从英伟达看国产 GPU 发展机遇与挑战

前瞻研究系列报告 99 | 2022.9.16

中信证券研究部

陈俊云
前瞻研究首席分析师
S1010517080001

许英博
科技产业首席分析师
S1010510120041

贾凯方
前瞻研究分析师
S1010522080001

联系人：刘锐

核心观点

借助灵活性、计算速度、软件生态等层面的综合优势，GPU 正在成为通用并行计算领域最理想的载体芯片。英伟达发展历程表明，作为通用计算芯片，产品技术、软件生态等构成 GPU 厂商的核心竞争壁垒。我们预计到 2025、2030 年，全球 GPU 市场规模将分别增长至 278（图形渲染）/245（数据中心）亿美元、568/828 亿美元，国内市场则有望增长至 47（图形渲染）/74（数据中心）亿美元、97/248 亿美元。当前国内本土 GPU 厂商正在快速崛起，大部分核心团队具有英伟达、AMD 工作经历，且企业目前已平均发布 1-3 款相关产品，产品核心参数约落后英伟达、AMD 1~2 代左右，正逐步从“可用”走向“好用”。尽管目前国内大部分本土 GPU 厂商仍处于早期阶段，短期仍需克服产品验证、量产落地等潜在挑战，但长期发展机遇值得关注。基于对下游应用场景的分析，我们判断本土厂商有望率先在 AI（训练、推理）领域实现突破，并可逐步向图形渲染、复杂科学计算等领域扩展。我们看好本土 GPU 厂商的长期投资机会，建议关注二级市场的头部企业，以及一级市场的摩尔线程、沐曦集成电路、瀚博半导体等。

- **报告缘起：1）**市场需求：AI、高性能计算、图形渲染等推动 GPU 等大算力并行计算芯片需求。近些年来，随着人工智能硬件、软件算法、应用场景丰富度的增加，算法模型参数不断增加，带动了对数据中心并行计算大算力的需求。Bloomberg 数据显示，2021 年全球数据中心逻辑芯片市场规模约 436 亿美元，其中 GPU 市场约 100 亿美元。**2）**近期市场对国产 GPU 关注度较高。复盘英伟达发展历史，公司在图形渲染&数据中心领域保持较高的市占率，并实现产业引领，核心原因在于：借助持续、高频迭代保持产品技术行业领先，并借助 CUDA 等实现软件生态构建，不断提升产品易用性等。**3）**本篇报告中，我们主要从全球市场出发，就 GPU 产业本身的产品特性、技术路线、市场空间，以及国内市场现状、演进路径、竞争格局等进行了系统的分析和讨论，力图针对国内 GPU 市场构建一个完整的产业&投资蓝图。

- **全球市场：GPU 是并行计算理想载体芯片，数据中心料将是中期需求增长主要场景。1）**GPU 应用场景正由早期的图形渲染，逐步拓展至高性能运算、科学计算等领域，并已超越 CPU、FPGA、ASIC 等，成为通用并行计算领域的理想载体芯片。**2）**图形渲染用 GPU：游戏为主，我们预计中期市场规模有望保持 10%~15%平稳增长。2021 年，英伟达游戏显卡业务营收 105 亿美元，专业视觉收入（图形工作站）21 亿美元。假设英伟达在此领域中的市占率分别为 80%，我们预计 2021 年全球图形渲染领域（游戏+专业视觉）市场规模约 158 亿美元。我们预计 2025、2030 年该领域市场规模将分别达到 278、568 亿美元，增长动力主要源于产品 ASP 提升、游戏玩家数量增长等。**3）**数据中心用 GPU：英伟达主导，我们预计中期蒋保持 25%以上的年均复合增速。依据 Liftr Insights & Top500.Org 数据，我们认为英伟达在数据中心独立 GPU 领域中的市占率约 80-90%，2021 年英伟达数据中心 GPU 营收约 80.3 亿美元，对应全球数据中心 GPU 芯片市场规模约 100 亿美元。展望未来，我们预计 2025 年、2030 年全球数据中心 GPU 芯片市场规模分别为 245、828 亿美元，增长动力主要源于产品 ASP 提升、应用场景不断拓展等。

- **国内 GPU 市场：中期潜在空间可观，本土厂商开始规模崛起&产品落地。1）**图形渲染领域：IDC 数据显示，2016-2021 年中国 PC 出货量约占全球 17%的比例。我们假设在图形渲染领域，国内 GPU 出货量占比亦和 PC 表现相对一致，并保持和全球市场相似的增速，以及应用场景分布等。参考全球 GPU 图形渲染领域市场规模，我们测算/预测 2021 年、2025 年、2030 年，国内 GPU（图形

渲染）的市场规模约为 27、47、97 亿美元。**2**）数据中心 GPU：主要用于 AI（训练&推理）、高性能计算等，我们测算 2021 年国内数据中心 GPU 市场规模约为 20 亿美元，对应全球市场比重约为 20%，我们预计 2030 年国内数据中心 GPU 市场规模有望增长至 250 亿美元，占全球市场 30%左右比重。**3**）国内 GPU 厂商开始快速崛起，大多数企业目前已发布 1-3 款相关产品，大部分企业核心团队亦具有英伟达、AMD 工作经历。从对外公布产品核心参数来看，国内厂商产品平均落后英伟达、AMD 1~2 代左右，但正逐步从“可用”走向“好用”。

- **本土 GPU 厂商：有望率先在 AI 场景实现落地，并扩展至图形渲染、复杂科学计算等领域**。源于 AI、高性能计算、复杂图形渲染等下游应用场景快速增长，我们看好国内 GPU 市场需求前景，国产 GPU 厂商未来发展可期。但客观而言，国内大部分本土 GPU 厂商当前仍处于早期阶段，短期仍需克服产品验证、量产落地等潜在挑战，结合市场需求、技术难度、产业生态等维度因素综合考量，我们判断本土厂商有望率先在 AI（训练、推理）领域实现突破，并可逐步向图形渲染、复杂科学计算等领域扩展。
- **风险因素：**欧美高通胀持续风险；欧美经济陷入衰退风险；反垄断及数据监管政策持续趋严的风险；中期个人用户消费不足、企业 IT 支出下滑超预期风险；国际贸易冲突持续加剧风险等；芯片半导体行业短缺的风险；公司关键产品研发进展不顺利的风险；公司产品用户验证不急预期的风险等。
- **投资建议：**当前国内本土 GPU 厂商正在快速崛起，大部分核心团队具有英伟达、AMD 工作经历，且企业目前已平均发布 1-3 款相关产品，并逐步从“可用”走向“好用”。参考英伟达发展历程，GPU 作为通用计算芯片，产品技术、软件生态等构成 GPU 厂商的核心壁垒，国内大部分本土 GPU 厂商当前仍处于早期阶段，短期仍需克服用户验证、产品落地等潜在挑战，但中长期发展前景值得期待。我们判断本土厂商有望率先在 AI（训练、推理）领域实现突破，并可逐步向图形渲染、复杂科学计算等领域扩展。我们看好本土 GPU 厂商的长期投资机会，建议关注二级市场的头部企业以及一级市场的摩尔线程、沐曦集成电路、瀚博半导体等。

目录

插图目录

表格目录

报告缘起

市场需求：AI、高性能计算、图形渲染等推动 GPU 等并行计算芯片需求

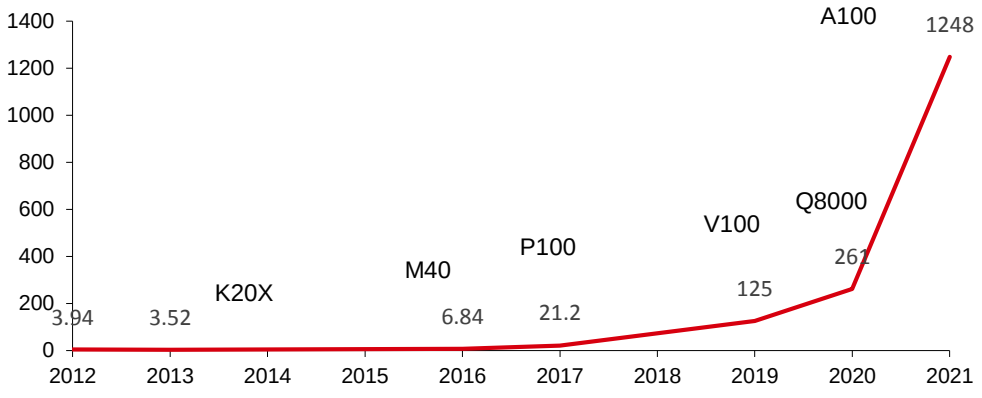
需求场景：AI 训练&推理、复杂科学计算、大规模图形渲染等，持续推动并行计算芯片需求。由于 GPU（Graphics Processing Uni，图形处理器）是由成百上千个阵列排布的运算单元 ALU 组成，使得 GPU 更适用于大规模并发运算，其在图形处理、计算加速等领域有着广泛的运用。**2）**由于 GPU 加速器强大的并行处理能力，超算中心工作人员可以更好地设计深度网络结构，使得其在超算领域&数据中心领域更具经济效益，导致 GPU 在 AI 训练&推理、科学计算等领域有着广泛的应用。

- GPU 用于 AI 训练&AI 推理领域。在典型 AI 模型卷积网络中，大量数据以图片形式导入，在进行运算过程中，数据均为矩阵形式，而矩阵运算通常适合并行，因此 AI 算法的特性，使得 GPU 的运算速度明显大于 CPU，使得 GPU 得以大量应用在 AI 的训练与推理当中。

- GPU 可用于复杂科学计算中。科学计算将物理、化学、生物、航空航天等领域的问题转化为数学模型，通过计算和求解模型用于实际产业。从计算数据来看，由于科学计算中所用数据多数以矩阵为形式，同时由于科学计算对误差有强制要求，因此在运算中需要在并行运算基础上保证一定的精度。而现代 GPU 在并行&矩阵运算的基础上，已经能够满足科学计算所需的精度要求。

近些年来，随着人工智能软件算法的发展，复杂科学计算的进步，以及图形渲染功能的增加，带动底层芯片并行计算能力需求的快速提升。以全球 AI 芯片领军者英伟达的发展状况来看，公司 AI 芯片算力由 2012 年的 4Tops 提升至 2021 年的 1248Tops，9 年时间提升了约 315 倍。

图 1：英伟达单芯片推理性能（Int8 Tops）

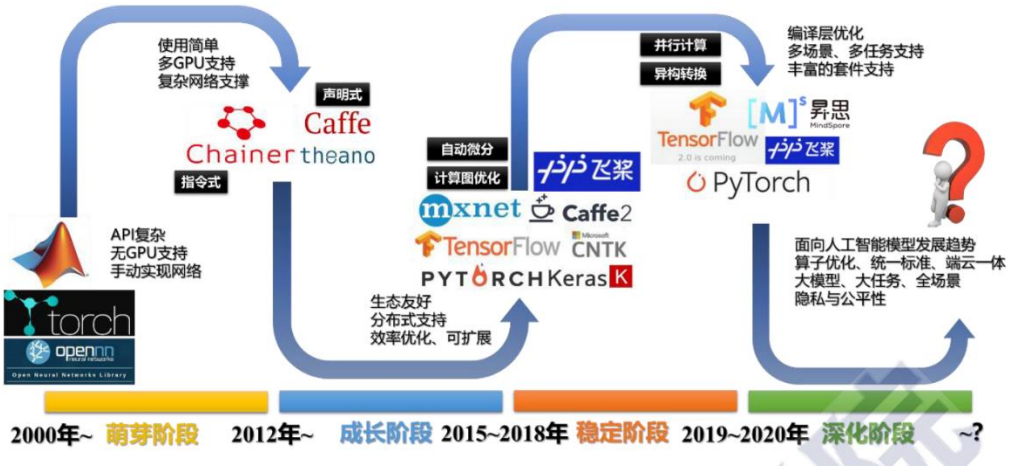


资料来源：英伟达官网，中信证券研究部

AI 框架、并行计算框架等引入&丰富，不断推动针对并行计算芯片软件开发门槛降低。1）从人工智能软件算法框架的发展历史来看，2015 年谷歌宣布开源 TensorFlow，2019 年 PFN 宣布将研究方向由 Chainer 转向 PyTorch。目前 AI 框架形成了 TensorFlow 和

