电子

证券研究报告
2023 年 06 月 14 日

AI 算力需求持续释放，重点看好 AI 服务器产业链

AI 服务器直接受益于需求端增长，Nvidia 高速互联助力 AI 运算。2018 年至今，模型升级带来参数量指数级增加，AI 训练+推理段模型需求催生 AI 服务器海量需求（训练阶段算力需求=6×模型参数数量×训练集规模），按照 A100 640GB 服务器参数假设，据测算，训练端需要服务器数量为 3423 台/日，推理端按照 1 亿活跃用户测算对应成本为 4000 万美元，并随着用户量和访问互动量上升，算力需求持续提升。此外，Nvidia 高速互联助力 AI 运算，800G 交换机陆续发布，下一代超宽互联蓄势待发。

AI 重塑服务器行业格局，ODM 模式/厂商受益于 AI 服务器增长。根据 Statista 数据，2021 年全球服务器市场规模达到 831.7 亿美元，同比增长 6.97%，其中 AI 服务器市场达到 156.3 亿美元，同比增长 39.1%。AI 服务器有望成为增速最快的细分板块，预计 AI 服务器市场将在 2026 年达到 347.1 亿美元，5 年 CAGR 达到 17.3%。ODM 模式/厂商受益于 CSP 客户增量，2022 年出货的服务器中 ODM 生产的市占率或达 50%。全球 ODM 厂商竞争格局集中度高，CR5 高达 94.4%，分别为鸿海/工业富联（43%）、广达（17%）、纬创（14%）、英业达（12.8%）和超微（7.6%）。

AI 服务器放量预期利好上游核心部件。AI 服务器销量增加将拓宽上游核心零部件的增量市场，尤其是 AI 芯片（GPU、ASIC 和 FPGA）、存储器、固态硬盘等。其中，国内上游环节的国产替代程度不一。我们认为国内处于乐观机遇期的有 AI 服务器制造厂商、服务器用 PCB 厂商、DRAM 厂商和散热厂商，PCB 和 DRAM 的服务器领域均是行业实现扩产和开拓市场的重点所在；制约瓶颈为人工智能芯片，仍被国外厂商垄断，有望实现突破的环节为先进封装 Chiplet，或成为我国算力困境的关键转折点。

投资建议：建议关注 AI 服务器及上游产业链相关标的：1)AI 服务器龙头：工业富联；2)服务器 PCB：鹏鼎控股；3)服务器线束与连接器：电连技术、兆龙互连；4)算力芯片：寒武纪、海光信息（天风计算机团队覆盖）、景嘉微（天风计算机团队联合覆盖）；5)存储供应链：兆易创新、北京君正、江波龙（天风计算机团队联合覆盖）、澜起科技、雅克科技、鼎龙股份（天风化工团队联合覆盖）、华懋科技（天风汽车团队联合覆盖）、华特气体；6)边缘 AI：瑞芯微、晶晨股份、全志科技、恒玄科技、富瀚微、中科蓝讯、乐鑫科技；7)AI to B/机器视觉：大华股份、海康威视、鼎捷软件（天风计算机团队覆盖）、凌云光、天准科技、舜宇光学、海康威视、奥普特（天风机械军工团队覆盖）；8)Chiplet：长电科技、通富微电、华天科技、长川科技（天风机械团队覆盖）、华峰测控（天风机械团队覆盖）、利扬芯片、芯碁微装、伟测科技

风险提示：中美贸易摩擦导致上游原材料断供、AI 服务器出货不及预期、技术瓶颈仍未摆脱

投资评级

行业评级	强于大市(维持评级)
上次评级	强于大市

作者

潘暕 分析师
SAC 执业证书编号：S1110517070005
panjian@tfzq.com

俞文静 分析师
SAC 执业证书编号：S1110521070003
yuwenjing@tfzq.com

行业走势图

资料来源：聚源数据

相关报告

1《电子行业点评：苹果 MR 发布在即，重点推荐相关产业链》 2023-05-31
2《电子行业深度研究：电子行业 23Q1 总结：有望进入复苏周期》2023-05-21
3《电子行业深度研究：AI 赋能制造业道路，传统安防龙头估值逻辑切换》2023-04-09

内容目录

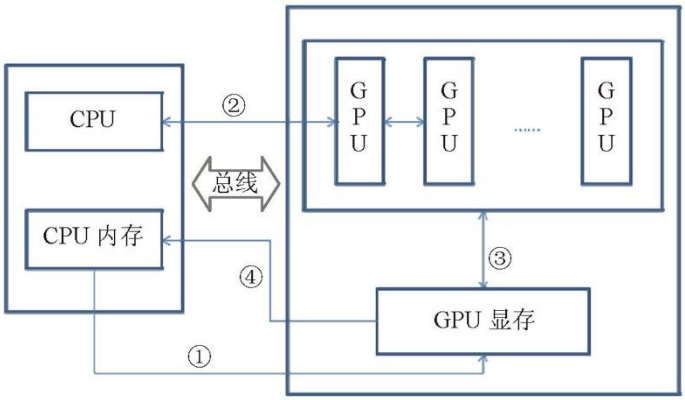
图表目录

1. 人工智能服务器为算力支撑，助燃 AI 产业化

1.1. 从通用服务器到 AI 服务器的过渡

AI 服务器在众多服务器中脱颖而出源于其架构优势。AI 服务器是承载智慧计算中 AI 计算的核心基础设施，是一种能够提供人工智能的数据服务器，既可以用于支持本地应用程序和网页，也可以为云和本地服务器提供复杂的 AI 模型和服务，通过异构形式适应不同应用范围以及提升服务器的数据处理能力，异构方式包括 CPU+GPU/TPU/ASIC/FPGA。

图 1：AI 服务器的 CPU+架构



资料来源：《人工智能服务器技术研究》王峰，天风证券研究所

AI 服务器的发展脱胎自通用服务器的性能提升需求。复盘主流服务器的发展历程，随着数据量激增、数据场景复杂化，诞生了适用于不同场景的服务器类型：通用服务器、云计算服务器、边缘计算服务器、AI 服务器。随着大数据、云计算、人工智能及物联网等网络技术的普及，充斥在互联网中的数据呈现几何倍数的增长，使得以 CPU 为主要算力来源的传统服务器承受着越来越大的压力，并且对于目前 CPU 的制程工艺而言，单个 CPU 的核心数已经接近极限，但数据的增加却还在继续，因此服务器数据处理能力必须得到新的提升，在这种环境下，AI 服务器应运而生。面对 ChatGPT 所引出的大规模预训练模型，AI 服务器以其架构优势带来的大吞吐量特点，有望在一众服务器中脱颖而出。

表 1：主流服务器类型（不完全统计）

	特点	配置	应用场景
通用服务器	物理服务器，独立存在，拥有完全管理员权限和独立 IP 地址，安全稳定性高。	CPU、硬盘、内存等。	
云计算服务器	通过虚拟化技术，将一台/多台服务器虚拟化成一个可以切分的资源池，客户按需灵活配置与扩展，管理便捷，费用相对低廉。	按客户需求配置 CPU、内存、数据盘等。	适合对业务弹性扩展需求和易用性的需求：电商、IT 行业、教育、移动应用、游戏等。
边缘计算服务器	承担 70%以上的计算任务，需支持 ARM/GPU/NPU 等异构计算，针对不同业务场景开发，远程控制运维。		工业互联网、车联网、医疗保健、AR/VR、智慧城市等。
AI 服务器	采用异构形式服务器，承担大量计算；大规模并行运算、多重向量/张量运算、计算效率高。	GPU/FPGA/ASIC 等加速芯片、CPU、内存等。	融合深度学习、机器视觉、知识图谱等人工智能技术的应用：医疗影像智能分析、人脸/语音/指纹识别、安防监控场景等。

资料来源：人工智能与创新公众号、皖云数科公众号、物联网世界官网、高升数据公众号、机智云物联网公众号、天风证券研究所

1.2. AI 服务器与普通服务器相比有着明显的性能优势

AI 服务器相较于传统服务器算力上大幅跃升。AI 服务器利用 CPU+的架构模式，CPU 仍作为 CPU 的数据处理主要模块，同时植入并行式计算加速部件，如 ASIC、FPGA、GPU 等，负责人工智能计算负载加速。总而言之，在 CPU+架构下，AI 服务器的技术选型和部件配置针对不同的业务场景做相应的调整优化，通过合理的负载分担实现计算能力的提升。

表 2：AI 服务器与普通服务器相比具有更好的技术优势

	AI 服务器	普通服务器
卡的数量	以加速卡为主导，基础要求为四块以上的 GPU 卡，甚至需要搭建外部服务器作为支持。	以 CPU 为主导，单卡/双卡 CPU。
P2P 通讯	GPU 卡间需要大量的参数通信，模型越复杂，通信量越大： SXM3 协议下，P2P 带宽高值 300GB/s； SXM2 协议下，P2P 带宽高值 50GB/s； PCI3.0 协议下，P2P 带宽高值 32GB/s。	普通 GPU 服务器一般只要求单卡性能。
特有设计	全面考虑对存储、通信、网络等相关领域的技术方案进行合理配置，使之与计算部件的计算能力相匹配，避免出现性能瓶颈。	-
专有技术	Purley 平台更大内存带宽； NVlink 提供更大的互联带宽； TensorCore 提供更强的 AI 计算力。	-

资料来源：南京锟前官网，天风证券研究所

1.3. AI 服务器的进步源于芯片，未来将关注更多功能性

我们认为 AI 服务器的技术迭代取决于硬件中 AI 芯片的选择。传统普通服务器数据处理核心单一，以 CPU 为主，AI 服务器则采取 CPU+的异构方式完成数据处理能力提升。参考英伟达 HGX 的产品迭代路径，核心差异在于搭载的 GPU，HGX-1、HGX-2 和 HGX A100 分别对应 Tesla P100 GPU、Tesla V100 和 A100，在深度学习和高性能计算领域都展现出更高的性能。因此，我们认为可以从 AI 芯片发展窥见 AI 服务器的技术迭代历程，主要以 CPU+GPU 为主的 AI 服务器，结合云服务趋势发展至 CPU+GPU/ASIC/FPGA 的多功能+优性能+灵活性+高性价 AI 服务器。

表 3：AI 服务器的技术迭代核心在于硬件架构

阶段性进展	硬件架构发展路径
汇聚主流 2010-2014	2010 年后随着云计算、大数据为代表的新一代信息技术高速发展并逐渐开始普及，云端采用“CPU+GPU”混合计算模式，推动人工智能算法的演进和人工智能芯片的广泛使用。
百花齐放 2015-2020	随着以深度学习为核心的人工智能技术得到全球范围内的关注，业界对人工智能算力的要求越来越高。更多高性能的定制化芯片开始出现，部署不同场景的服务器中，2015 年 Google 推出基于 ASIC 架构的 TPU 用于云端训练和推理，2017 年微软发布基于 FPGA 芯片组建的 Project Brainwave 低时延深度学习系统，2018 年亚马逊发布高性能推理芯片 AWS Inferentia 等。
融合发展 2021-至今	上云需求+数据规模膨胀，云原生+人工智能服务器逐渐成为一种趋势，既能够应对人工智能计算密集型工作，也能兼顾高弹性、高敏捷和高性价比的优势。

资料来源：《人工智能芯片产业技术发展研究》商惠敏，澎湃，驱动科技公众号，Kyligence 公众号，天风证券研究所

结合整个人工智能技术和服务的发展，未来人工智能服务器将重点发展软硬件平台融合、低功耗设计、智能边缘计算等领域。特别是随着量子计算、类脑芯片等新一代人工智能计算加速技术的兴起，人工智能服务器的设计与实现可能会产生颠覆性变化。总之，人工智能服务器作为提供计算能力的核心要素，都是人工智能研发应用体系中不可或缺的组成部分，它的良性发展势必会为整个人工智能产业的成长奠定坚实的基础。

表 4：AI 服务器未来发展方向

发展趋势	主要内容
软硬件融合	通过更多考虑软硬件协同，以系统应用为导向驱动芯片等硬件设计，让用户得到更好的体验。
低功耗发展	在节能减排要求下，重点探索液冷技术以实现更高的散热效率和更少的电能使用。
推理工作负载	伴随企业人工智能应用成熟度逐步递增，企业将人工智能训练负载转移到推理工作负载，这意味着人工智能模型将逐步进入广泛投产模式。

资料来源：半导体行业观察公众号，《2022-2023 中国人工智能计算力发展评估报告》IDC 与浪潮信息，天风证券研究所

2. 庞大算力需求是 AI 服务器未来放量的关键驱动力

2.1. AI 模型优化迭代引发算力缺口，未来推理功能将需更多算力

计算力是人工智能研究的生产力。在当今以深度学习为中心的人工智能发展中，AI 模型的进步主要依赖于模型的规模化扩展。在 AlexNet 网络模型出现后，ResNet、Transformer、BERT 等优秀模型也在不断涌现，尤其在图像、语音、机器翻译、自然语言处理等领域带来了跨越式提升。AI 模型智能程度不断发展的同时，AI 模型的数据量、结构的复杂程度也在不断增加，其带来了模型的参数量增长，模型尺寸呈指数级增加。随着模型尺寸不断膨胀，实现高效 AI 模型训练的重要支撑即更快的算力，即在短时间内完成大规模 AI 计算。

