究中心

(七) 风险提示

AIGC 发展不及预期。 算力、数据、算法是AIGC产业发展的重要基石, 若其中任一因素发展不及预期, 则相关因素以及模型和应用发展都有可能不及预期。

AI 服务器出货量不及预期。 AI服务器出货量一方面受到企业IT支出, 尤其是

云计算厂商的CapEx支出影响，因此若企业对算力设施的投入不及预期，可能影响AI服务器的出货量；另一方面AI服务器价值量占比较高的GPU和CPU若出现供应紧张等问题，也会对AI服务器的出货量级和节奏造成一定消极影响。

国产厂商技术和产品进展不及预期。目前国内厂商不断开发模型以及相关应用，若在算力、数据、算法等某一因素上支撑不及预期，则可能导致国内厂商技术和产品进展不及预期。

广发投资策略研究小组

- 戴康：CFA，首席分析师，中国人民大学经济学硕士，12年A股策略研究经验。
- 郑恺：联席首席分析师（行业比较），华东师范大学金融学硕士，8年A股策略研究经验。
- 曹柳龙：CPA，资深分析师（行业比较），华东师范大学管理学硕士，7年A股策略研究经验。
- 韦冀星：资深分析师（大势研判、行业比较），美国杜兰大学金融学硕士，4年A股策略研究经验。
- 倪康：资深分析师（主题策略），中山大学经济学硕士，4年A股策略研究经验。
- 吴迪：资深分析师（海外及港股策略），复旦大学经济学硕士，2023年加入广发证券发展研究中心。
- 李学伟：高级分析师（主题策略，专题研究），复旦大学硕士，2022年加入广发证券发展研究中心。
- 杨泽蓁：研究员（行业比较），上海财经大学硕士，2022年加入广发证券发展研究中心。
- 徐嘉奇：研究员（行业比较），伦敦国王学院硕士，2022年加入广发证券发展研究中心。
- 侯蕾：研究员（主题策略），复旦大学硕士，2022年加入广发证券发展研究中心。
- 杨藤：研究员（海外策略），新加坡国立大学硕士，2023年加入广发证券。

广发证券—行业投资评级说明

- 买入：预期未来12个月内，股价表现强于大盘10%以上。
- 持有：预期未来12个月内，股价相对大盘的变动幅度介于-10%~+10%。
- 卖出：预期未来12个月内，股价表现弱于大盘10%以上。

广发证券—公司投资评级说明

- 买入：预期未来12个月内，股价表现强于大盘15%以上。
- 增持：预期未来12个月内，股价表现强于大盘5%-15%。
- 持有：预期未来12个月内，股价相对大盘的变动幅度介于-5%~+5%。
- 卖出：预期未来12个月内，股价表现弱于大盘5%以上。

联系我们

	广州市	深圳市	北京市	上海市	香港
地址	广州市天河区马场路 26号广发证券大厦 47楼	深圳市福田区益田路 6001号太平金融大厦 31层	北京市西城区月坛北 街2号月坛大厦18 层	上海市浦东新区南泉 北路429号泰康保险 大厦37楼	香港德辅道中189号 李宝椿大厦29及30 楼
邮政编码	510627	518026	100045	200120	-
客服邮箱	gfzqyf@gf.com.cn				

法律主体声明

本报告由广发证券股份有限公司或其关联机构制作，广发证券股份有限公司及其关联机构以下统称为“广发证券”。本报告的分销依据不同国家、地区的法律、法规和监管要求由广发证券于该国家或地区的具有相关合法合规经营资质的子公司/经营机构完成。