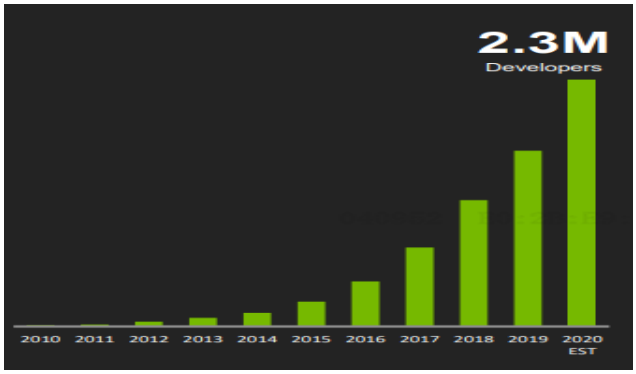
PyTorch 双寡头垄断的竞争格局。其中，谷歌开源 TensorFlow 项目，在很大程度上降低了人工智能的开发门槛和难度。**2）**TensorFlow 主要用于处理机器学习中的计算机视觉、推荐系统和自然语言处理（NLP）的模型训练和推理，涉及模型隐藏层相对较多，模型量相对较大，基本上均需要 CUDA 的加速处理。随着 TensorFlow 的开源，涉及到的开发开发者快速增加，CUDA 软件下载量也呈现陡增趋势。据英伟达在 2021GTC 大会上宣布，截至 2020 年底，CUDA 累计下载量超过 2000 万次，其中 2020 年下载量超过 600 万次。涉及到的开发人员约 230 万人（2020 年新增超过 60 万人）。

图 2：人工智能框架发展史



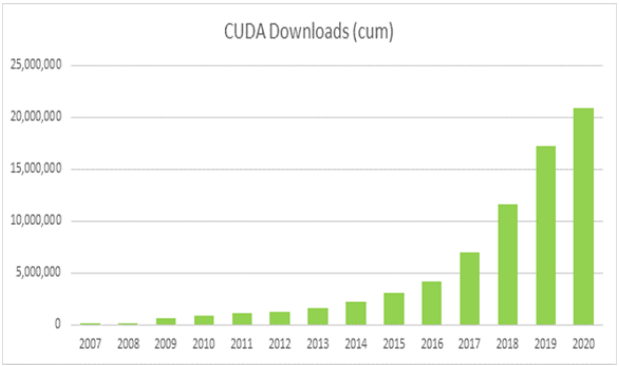
资料来源：中国信通院官网；注：Logo 来自各公司官网

图 3：英伟达 CUDA AI 开发者人数



资料来源：NVIDIA 2021GTC 大会

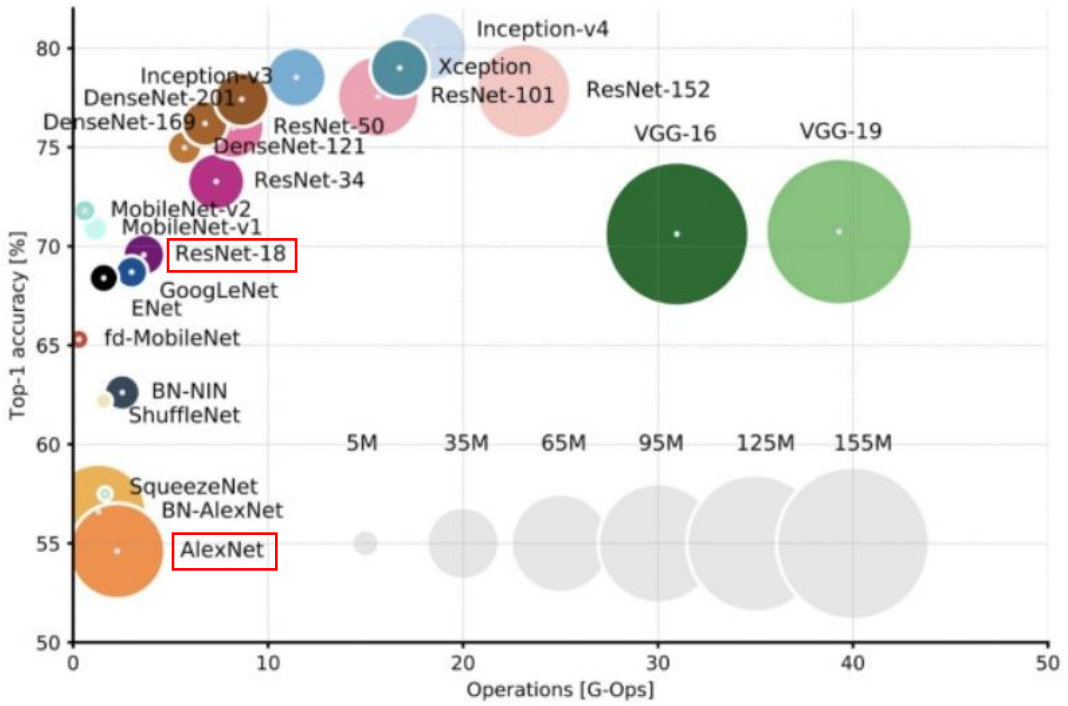
图 4：英伟达 CUDA 累计下载次数



资料来源：NVIDIA 2021GTC 大会

算法丰富、算法复杂度提升等，亦成为市场需求的重要驱动力。1）如前所述，过去 9 年，AI 芯片的算力大幅提升，也带动 AI 算法模型参数的大幅增加。从 Alexnet、ResNet 开始，到 BERT 网络模型，参数量已超过 3 亿规模，随后 GPT-3 模型超过百亿，Switch Transformer 的问世还一举突破万亿规模。**2）**英伟达 2020 年发布的 Megatron-LM 模型，参数量达到了 83 亿，相比于在 2018 年以参数量震惊世界的 BERT 模型又提升了 5 倍。模型体积几何倍数的增长也带了更多数据中心侧的需求，只有依靠上千块 GPU 并行运算才能在以天为单位的训练时长中完成对 Transformer 模型的训练。

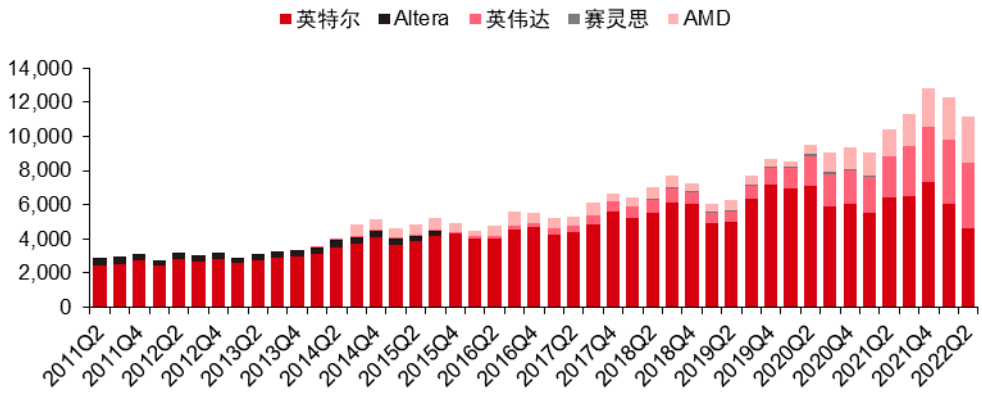
图 5：深度学习初期模型越来越大



资料来源：Purdue University，中信证券研究部

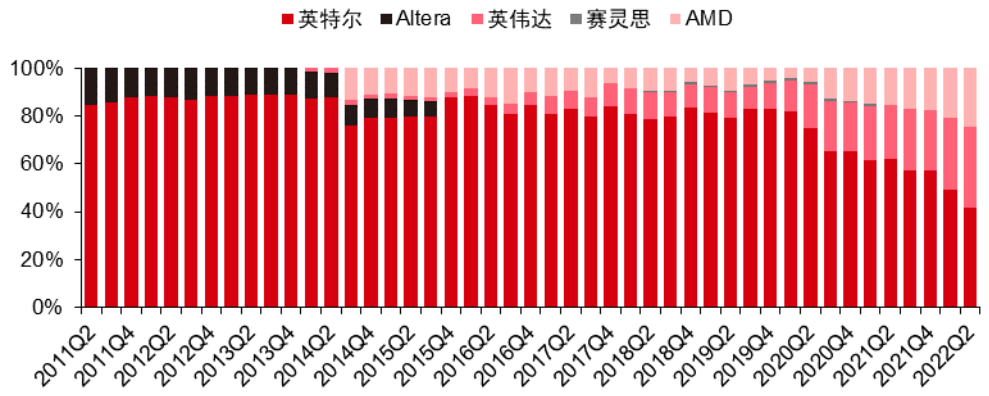
2021 年，全球数据中心逻辑计算芯片市场规模高达 436 亿美元。1）在过去数年，全球数据中心芯片市场保持高速增长，由 2012 的 122 亿美元增长至 2021 的 436 亿美元，符合增长率约 15%。2）从市占率来看，早期英特尔和 Altera 几乎垄断数据中心约市场份额，伴随着 AMD 和英伟达产品矩阵的增加，AMD 和英伟达在数据中心领域中的的市占率不断提升。截至 2022Q2，英特尔全球数据中心芯片市占率约 41.5%、英伟达市占率为 34.0%、AMD 市占率为 24.5%。

图 6：全球数据中心芯片市场营收规模（百万美元）



资料来源：Bloomberg，中信证券研究部。注：1）Altera 于 2015 年被英特尔收购；2）赛灵思于 2022 年被 AMD 收购

图 7：全球数据中心芯片市场市占率



资料来源：Bloomberg，中信证券研究部。注：1）Altera 于 2015 年被英特尔收购；2）赛灵思于 2022 年被 AMD 收购

英伟达历史借鉴：产品技术、软件生态等构筑 GPU 核心壁垒

近期，英伟达最新两则公告，导致市场对国产 GPU 的关注度提升。1）8 月 31 日，英伟达发布公告称：（a）8 月 26 日，美国政府对英伟达未来出口到中国(包括香港)和俄罗斯的 A100 和即将推出的 H100 芯片实施了新的许可证要求，该许可立即生效。新的许可证要求将解决涉及的产品可能用于或转移到中国和俄罗斯的“军事最终用途”或“军事最终用户”的风险。（b）该许可涉及到的芯片主要包括：英伟达 A100 和即将出货的 H100 两款芯片、基于 A100/H100 打造的 DGX 产品、以及未来实现峰值性能和芯片对芯片 I/O 性能均等于或大于大致相当于 A100 的阈值的任何 NVIDIA 芯片。目前来看，美国政府对中国以及俄罗斯出口限制的主要是针对数据中心的高端独立 GPU 芯片及相关产品。（c）公司于 2022 年 8 月 24 日提供的 FY2023Q3 展望（对应 CY2022 年 8 月-CY2022 年 10 月），其中有对中国大约 4 亿美元的潜在销售可能会受到新的许可证要求的限制。**2）**9 月 1 日，公司发布公告称，公司已美国政府新的授权审批，具体内容包括：（a）美国政府已批准英伟达继续开发 H100 芯片所需要的出口、在出口和国内转移。(b)允许英伟达在 2023 年 3 月 1 日前，为 A100 的美国客户提供所需的出口支持。目前，公司 A100 的美国客户包括戴尔、思科等服务器设备厂商，以及终端客户亚马逊、谷歌等。(c)美国政府授权 A100 和 H100，在 2023 年 9 月 1 日之前通过英伟达在中国香港的工厂履行订单和物流。（d）TAIPEI TIMES 报道，美国政府放宽许可授权的主要原因是，A100 的部分开发工作是依赖中国工程师&中国运营部门进行。若 A100 无法完成开发，对英伟达的业绩影响相对较大。

图 8：英伟达 8 月 31 日公告

On August 26, 2022, the U.S. government, or USG, informed us that it has imposed a new license requirement, effective immediately, for any future export to China (including Hong Kong) and Russia of our A100 and forthcoming H100 integrated circuits. DGX or any other systems which incorporate A100 or H100 integrated circuits and our A100X are also covered by the new license requirement. The license requirement also includes any future NVIDIA integrated circuit achieving both peak performance and chip-to-chip I/O performance equal to or greater than thresholds that are roughly equivalent to the A100, as well as any system that includes those circuits. A license is required to export technology to support or develop covered products. The USG indicated that the new license requirement will address the risk that the covered products may be used in, or diverted to, a 'military end use' or 'military end user' in China and Russia. We do not sell products to customers in Russia.

The new license requirement may impact our ability to complete our development of H100 in a timely manner or support existing customers of A100 and may require us to transition certain operations out of China, which could be costly and time consuming, and adversely affect our research and development and supply and distribution operations, as well as our revenue, during any such transition period. We are engaged with the USG and are seeking exemptions for our internal development and support activities.

We are engaging with customers in China and are seeking to satisfy their planned or future purchases of our Data Center products with products not subject to the new license requirement. To the extent that a customer requires products covered by the new license requirement, we may seek a license for the customer but have no assurance that the USG will grant any exemptions or licenses for any customer, or that the USG will act on them in a timely manner. The new requirement may have a disproportionate impact on NVIDIA and may disadvantage NVIDIA against our competitors, who are not subject to the same restrictions.