表 5：2018-2020 年模型升级带来的参数量激增

时间	公司	模型	参数量（亿）
2018	Open AI	GPT-1	1.5
2018	Google	Bert-Large	3.4
2019	微软	MT-DNN	3.3
2019	Open AI	GPT-2	15.42
2019	阿里巴巴	PERSEUS-BERT	1.1
2019	NVIDIA	Pmojert Megatron	83
2019	Facebook	RoBERTa	3.35
2019	Facebook	XILM	6.65
2019	NVIDIA	Megaltron-Scaled Version of OpenAI GPT-2	83
2020	微软	NLG	172
2020	Open AI	GPT-3	1750

资料来源：ittbank 公众号，天风证券研究所

AI 训练模型的需求或将释放 AI 服务器海量产能。根据 OpenAI 在 2020 年发表的论文，训练阶段算力需求与模型参数数量、训练数据集规模等有关，且为两者乘积的 6 倍：训练阶段算力需求=6 × 模型参数数量 × 训练集规模。

GPT-3 模型参数约 1750 亿个，预训练数据量为 45 TB，折合成训练集约为 3000 亿 tokens。即训练阶段算力需求=$6 \times 1.75 \times 10^{11} \times 3 \times 10^{11}=3.15 \times 10^{23}$ FLOPS=3.15×10^{8} PFLOPS

依据谷歌论文，OpenAI 公司训练 GPT-3 采用英伟达 V100 GPU，有效算力比率为 21.3%。GPT-3 的实际算力需求应为 1.48×10^{9} PFLOPS(17117 PFLOPS-day)。

假设应用 A100 640GB 服务器进行训练，该服务器 AI 算力性能为 5 PFLOPS，最大功率为 6.5 kw，则我们测算训练阶段需要服务器数量=训练阶段算力需求÷服务器 AI 算力性能=2.96×10^{8} 台（同时工作 1 秒），即 3423 台服务器工作 1 日。

表 6：训练所需服务器测算

项目	值

参数量	1750 亿
预训练数据量	45TB
算力需求	$3.15*10^8$ PFLOPS
有效算力比率	21.3%
实际算力需求	$1.48*10^9$ PFLOPS
A100 服务器算力性能	5 PFLOPS
工作 1 天所需服务器（台）	3423

资料来源：英伟达公司官网，天翼智库微信公众号，天风证券研究所

后期推理侧算力需求将成为主力。据 IDC 数据，2021 年中国人工智能服务器工作负载中，57.6%的负载用于推理，42.4%用于模型训练，预计至 2026 年 AI 推理的负载比例将进一步提升至 62.2%。具体落到 ChatGPT 的算力需求场景中，可以将预训练、Finetune 归类为训练，日常运营归类为推理。ChatGPT 的更新迭代已展示出对算力的高需求：从训练端来看，GPT1-3 经历了从 1.5 亿到 1750 亿参数的增长过程，预计需要 3640PFlop/s-day；从推理端来看，据 SimilarWeb 数据，2023 年 1 月 ChatGPT 官网日访问量突破千万级别，1 月 31 日达到 2800 万访问量，以 3.4%的日增速度持续扩展，其中独立访客数量为 1570 万。据 CNBC，达到 1 亿级别活跃用户将产生成本约 4000 万美元，则该 1570 万用户所耗费互动成本约为 628 万美元，推算得相对算力处理量为 5714.8 PFlop/s-day，随着新用户不断进入并多次访问，该算力需求将不断推高，是未来算力的重点需求端。

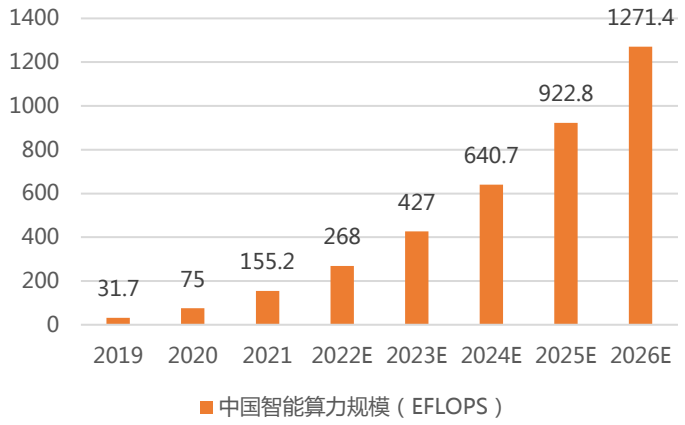
表 7：以 ChatGPT 产品测算所需算力

应用场景	ChatGPT 具体应用	功能	算力要求（PFlop/s-day）（以 GPT-3 模型为例）
训练	预训练	通过大量无标准的纯文本数据，训练模型基础语言能力，得到 GPT-1/2/3 基础大模型。	3640
	Finetune	在预训练的大模型基础上，进行监督学习、强化学习等，实现对模型参数量的优化调整。	-
推理	日常运营	基于用户的输入进行推理计算，并输出反馈结果。	5714.8

资料来源：《2022-2023 中国人工智能计算力发展评估报告》IDC 与浪潮信息，Similarweb，ittbank 公众号，蓝海大脑官网，天风证券研究所

国内大厂在 AI 元年坚定入局+国内算力基建化，或将成为国内 AI 服务器需求来源。在 ChatGPT 带动的浪潮中，国内大厂开始密切发布类 ChatGPT 产品，3 月 16 日百度正式推出国内首款生成式 AI 产品“文心一言”、3 月 30 日腾讯推出“混元”AI 大模型、4 月 7 日阿里的“通义千语”开启内测邀请，华为 4 月 8 日推出盘古大模型等。据 IDC 数据，2021 年中国智能算力规模为 155.2 EFLOPS，随着 AI 模型日益复杂、计算数据量快速增加、人工智能应用场景不断深化，未来国内智能算力规模将持续增长，预计 2026 年智能算力规模将达 1271.4 EFLOPS，5 年 CAGR 达到 52.3%。对于国内人工智能算力需求的高增长，我国正牵头东数西算、智能计算中心，通过算力基础设施完成从点到网的升级，据 IDC 观点，中国市场的人工智能硬件支出占人工智能总支出的比例将在未来 5 年保持在 65%左右，其中服务器是硬件中的重要部分。我们认为，AI 服务器作为智能算力运算的主要载体，看好未来持续放量的高预期走势。

图 2：中国智能算力规模

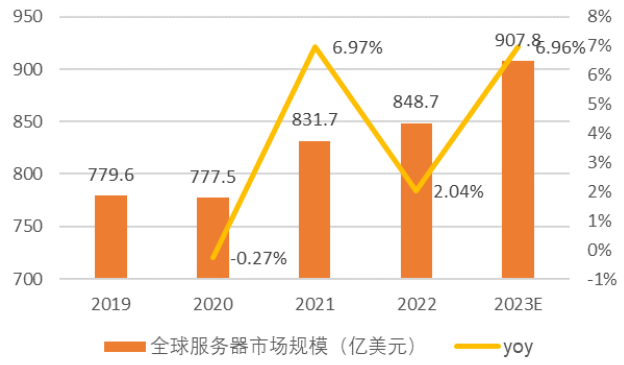


资料来源：《2022-2023 中国人工智能计算力发展评估报告》IDC 与浪潮信息，贵州省大数据发展管理局，天风证券研究所

2.2. AI 服务器作为承载算力主体将受益，成服务器市场的增长主力

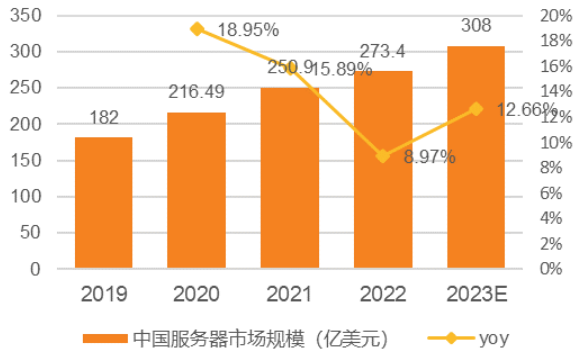
AI 服务器市场扩张表现亮眼，中国市场领跑全球。根据 Statista 数据，2021 年全球服务器市场规模达到 831.7 亿美元，同比增长 6.97%，其中 AI 服务器市场达到 156.3 亿美元，同比增长 39.1%，在整体服务器市场中占比 18.79%，同比提升 4.34pct，这主要系随着人工智力所需算力扩大，AI 服务器作为新型算力基础设施的主体将直接影响 AI 创新迭代和产业落地，我们认为市场将需更多、更强算力的 AI 服务器作为核心解决方案，据 IDC 与浪潮信息，预计 AI 服务器市场将在 2026 年达到 347.1 亿美元，5 年 CAGR 达到 17.3%。再看中国市场，2021 年国内服务器市场规模达到 250.9 亿美元，同比增长 15.9%，高于全球增长速度；其中 AI 服务器市场达到 59.2 亿美元，同比增长 68.2%，在国内服务器市场中占比 23.6%，这主要得益于国内人工智能应用的加速落地，浪潮信息、新华三、宁畅等厂商助推人工智能基础设施产品的优化更新，IDC 调研显示，超过 80%的中国厂商表示在未来将增加人工智能服务器的投资规模，预计在 2026 年中国 AI 服务器市场规模将达到 123.4 亿美元，5 年 CAGR 为 15.82%。

图 3：全球服务器市场规模及增速



资料来源：Statista，天风证券研究所

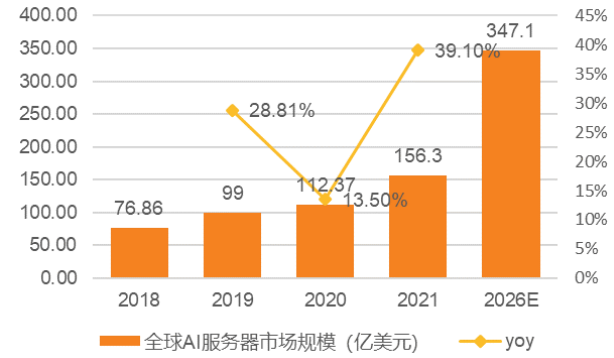
图 4：中国服务器市场规模及增速



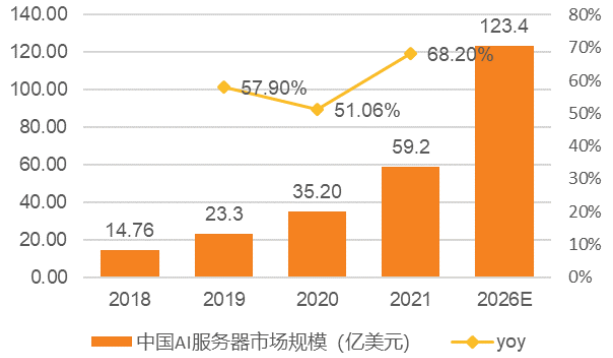
资料来源：中商产业研究所公众号，天风证券研究所

图 5：全球 AI 服务器市场规模及增速

图 6：中国 AI 服务器市场规模及增速



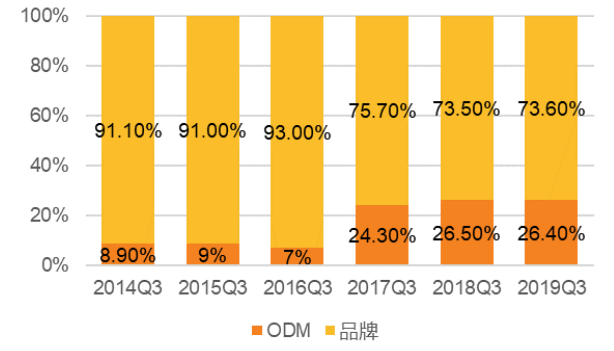
资料来源：量子位公众号，《2022-2023 中国人工智能计算力发展评估报告》IDC 与浪潮信息，天风证券研究所



资料来源：量子位公众号，《2022-2023 中国人工智能计算力发展评估报告》IDC 与浪潮信息，天风证券研究所

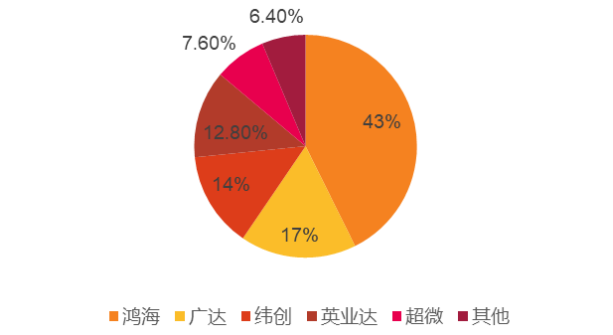
全球服务器行业生产模式趋向白牌化，ODM 厂商市占率集中。服务器厂商可分为 ODM 厂商和品牌厂商两种类型。品牌服务器厂商以浪潮、华为、新华三为代表，ODM 厂商以超微、广达为代表，根据品牌服务器厂商的委托完成硬件生产，加贴委托方商标并交付给品牌持有者进行销售，近年不少客户会绕过品牌商向 ODM 厂商直接订购服务器成品，白牌服务器生产模式的兴起对传统品牌服务器厂商造成一定冲击。在下游业务业务+成本+时间的需求考虑下，Facebook 在 2011 年率先主导成立 OCP 联盟，向内部成员统一硬件标准与硬件开源，成员中的白牌厂商开始获得服务器设计方案，为白牌厂商绕过品牌厂商提供技术支持，这将进一步推高服务器白牌厂商的市占率。据重磅数据，ODM 市场份额从 2014Q3 的 6.6%快速攀升至 2019Q3 的 26.4%，根据 2022 年全球服务器销售量（848.7 亿美元）与全球服务器代工市值（4557 亿元）测算，2022 年出货的服务器中 ODM 生产市占率或达 50%，我们认为这主要得益 CSP 客户（如微软、谷歌）的发展，服务器需求随着业务成长而持续增长，采购量不断增加。根据 Digitimes Research，全球 ODM 厂商竞争格局集中度高，CR5 高达 94.4%，分别为鸿海（43%）、广达（17%）、纬创（14%）、英业达（12.8%）和超微（7.6%），主要客户涵盖惠普、戴尔、联想、亚马逊、微软、谷歌等国际巨头。随着超大型资料中心需求强劲+下游云计算领域 Capex 扩大，造就白牌厂商的强劲发展势头。