广发证券股份有限公司具备中国证监会批复的证券投资咨询业务资格，接受中国证监会监管，负责本报告于中国（港澳台地区除外）的分销。

广发证券（香港）经纪有限公司具备香港证监会批复的就证券提供意见（4号牌照）的牌照，接受香港证监会监管，负责本报告于中国香港地区的分销。

本报告署名研究人员所持中国证券业协会注册分析师资质信息和香港证监会批复的牌照信息已于署名研究人员姓名处披露。

重要声明

广发证券股份有限公司及其关联机构可能与本报告中提及的公司寻求或正在建立业务关系，因此，投资者应当考虑广发证券股份有限公司及其关联机构因可能存在的潜在利益冲突而对本报告的独立性产生影响。投资者不应仅依据本报告内容作出任何投资决策。投资者应自主作出投资决策并自行承担投资风险，任何形式的分享证券投资收益或者分担证券投资损失的书面或者口头承诺均为无效。

本报告署名研究人员、联系人（以下均简称“研究人员”）针对本报告中相关公司或证券的研究分析内容，在此声明：（1）本报告的全部分析结论、研究观点均精确反映研究人员于本报告发出当日的关于相关公司或证券的所有个人观点，并不代表广发证券的立场；（2）研究人员的部分或全部的报酬无论在过去、现在还是将来均不会与本报告所述特定分析结论、研究观点具有直接或间接的联系。

研究人员制作本报告的报酬标准依据研究质量、客户评价、工作量等多种因素确定，其影响因素亦包括广发证券的整体经营收入，该等经营收入部分来源于广发证券的投资银行类业务。

本报告仅面向经广发证券授权使用的客户/特定合作机构发送，不对外公开发布，只有接收人才可以使用，且对于接收人而言具有保密义务。广发证券并不因相关人员通过其他途径收到或阅读本报告而视其为广发证券的客户。在特定国家或地区传播或者发布本报告可能违反当地法律，广发证券并未采取任何行动以允许于该等国家或地区传播或者分销本报告。

本报告所提及证券可能不被允许在某些国家或地区内出售。请注意，投资涉及风险，证券价格可能会波动，因此投资回报可能会有所变化，过去的业绩并不保证未来的表现。本报告的内容、观点或建议并未考虑任何个别客户的具体投资目标、财务状况和特殊需求，不应被视为对特定客户关于特定证券或金融工具的投资建议。本报告发送给某客户是基于该客户被认为有能力独立评估投资风险、独立行使投资决策并独立承担相应风险。

本报告所载资料的来源及观点的出处皆被广发证券认为可靠，但广发证券不对其准确性、完整性做出任何保证。报告内容仅供参考，报告中的信息或所表达观点不构成所涉证券买卖的出价或询价。广发证券不对因使用本报告的内容而引致的损失承担任何责任，除非法律法规有明确规定。客户不应以本报告取代其独立判断或仅根据本报告做出决策，如有需要，应先咨询专业意见。

广发证券可发出其它与本报告所载信息不一致及有不同结论的报告。本报告反映研究人员的不同观点、见解及分析方法，并不代表广发证券的立场。广发证券的销售人员、交易员或其他专业人士可能以书面或口头形式，向其客户或自营交易部门提供与本报告观点相反的市场评论或交易策略，广发证券的自营交易部门亦可能会有与本报告观点不一致，甚至相反的投资策略。报告所载资料、意见及推测仅反映研究人员于发出本报告当日的判断，可随时更改且无需另行通告。广发证券或其证券研究报告业务的相关董事、高级职员、分析师和员工可能拥有本报告所提及证券的权益。在阅读本报告时，收件人应了解相关的权益披露（若有）。

本研究报告可能包括和/或描述/呈列期货合约价格的事实历史信息（“信息”）。请注意此信息仅供用作组成我们的研究方法/分析中的部分论点/依据/证据，以支持我们对所述相关行业/公司的观点的结论。在任何情况下，它并不（明示或暗示）与香港证监会第5类受规管活动（就期货合约提供意见）有关联或构成此活动。

权益披露

(1) 广发证券（香港）跟本研究报告所述公司在过去12个月内并没有任何投资银行业务的关系。

版权声明

未经广发证券事先书面许可，任何机构或个人不得以任何形式翻版、复制、刊登、转载和引用，否则由此造成的一切不良后果及法律责任由私自翻版、复制、刊登、转载和引用者承担。