Our outlook for our third fiscal quarter provided on August 24, 2022 included approximately $400 million in potential sales to China which may be subject to the new license requirement. Our future revenue and profitability may be substantially reduced relative to this outlook, and our competitive position may be harmed, if customers do not want to purchase our alternative product offerings or if the USG does not grant licenses in a timely manner or denies licenses to significant customers. Even if the USG grants the requested licenses, the new requirement may benefit our competitors, as the licensing process will make our sales and support efforts more cumbersome, less certain, and encourage customers in China to pursue alternatives to our products, including semiconductor suppliers based in China, Europe, and Israel.

资料来源：Wind

图 9：英伟达 9 月 1 日公告

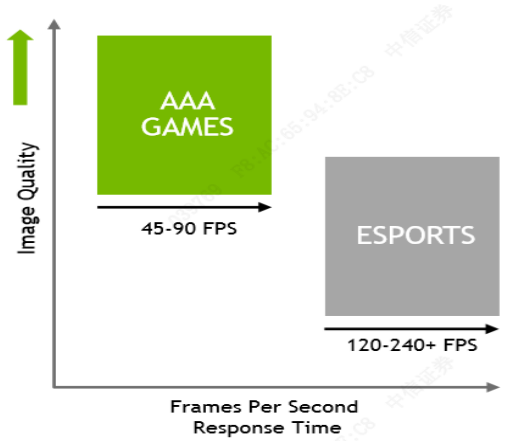
Item 8.01 Other Events

The U.S. government has authorized exports, reexports, and in-country transfers needed to continue NVIDIA Corporation's, or the Company's, development of H100 integrated circuits after the Company filed its Current Report on Form 8-K with the U.S. Securities and Exchange Commission on August 31, 2022. The authorization also allows the Company to perform exports needed to provide support for U.S. customers of A100 through March 1, 2023. Additionally, the U.S. government authorized A100 and H100 order fulfillment and logistics through the Company's Hong Kong facility through September 1, 2023.

资料来源：Wind

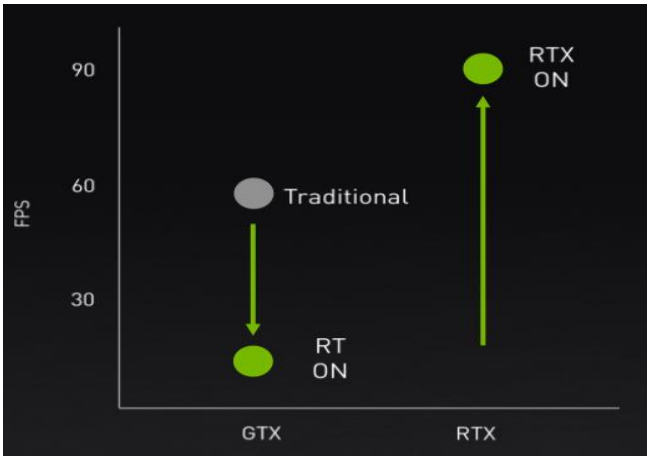
英伟达图形渲染领域：保持稳定、高频技术迭代，不断实现技术领先，例如 RTX&DLSS 等技术，并和开发者、应用厂商构成稳固的合作同盟。1）2020 年安培架构产品中，RTX 技术升级到第二代，并逐步向第三代 Tensor Core 技术推进，带动 RTX 系列显卡图像运算能力的全面提升，而 DLSS、Reflex 等能力带动游戏体验的提升，DLSS 2.0 将 FPS 提升近 30，Reflex 降低 50%的游戏延迟。对于超大型以及精品游戏的运行，大幅提升体验能力。对于大型 3A 游戏，在高画质条件下需满足 45-90FPS，电竞场景下需要 120-140FPS。在 GTX 的传统产品线中，开启 RT（光线追踪）之后，游戏帧数从 60 掉至不足 30 帧。但在 RTX 产品中，可提升至 90FPS 以上。**2）**鉴于英伟达 GPU 在软件领域的优势显著，公司 PC 用独显 GPU ASP 亦显著高于竞争对手 AMD。2016 年，英伟达 PC 用独显 GPU ASP 为 81.3 美元/个，AMD 对应 ASP 为 31.0 美元/个。2021 年，英伟达 PC 用独显 GPU ASP 为 163.2 美元/个，AMD 对应 ASP 为 86.6 美元/个。

图 10：不同类型游戏场景所需的帧数



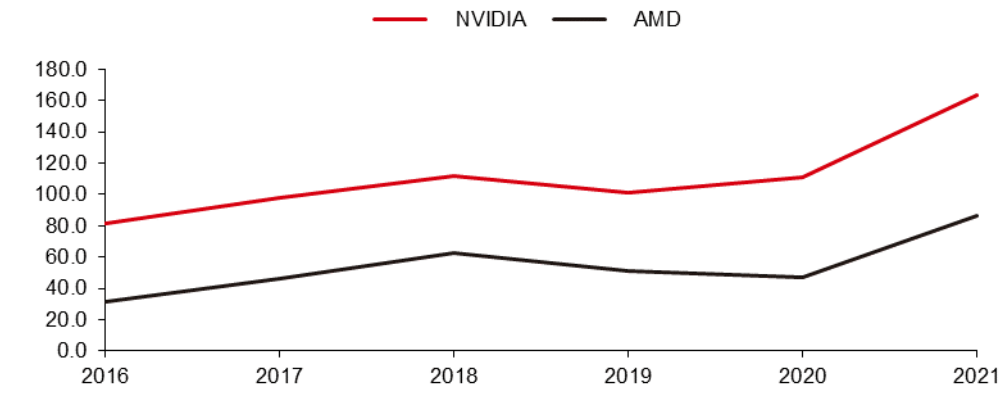
资料来源：2021GTC 大会

图 11：RTX 帧数大幅领先传统架构



资料来源：2021GTC 大会

图 12：英伟达&AMD PC 用独显 ASP（美元/个）



资料来源：IDC，中信证券研究部。

英伟达数据中心领域：借助 CUDA 实现 GPU 从图形显示到通用计算的跨越，以及产业生态壁垒的构建，并借助 DSA、NVlink 等架构创新、优化等实现持续性能领先。1）沿着技术层面的核心差异，我们按照训练&推理、边缘&数据中心两个维度，梳理目前全球主要的 AI 芯片参与者，整体而言，相较于全球其他主要竞争对手，英伟达在产品完整度、存量市场份额等层面实现领先，同时我们判断这种领先优势长周期亦将大概率维持。**2）**从公司的软件生态布局来看，英伟达构建了从底层到上游细分领域的应用开发软件，可大幅降低开发者的开发周期。

表 1：全球 AI 芯片主要参与者及下游应用场景

部署位置	芯片类型	训练（Training）	推理（Inference）
数据中心云端（Cloud）	GPU	英伟达、AMD	英伟达
	FPGA	英特尔、赛灵思	英特尔、赛灵思、亚马逊、微软、百度、阿里、腾讯
	ASIC	谷歌、华为	谷歌、寒武纪、比特大陆、Groq、Habana
边缘及终端（Device）	GPU	-	英伟达、ARM
	FPGA	-	深鉴科技

ASIC	-	寒武纪、地平线、华为海思、高通、ARM

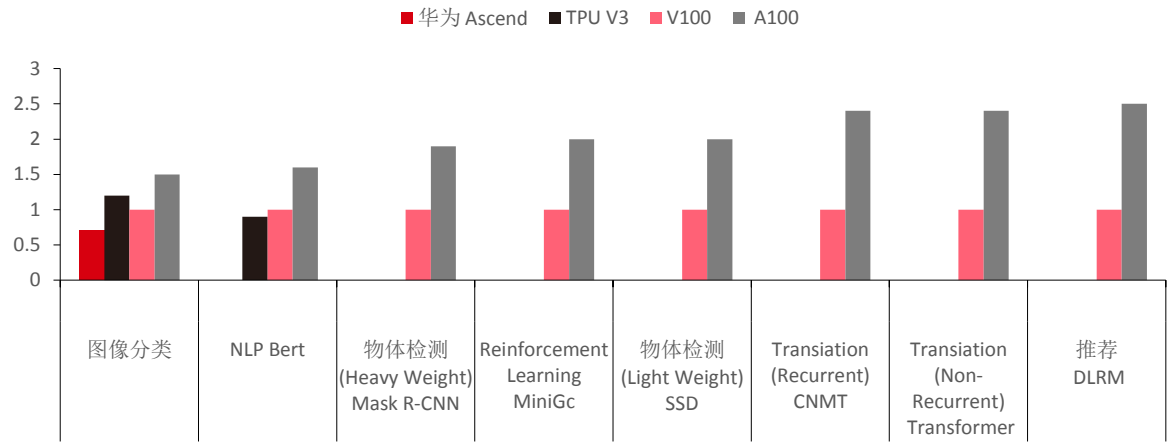
资料来源：各公司官网，中信证券研究部

- **产品丰富度&技术竞争力**：英伟达系统级产品布局、在训练环节的突出表现&领先优势已基本成为市场的共识，而在推理领域，伴随新一代安培架构、Hopper 架构的升级，以及由此实现的训练、推理的统一，以及对稀疏矩阵运算问题的良好支持，目前在推理方面，英伟达最新的 A100 芯片的 Int 8 Tops 已经达到 1248，较上一代提升超过 5X。同时在训练环节，根据 Mlperf 的评测，在图像识别、对话式 AI、推荐系统等多个模型的对比评测中，英伟达芯片训练性能全面领先华为、谷歌等主要竞争对手。基于技术层面的全面分析，我们判断英伟达有望在企业对外服务（训练、推理）、企业内部服务（训练）环节保持持续领先，但在企业内部服务（推理）仍面临延迟、功耗等层面的明显短板。而我们看到，英伟达在数据中心市场的产品迭代节奏继续延续既有的习惯，即继续保持对芯片性能的狂热追逐，以及每两年升级一次产品（CPU、DPU、GPU）的频率。

图 13：全球 AI 芯片市场主要参与企业（按主要场景划分）

资料来源：中信证券研究部绘制

图 14：训练相对加速倍数 Mlperf 评测



资料来源：英伟达官网，中信证券研究部

表 2：英伟达在 AI 训练、推理环节优劣势分析

	AI 训练				AI 推理			
外部服务	判断：芯片性能、软件堆栈等支撑英伟达持续领先				判断：有灵活性优势，但延迟、功耗仍是潜在的不足			
	谷歌	微软	亚马逊	Facebook	谷歌	微软	亚马逊	Facebook
	TPU v2/3	N/A	N/A	N/A	TPU v2/3	Brainwave	Inferentia	N/A
内部使用	判断：英强大芯片性能能在一定程度上降低成本端劣势，但面临着内部解决方案和 ASIC 的竞争				判断：延迟、功耗将成为主要短板，竞争力一般			
	谷歌	微软	亚马逊	Facebook	谷歌	微软	亚马逊	Facebook
	TPU v2/3	N/A	N/A	N/A	TPU v1/2/3	Brainwave	N/A	N/A

资料来源：各公司官网等，中信证券研究部

- **英伟达基于 CUDA 构建了丰富的软件生态，显著提升 GPU 的易用性**。从软件技术分类来看，公司在软件领域中的产品布局主要分为：基础架构、游戏与娱乐、应用工具、应用框架四大部分。具体内容如下：（a）在基础架构方面，公司软件产品主要围绕 AI&通用能力布局。其中 AI 主要包括边缘 AI、AI 垂直领域解决方案、AI 推理等；通用领域则围绕 IO 传输、vGPU 等。（b）在游戏娱乐方面，公司的产品布局主要包括 Geforce 云游戏平台、直播领域的 Broadcast App 和元宇宙领域中的 Omniverse Machinima；（c）在应用工具方面，公司可面向不同的应用场景（AI、数据分析、元宇宙等领域），提供不同的开发工具。如：在 AI 领域，可提供加速 AI 部署与工作流程的 NGC 产品；在元宇宙领域，可提供 3D 虚拟协作的 Omniverse 产品。（d）在具体应用框架方面，主要凭借公司 AI 与数据分析能力，在自动驾驶、视频分析、推荐系统等各垂直领域提供具体应用框架，帮助提高各行业运营效率。

表 3：英伟达软件产品布局一览

一级分类	二级分类	三级分类	软件名称
基础架构	AI	AI 垂直领域解决方案	AI Enterprise
	AI	边缘 AI 计算	EGX
	AI	简易边缘 AI 部署与运维	Fleet Command
	AI	AI 推理	Triton