图 7：ODM 和品牌服务器厂商市占率



资料来源：重磅数据，天风证券研究所

图 8：预计 2022 年 ODM 厂商竞争格局

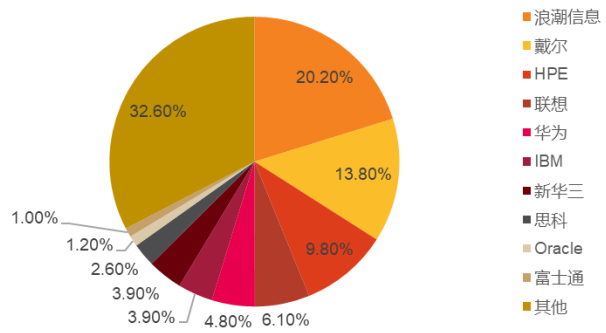


资料来源：芯智讯，天风证券研究所

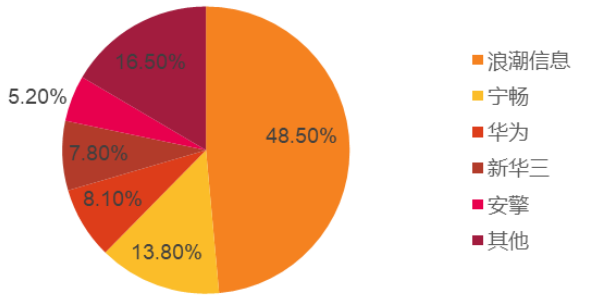
国内服务器品牌厂商在 AI 服务器市场占优，未来有望放大优势。从全球 AI 服务器市场格局来看，2021 年上半年数据显示，中国厂商在 TOP5 厂商中占据过半席位（浪潮、联想、华为），累计占比为 31.1%，其中浪潮信息以 20%的市占率占据榜首。从国内 AI 服务器市场竞争格局来看，市场集中度较高，浪潮信息占据近半市场，CR5 超过 80%。未来，随着国产厂商在 AI 服务器的持续深耕，有望在既有市场优势中吸收更大来自 AI 产业的发展机遇。对比全球/中国 AI 服务器竞争格局，我们认为国产服务器厂商在 AI 服务器市场优势较为明显，未来有望利用既有市场优势+推出具有竞争力的 AI 服务器进一步拓宽市场份额，吸收来自 AI 产业的发展机遇，有望以技术优势打造盈利护城河，在国际市场上争取更高话语权。

图 9：2021 年 H1 全球 AI 服务器竞争格局

图 10：2021 年 H1 中国 AI 服务器竞争格局

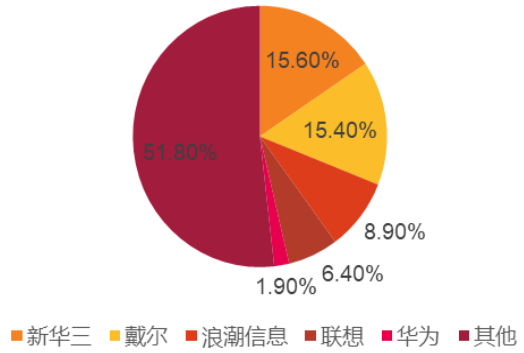


资料来源：机器人技术与应用公众号，天风证券研究所



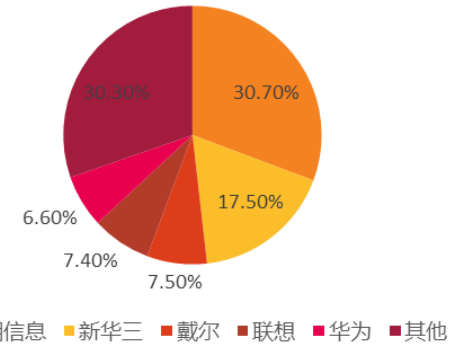
资料来源：IDC 公众号，天风证券研究所

图 11：2021 年全球服务器竞争格局



资料来源：智研咨询公众号，天风证券研究所

图 12：2021 年中国服务器竞争格局



资料来源：智研咨询公众号，天风证券研究所

表 8：AI 服务器厂商

公司	主要情况
浪潮信息	浪潮信息是全球领先的 IT 基础设施产品、方案和服务提供商，通过场景优化设计形成了丰富的产品线，涵盖计算型、存储型、多节点、关键应用、整机柜等各类服务器，支持全场景高效计算。2022 年上半年，浪潮信息服务器出货量保持全球第二，中国第一。
工业富联	工业富联在云计算服务器出货量持续全球第一，与全球主要服务器品牌商、国内外 CSP 客戶深化合作，推出新一代云计算基础设施解决方案，包括模块化服务器、高效运算 HPC 等，重点解决因 ChatGPT 持续升温而引发 AIGC 算力井喷需求。公司多年来为数家第一梯队云服务商 AI 服务器（加速器）与 AI 存储器供应商，产品已经开发至第四代。伴随着 AI 硬件市场迅速成长，公司相关产品 2022 年出货加倍，AI 服务器及 HPC 出货增长迅速，在 2022 年云服务商产品中，占比增至约 20%，持续维持增长态势。
纬颖科技	纬颖科技专注于提供超大型数据中心及云端基础构架各项产品及系统的解决方案，藉由传承纬创资通丰富的创新设计技术，以及多年的制造经验，纬颖科技的顶尖团队为客户提供巨型数据中心、运营商及企业的云计算解决方案，其中包括高性价比的服务器、储存设备、网络系统、机房基础设备及软硬体整合的私有云解决方案。
中兴通讯	中兴通讯服务器产品拥有该领域多项自主知识产权及专利，涵盖机架服务器、刀片服务器和云计算节点服务器，满足互联网、高性能计算、数据库、数据中心等多种场景需求，并配合中兴通讯其它产品形成丰富的产品组合，满足客户的各种业务需要。服务于政府、金融、交通、教育、医疗、电信运营商和互联网等多个行业。
中科曙光	中科曙光以 IT 核心设备研发、生产制造为基础，对外提供高端计算机、存情、安全数据中心等 ICT 基础设施产品，大力发展云计算、大数据、人工智能、边缘计算等先进计算业务，为用户提供全方位的信息系统服务解决方案。主要产品为高端计算机、存储产品、软件开发、系统集成、技术服务。
超聚变	超聚变聚焦发展算力与生态，致力成为全球领先的算力基础设施与服务提供者，凭借在多样性算力、液冷技术等方面的优势，成功入选 Gartner 全球服务器代表性厂商名录。
新华三	新华三集团作为数字化解决方案领导者，致力于成为客户业务创新、数字化转型值得信赖的合作伙伴。作为紫

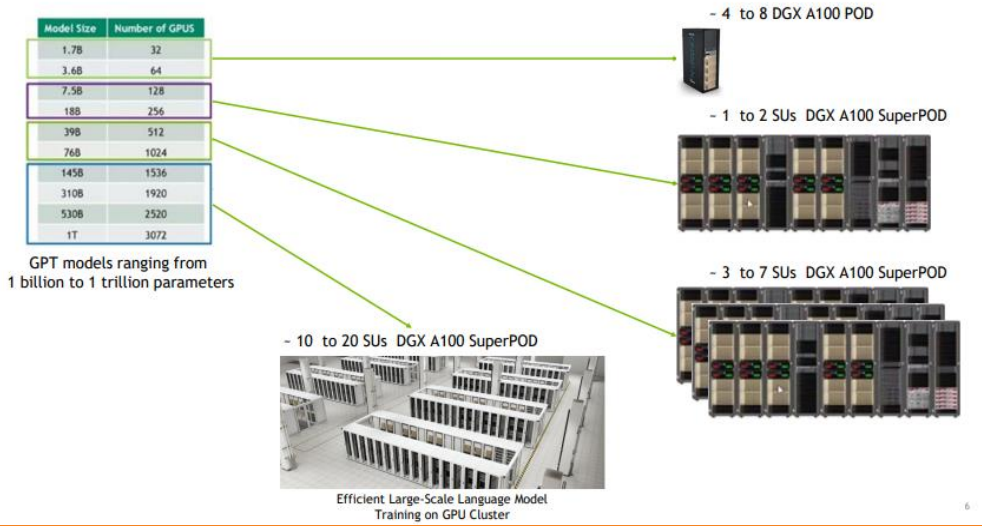
	光集团旗下的核心企业，新华三通过深度布局“云-网-算-存-端”全产业链，不断提升数字化和智能化赋能水平。新华三拥有计算、存储、网络、5G、安全、终端等全方位的数字化基础设施整体能力，提供云计算、大数据、人工智能、工业互联网、信息安全、智能联接、边缘计算等在内的一站式数字化解决方案，以及端到端的技术服务。同时，新华三也是 HPE®服务器、存储和技术服务的中国独家提供商。
宁畅	宁畅信息产业（北京）有限公司是集研发、生产、部署、运维一体的服务器厂商，及 IT 系统解决方案提供商，为全行业客户提供基于 X86 架构通用机架、人工智能、多节点、边缘计算及 JDM 全生命周期定制等多类型服务器及 IT 基础设施产品。
英业达	英业达自 1975 年成立以来，从早期制造计算机、电话机，而后制造笔记本电脑与服务器，奠定了公司扎实稳固的基础。英业达企业电脑事业群(EBG)成立于 1998 年，专注服务器研发制造。英业达数据中心解决方案传承 EBG，专注于提供超大型数据中心、运算密集行业包括互联网及电信运营商最佳解决方案，领先研发制造能力深获客户信任。
广达	广达电脑拥有完整及领先业界的数据中心产品线，提供各种数据中心所需的 IT 硬件设备，包括服务器、存储设备及网络交换机等，从产品研发、制造、整合到优化，为客户提供最全面的一站式服务，以丰富经验协助客户解决下一代数据中心于设计及运营上的挑战，并为云端服务运营商、电信运营商以及企业用户建构公有云、混合云与私有云。
拓维信息	拓维信息作为中国领先的软硬一体化产品及解决方案提供商，依托自身 10 多年的行业云服务能力，基于自主研发的行业 PaaS 平台，为政企客户提供 SaaS 化产品、云化解决方案、智能硬件、咨询规划、项目设计、运维服务、信息安全等一站式拓维云服务，目前业务已覆盖全国近 20 个省市，覆盖了政府、教育、通信、金融、制造、交通等十多个领域。
中国长城	中国长城科技集团股份有限公司是中国电子信息产业集团有限公司旗下“安全、先进、绿色自主计算产业专业子集团”，是中国“PKS”自主计算体系建设主力军和网信科技自主创新生力军。中国长城持续聚焦自主计算产业、系统装备核心主业。中国长城坚持“芯端一体，双核驱动”的发展战略，致力于构建以“芯-端”为核心的自主计算产品链，全面带动“网-云-数-智”自主计算产业生态发展。自重组以来成功突破高端通用芯片（CPU）、固件等关键核心技术，依托“PKS”自主计算体系，构建了从芯片、台式机、笔记本、服务器、网络交换设备到应用系统等具有完整自主知识产权的产品谱系，成功赋能党政办公及金融、能源、电信、交通等重点信息化领域的数字化转型。
宝德科技	宝德计算机系统股份有限公司是以服务器和 PC 整机的研发、生产、销售及提供相关的综合解决方案为主营业务的国家级高新技术企业和国家专精特新“小巨人”企业，致力于成为中国领先的计算产品方案提供商，为政府、互联网、教育、金融、电力、交通、医疗、运营商、安平等行业客户持续提供先进的算力产品、解决方案和全栈服务。多年来，凭借不断创新的产品技术和独特的软硬件综合实力，宝德计算勇夺信创市场 NO.1，稳居 X86 服务器国内品牌 TOP5 和全球 TOP9、中国 AI 服务器 NO.3。

资料来源：各公司官网、公告、东莞台博会公众号，ifind，wind，天风证券研究所

2.3. Nvidia 高速互联助力 AI 运算，多 GPU 通信成为关键技术

DGX POD 是 NVIDIA 的 AI 基础架构，主要由运算、HCA（网络适配器）、交换机组成。1）运算 NVIDIA DGX Systems：主要使用 DGX A100 和 DGX H100 Systems；2）HCA：NVIDIA DGX H100 系统配备了 NVIDIA ConnectX-7 HCA。NVIDIA DGX A100 系统配备 ConnectX-7 或 ConnectX-6 HCA；3）交换器：在系统之间提供多条高带宽、低延迟的路径，DGX BasePOD 配置可以配备四种类型的 NVIDIA 网络交换机。配置需求以 GPT 模型为例，参数范围从 10 亿到 1 万亿个，会有不同的配置需求；17~36 亿个参数预计需要 32~64 颗 GPU，大约需要 4 到 8 个 DGX A100 POD，若达到 1450 亿个参数预计需要 1536 颗 GPU，需要在 GPU 集群上进行高效的大规模语言模型训练，需要 10 到 20 SUs 的 DGX A100 SuperPOD，其中每个 SU 由 20 个 DGX A100 系统组成。