一级分类	二级分类	三级分类	软件名称
	通用	IO 传输加速	Magnum IO
	通用	虚拟 GPU	vGPU
	通用	软件	软件
游戏与娱乐	游戏	游戏参数自适应	GeForce Experience
	直播	优化视频质量	Broadcast App
	元宇宙	创建 3D 角色与场景	Omniverse Machinima
应用工具	AI	加速 AI 部署与工作流程	NGC 目录
	数据科学	数据分析	NVIDIA 工作台
	元宇宙	线上 3D 虚拟协作	Omniverse
	通用	数据中心 GPU 监控	DCGM
	通用	高质量录屏	RTX Experience
	通用	桌面窗口管理	RTX 桌面管理器
具体应用框架	AI	自动驾驶	NVIDIA DRIVE
	AI	云端 AI 视频流	NVIDIA Maxine
	AI	对话 AI	NVIDIA Riva
	AI	医疗	NVIDIA Clara
	AI	智能视频分析	Metropolis
	AI	机器人	Isaac
	通信	5G	Aerial
	数据科学	推荐系统	Merlin
	数据科学	数据分析	RAPIDS
	通用	高性能运算	NVIDIA HPC SDK

资料来源：公司官网，中信证券研究部

小结：伴随 AI、高性能计算、大规模图形渲染等应用场景的不断拓展和丰富，市场对大算力并行计算芯片的需求快速增长，截止目前，全球数据中心领域逻辑芯片市场规模已经超过 400 亿美元。同时近期市场对国产 GPU 领域的关注度提升。基于英伟达的历史复盘，可以看出公司在图形渲染&数据中心领域保持较高的市占率，并实现产业引领。我们认为核心原因在于：借助持续、高频迭代保持产品技术行业领先，并借助 CUDA 等实现软件生态构建，不断提升产品易用性等。GPU 作为大算力并行计算芯片领域最为可行的承载者，在本篇报告中，我们将从全球市场出发，就 GPU 产业本身的产品特性、技术路线、市场空间，以及国内市场现状、演进路径、竞争格局等展开系统的分析和讨论，力图针对国内 GPU 市场构建一个完整的产业&投资蓝图。

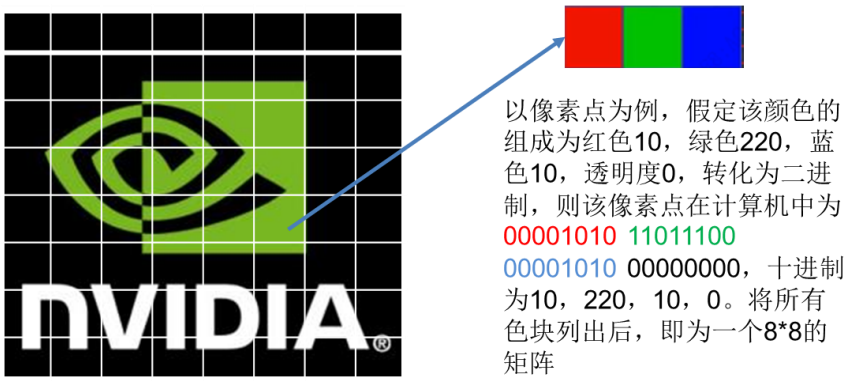
全球 GPU 市场：并行计算理想载体芯片，数据中心为中期需求增长主要场景

GPU：通用并行计算理想载体芯片，从图形处理向 AI、高性能计算等领域扩展

GPU 定义：图形处理器，但承载功能已在早期定义上明显泛化。1）发展早期，更多

称为图形处理器（GPU），又称显示核心、视觉处理器、显示芯片，是一种专门在个人电脑、工作站、游戏机和一些移动设备（如平板电脑、智能手机等）上做图像和图形相关运算工作的微处理器。**2）**由于计算机只能识别二进制数字，因此在进行图形运算时，要把图片转换成计算机能够理解的二级制数组（见下图图示），因此 GPU 在进行运算时，所针对的都是矩阵数据，因此 GPU 的大部分计算是并行的。这意味着 GPU 更加适合并行计算与矩阵运算。

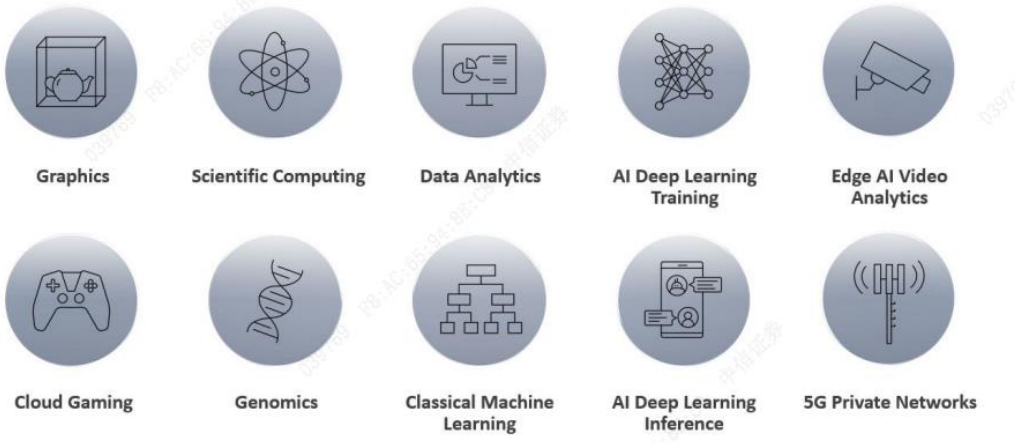
图 15：以红绿蓝三原色为例，计算机如何表示图像



资料来源：Techweb

GPU 应用场景：由早期的图形渲染，逐步拓展至高性能运算、科学计算等领域，GPU 是通用并行计算的理想载体。1）由于计算机以及图形运算的特性，GPU 所进行的运算多数为矩阵运算、并行运算，这些特征使得 GPU 更加适合当前以 AI 为代表的高性能计算、科学计算等领域，GPU 的使用范围也由早期的图形渲染领域，逐步拓展至高性能运算&科学运算领域。**2）**与其他逻辑计算芯片相比，GPU 在通用性、计算速度、规模化部署经济性等核心指标上面，能够做到较好的平衡，因此在目前 AI、复杂科学计算等并行计算领域，逐步形成了 GPU 主导，FPGA、ASIC、CPU 为辅的稳定局面。

图 16：GPU 可适用的计算范围



资料来源：英伟达官网

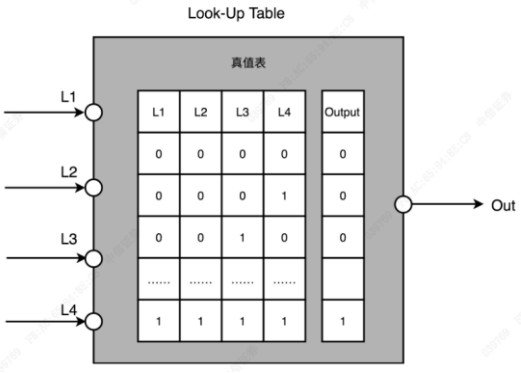
- **CPU：适合处理复杂的串行计算和逻辑控制，并行运算性能显著弱于 GPU**。由于功能与设计架构的不同，CPU 与 GPU 的计算能力也存在差异，CPU 的架构使得其适合流水线式的串行计算与复杂计算，而 GPU 的架构使得其适合运算逻辑简单但可以同步进行的并行计算。因此在参数上，我们会看到 CPU 具有更高的频率与缓存，而 GPU 具备更多的核心。

图 17：CPU 与 GPU 架构

资料来源：赛迪网。注：绿色的是计算单元，橙红色的是存储单元，橙黄色的是控制单元。

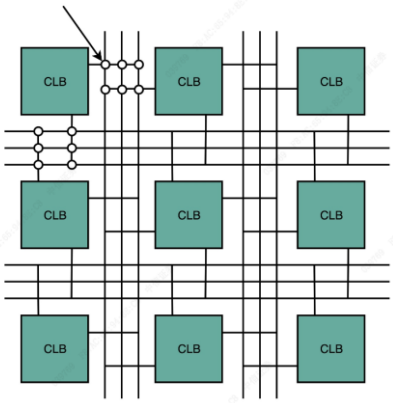
- **FPGA：灵活性突出，但易用性、计算速度、经济性较 GPU 欠佳**。FPGA 是一种偏向于硬件的可编程芯片，FPGA 中使用了大量逻辑门（数字电路中的基础部件，通过电压高低以及组合，将输入的命令转化为 0 或 1），建立真值表（输入不同代码，输出不同结果的查询表），通过可编程逻辑布线（可以理解为电路开关，编程即是对开关调整，实现门之间的电路组合）来实现算法。由于直接对硬件编程，相较于 GPU 的平均计算效率与可编程性更高，但由于需要直接对硬件进行编程以及较高的成本（为满足编程要求通常晶体管冗余设计），通用性、大规模部署成本以及最高计算能力不如 GPU。

图 18：逻辑门组合为真值表以及 CLB



资料来源：servetheh

图 19：CLB 与可编程逻辑布线构成 FPGA

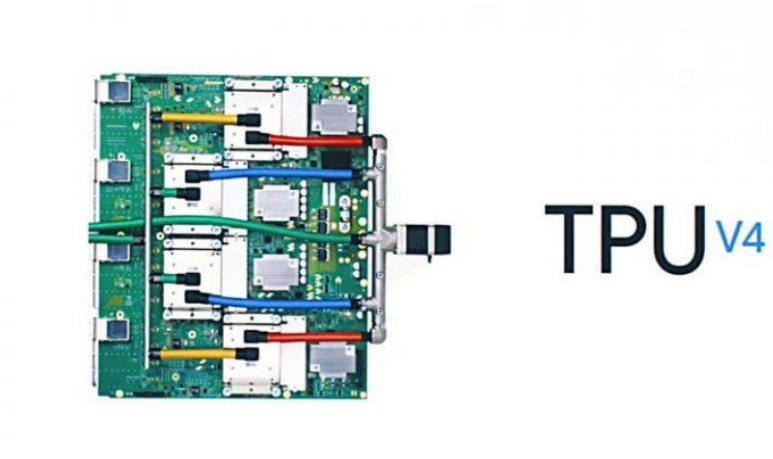


资料来源：servetheh

- **ASIC：特定场景性能最优，但通用性不足**。ASIC 芯片是针对某一特定场景所研制的专用芯片，优势在于运算效率极高、部署成本较低。但对于实际应用而言，

如果算法出现迭代升级或数据结构发生改变，ASIC 的效率将会大幅下降，因此相较于 GPU 而言，ASIC 更多用于挖矿、音视频解码等专用场景。因此 ASIC 的平均算力会更强，但在通用场景下以及最高运算能力上，GPU 优势更大。

图 20：谷歌云专用 AI 处理器 TPU v4 为 ASIC 芯片



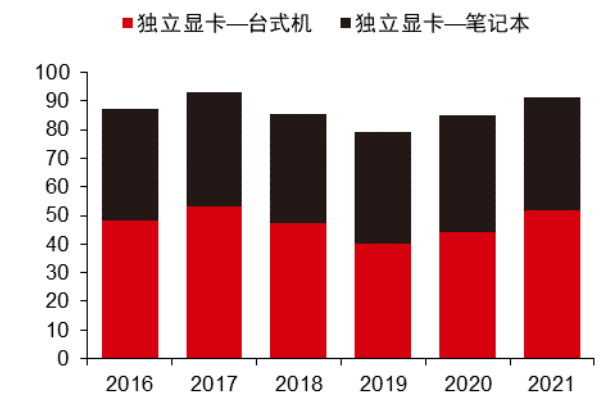
资料来源：智东西

小结：整体而言，正是基于 GPU 本身的优异特性，以及英伟达等企业在芯片架构、软件生态等层面的不断努力，叠加 AI、高性能计算、大规模图形渲染等应用场景的快速崛起，GPU 逐步成为全球大算力并行计算领域的主导者。而在产品端，我们也总结发现，GPU 厂商亦结合下游的应用场景，在一个大的体系结构下，针对计算单元、缓存、总线带宽等技术点的优化和组合。在下文内容中，我们主要讨论当下最主流的应用场景&产品：用于游戏等场景中图形渲染的显卡，以及用于数据中心 AI、高性能计算等场景的 GPGPU（通用计算 GPU）。

图形渲染：游戏为主，中期有望保持 10%~15%平稳增长

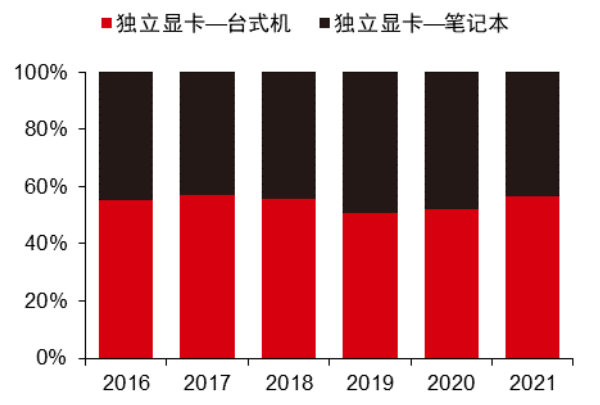
目前在图形渲染领域，游戏画面渲染为主要应用场景，同时亦包括图形工作站等场景，独立显卡为主要硬件载体。IDC 数据显示，目前全球独立显卡出货量，近 5 年稳定在 8000-9300 万部。按独立显卡的类型划分，其中台式机用独立显卡比例约 40%-53%，笔记本&工作站独立显卡比例约 47%-60%。按照品牌商来看，英伟达独立显卡近 5 年市占率一直稳步提升，市占率由 2018 年的 58.8%提升至 2021 年的 74.3%，AMD 市占率由 2018 年的 31%降低至 2021 年的 19%。

图 21：独显 GPU——出货量（百万个，按类型类型划分）



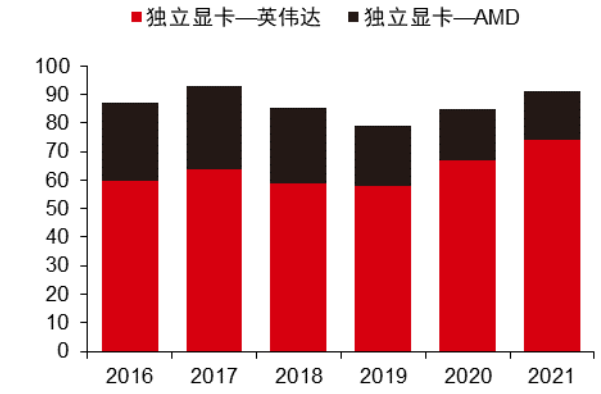
资料来源：IDC，中信证券研究部

图 22：独显 GPU——出货量占比（%，按类型划分）



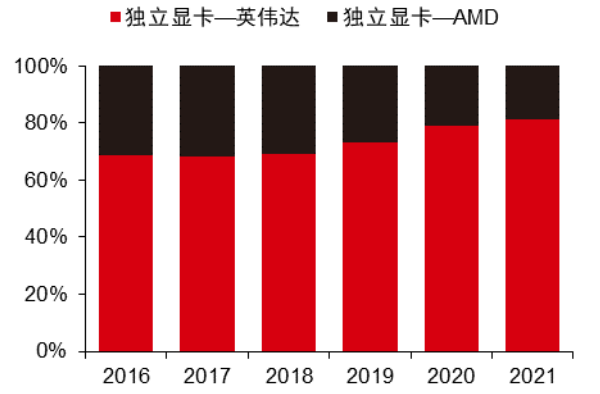
资料来源：IDC，中信证券研究部

图 23：独显 GPU——出货量（百万个，按品牌划分）



资料来源：IDC，中信证券研究部

图 24：独显 GPU——出货量占比（%，按品牌划分）



资料来源：IDC，中信证券研究部

市场规模判断：预计 2025、2030 年将分别达到 278、568 亿美元。2021 年，英伟达游戏显卡业务实现销售收入 105 亿美元，专业视觉收入（图形工作站）21 亿美元。我们假设英伟达在全球游戏显卡领域收入占比 80%，专业视觉领域收入占比 80%，则 2021 年，在图形渲染（含游戏、专业视觉等）领域，全球 GPU 市场规模为 158 亿美元。同时为了测算该领域中期市场规模，我们作出如下简化假设：1）假设图形渲染领域，中期应用场景仍主要由游戏画面渲染、专业视觉构成，其他长尾场景忽略；2）显卡 ASP，参考英伟达产品价格走势，考虑到产品性能、制造成本等因素，预计显卡 ASP 年复合增速在 10%~15%之间，取中位值 12.5%；3）游戏用户，疫情期间，全球高端游戏玩家出现大幅增长（预计增幅 1 亿人），中期预计保持平稳增长，每年增速 0~5%，取中位值 2.5%；4）假设专业视觉的市场规模占游戏比例维持在 20%左右。综合上述假设，中性情形下，我们预计全球 GPU（图形渲染）在 2025、2030 年的市场规模有望分别达到 278、568 亿美元。

表 4：全球 GPU（图形渲染）市场规模预测

类别（亿美元）	指标
游戏市场规模（2021 年）	131

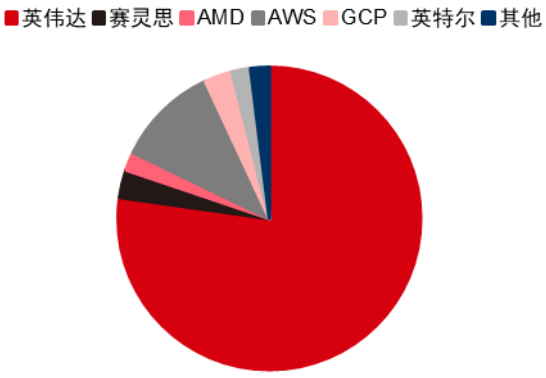
类别（亿美元）	指标
专业视觉市场（2021 年）	26
显卡市场合计（2021 年）	158
游戏用户-中期复合增速	2.5%
游戏显卡 ASP-中期复合增速	12.5%
游戏市场-年复合增速	15.3%
专业视觉/游戏显卡收入比重	20%
游戏市场规模（2025 年）	232
专业视觉市场（2025 年）	46
显卡市场合计（2025 年）	278
游戏市场规模（2030 年）	473
专业视觉市场（2030 年）	95
显卡市场合计（2030 年）	568

资料来源：IDC、彭博，中信证券研究部预测

数据中心：AI&高性能计算等，预计中期保持 25%以上年均复合增速

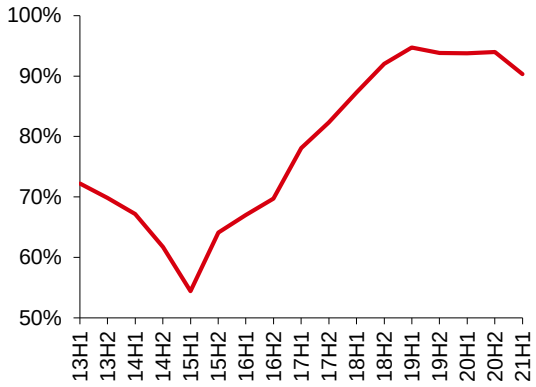
市场格局：英伟达 GPU 在 AI 训练、高性能计算领域占据主导地位。作为图形渲染之后另一主要应用场景，目前客户主要通过在数据中心部署英伟达、AMD 的 GPU 芯片，实现 AI 训练、高性能计算等，同时辅以自研加速卡等，服务于特定场景的 AI 训练、推理等。**1）**根据 Liftr Insights 数据显示，2021Q1，在全球 TOP 云厂商数据中心 AI 加速芯片市场，英伟达份额占比为 78%，近年来基本稳定在 80%附近，市场领先地位稳固。同时根据 Lifter 2019 年 5 月的数据显示，全球四大云计算平台阿里云、AWS、Azure 和谷歌云（GCP）中，英伟达 TESLA 系列 GPU 产品的市场占有率大幅领先。其中，阿里云采用英伟达 TESLA 系列 GPU 比例为 81%，AWS、Azure 和 GCP 使用比例分别为 89%、100%和 100%，市场份额绝对领先。**2）**另外据 Top500.Org 数据显示，英伟达 GPU 产品在全球 Top 500 超算中心的渗透率逐年提高，由 2013H1 的 72.2%提升至 2021H2 的 90.3%，几乎处于垄断地位。

图 25：全球 TOP 云厂商数据中心部署并行计算芯片份额结构（2021）



资料来源：Liftr Insights，中信证券研究部

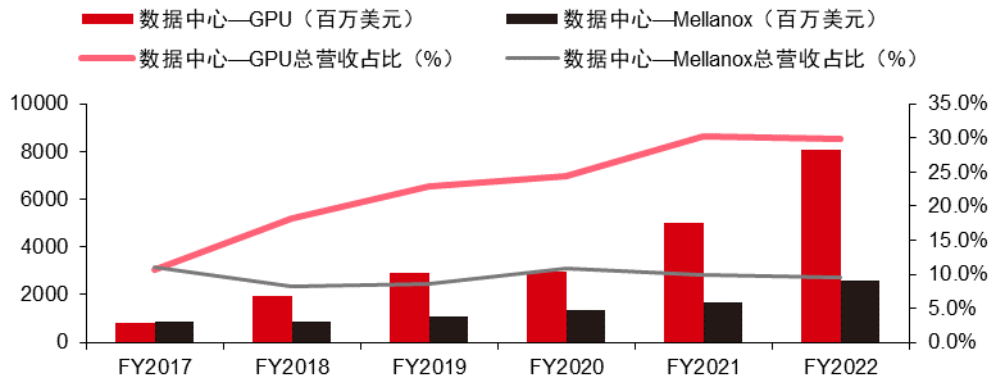
图 26：英伟达 GPU 产品在全球 Top 500 超算中心市场占有率



资料来源：Top 500.Org，中信证券研究部

市场规模：我们测算全球数据中心 GPU 芯片市场规模，2021 年约为 100 亿美元左右。FY2022（对应 CY2021）英伟达数据中心营收约 106 亿美元，其中 Mellanox 营收约 25.7 亿美元，则英伟达数据中心 GPU 相关产品营收约 80.3 亿美元。在市场竞争段落中提到，英伟达在数据中心领域中的市占率约 80%，依次测算，2021 年，全球数据中心 GPU 芯片市场规模约为 100 亿美元左右。

图 27：英伟达数据中心营收构成及占比：按不同业务划分



资料来源：Bloomberg，中信证券研究部测算。注：Mellanox 于 FY2021 并表，FY2017-FY2020 计算营收占比时，将 Mellanox 的收入计入英伟达总营收中，方便前后对比

GPU 数量：我们测算 2021 年，全球数据中心 GPU 芯片市场出货量约 200 万个。依据英伟达在数据中心领域中 GPU 产品的价格测算，假设对应产品的 ASP 约 5000 美元/个，对应 FY2022 年（对应 CY2021 年）英伟达 GPU 产品出货量约 160 万个。在市场竞争段落中提到，英伟达在数据中心领域中的市占率约 80%，依次计算，全球数据中心 GPU 市场出货量约 200 万个。

图 28：2020Q1，阿里云、亚马逊云、微软云 GPU 加速卡市占率

资料来源：Liftr Insights

表 5：公司数据中心主要产品参数及售价

产品分类	产品名称	发布时间	售价（万美元）	主要参数及性能
GPU Tesla 系列（计算显卡）	P100	2016	0.75	**Pascal 架构**，3584 个 CUDA cores，单精度 10.6T，显存 16GB，显存带宽 720GB/s
	P4	2016	0.25	**Pascal 架构**，3584 个 CUDA cores，单精度 8T，显存 8GB，显存带宽 192GB/s
	V100	2017	1.15	**Volta 架构**，5120 个 CUDA cores，单精度 15.7T，显存 32GB 或 16GB，显存带宽 900GB/s
	T4	2018	0.25	**Turing 架构**，2560 个 CUDA cores，单精度 8.1T，显存 16GB,带宽 320GB/s
	A100	2020	1.5-2.7	**Ampepre 架构**，6912 个 CUDA cores，单精度 19.5T，显存 40GB，显存带宽 1.6 TB/s
	H100	2022	3.65	**Hopper 架构**，7296 个 CUDA cores，单精度 60T，显存 80GB，显存带宽 3 TB/s
DGX（主要用于 AI）	DGX-1	2017	14.9	8 个 Tesla V100 GPU，512Gb DDR4；2 个 Intel E5-2698 CPU
	DGX-2	2018	39.9	16 个 Tesla V100 GPU， 1.5TB 内存，2 个 Intel 8168 CPU
	DGX-A100	2020	19.9	8 个 Tesla A100 GPU，1TB 内存，2 个 AMD 7742CPU
	DGX-H100	2022	NA	8 个 Tesla H100 GPU，2TB 内存， 2 个 X86 架构服务器，4 个 NVSwitch
HGX（用于 AI 和超大型数据中心加速器）	HGX-1	2017	14.9	8 个 Tesla V100，显存 256GB
	HGX-2	2018	39.9	16 个 Tesla V100，显存 512GB
	HGX-3	2020	NA	共三个版本，分别搭载 4/8/16 个 Tesla A100，显存分别为 160/320/640GB
	HGX-H100	2022	NA	共四个版本，分别搭载 4/8/16/32 个 Tesla A100，显存分别为 320GB/640GB/10TB/20TB

资料来源：公司官网，中信证券研究部

数据中心 GPU 市场规模：预计 2025、2030 年将分别达到 245、828 亿美元。结合既有的认知和判断，我们做出如下简化假设：1）假设中期全球数据中心大算力逻辑芯片市场增速和过去相似（2014~2021 年），年市场规模复合增速维持在 15%~20%之间，取中位值 17.5%；2）数据中心领域，并行计算需求占比持续提升，预计每年相对份额提升 3%左右。基于上述简化假设，我们中性预计，全球数据中心 GPU 市场规模将在 2025、2030 年分别达到 245、828 亿美元，同时若中期 AI 技术进步、高性能计算需求超出我们的预期，则最终市场规模将显著高于我们当前的预测。