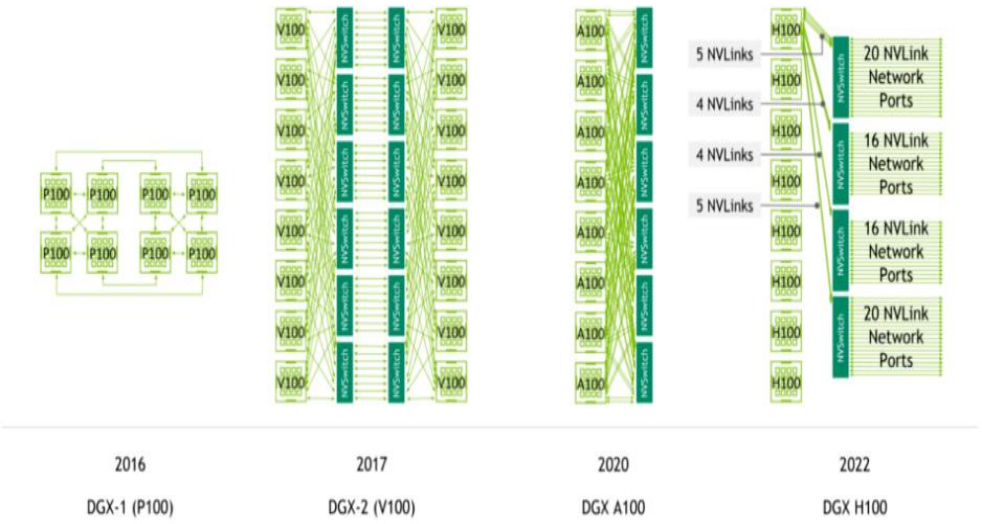
图 13：GPT 模型参数对应 GPU 数量与 DGX A100 POD 配置



资料来源：联想凌拓公司官网，天风证券研究所

NVIDIA DGX Systems 实现服务器中所有 GPU 之间的高带宽、任意连接。2016 年 Nvidia 在 GPU 技术大会上推出全球首款深度学习超级计算机 NVIDIA DGX-1，实现了与硬件、深度学习软件和开发工具的全面集成，为深度学习提供每秒高于 21 万亿次浮点运算峰值性能的新型半精度指令技术。2017 年 NVIDIA NVSwitch 与 NVIDIA V100 Tensor Core GPU 和第二代 NVLink 一起推出。2020 年 NVIDIA A100 Tensor Core GPU 引入了第三代 NVLink 和第二代 NVSwitch，使每 CPU 带宽和减少带宽都增加了一倍。2022 年使用第四代 NVLink 和第三代 NVSwitch，具有八个 NVIDIA H100 Tensor Core GPU 的系统具有 3.6 TB/s 的二等分带宽和 450 GB/s 的缩减操作带宽;与上一代相比,这两个数字分别增加了 1.5 倍和 3 倍。此外，使用第四代 NVLink 和第三代 NVSwitch 以及外部 NVIDIA NVLink 交换机，现在可以实现 NVLink 速度跨多台服务器进行多 GPU 通信。

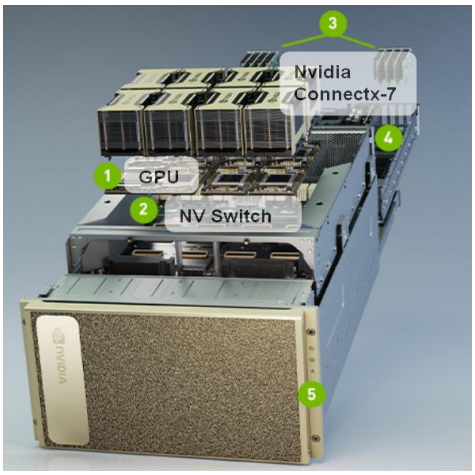
图 14：NVIDIA DGX Systems



资料来源：英伟达公司官网，天风证券研究所

NVIDIA DGX A100 配置：8 个 NVIDIA A100 GPU 搭载共 640 GB GPU 内存，每个 GPU 使用 12 个 NVLink，GPU 至 GPU 带宽每秒 600 GB，6 个第二代 NVSwitch 双向带宽每秒 4.8 TB 比前一代高出 2 倍。10 个 NVIDIA ConnectX-7 每秒 200GB 网络接口每秒 500 GB 的双向带宽峰值。

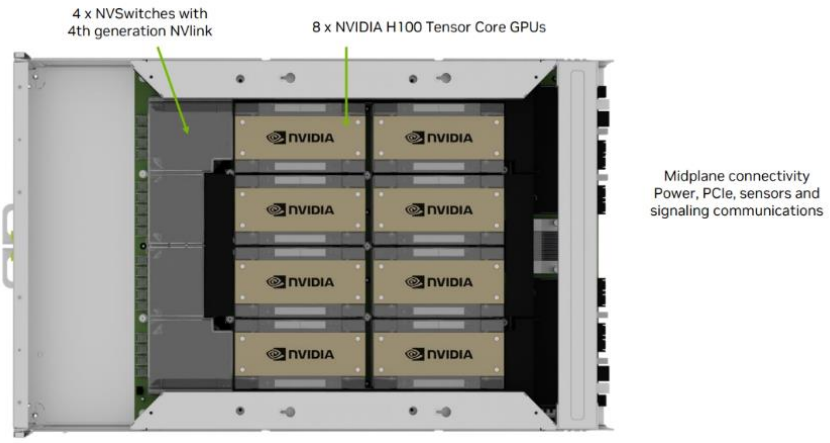
图 15：NVIDIA DGX A100 配置



资料来源：英伟达公司官网，天风证券研究所

NVIDIA DGX H100 配置：8 个 NVIDIA H100 GPU，总 GPU 内存高达 640GB，每个 GPU 都拥有 18 个 NVIDIA NVLink，提供每秒 900GB 的 GPU 至 GPU 双向带宽，4 个 NVIDIA NVSwitch，每秒 7.2TB 的 GPU 双向带宽，比前一代快 1.5 倍。8 个 NVIDIA ConnectX-7 和 2 个搭载每秒 400Gb 网络适配器的 NVIDIA BLUEFIELD DPU。

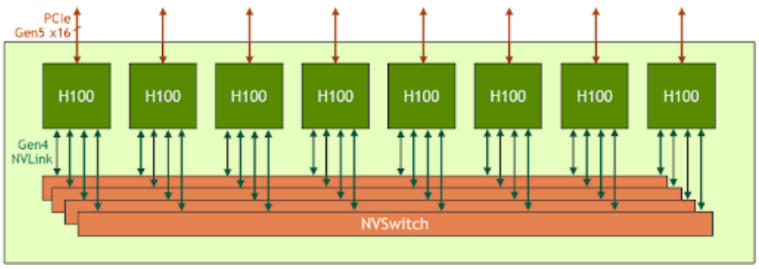
图 16：NVIDIA DGX H100 配置



资料来源：gdep 官网，天风证券研究所

DGX H100 中 8 个 GPU+NVLink+NVSwitch 是关键的组成部分。DGX H100 拥有八个 H100 张量核 GPU 和四个第三代 NV 交换机。每个 H100 GPU 都有多个第四代 NVLink 端口，并连接到所有四个 NVLink 交换机。每个 NVSwitch 都是一个完全无阻塞的交换机，完全连接所有八个 H100 Tensor Core GPU。

图 17：NVIDIA DGX H100（H100、NVLink、NVSwitch 配套）

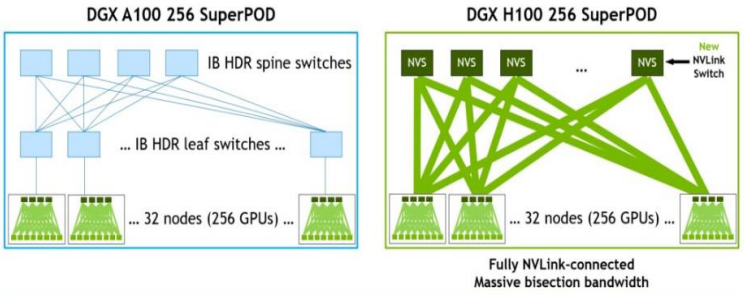


资料来源：英伟达公司官网，天风证券研究所

DGX H100 可以支持 256 个 GPU 连接，并提供 57.6TB 的带宽。将新的 NVLINK Network 技术与新的第三代 NVSwitch 结合，而每一个 GPU 节点都会公开节点中 GPU 之所有 NVLink 带宽的 2:1 锥形层级。节点是通过 NVLink Switch 模块中包含的第二层 NVSwitch 连接在一

起，这些模块常驻于运算节点外部，并将多个节点连接在一起。DGX H100 SuperPod 可以横跨多达 256 个 GPU，连接之节点可以提供 57.6 TB 的全部对全部带宽，使用以第三代 NVSwitch 技术为基础的新型 NVLink Switch，通过 NVLink Switch System 完全连接。

图 18：DGX A100 与 DGX H100 32 节点 256 GPU NVIDIA SuperPOD 架构比较

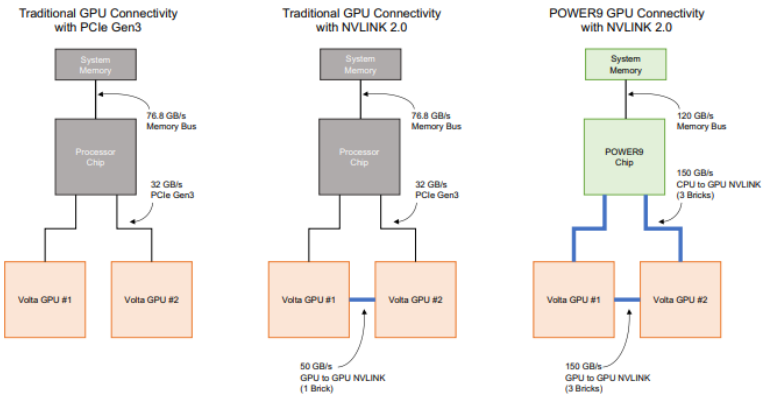


	A100 SuperPod			H100 SuperPod			Speedup	
	Dense PFLOP/s	Bisection [GB/s]	Reduce [GB/s]	Dense PFLOP/s	Bisection [GB/s]	Reduce [GB/s]	Bisection	Reduce
1 DGX / 8 GPUs	2.5	2,400	150	16	3,600	450	**1.5x**	**3x**
32 DGXs / 256 GPUs	80	6,400	100	512	57,600	450	**9x**	**4.5x**

资料来源：英伟达公司官网，天风证券研究所

NVLink 相较于 PCIe 能提供更大的带宽。以第二代 Volta NVLink 为例，通道带宽为 300GB/s，PCIe 3.0 带宽为 16GB/s；以第四代 NVLink 为例，每个通道的带宽为 100 Gbps，是 PCIe Gen 5 的 32 Gbps 带宽的三倍多；此外通过组合多个 NVLink 以提供更高的聚合通道数，从而产生更高的吞吐量。第四代 NVLink 技术为多 GPU 系统配置提供 1.5 倍带宽，并改善可扩充性。单个 NVIDIA H100 Tensor 核心 GPU 最高可支持 18 个 NVLink 联机，总带宽可达每秒 900 GB(GB/秒)，将近是第 5 代 PCIe 带宽的 7 倍。

图 19：PCIe 与 NVLink 架构比较



资料来源：IBM Redbooks 官网，天风证券研究所

NVLink 高速通信不断迭代，第四代相较于第三代带宽提升 0.5 倍。为了将训练时间从数个月压缩至数天，需要在服务器集群中的每一个 GPU 之间进行高速无缝通讯。而 PCIe 因带宽有限而造成瓶颈，因此需要更快速、更具扩充性的 NVLink 互连。NVIDIA A100 Tensor 核心 GPU 采用的第三代 NVLink（包含 12 个第三代 NVLink 链路，提供每秒 600 GB 的总带宽），H100 GPU 采用新的第四代 NVLink（H100 包含 18 个第四代 NVLink 链路，提供每秒 900 GB 的总带宽），H100 GPU 相较于 A100 提供 1.5 倍的通讯带宽。