表 6：全球数据中心 GPU 芯片市场规模测算/预测（亿美元）

类别	数值
全球数据中心逻辑芯片市场（2021 年，亿美元）	436
全球数据中心 GPU 市场（2021 年，亿美元）	100
GPU/逻辑芯片份额占比（2021）	23%
数据中心逻辑芯片市场-年复合增速	17.5%
数据中心 GPU 年份额相对增幅	3%
全球数据中心逻辑芯片市场（2025 年，亿美元）	831
全球数据中心逻辑芯片市场（2030 年，亿美元）	1861
GPU/逻辑芯片份额占比（2025）	30%

类别	数值
GPU/逻辑芯片份额占比（2030）	45%
全球数据中心 GPU 市场（2025 年，亿美元）	245
全球数据中心 GPU 市场（2030 年，亿美元）	828
全球数据中心 GPU-复合增速（2021~2030）	26%

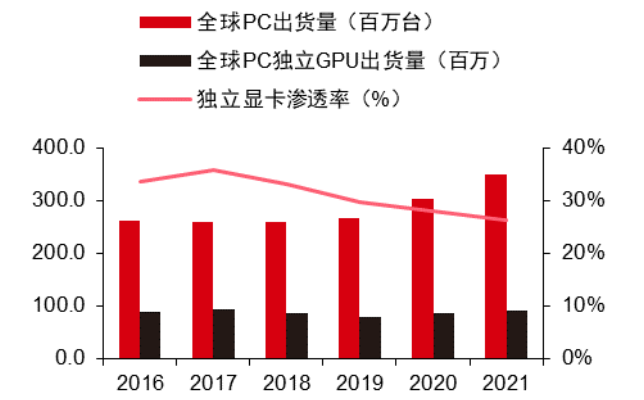
资料来源：IDC、彭博，中信证券研究部测算/预测

国内 GPU 市场：中期潜在空间可观，本土厂商开始规模崛起&产品落地

国内市场现状：和全球市场同步，预计 2030 年规模将突破 300 亿美元

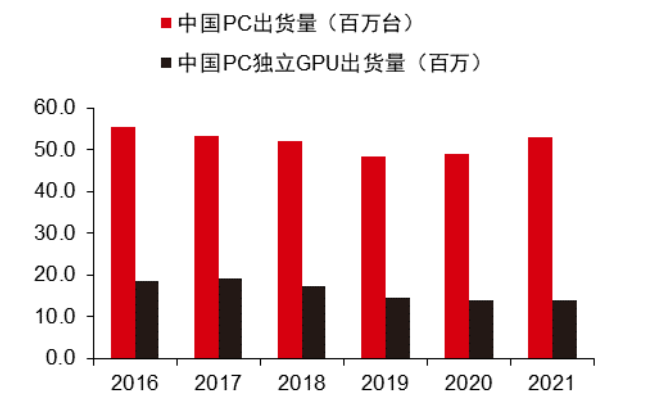
图形渲染：当前国内市场规模约 27 亿美元，预计 2025、2030 年将分别达到 47、97 亿美元。由于缺乏直接的统计数据，我们做出如下简化假设：1）IDC 数据显示，2016-2021 年，全球 PC 出货量为 2.6-3.5 亿台，同期国内 PC 销量占全球销量比重约在 17%左右，我们假设在图形渲染领域，国内 GPU 出货量占比亦和 PC 表现相对一致，并保持和全球市场相似的增速，以及应用场景分布等。参考我们在上文中的测算，我们测算、预测 2021 年、2025 年、2030 年，国内 GPU（图形渲染）的市场规模约为 27、47、97 亿美元。当然，若考虑到国内庞大的游戏用户数，以及专业视觉等领域的旺盛需求等，最终的实际数据料将大幅优于我们当前的测算&预测。

图 29：全球 PC 用独立显卡 GPU 渗透率测算



资料来源：IDC，中信证券研究部测算

图 30：中国 PC 用独立显卡 GPU 出货量（百万）

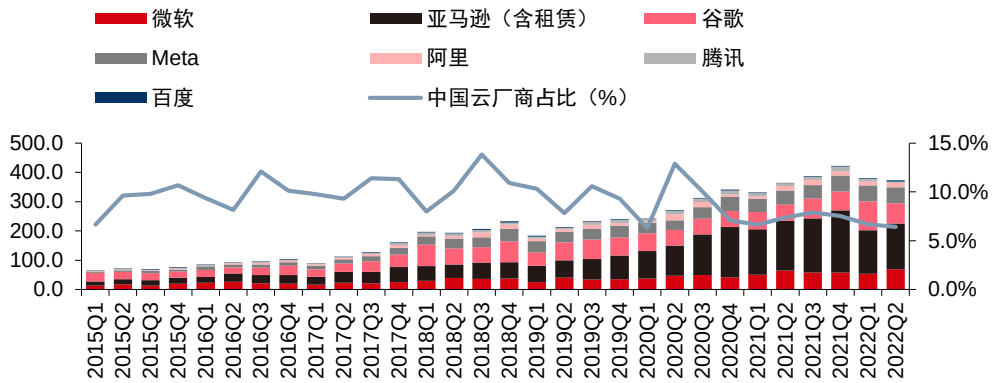


资料来源：IDC，中信证券研究部测算

数据中心：我们测算国内数据中心 GPU 市场约占全球 20%左右比重，对应 2021 年整体出货量约 40 万个，对应市场规模约 20 亿美元。1）从互联网云厂商 Capex 支出来看，阿里巴巴+腾讯+百度三家互联网厂商的 Capex 占全球主要互联网云商场（微软、亚马逊（含租赁）、谷歌、Meta）总 Capex 比例的 7%-13%。若扣除亚马逊在租赁领域中的 Capex 支出，我们预计中国三家互联网厂商的 Capex 占比将超过 10%。**2）**Top 500.Org 网站显示，截至 2021 年 11 月，全球 Top 500 超算中心，中国拥有 173 个超算中心，为全世界最多的超算中心国家，占有率约 34.6%。**3）**综合考虑中国互联网云厂商 Capex 占比约 10%，Top 500 超级计算机个数市占率约 34.6%，我们认为中国数据中心 GPU 需求量约占全球

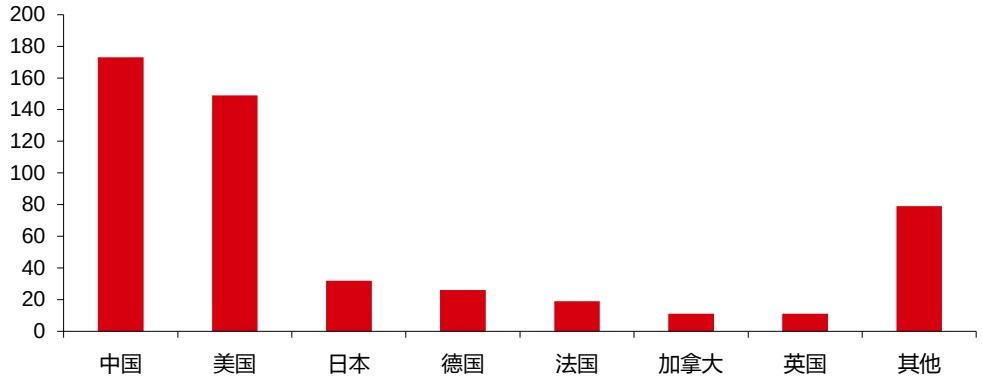
数据中心总 GPU 需求量的 20%左右。**4）**如前面章节测算，我们判断 2021 年全球数据中心 GPU 加速器市场出货量约 180-200 万个，2021 年全球数据中心 GPU 加速市场规模约 100 亿美元。按照 20%市占率计算，我们预计 2021 年中国数据中心 GPU 加速器市场出货量约 40 万个，对应市场规模约 20 亿美元。

图 31：全球主要云厂商 capex 支出（亿美元）



资料来源：Bloomberg，中信证券研究部

图 32：全球 Top500 超算中心分布（按地区）



资料来源：Top 500.Org，中信证券研究部

中期展望：我们预计 2030 年国内数据中心 GPU 芯片市场规模有望增长至 250 亿美元，对应 CAGR 为 32%。如前所述，我们预计全球数据中心 GPU 加速市场规模有望由 2021 年的 100 亿美元增长至 2030 年的 828 亿美元（对应 CAGR 为 26%）。综合考虑国内 AI、高性能产业的发展，以及头部科技公司的资本开支，Top 500 超级计算机数量等，我们认为未来中国数据中心 GPU 芯片需求量将占到全球数据中心总 GPU 需求量的 25%-30%左右。依此计算，我们预计中国数据中心 GPU 芯片市场规模有望由 2021 年的 20 亿美元增长至 2030 年的 250 亿美元（对应 CAGR 为 32%）。当然考虑到国内企业在 AI、高性能计算领域的积极努力和进展，最终实际数字大概率会好于我们当前的中性预期。

国内市场格局：本土厂商快速崛起，产品亦逐步上市

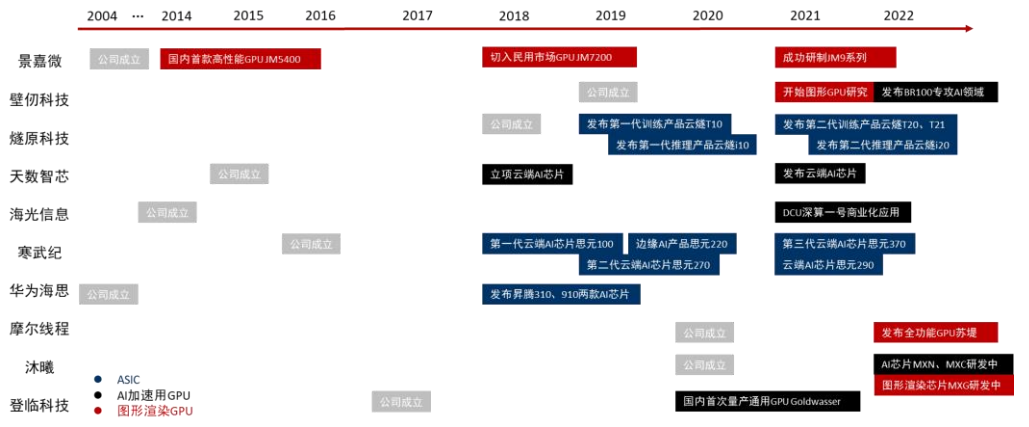
国内 GPU 厂商：开始快速崛起，大多数企业目前已发布 1-3 款相关产品，大部分核心团队具有英伟达、AMD 工作经历。1）2014-2020 年，国内成立若干 GPU&云端 AI 芯片相关企业，目前此类企业已发布 1-3 代产品，产品落地进度不断加快。**2）**从国产 GPU 相关企业创始人的团队背景来看，大部分企业创始人团队均有在英伟达、AMD 等企业有多年的工作经验。

表 7：国内 GPU 相关部分企业梳理

企业名称	成立时间	代表性产品	应用场景
景嘉微	2006	JM9 系列	图形渲染、AI 计算
摩尔线程	2020	苏堤	图形渲染、AI 计算
芯动科技	2007	风华二号	图形渲染、AI 计算及科学计算
龙芯中科	2001	7A2000	图形渲染
芯瞳半导体	2019	Genbu01	图形渲染
海光信息	2014	深算一号	AI 训练、推理及科学计算
壁仞科技	2019	BR100	云端推理、训练及科学计算
寒武纪	2016	思元 370/220	云、边的推理和训练
登临科技	2017	Goldwasser	云、边的推理和训练
鲲云科技	2016	CAISA 芯片	云、边数量流推理
瀚博半导体	2018	SV100	云、端 AI 推理
天数智芯	2015	BI	云端训练
燧原科技	2018	云燧 i20	云端推理
阿里平头哥	2018	含光 800	云端推理
墨芯	2021	ANTOUM	云端推理
沐曦科技	2020	MXN	云端推理
华为海思	2004	昇腾 910	边缘计算 AI
昆仑芯	2011	2 代加速芯片	AI 计算
华夏芯	2014	可编程 AI 加速卡	AI 计算