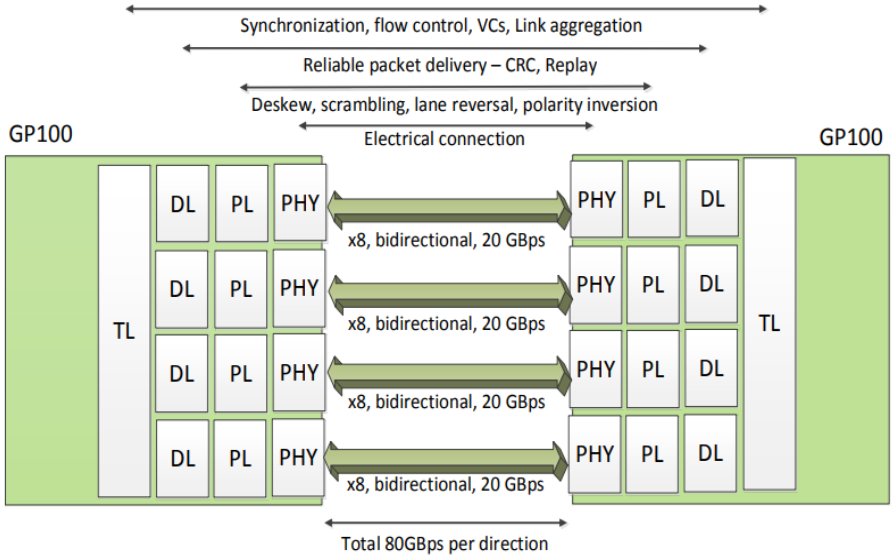
图 20：NVIDIA NVLink 迭代规格

	第二代	第三代	第四代
NVLink 頻寬大小總計	每秒 300GB	每秒 600GB	每秒 900GB
每 GPU 連結數量上限	6	12	18
支援的 NVIDIA 架構	NVIDIA Volta™ 架構	NVIDIA Ampere 架構	NVIDIA Hopper™ 架構

资料来源：英伟达公司官网，天风证券研究所

NVLink 由物理层（PHY）、数据连接层（DL）以及交易层（TL）组成。1）物理层 PL 与 PHY，负责跨所有八个信道并确保原始的数据可在各种物理媒体上传输；2）DL 数据连接层为位于物理层与网络层之间，在两个网络实体之间提供数据链路连接的建立、维持和释放管理；3）TL 交易层请求和响应信息形成的基础。

图 21：以 P100 为例，NVLink 组成架构



资料来源：英伟达公司官网，天风证券研究所

NVSwitch 是 NVLink 交换系统的关键，实现 GPU 高速跨节点连接。新的第三代 NVSwitch 技术包含常驻于节点内部和外部的交换器，可以在服务器、丛集和数据中心环境中连接多个 GPU。节点内部每一个新的第三代 NVSwitch 皆提供 64 个第四代 NVLink 链路端口，以加快多 GPU 联机能力。第三代 NVSwitch 是 NVLink 交换机系统的关键，以 NVLink 速度实现 GPU 跨节点的连接。

图 22：NVIDIA NVSwitch 参数

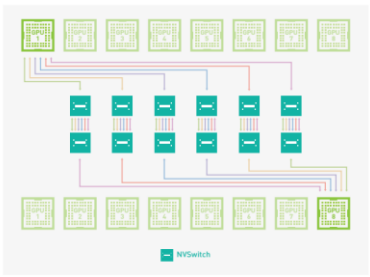
	第一代	第二代	第三代
直接互連/有節點的 GPU 數量	高達 8 個	高達 8 個	高達 8 個
NVSwitch GPU 至 GPU 頻寬	每秒 300GB	每秒 600GB	每秒 900GB
彙總頻寬大小總計	每秒 2.4TB	每秒 4.8TB	每秒 7.2TB
支援的 NVIDIA 架構	NVIDIA Volta 架構	NVIDIA Ampere 架構	NVIDIA Hopper 架構

资料来源：英伟达公司官网，天风证券研究所

NVSwitch 为定制工艺构建，并行运行增加互联传输效率。NVSwitch 芯片并行运行，以支持数量日益增加的 GPU 之间的互连，核心逻辑是让端口逻辑模块中的数据包转换，进出多

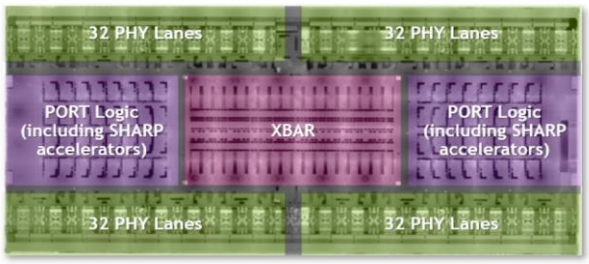
GPU 的流量看似是通过单一 GPU。随着 NVLink 交换机系统提供的带宽是 InfiniBand 的 4.5 倍，大规模模型培训变得更加实用；例如，当使用 14 TB 嵌入表训练推荐引擎时，与使用 InfiniBand 的 H100 相比，预计使用 NVLink 交换系统的 H100 在性能上会有显着提升。第三代 NVSwitch 它使用为 Nvidia 定制的 TSMC 4N 工艺构建，该芯片包含 251 亿个晶体管，比NVIDIA V100 Tensor Core GPU 的晶体管多，面积为 294 平方毫米封装尺寸为 50 mm x 50 mm，共有 2645 个焊球。

图 23：NVSwitch 拓扑图（以 16 个 GPU 为例）



资料来源：英伟达公司官网，天风证券研究所

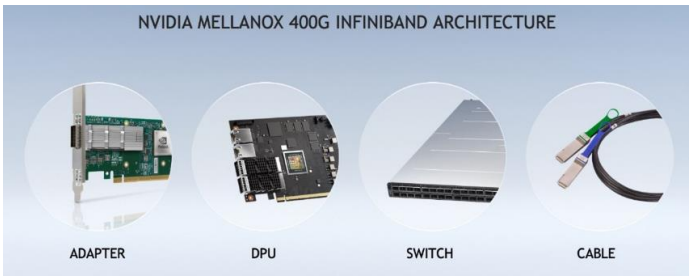
图 24：NVSwitch 芯片



资料来源：英伟达公司官网，天风证券研究所

NVIDIA Quantum InfiniBand 提供高效能运算网络解决方案。InfiniBand 和以太网持续竞争网路通信世界的主导地位。作为被最广泛使用的网路通讯协议，以太网拥有优秀的性价比以及和多数装置兼容的优势。然而，现今网路早已发展成一个更为庞大复杂的系统，大量的数据运算需求让人们开始关注 InfiniBand 架构的优势。InfiniBand 主机信道适配卡(HCA) 提供了超低延迟、极高传输量和创新的 NVIDIA 网络内运算引擎，以提供现代工作负载所需的加速、可扩充性和功能丰富的技术。2020 年 Nvidia 于 SC20 大会上，宣布推出 NVIDIA Mellanox 400G InfiniBand，这是全球首个 400Gb/s 网速的端到端网络解决方案，可为全球的人工智能(AI)和高效能运算用户提供最快的网络互连效能，同时成功将运算、可程序化和软件定义三种技术结合，成为业界领先的软件定义、硬件加速的可程序设计网络。

图 25：NVIDIA Mellanox 400G InfiniBand 组成



资料来源：英伟达公司官网，天风证券研究所

NVIDIA ConnectX InfiniBand 智能适配卡可运用更快的速度和创新的网络内运算。NVIDIA ConnectX 能降低营运成本，提升投资报酬率，实现高效能运算、机器学习、进阶储存空间、丛集数据库、低延迟嵌入式 I/O 应用程序等强大功能。ConnectX-7 智能主通道配接器 (HCA)采用 NVIDIA Quantum-2 InfiniBand 架构，可提供每秒 400GB 的吞吐量；ConnectX-6 HDR 智能主通道配接器(HCA)采用 NVIDIA Quantum InfiniBand 架构，可提供每秒 200GB 的吞吐量。

图 26：NVIDIA ConnectX-7 智能主通道配接器 (HCA)示意图

资料来源：深博公司官网，天风证券研究所

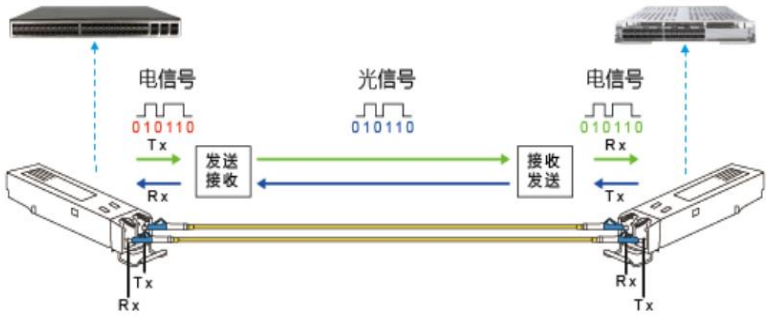
NVIDIA Quantum InfiniBand 交换器，提供庞大吞吐量、网络内运算的架构。QM8700 InfiniBand 系列，具备高达每秒 16Tb 的无阻塞带宽，提供多达 40 埠、每埠每秒 200Gb 的完整双向带宽。QM9700 InfiniBand 系列，拥有 64 个 400Gbps 端口或 128 个 200Gbps 端口，能以多种切换器系统的设置供货，在 400Gbps 下最高搭载 2,048 个端口，或于 200Gbps 下最高搭载 4,096 个端口。

图 27：NVIDIA Quantum-2 QM9700 Series 示意图

资料来源：Nvidia，天风证券研究所

光模块是实现光信号传输过程中光电转换和电光转换功能的光电子器件。光模块工作在 OSI 模型的物理层，是光纤通信系统中的核心器件之一。它主要由光电子器件（光发射器、光接收器）、功能电路和光接口等部分组成，主要作用就是实现光纤通信中的光电转换和电光转换功能。光模块工作原理图所示，发送接口输入一定码率的电信号，经过内部的驱动芯片处理后由驱动半导体激光器（LD）或者发光二极管（LED）发射出相应速率的调制光信号，通过光纤传输后，接收接口再把光信号由光探测二极管转换成电信号，并经过前置放大器后输出相应码率的电信号。

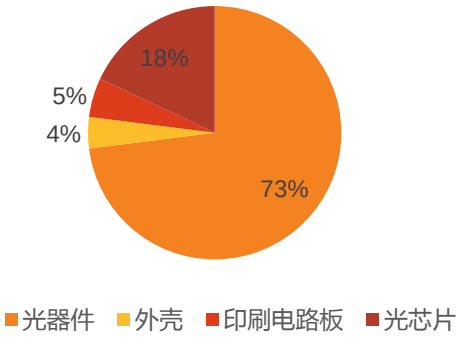
图 28：光模块工作原理



资料来源：Huawei，天风证券研究所

光模块中光芯片成本占比约 18%。光器件可分为有源、无源，其中光无源器件不需要外加能源驱动工作，是光传输系统的关节，光有源器件是光通信系统中将电信号转换成光信号或将光信号转换成电信号的关键器件。根据华经产业研究院数据，光模块成本中，光器件占 73%、外壳占 4%、印刷电路板占 5%、光芯片占 18%。

图 29：光模块成本结构



资料来源：华经产业研究院公众号，天风证券研究所

光芯片按功能可以分为激光器芯片和探测器芯片。光芯片采用光波（电磁波）来作为信息传输或数据运算的载体，一般依托于集成光学或硅基光电子学中介质光波导来传输导模光信号，将光信号和电信号的调制、传输、解调等集成在同一块衬底或芯片上。按功能可以分为：1）激光器芯片（发射信号），主要将电信号转化为光信号，按出光结构可进一步分为面发射芯片和边发射芯片，面发射芯片包括 VCSEL 芯片，边发射芯片包括 FP、DFB 和 EML 芯片；2）探测器芯片（接收信号），主要将光信号转化为电信号，主要有 PIN 和 APD 两类。

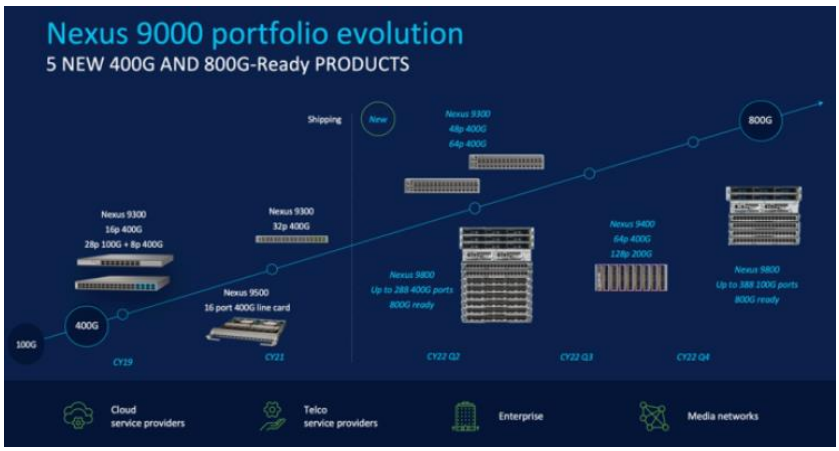
表 9：光芯片可分为激光器芯片和探测器芯片

种类	产品类别	工作波长	应用场景
激光器芯片	VCSEL	800-900nm	500 米以内短距离传输
	FP	1310-1550nm	中低速无线接入短距离市场
	DFE	1270-1610nm	中长距离传输
	EML	1270-1610nm	长距离传输
探测器芯片	PIN	830-860/1100-1600nm	中长距离传输
	APD	1270-1610nm	长距离单模光纤

资料来源：源杰科技招股书、中商产业研究院公众号，天风证券研究所

800G 交换机陆续发布，下一代超宽互联蓄势待发。2022 年 10 月思科在 OCP 全球峰会上发布了两款新的 800G 交换机系列及新的光模块，以支持超级数据中心运营商和电信运营商对更大的交换容量、灵活性和提升功效的要求。2023 年 1 月 Juniper 宣布在线媒体连接的 Virgin Media O2 成功使用 Juniper 网络升级了其 IP 核心骨干网络，Virgin Media O2 通过 Juniper 网络 PTX10008（PTX10008 路由器支持 400G，未来还可以通过硅创新和机箱中易于交换的线卡升级到 800G）分组传输路由器成功地将其在英国的六个骨干站点的所有核心流量迁移。2023 年 4 月新华三重磅发布 S12500G-EF 新一代绿色智能交换机，支持超宽 400G,未来可无缝升级 800G,为下一代超宽互联就绪。

图 30：思科第一款 800G Nexus

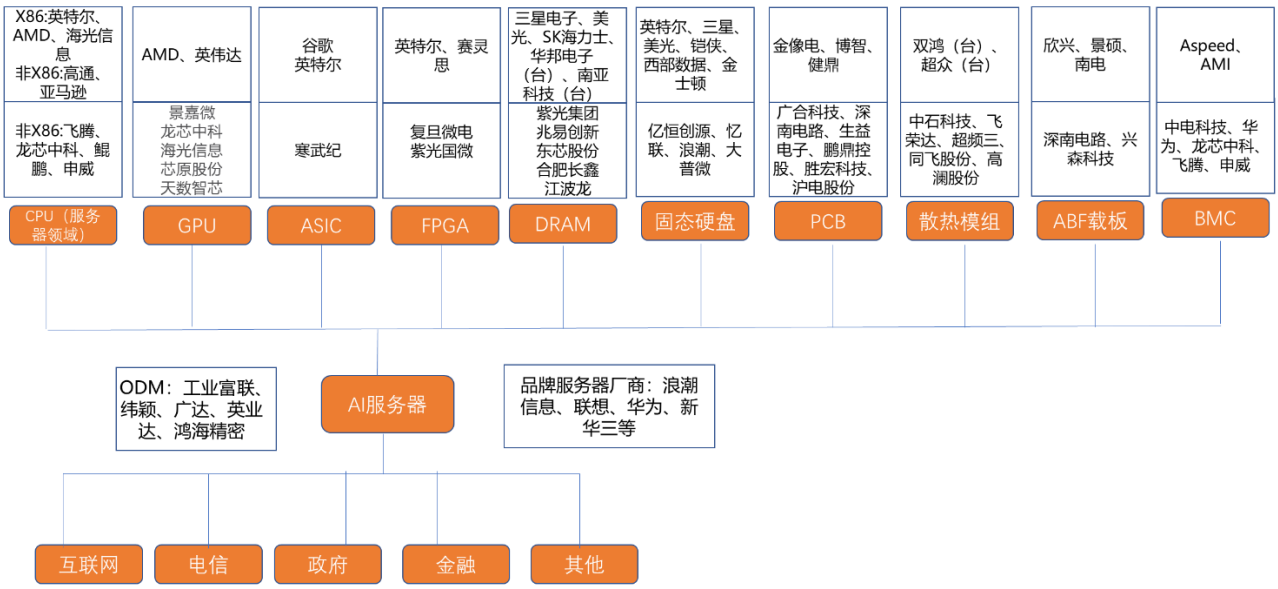


资料来源：思科、SDN LAB，天风证券研究所

3. AI 服务器放量预期利好上游核心部件，挑战机遇共存

各环节国产发展程度不一，机遇与挑战并存。我们认为，在算力和数字时代的背景下，AI 服务器作为算力载体为数字经济提供发展动力，更加彰显其重要性。纵观 AI 服务器的全景产业链，我们认为当下可以把握的机遇有 AI 服务器上游中部分电子元件环节（PCB/存储器），制约 AI 服务器发展的瓶颈主要在上游的 GPU 环节，未来有望突破的转折为 Chiplet 工艺。

图 31：相关产业链

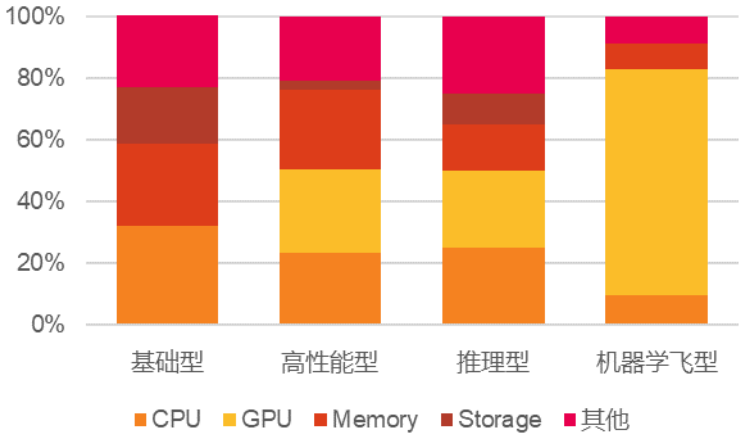


资料来源：富果研究部，中商产业研究院公众号，威尔克通信实验室公众号，果壳硬科技公众号，Wind，各公司公告等，天风证券研究所

3.1. 危机四伏：上游供应危机尚未解除，国产替代必需提上日程

拆解服务器的成本，芯片成本与性能高低成正比。据 IDC 2018 年关于服务器成本机构数据，芯片成本在基础性服务器中约占总成本的 32%，该比重随着性能和运算能力的要求而逐步攀升，在高性能或具有更强运算能力的服务器中，芯片相关成本可以高达 80%。

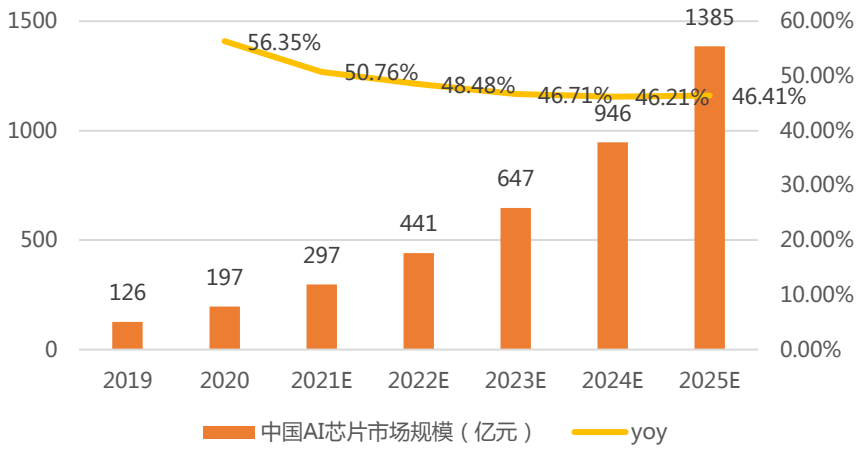
图 32：2018 年 IDC 公布的服务器成本拆解



资料来源：IDC，智研咨询，天风证券研究所

AI 产业化落地将推动人工智能芯片快速放量。由于 CPU 并不适合 AI 服务器中的大规模并行计算，因此在 AI 服务器中，主要是用 GPU、FPGA、ASIC 等计算芯片补齐 CPU 中人工智能负载处理的短板。在国家政策支持+资本推动+产业链和应用场景持续完善扩张的驱动下，中国 AI 芯片需求有望持续上涨。据艾瑞咨询数据，2021 年中国 AI 芯片市场规模将达到 297 亿元，未来在人工智能、云计算、数据中心、边缘计算等领域的广泛应用，中国市场规模预计到 2025 年达到 1385 亿元，2021-2025 年 CAGR 预计达到 47%。

图 33：中国 AI 芯片市场规模（亿元）



资料来源：艾瑞咨询公众号，天风证券研究所

人工智能芯片搭载率将持续增高，GPU 仍为主流方案。据 IDC 调研显示，当前每台人工智能服务器上配备 2 个 GPU、3 个 FPGA 或 3 个 ASIC 的比例最高，未来 18 个月，比例最高的服务器有望配备 4 个 GPU、7 个 FPGA 或 5 个 ASIC，普遍搭载率均呈上升趋势。再看中国市场，目前在国内市场主要是用以 GPU 为主实现数据中心计算加速，市场占有率近 90%，这主要是因为 GPU 可以较好支持高度并行的工作负载。ASIC、FPGA、NPU 等非 GPU 芯片市场占有率超过 10%，得益于近年智慧城市建设、无人驾驶载具、智慧医疗系统构建、智能家居等成为热门领域，应用于该类领域的非 GPU 芯片也得到发展。我们认为，未来面对需求的多元增长，AI 芯片将呈现百花齐放的发展空间。

表 10：主流 AI 芯片的对比

AI 芯片分类	功能	优势	缺点	竞争格局
GPU	主要用于处理图形、图像领域的海量数据运算，适用于高级复杂算法和通用性人工	更强大的浮点运算能力和更快的并行计算速度；通用型更强	性能功耗比较低	寡头垄断：技术壁垒极高，硬件结构精密复杂，图形算法规模庞大、软件生态相对

	智能平台；			封闭。 海外厂商：英伟达、AMD、Intel 等；
FPGA	现场可编程门阵列，使用者可以根据不同的应用需求，使用硬件描述语言对 FPGA 芯片上集成的基本门电路和存储器进行重新定义，适用于芯片功能尚未定型、算法仍需不断完善的情况下使用。	灵活性更强，速度快、功耗低、可编程性强。	对使用者技术水平要求较高。价格高。整体运算能力不高。	寡头垄断：目前国内 FPGA 芯片集中在 40mm、55mm 工艺水平，与国外的工艺仍存在一定的差距。 国外厂商：英特尔、赛灵思等， 国内厂商：复旦微电、紫光国微等；
ASIC	是一种根据特殊应用场景要求进行全定制化的专用人工智能芯片	功耗低、可扩展性强，算存一体	电路需要定制，开发周期长，难以扩展，风险高	目前处于技术发展初期，市场竞争格局分散，我国生产水平与世界领先水平差距较小。 海外：谷歌、英特尔等， 国内：寒武纪等；

资料来源：商惠敏《人工智能芯片产业技术发展研究》，刘晨《我国人工智能产业竞争力分析与建议》，天风证券研究所

GPU 海外寡头垄断格局+禁运风险或为国产 AI 服务器的主要瓶颈。GPU 主要分为独立 GPU 和集成 GPU，前者用于 AI 服务器、高性能电脑中，后者则主要用于移动端设备。目前 Nvidia 和 AMD 垄断独立 GPU 市场，其中 Nvidia 优势更为明显，2021Q1 市占率达到 83%。同时，据电子发烧网，Nvidia 的 GPU 芯片是 AI 大模型的关键，在大模型训练市场的市占比近 100%，而 GPT-3.5 大模型需要高达 2 万枚 GPU，未来商业化后或将超过 3 万枚。同时国内如浪潮、宁畅等国内品牌厂商的 AI 服务器中同样配置 Nvidia 的芯片。受到中美脱钩的持续影响，部分供应 AI 服务器的 GPU 成为限制出口的产品，直接影响国内 AI 服务器的出货量。根据美国商务部工业与安全局宣布的针对中国出口先进芯片的管制新规声明，凡输入/输出（I/O）双向传输速度高于 600GB/s，同时每次操作的比特长度乘以 TOPS 计算出的处理性能大于或等于 4800 的产品，将无法出口至中国，英伟达的 A100 即属于限制范围之内。从 AI 芯片行业投融资来看，目前国内 AI 芯片产业热度持续高涨，根据 IT 桔子的数据，2022 年中国 AI 芯片行业投融资额达到 179.5 亿元，资本的持续进入有望加速国内 GPU 国产化进程，逐步切入 AI 服务器的供应链中。

表 11：国内 AI 服务器所搭载的 GPU 厂商主要以英伟达为主

品牌厂商	型号	GPU
浪潮信息	NF5688M6	8 个 NVIDIA 最新的 NVSwitch 全互联 500W Ampere 架构 GPU
宁畅	X660 G45 LP	8 个 NVIDIA 的 A800
拓维信息	兆瀚 AI 推理服务器	8 张 Atlas300I（国产芯片）
宝德科技	PR4910W	8 个 NVIDIA A800/10 个 NVIDIA 的 A40/30

资料来源：各公司官网，华为计算公众号，天风证券研究所

表 12：英伟达 A100 与 A800GPU 性能相差不大

性能	A100	A800
数据传输速率	600GB/s	400GB/s
显存带宽（最高）	2TB/s	2TB/s
显存容量（最高）	80GB	80GB
FP64	9.7TFLOPS	9.7TFLOPS

FP32	19.5TFLOPS	19.5TFLOPS
Tensor Float 32（TF32）	156TFLOPS/312 TFLOPS	156TFLOPS/312 TFLOPS

资料来源：英伟达公司官网，51CTO，新智元，天风证券研究所

国内厂商正在陆续推出 GPU 产品进行市场检验，国产替代提上日程。伴随资本和政策的持续加码，一批国内 GPU 厂商逐渐崭露头角（见下表）。然而，我们仍需看到在芯片设计制造领域，我国仍缺乏设计软件、先进制程及设备与世界领先水平之间仍有差距，该领域部分产品及装备仍十分依赖进口，国产 GPU 之路仍是路漫漫其修远兮。