资料来源：IT 橙子，中信证券研究部

图 33：中国国产 GPU 企业发展历史



资料来源：各公司官网，中信证券研究部绘制

表 8：中国 GPU 厂商创始人团队背景

	公司	创始人	简介
AMD 背景	天数智芯（2015）	郑金山	供职于 Trident、XGI、ATI Technologies、AMD 和酷芯微电子，任高级经理、PMTS 和架构师等职位。
	燧原科技（2018）	赵立东 CEO	犹他州立大学电子与计算机硕士学位、清华大学电子工程学士学位
		张亚林 COO	Juniper Networks/AMD 产品工程部高级总监/紫光通信副总裁，曾任职 AMD。
	壁仞科技（2019）	张文	跨界投资者、曾任商汤科技并担任总裁。
		李新融 CEO	AMD 全球副总裁、中国研发中心总经理。
		陈文中 高级副总裁	在 AMD 带领 500 人的技术团队，在 8 年内实现了 9 款芯片的流片与量产，其中包括首款采用 HBM 技术的 GPU 芯片。
	沐曦（2020）	陈伟良 CEO	清华大学微电子学硕士，AMD 近 14 年工作经验。
		杨建 CTO	浙江大学毕业，AMD 大中华区第一位科学家，曾参与及主导数十款 GPU 产品量产及交付全流程，并作为三维图形与科学计算生态专家，拥有多项发明专利。历任 Trident、S3、ATI/AMD、海思等公司芯片架构师、软件架构师、首席架构师等职位。
		彭莉 硬件架构师	15 年高性能 GPU 芯片设计经验，历任 AMD 首席 SOC 架构师、系统架构师、GFXIP 架构师。
英伟达背景	摩尔线程（2020）	张建中	前英伟达全球副总裁。
其他	景嘉微（2004）	曾万辉	国防科学技术大学微波与毫米波技术硕士。
	登临科技（2017）	李建文	清华大学微电子所/在 GPU 领域有二三十年的从业经历，曾在图芯科技（2004 年创立）担任副总裁，由他负责的 GPU/GPGPU IP 产品。

资料来源：IT 橙子，中信证券研究部

产品竞争力：国内厂商产品核心参数约落后英伟达、AMD 1~2 代左右，正逐步从“可用”走向“好用”。1）通过对比海外 GPU 厂商和国内 GPU 厂商相关产品的参数，可以看出国内 GPU 厂商在半精度&单精度领域中的计算能力，相差约 1 代差距；国内 GPU 厂商在双精度（64 位）计算领域能力近乎空白，但双精度运算更多应用于复杂科学计算。**2）**考虑到英伟达、AMD 在 GPU 架构中加入了张量核 TensorCore 或 Matrix Core（可用于执行融合乘法加法运算），这种计算单元层面的 DSA 架构设计，亦使得他们在 AI 训练、推理环节具有更高的计算效率：

- **英伟达 Tensor Core：**2017 年公司发布的 Volta 架构首次引入了张量核 Tensor Core 模块，用于执行融合乘法加法，支持 INT32 计算；2018 年公司发布的 Turing 架构对 Tensor Core 进行了升级，并增加了对 INT8、INT4、Binary(INT1)的计算能力；2020 年公司发布的 Ampere 架构对 Tensor Core 再次升级，增加了 TF32 和 BF16 两种数据格式的支持，也增加了对稀疏矩阵计算的支持。2022 年公司发布的 Hopper 架构对 Tensor Core 再次升级，增加了 TF8 数据格式的支持。

- **AMD Matrix Core：**2020 年英伟达推出张量核 Matrix Core，对标英伟达 Tensor Core，并用于 MI100 加速器（可支持 FP64、FP32 计算格式）；2021 年底，AMD 发布 MI250/250X 加速卡，基于 Matrix Core 的加持下，FP64/FP32 计算能力可提升一倍。

表 9：中国 GPU 厂商与海外 GPU 厂商产品参数对比

一、中国 GPU 厂商产品及参数

品牌	昆仑芯	壁仞科技	燧原科技	海光	寒武纪	华为海思
产品	R200	BR100	i20	深算一号	MLU370-X8	昇腾 910
发布日期	2021	2022	2021	2021	2022	2018
工艺	7nm	7nm	12nm	7nm	7nm	7nm
半精度（FP16）	128 TFLOPS	NA	128 TFLOPS	NA	96 TFLOPS	320 TFLOPS
单精度（FP32）	32 TFLOPS	256 TFLOPS	32 TFLOPS	NA	24 TFLOPS	NA
双精度（FP64）	NA	NA	NA	10.8 TFLOPS	NA	NA
INT8	256 TOPS	2048 TOPS	256 TOPS	NA	256 TOPS	640 TOPS
CUDA 兼容	NA	是	否	NA	否	NA

二、海外 GPU 厂商产品及参数

品牌	AMD	AMD	NVIDIA	NVIDIA	NVIDIA	NVIDIA	NVIDIA
产品	INSTINCT MI100	INSTINCT MI250	P100	V100 SXM2	T4	A100 80GB PCIe	H100 PCIe
发布日期	2020	2021	2016	2017	2018	2020	2022
工艺	7nm	6nm	16nm	12nm	12nm	7nm	4nm
半精度（FP16）	184.6 TFLOPS	362.1 TFLOPS	21.2 TFLOPS	125 TFLOPS	65 TFLOPS	312 TFLOPS*	NA
半精度（FP16 Tensor Core）	NA	NA	不支持	不支持	不支持	不支持	1600 TFLOPS*
单精度（FP32）	23.1 TFLOPS	45.3 TFLOPS	10.6 TFLOPS	15.7 TFLOPS	NA	19.5 TFLOPS	48 TFLOPS
单精度（FP 32 Tensor Float）	46.1 TFLOPS（AMD 为 Matrix Core）	90.5 TFLOPS（AMD 为 Matrix Core）	不支持	不支持	不支持	156 TFLOPS	800 TFLOPS
双精度（FP64）	11.5 TFLOPS	45.3 TFLOPS	5.3 TFLOPS	7.8 TFLOPS	8.1 TFLOPS	9.7 TFLOPS	24TFLOPS
双精度（FP 64 Tensor Core）	不支持	90.5 TFLOPS（AMD 为 Matrix Core）	不支持	不支持	不支持	19.5 TFLOPS	48 TFLOPS
INT8	184.6 TOPs	362.1 TOPs	NA	NA	130 TOPs	624 TOPs*	NA
INT8 （Tensor Core）	不支持	NA	不支持	不支持	不支持	不支持	3200 TOPs
CUDA 兼容	否	否	是	是	是	是	是

资料来源：各公司官网，中信证券研究部

本土 GPU 厂商：有望率先在 AI 领域实现落地，并逐步扩展至图形渲染、复杂科学计算等场景

市场机遇：基于上文对英伟达历史的复盘和分析，作为典型的通用芯片，产品技术、软件生态是 GPU 厂商不断做大做强的核心基础和支撑。同时在 GPU 实际落地应用中，需要将硬件、软件应用、游戏引擎、操作系统、OEM 等众多环节匹配到一起，才能更好地发挥性能作用。目前国产 GPU 厂商正处于起步阶段，市场需求、产业政策均有利于其发展&壮大：

1）国产 GPU 厂商开始切入相关客户产品中**：**英伟达最新公告背景下，倒逼国内相关客户开始使用国产 GPU 产品，在一定程度上能够帮助相关企业与客户建立密切联系，进而帮助相关企业进行快速的技术和产品迭代。

2）市场需求：依据我们前文预测，2030 年全球 GPU（图形渲染）市场规模为 568 亿美元，中国市场规模约 97 亿美元；2030 年全球数据中心 GPU（AI、高性能计算等）市场规模为 828 亿美元，中国市场规模约 250 亿美元。

面临挑战：目前国产 GPU 厂商大多仍处于早期发展阶段，仍需要在技术、产品商业化落地等方面不断努力：

1）核心技术人才招募。（a）从英伟达 GPU&AMD 的发展历史来看，公司 GPU 架构基本可以做到两年更新一代，这对于架构师对于芯片研发的理解和应用场景的全判断要求较高。如：Jim Keller 于 2012 年左右加入 AMD，帮助涉及了 Zen 微架构，大幅提升公司产品在数据中心领域的竞争力。（b）GPU 下游应用领域，并非是单纯的硬件算力比拼，对于软件开发及软件生态的建设亦相对重要。未来如何招聘大量的软件&AI 人才，仍是国产 GPU 厂商目前需要面临的重要问题。

2）产品设计、流片、客户验证，再到量产交付的全流程跑通。（a）GPU 是一种技术门槛极高的细分赛道领域，前期投入资金成本相对较高，这对于企业的融资能力要求相对较高。（b）从 GPU 的开发及使用流程来看，GPU 从最初设计到制造、流片、量产，周期通常不会低于 18－24 个月。从产品点亮到推出，再到后续的大量出货和用户验证，再到后续找到可持续落得的应用场景，仍面临着较多的挑战。

技术路线选择：AI 为中短期最可能突破&落地场景，并可逐步向图形渲染、复杂科学计算等领域扩展。目前 GPU 的应用场景，主要应用于图形渲染、AI 训练&推理、复杂科学计算等领域，结合市场规模、客户结构、技术特性等要素，对于本土 GPU 厂商而言，我们判断，AI 将是最可能率先获得突破的领域，并在此基础上，不断向图形渲染、复杂科学计算等领域进行延伸：

- **AI 训练**：大模型逐步成为 AI 领域的主流，叠加下游自然语言理解、计算机视觉、推荐系统等应用场景的不断扩展，AI 训练料将成为中期国内 GPU 最大的需求领域。同时 AI 模型更多基于神经网络结构，因此对计算精度要求并不严苛，亦使得本土 GPU 厂商面临的技术门槛相应降低，我们预计这将是本土 GPU 厂商最容易实现突破的领域。

- **AI 推理**：从英伟达&谷歌等科技巨头的产品参数来看，AI 推理环节对计算精度的要求显著低于 AI 训练环节，一般 4~8 位即可满足，但 AI 推理本身对实时性要求较高，且下游场景过于碎片化，如何实现灵活性、细分场景之间的有效平衡，是当前面临的主要难题，因此初创企业更多在自动驾驶、安防等领域寻找市场机遇。

- **图形渲染**：主要场景包括游戏画面渲染，以及专业图形创作&渲染等领域，作为典型的 2C 市场，客户更专注产品的性价比、品牌、生态支持等，且 GPU 图形管线设计复杂度相对更高。

- **复杂科学计算**：主要应用场景包括国防、航天、气象等高性能计算领域，为控制累计误差，需要较高的计算精度，一般需要 64 位双精度运算，整体技术架构复杂性远大于 AI 训练、推理环节。

表 10：各类别场景对 GPU 特性需求分析

场景	技术特性
图形渲染	图形管线设计相对复杂
AI 训练	计算精度要求不高，一般 8~32 位计算精度
AI 推理	计算精度要求最低，一般 4~16 位计算精度
复杂科学计算	计算精度要求最高，一般需要 64 位双精度

资料来源：中信证券研究部整理

风险因素

全球核心技术跨境流动受阻风险；地缘政治冲突不断加剧风险；本土企业在核心技术环节进展不及预期风险；企业核心技术人才流失、难于招募风险；下游应用场景落地不及预期风险；生产制造等芯片关键技术环节受阻风险等。

投资建议

当前国内本土 GPU 厂商正在快速崛起，大部分核心团队具有英伟达、AMD 工作经历，且企业目前已平均发布 1-3 款相关产品，并逐步从“可用”走向“好用”。参考英伟达发展历程，GPU 作为通用计算芯片，产品技术、软件生态等构成 GPU 厂商的核心壁垒，国内大部分本土 GPU 厂商当前仍处于早期阶段，短期仍需克服用户验证、产品落地等潜在挑战，但长期前景值得期待。我们判断本土厂商有望率先在 AI（训练、推理）领域实现突破，并可逐步向图形渲染、复杂科学计算等领域扩展。我们看好本土 GPU 厂商的长期投资机会，建议关注二级市场的头部企业以及一级市场的摩尔线程、沐曦集成电路、瀚博半导体等。

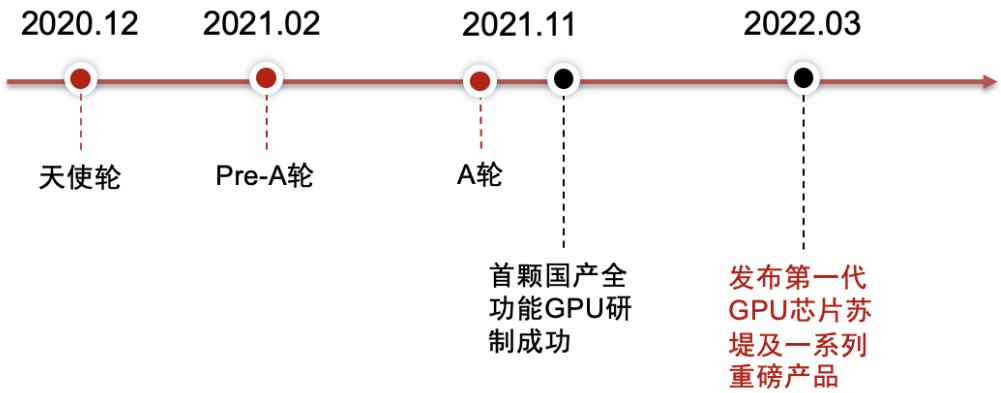
附录：国内部分重点 GPU 企业介绍

摩尔线程：专注于研发设计全功能 GPU 芯片及相关产品

英伟达背景出身，打造研运一体 GPU 公司。成立于 2020 年 10 月，致力于构建视觉计算及人工智能领域计算平台，研发全球领先的 GPU，建立高性能计算生态系统。摩尔线程拥有能够覆盖 GPU 研发设计、生产制造、市场销售、服务支持等完整成熟的团队，逐步成为国产现代全功能 GPU 实现的核心力量。创始人张建中是前英伟达全球副总裁，中国区总经理，英伟达中国公司创始人，曾任惠普、戴尔公司高管。