表 13：国内生产供应服务器的 GPU 厂商

公司	进展	性能
景嘉微（300474.SZ）	第三代 GPU 芯片 JM9 系列的两款产品分别于 2021 年 11 月 16 日和 2022 年 6 月 28 日完成阶段性测试工作	用于地理信息系统、媒体处理、CAD 辅助设计、游戏、虚拟化等高性能显示和人工智能计算领域。
龙芯中科（688047.SH）	2022 年推出 16 核芯片产品 3C5000，32 核 3D5000 研制成功	主要用于存储服务器，取得石油、石化等客户的突破；
海光信息（688041.SH）	海光深算一号 DCU 实现商业化应用，深算二号正在研发中	广泛用于大数据处理、人工智能、商业计算等计算密集类应用领域，主要部署在服务器集群或数据中心，为应用程序提供高性能、高能效比的算力，支撑高复杂度和高吞吐量的数据处理任务。
芯原股份（688521.SH）	推出 Vivante3DGPGPUIP	提供从低功耗嵌入式设备到高性能服务器的计算能力，满足广泛的 AI 计算需求。
天数智芯	2021 年 3 月发布了首款通用 GPU 训练芯片天垓 100	广泛支持传统机器学习、数学运算、加解密及数字信号处理等领域；
	2022 年 12 月发布通用 GPU 智铠 100	支持国内外主流深度学习框架，拥有丰富编程接口拓展和高性能函数库，广泛适用于智慧城市、智慧港口、智慧交通等众多领域。
壁仞科技	2022 年 9 月发布首款通用 GPU 芯片 BR100	算力创下全球记录，率先采用 Chiplet 技术。

资料来源：Wind，电子发烧网公众号，天天 IC 公众号，天数智芯公众号，天风证券研究所

3.2. 柳暗花明：以 Chiplet 工艺打开我国对算力的想象空间

Chiplet 有望成为我国先进制程及算力受限的困境的突破口。在 AI 时代浪潮的裹挟下，以 ChatGPT 为代表的是大数据+大模型+大算力的产物，每一代 GPT 模型的参数量高速增长，随着科技头部企业类 ChatGPT 项目入局，整体在算力提升、数据存储及数据传输端需求迭起。我们认为采用 Chiplet，即将模块化设计引入半导体制造和制造，协同计算助力 HPC 芯片算力突破及性能提升，从而满足大型模型的训练需求。国外厂商 AMD 于 2021 年 6 月发布基于台积电 3D Chiplet 封装技术的第三代服务器处理芯片，国内华为于 2019 年推出基于 Chiplet 技术的 7nm 鲲鹏 920 处理器，实现多核高并发和资源调度优化，计算性能提高 20%。在美国对我国半导体产业持续封锁的状态下，国产算力和先进制程瓶颈有望在 Chiplet 助力下实现突破，甚至在该领域实现弯道超车，如壁仞科技和寒武纪推出采用 Chiplet 工艺的芯片（见下表），在部分性能上可以达到甚至超越英伟达供应 AI 服务器的 A100。国内这一领域的优势一方面是由于 Chiplet 开发模式将芯片工艺转向系统集成，因而能够以我国在应用创新的优势换取光刻机受限的缓冲期；另一方面则是 Chiplet 的核心为“先进封装”技术，国内 Chiplet 封装产业技术积累深厚，国际竞争力强，如长电科技、

通富微电和华天科技具备 Chiplet 量产能力，并据 ittbank 数据，长电科技、通富微电和华天科技均位列 2021 年全球营收前十的封测厂商排名中。

表 14：壁仞科技的 BR100、寒武纪的思元 370 与 NVIDIA A100 性能比较

性能比较	壁仞 BR100（2022 年）	思元 370-X8（2021 年）	NVIDIA A100（2020 年）
晶体管数量	770 亿颗	390 亿颗	542 亿颗
芯片面积	1074mm^2	-	828 mm^2
工艺节点	台积电 7nm	7nm	台积电 7nm
显存容量	64GB HBM2E	48GB	80GB HBM2E
峰值算力	2048TOPS（INT8）	256TOPS（INT8）	624TOPS（INT8）
最大热设计功耗 TDP	550W	250W	400W

资料来源：芯东西公众号，寒武纪官网，寒武纪公司公告，天风证券研究所

Chiplet 或将带来全产业链投资机遇。Chiplet 工艺有望突破国产芯片的算力瓶颈，成为半导体发展核心，2022 年 12 月中国工信部中国电子工业标准化技术协会审定并发布了《小芯片接口总线技术要求》，中国迎来了首个原生 Chiplet 小芯片标准。该推广将推动本土半导体芯片这一领域的发展。

表 15：Chiplet 实现量产的国内公司

功能	企业	主要工艺
封测	长电科技	采用通过 Chiplet 异构集成技术完成的 XDFOI ™Chiplet 高密度多维异构集成系列工艺，已按计划进入量产计划。
	通富微电	公司在多芯片组件、集成输出封装、2.5/3D 等先进封装技术方面均提前布局，为客户提供多样化的 Chiplet 封装解决方案，并且已经开始大规模生产 Chiplet 产品。
	华天科技	公司开发了 3D FO SiP 封装技术，完成用于高性能计算的大尺寸 HFCBGA 产品，公司已量产 Chiplet 产品，主要应用于 5G 通信、医疗等领域。

资料来源：各公司官网、公告，天风证券研究所

3.3. 适逢其会：为上游部分元件打开增量空间

3.3.1. AI 服务器或将打开 PCB 增量市场

AI 服务器价值量提升为 PCB 市场带来发展空间。PCB 主要参与服务器内部的主板、电源背板、硬盘背板、网卡、Riser 卡等，随着服务器对运算及传输速率的要求不断提升，对 PCB 提出更严苛的电性能要求，同时也给对应的 PCB 市场带来良好机会。从 PCB 价值量来看，传统通用服务器 PCB 以 8-10 层 M6 板为主，价值量约为 3400 元。AI 服务器分为训练服务器和推理服务器。训练服务器 PCB 以 18-20 层 M8 板为主，价值量约 10350 元。推理服务器 PCB 以 14-16 层 M6 为主，价值量约为 7140 元。总体来看，AI 服务器 PCB 价值量约为普通服务器 PCB 的 2-3 倍。从 PCB 下游市场来看，根据 Prismark 数据，2021 年全球服务器用 PCB 的产值为 78.04 亿美元，预计 2026 年产值达到 124.94 亿美元，5 年 CAGR 为 9.9%，增速快于其他 PCB 品类。服务器行业发展空间广阔+消费电子/PC 领域呈现疲态，众多 PCB 企业在积极布局服务器用 PCB 领域。

表 16：PCB 领域的国内厂商

企业	服务器用 PCB 产品情况
鹏鼎控股	PCB 行业的龙头企业，对于 AI 服务器对主板的高要求，公司表示已做好应对市场放量的准备。
深南电路	从事高中端印刷电路板的设计、研发及制造等相关工作，正重点布局数据中心（含服务器）领域。
生益电子	专注于各类印刷电路板的研发、生产与销售业务。公司印制电路板产品定位于中高端应用市场，服务器是其应用领域之一。
广合科技	2020-2022 年服务器领域 PCB 营收占比超过 60%，公司产品应用于 Intel 发布的 Purley、Whitley 及 Eagle Stream 各世代芯片平台的服务器，正与客户针对下一代芯片平台开展研究。公司在 AMD 发布的各世代芯片平台也有对应的 PCB 产品向服务器厂商供应，在 ARM 架构芯片平台也有供货和储备。

胜宏科技	应用于 Eagle Stream 级服务器领域的产品实现规模化量产；基于 AI 服务器的加速模块的高阶 HDI 及高多层产品，已实现 4 阶 HDI 及高多层的产品化，6 阶 HDI 产品已在加速布局中。
沪电股份	2021 年次世代服务器平台印制电路板已进入客户样品打样阶段，已具备新一代服务器平台用 PCB 的批量生产能力。

资料来源：各公司公告、各公司与投资者会议纪要、天风证券研究所

3.3.2. 高算力使服务器芯片散热成为难题，散热模组应需求提高散热效率

服务器的高功率或将引导散热模组转型，开拓液冷新市场。以 NVIDIA DGX A100 为例，AI 服务器系统功耗达到了 6.5KW。散热效率一直是服务器厂商需要解决的问题之一，当前主流模式是利用散热鳍片、热导管、风扇、空调等组成的风冷模式。随着大数据、云计算带来天量的数据处理等高通量的计算业务，使得服务器芯片的散热收到严重挑战，芯片热封装壳温也在不断提高，达到了风冷的极限。同时风冷的耗电量极高，数据中心的制冷空调系统用电量占整个数据中心的 30-50%。相比之下，液体比热容为空气的 1000-3500 倍，导热性能是空气的 15-25 倍，利用自然冷却显著降低耗电量，使得液冷成为风冷的不二选择，在未来或将全面替代风冷，成为 AI 服务器乃至数据中心的标配。目前已有不少专注于热功能的厂商，在加大对液冷模式的研发和布局，如老牌散热模组供应商双鸿、超众已切入液体循环散热的领域，中石科技能够提供全方位热管理综合解决方案，飞荣达针对服务器的散热需求开发轴流风扇、特种散热器、单相液冷冷板模组、两相液冷模组等产品。我们认为，数据量增长给服务器带来更迫切的散热需求，散热组件的重要性将持续凸显，尤其看好未来液冷对气冷替代所产生的需求。

表 17：散热模组领域的国内厂商

企业	主要情况
中石科技	为客户提供热模组产品，新基建（服务器、数据中心等高性能计算需求）是公司面向的四大高成长行业之一，为其提供高端热管理综合解决方案；覆盖的产品包括但不限于导电垫、导热硅脂、导热凝胶、导热相变材料 PCM、EMI 材料、热模组等。
飞荣达	散热模组是公司的主营产品之一，应用于通信设备、计算机、游戏机、智慧屏、工控设备等领域。飞荣达针对服务器的散热需求开发轴流风扇、特种散热器、单相液冷冷板模组、两相液冷模组等产品
超频三	为客户提供热模组产品，积极导入四大高成长性行业，如新消费电子（智能家居/医疗设备等）、数字基建（服务器/数据中心）、智能交通、清洁能源等。
同飞股份	公司目前已主要形成了液体恒温设备、电气箱恒温装置、纯水冷却单元和特种换热器四大类产品，应用涵盖多个工业制冷领域。
高澜股份	公司是 IDC 领域中从事液冷醉酒的解决方案提供商之一，多年前边开始关注 IDC 液冷赛道。
双鸿（台）	全方位热流方案提供者，为全球第一大笔记型计算机散热模块设计及制造厂，同时公司已经切入液体循环散热的领域。
超众（台）	台湾笔记本电脑散热模组供应商，并与 Intel 及台湾笔记型电脑大厂同步开发 CPU 散热设计。

资料来源：Wind，各公司投资者调研纪要，富果研究，天风证券研究所

3.3.3. 数据存力作为算力进阶需求迭起，看好国产存储器后续发展

服务器内的数据存力重要性成为存储器发展动力。在服务器行业，既需要高频宽的 DRAM 作为暂存，提供处理器技术时即时所需的资料；也需要 NAND Flash 用以存放资料，通常以固态硬盘（SSD）的形态存在。

·DRAM 存储器：目前我们生活中接触各种内存概念产品多为 DRAM，领导标准机构 JEDEC 将 DRAM 定义为标准 DDR、移动 DDR、图形 DDR 三个类别，分别指代为电脑内存、手机运存、显卡显存，其中图形 DDR 能提供极高吞吐量，适合面向图形应用程序、数据中心加速以及 AI 的数据密集型应用程序，并且将很多 DDR 芯片堆叠后与 GPU 封装在一起，就构成了另一种形式的显存，即 HBM。我们认为，ChatGPT 催生对更高性能存储的需求，HBM 技术有望随着人工智能快速发展而发展，从而成为适配 AI 服务器的高阶选择。这从 2023 年初 ChatGPT 带动三星电子和 SK 海力士 HBM 订单激增可见一斑，AI 服务器中所需的英特尔 A100、V100 均搭载了 HBM2。与其他 DRAM 相比，HBM 通过垂直连接多个 DRAM

显著提高数据处理速度，售价是普通 DRAM 的五倍，但由于其生产的复杂性和技术的先进性，国内外市场均由三星和 SK 海力士主导。HBM 当前在 DRAM 渗透率低，同时出于性价比选择，传统服务器仍多选择 DDR 和 GDDR 以提高内存性能。根据前瞻产业研究所，2020 年 DRAM 市场由海外厂商三星、海力士和美光所主导，CR3 达到 94.5%。根据清枫资本公众号，2020 年服务器用 DRAM 占比 34.9%。我们看好中长期内高算力需求增长将成为存储器发展的成长驱动力。

表 18：国内企业的存储器产品

企业	代表产品
紫光国芯	DDR/DDR2/DDR3/DDR4 DDR5（试产）
合肥长鑫（未上市）	DDR4/LPDDR4/LPDDR4X/DDR5/LPDDR5 GDDR6（规划） GDDR6（试产）
兆易创新	DDR3/DDR4/LPDDR4
江波龙	主要从事 Flash 及 DRAM 存储器的研发、设计和销售，目前提供消费级、工规级、车规级存储器以及行业存储软硬件应用解决方案
东芯半导体	DDR3/LPDDR2
华邦电子（台）	DDR3/DDR4/DDR4X
南亚科技（台）	DDR5/LPDDR4X

资料来源：果壳硬科技公众号，Wind，天风证券研究所

·SSD 固态硬盘：主要分为企业级和消费级，企业级 SSD 即应用于高性能计算、边缘计算、高端存储、数据中心等各种企业级场景中的固态硬盘，具备不间断工作能力，能够处理 I/O 密集型工作负载。海量数据处理是人工智能计算负载的典型特点，当前 SSD 成本不断下降，已经成为高性能服务器的必须，其中也包括人工智能服务器。并且，X86 处理器的发展也对周边存储设备起到了带动作用，PCIe 4.0 SSD 在数据中心占比大幅增长，满足数据密集型应用高速吞吐的需求。随着未来持续增加的数据存储、传输需求，只有性能更强、容量更大、更稳定耐用的存储设备才能支撑数字经济发展，而企业级 SSD 面向企业级客户，比消费级 SSD 具备更强性能、更高可靠性和更强耐用性，我们认为有望成为 AI 服务器的刚性需求。当前的 SSD 市场的领先企业包括英特尔、西部数据、美光、东芝，国内 SSD 企业仍处于追赶期，包括忆联、忆恒创源、浪潮、大普微等均推出企业级 SSD，从产品性能和产能方面逐渐对标国际领先企业。当前企业级 SSD 下游客户主要来自云服务器&互联网，我们预计在未来几年内，云计算与互联网仍是国内企业级 SSD 硬盘的购买主力，尤其是目前处于人工智能元年，对 AI 服务器的增量需求或将稳固企业级 SSD 硬盘的增长。

表 19：提供企业级 SSD 的企业

企业	代表产品
铠侠	企业级 PM7 SSD 系列主打高性能计算、人工智能、缓存层、金融交易与分析等用例，读取性能较上一代提升约 20%，容量水平可达到 30.72TB。
西部数据	推出了面向热存储的 Ultrastar DC SN840 高性能 SSD，通过双接口、1 和 3 DW/D 耐久性以及 TCG 加密等设计，适应 HPC、数据库、虚拟化、人工智能、5G 通讯、自动驾驶和机器学习等不同应用。
金士顿	DC1500M 采用高性能 Gen 3.0 x4 PCIe NVMe 设计，可预测的随机读写性能和延迟，适应高性能云服务、媒体采集、传输以及大数据应用要求。
长江存储	提供 3D NAND 闪存晶圆和颗粒以及企业级固态硬盘等产品，应用于服务器、数据中心等领域。
亿恒创源	国内企业级 NVM SSD 产品提供商，企业级 SSD 在数据库、大数据、云计算、人工智能等领域广泛应用。

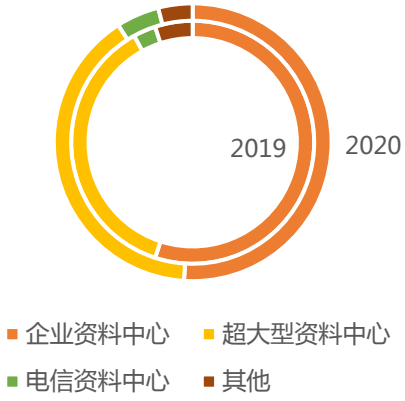
资料来源：千山资本，全球 SSD 公众号，天风证券研究所

3.4. 下游指明服务器发展方向，人工智能渗透率提高将扩展 AI 服务器应用

全球服务器需求以企业资料中心为主，超大型资料中心（HDC）占比稳步增长。企业资料中心主要是由拥有庞大数据存储需求的公司建构，如 Netflix、Zoom 等互联网公司和部分

金融企业等，市场份额正在被超大型资料中心所挤压，这主要是对存储需求大的企业数量有限。超大型资料中心指的是云服务提供商（CSP），如亚马逊的 AWS，微软的 Azure、阿里的阿里云，主要为有云端需求的企业提供服务，解决中小企业无法负担直接搭建企业资料中心的成本困境，我们认为随着中小产业持续导入，或将成为云服务的重要增量市场，从而拉动服务器的增长。

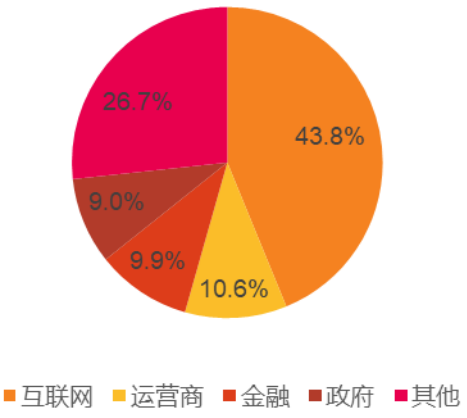
图 34：2019/2020 年全球服务器应用场景



资料来源：TrendForce，富果研究部，天风证券研究所

国内服务器下游重点关注互联网云服务商与电信运营商。中国服务器下游用户主要分布在互联网、运营商、政府、金融等多个领域。互联网以超过 40%份额成为主流应用，BAT 和第三方 IDC 服务器公司成为服务器行业的主要购买力来源，同时近年来云计算的支出也为服务器出货量贡献力量。国内电信领域同样值得关注，国内三大运营商提出算网一体，算力网络按照算网协同、算网融合和算网一体的三阶段演进路径已经成为业界共识，2022 年运营商进一步加大对算力底座的投资落地，2022 年三大运营商服务器相关招标中，共涉及近 60 万台服务器，金额超过 200 亿元，预计 2023 年运营商的 capex 将维持扩张，其中算力支出为最大支出。2022 年末在 AI 浪潮回卷趋势下，互联网巨头争先布局人工智能产品，AI 服务器有望成为服务器产品的新兴业务负载。

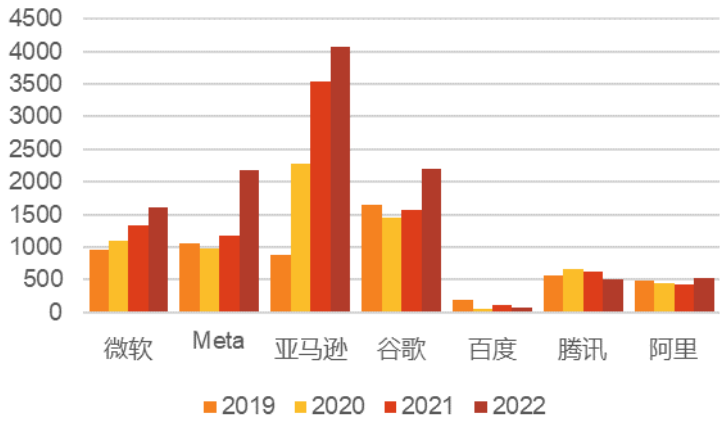
图 35：2021 年我国服务器下游应用场景



资料来源：IDC，中商产业研究院公众号，天风证券研究所

下游需求预期饱满，持续受益于 AI 应用和模型逐步落地。根据 Trendforce 报告指出，2022 年 AI 服务器采购中，北美四大云端厂商谷歌、亚马逊 AWS、Meta、微软合计占比 66%，国内字节跳动（6%）、腾讯（2.3%）、阿里巴巴（1.5%）和百度（1.5%）紧随其后。AI 服务器需求和云计算及互联网客户的 Capex 直接挂钩，下游客户 capex 企稳回暖使得 AI 服务器产能有望持续受益扩张，尤其是在 ChatGPT 的应用之下，AI 服务器采购商纷纷布局人工智能应用，AI 服务器行业预计将迎来高景气。

图 36：主要的 AI 服务器采购商 2019-2022 年 Capex（亿元）



资料来源：Wind，天风证券研究所

4. 投资建议

建议关注 AI 服务器及上游产业链相关标的：1)AI 服务器龙头：工业富联；2)服务器 PCB：鹏鼎控股；3)服务器线束与连接器：电连技术、兆龙互连；4)算力芯片：寒武纪、海光信息（天风计算机团队覆盖）、景嘉微（天风计算机团队联合覆盖）；5)存储供应链：兆易创新、北京君正、江波龙（天风计算机团队联合覆盖）、澜起科技、雅克科技、鼎龙股份（天风化工团队联合覆盖）、华懋科技（天风汽车团队联合覆盖）、华特气体；6)边缘 AI：瑞芯微、晶晨股份、全志科技、恒玄科技、富瀚微、中科蓝讯、乐鑫科技；7)AI to B /机器视觉：大华股份、海康威视、鼎捷软件（天风计算机团队覆盖）、凌云光、 天准科技、舜宇光学、海康威视、奥普特（天风机械军工团队覆盖）；8)Chiplet：长电科技、通富微电、华天科技、长川科技（天风机械团队覆盖）、华峰测控（天风机械团队覆盖）、利扬芯片、芯碁微装、伟测科技

表 20：相关公司盈利预测与估值

产业链环节	企业	营业收入（亿元）			EPS（元/股）			PE		
		2021A	2022A	2023E	2021A	2022A	2023E	2021A	2022A	2023E
服务器	工业富联	4395.57	5118.50	5745.00	1.01	1.01	1.20	11.83	9.08	19.03
	浪潮科技	670.58	695.25	826.86	1.38	1.42	1.80	26.01	15.14	29.18
	拓维信息	22.30	22.37	30.92	0.07	-0.81	0.10	124.76	-8.17	194.58
	中国长城	177.90	140.27	189.02	0.20	0.04	0.17	69.63	273.55	84.88
	大华股份	328.35	305.65	347.14	1.13	0.77	1.03	20.82	14.76	19.84
GPU	景嘉微	10.93	11.54	17.21	0.97	0.64	0.90	156.62	85.84	114.11
	龙芯中科	12.01	7.39	16.59	0.66	0.13	0.52	-	662.11	276.31
	海光信息	23.10	51.25	74.46	0.16	0.35	0.56	-	116.05	145.46
	芯原股份	21.39	26.79	33.88	0.03	0.15	0.30	2881.83	297.18	294.27
FPGA	复旦微电	25.77	35.39	43.27	0.63	1.32	1.73	79.70	52.94	32.09
	紫光国微	53.42	71.20	92.27	3.22	3.10	4.05	69.88	42.55	21.34
ASIC	寒武纪	7.21	7.29	10.75	-2.06	-3.13	-1.95	-46.12	-17.40	-115.19
先进封装	长电科技	305.02	337.62	337.61	1.66	1.82	1.52	18.66	12.70	21.35
	通富微电	158.12	214.29	246.29	0.72	0.33	0.50	26.99	49.68	51.00
	华天科技	120.97	119.06	133.80	0.44	0.24	0.25	28.77	35.23	38.05
DRAM	紫光股份	676.38	740.58	850.81	0.75	0.75	0.94	30.43	25.86	36.45
	兆易创新	85.10	81.30	71.53	3.50	3.08	2.00	50.23	33.30	56.00
	东芯股份	11.34	11.46	13.45	0.59	0.42	0.55	76.32	62.41	67.04
	江波龙	97.49	83.30	101.16	2.73	0.18	0.71	-	334.50	159.21
PCB	鹏鼎控股	333.15	362.11	409.00	1.43	2.16	2.23	29.69	12.71	11.13
	广合科技	20.76	24.12	-	0.27	0.74	-	-	-	-
	深南电路	139.43	139.92	157.82	3.03	3.20	3.60	40.25	22.57	22.00

	生益电子	36.47	35.35	38.97	0.32	0.38	0.44	43.94	24.75	28.66
	胜宏科技	74.32	78.85	93.33	0.78	0.92	1.09	38.97	14.13	24.49
散热模组	中石科技	12.48	15.92	20.12	0.47	0.69	0.92	46.81	21.50	27.26
	飞荣达	30.58	41.25	54.70	0.06	0.19	0.46	422.50	75.99	37.04
	超频三	5.8	11.50	22.81	-0.38	0.04	0.21	-21.79	190.65	34.54
	同飞股份	8.29	10.08	18.12	2.31	1.37	1.40	55.39	67.28	35.78
	高澜股份	16.79	19.04	12.85	0.23	0.93	0.37	68.65	10.68	59.75

资料来源：截止于 2023 年 6 月 14 日的 Wind 一致预期，天风证券研究所

5. 风险提示

中美贸易摩擦导致上游原材料断供：若美国加大对于中国的制裁范围和强度，国内厂商芯片等原材料可能出现断供风险。

AI 服务器出货不及预期：若出现 GPU 供应不足等情况，可能导致 AI 服务器出货承压。

技术瓶颈仍未摆脱：若受到硬件架构等条件限制，AI 服务器技术迭代可能受阻。

分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告所表述的所有观点均准确地反映了我们对标的证券和发行人的个人看法。我们所得报酬的任何部分不曾与，不与，也将不会与本报告中的具体投资建议或观点有直接或间接联系。

一般声明

特别声明

投资评级声明

类别	说明	评级	体系
股票投资评级	自报告日后的 6 个月内，相对同期沪深 300 指数的涨跌幅	买入	预期股价相对收益 20%以上
		增持	预期股价相对收益 10%-20%
		持有	预期股价相对收益-10%-10%
		卖出	预期股价相对收益-10%以下
行业投资评级	自报告日后的 6 个月内，相对同期沪深 300 指数的涨跌幅	强于大市	预期行业指数涨幅 5%以上
		中性	预期行业指数涨幅-5%-5%
		弱于大市	预期行业指数涨幅-5%以下

天风证券研究

北京	海口	上海	深圳
北京市西城区佟麟阁路 36 号	海南省海口市美兰区国兴大道 3 号互联网金融大厦 A 栋 23 层 2301 房	上海市虹口区北外滩国际客运中心 6 号楼 4 层	深圳市福田区益田路 5033 号平安金融中心 71 楼
邮编：100031	邮编：570102	邮编：200086	邮编：518000
邮箱：research@tfzq.com	电话：(0898)-65365390	电话：(8621)-65055515	电话：(86755)-23915663
	邮箱：research@tfzq.com	传真：(8621)-61069806	传真：(86755)-82571995
		邮箱：research@tfzq.com	邮箱：research@tfzq.com