全功能 GPU 苏堤问世。公司成立不到 300 天的时间，于 2021 年 11 月公布首颗国产全功能 GPU 芯片研制成功，开创国产 GPU 研发速度先河。2022 年 3 月 30 日，公司推出基于其统一系统架构 MUSA 的首款 GPU 苏堤、基于苏堤的首款台式机显卡 MTT S60、首款数据中心级产品 MTT S2000，开拓 GPU 在中国市场的生态系统，助力驱动数字经济的发展。

图 34：摩尔线程及产品发展历程



资料来源：企查查，公司官网，中信证券研究部

表 11：摩尔线程产品参数

桌面级显卡 MTT S60（搭载苏堤芯片）	MUSA 核数量	2048 个
	单精度浮点算力	6TFLOPS
	像素填充率	192GPixels/s
	显存容量	8G
	超清显示	4K/8K
数据中心级产品 MTT S2000 性能（搭载苏堤芯片）	MUSA 核数量	4096 个
	单精度浮点算力	12TFLOPS
	显存容量	32G

资料来源：摩尔线程春季发布会，中信证券研究部

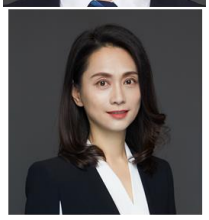
沐曦集成电路：国产高性能 GPU 芯片解决方案领先公司

公司概述：沐曦集成成立于 2020 年 9 月。公司专注于设计具有完全自主知识产权，针对异构计算等各类应用的高性能通用 GPU 芯片，致力于打造国内具有商用价值的 GPU 芯片，产品主要应用方向包括人工智能、云计算、数据中心等高性能异构计算领域。

公司创始人团队背景。公司汇聚顶尖技术、量产经验、管理能力人才，创始人陈维良曾任 AMD GPU 设计高级总监、AMD 全球 GPU SOC 设计总负责人、AMD 全球通用 GPUMI 产品线(高性能计算、云计算)设计总负责人。公司拥有国内最完整的 GPU 设计研发团队，参与过 AMD 从图像到高性能计算应用 GPU 的架构设计和量产，团队构建完整，且有多年合作共事基础。

图 35：沐曦集成电路创始团队背景

陈维良，创始人/董事长/CEO
清华大学微电子学研究所硕士，拥有超过20年的GPU芯片设计经验，曾担任世界顶尖GPU芯片公司高管，负责全球通用计算GPU产品线的整体设计与管理，曾在AMD有近20年工作经验。



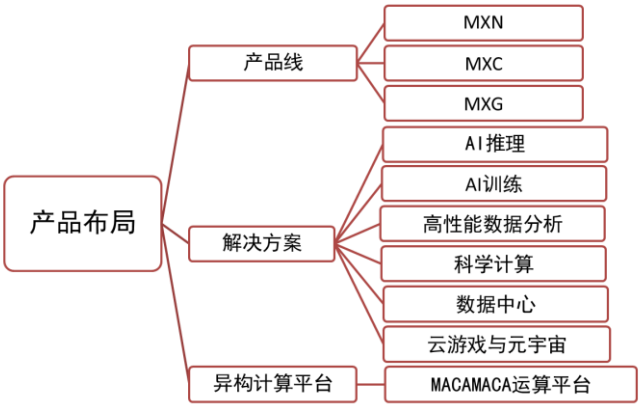
彭莉，CTO/首席硬件架构师
上海交通大学电子工程系硕士，AMD全球首位华人女科学家(Fellow)，拥有超过18年高性能GPU芯片设计经验，历任AMD首席SOC架构师、系统架构师、GFXIP架构师等职务，主导过多款GPU产品从架构到量产的全流程。

杨建，CTO/首席软件架构师
浙江大学博士，拥有21年大规模芯片及GPU软硬件架构设计经验，AMD大中华区第一位科学家(Fellow)，参与及主导数十余款GPU产品(55nm至7nm)量产全流程三维图形与高性能计算生态专家，丰富的软件及系统设计经验，拥有多项发明专利。

资料来源：公司官网，中信证券研究部

目前公司有两款产品，MXN 系列的 MXN 100 和 MXC 系列的 MXC 500。（1）MXN 系列是面向云端数据中心应用的人工智能推理产品，采用先进工艺结合高带宽内存，提供强大的 AI 算力和领先的视频编解码能力，可广泛应用于智慧城市、公有云计算、智能视频处理、云游戏等场景。目前的 MXN 100 是一款 7nm 芯片，于 2022 年 8 月已经流片，成功点亮。目前在正常测试软硬件，公司预计年底送达客户侧测试。（2）MXC 系列通用 GPU(GPGPU)芯片是针对 AI 训练和推理及科学计算的完美解决方案，沐曦自主知识产权架构提供强大高精度及多精度混合算力，可广泛应用于人工智能、数据中心以及科学计算、教育和科研等场景。MXC 500 是一款 6nm 芯片，公司计划 2022 年 10 月流片，2023 年上半年回片。

图 36：沐曦集成电路产品矩阵图



资料来源：公司官网，中信证券研究部

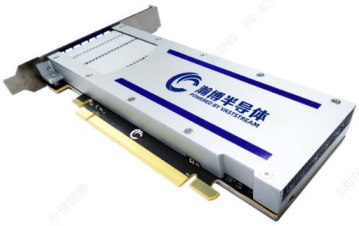
瀚博半导体：从 AI 与视频转向更广阔的通用计算市场

公司概览：专注于高性能通用加速芯片的 AI 与视频芯片厂商。公司成立于 2018 年 12 月，创始人钱军曾在思科、AMD 担任高管，具备 25 年以上的芯片设计经验。公司曾于 2020 和 2021 年间完成 A 轮、A+ 轮和 B 轮融资，总募资额超过人民币 24 亿元。其中最近一笔融资发生于 2021 年 12 月，由阿里巴巴集团、人保资本、经纬创投和五源资本联合领投，包含 B-1 和 B-2 轮，共计人民币 16 亿元。

产品布局：从加速卡向 GPU 迈进。目前公司拥有 VA1 通用 AI 推理加速卡与 SV100 系列芯片。VA1 加速卡具备高效的 AI 推理能力，INT8 峰值算力超 2000TOPS，并能够满足高密度视频的解码，支持 FP16 的浮点数运算。SV100 芯片则聚焦云端的推理，支持深度学习与计算机视觉等场景。根据公司在 2022 年世界人工智能大会的披露，公司发布了瀚博统一计算架构、全新数据中心（云端）AI 推理卡载天 VA10、边缘 AI 推理加速卡载天 VE1、以及瀚博软件平台 VastStream 扩展版等产品，并将继续整合统一计算架构，在边缘计算、云计算以及软件平台上持续进行投入，并预览了云端 GPU 芯片 SG100，正式进入到 GPU 市场。

商业化：签约多家政企客户，并与快手等互联网厂商建立合作。根据公司在 2022 世界人工智能大会的披露，2022 年以来，公司先后与福建大数据集团、国宁瑞能，高新兴、超聚变等行业领先企业，在智慧城市、智慧政务、智慧交通、智慧园区、智慧能源等多元场景，开展深入合作，为企业智能化、数据化提供国产 AI 算力解决方案。而公司依靠在视频领域的特色，亦与快手等互联网厂商建立合作关系。

图 37：公司 VA1 通用推理卡



资料来源：公司官网

图 38：公司 SV100 云端推理芯片

资料来源：公司官网

图 39：云端 GPU 芯片 SG100

资料来源：公司官网

表 12：公司两大产品主要能力

产品	能力	参数
VA1 通用推理卡	算力	INT8 峰值算力 200TOPS
	视频处理	H.264/AVS2 1080p
	通用性与拓展性	支持 FP16\BF16/INT8，满足主流神经网络部署
	其他	标准半高半长 75WPCIe Gen4 卡
SV100 云端推理芯片	算力	INT8 峰值算力 200TOPS
	场景	深度学习推理，支持计算机视觉、视频处理、自然语言处理、搜索推荐等
	视频处理	高密度视频解码

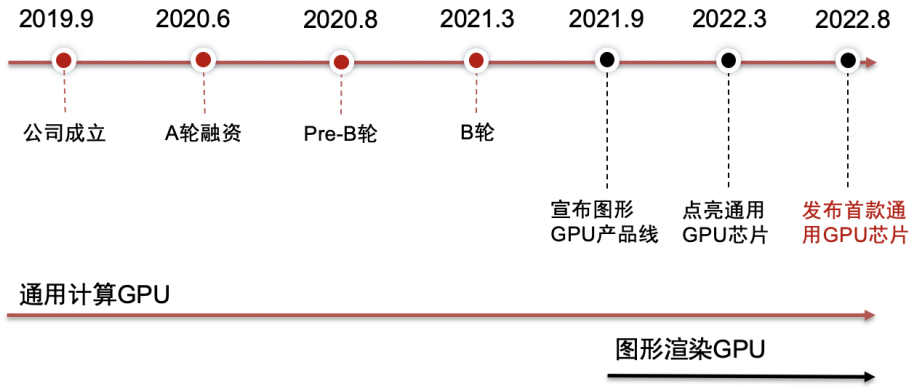
资料来源：公司官网，中信证券研究部

壁仞科技：专研通用计算体系，向图形渲染进发

公司概览：聚焦高性能算力芯片，专研通计算体系。GPU 壁仞科技创立于 2019 年，主要从事 GPU、DSA（专用加速器）的研发和销售，致力于开发原创通用计算体系，提供智能计算领域一体化解决方案。创始人张文曾任商汤科技总裁，具有哈佛大学法学博士及哥伦比亚工商管理硕士学位；联合创始人焦国方是图形 GPU 产品线总经理，具有超过 25 年的 GPU 产品架构及研发经验，曾任高通 GPU 团队负责人；联席 CEO 李新荣曾任 AMD 全球副总裁、中国研发中心总经理。

由通用计算向图形渲染全功能发力，补齐 GPU 全领域能力。1）公司聚焦云端通用智能芯片，并逐步扩展产品线至人工智能训练和推理、图形渲染等多个领域，实现 GPU 芯片的全功能全领域覆盖。**2）**目前公司产品线主要为 BR100 系列的通用 GPU，针对人工智能（AI）训练、推理，及科学计算等更广泛的通用计算场景开发，包含 BR100 与 BR104 两款产品。其中 BR100 产品形态为 OAM 模组，搭载一颗 BR100 GPU 芯片，制程为 7nm，在 FP32 精度下能够实现 256TFLOPS 的计算峰值。BR104 产品形态为 PCIe 板卡，搭载一颗 BR104 GPU 芯片，用于数据中心 GPU 服务器，采用 7nm 制程，FP32 精度下可达到 128TFLOPS 计算峰值。此外，公司提供 BIRENSUPA 软件开发平台，为旗下硬件提供完整功能架构的软件开发平台。后续看，公司将继续围绕通用计算芯片，进行硬件与软件的开发。

图 40：壁仞科技发展时间线



资料来源：壁仞科技官网，中信证券研究部

表 13：壁仞科技 AI 加速产品壁砺 100 参数

制程	FP32	BF16	INT8	内存容量	接口位宽	带宽	产品形态
7nm	256 TFLOPS	1024 TFLOPS	2048 TFLOPS	64GB HBM2E	4096 bit	1.64 TB/S	OAM 模组

资料来源：壁仞科技官网，中信证券研究部

商业化：GPU 芯片已经点亮，客户拓展进行时。2022 年 3 月，公司点亮了国内算力最大通用 GPU 芯片，2022 年 8 月发布首款通用 GPU 芯片，产品线逐步进入到落地阶段。在客户资源方面，根据公司在 2022 年世界人工智能大会上的披露，公司正在积极布局 BR100 商业化落地，目前已有平安科技、浪潮信息、万国数据等建立合作。

阿里平头哥：专注云与 AI 的芯片研发厂商

技术驱动产品创新，打造智联网芯片平台。平头哥半导体有限公司成立于 2018 年 9 月 19 日，是阿里巴巴集团的全资半导体芯片业务主体，由中天微和达摩院合并而来。平头哥拥有端云一体全栈产品系列，涵盖数据中心人工智能芯片、处理器 IP 授权等，实现芯片端到端设计链路全覆盖。平头哥坚持以技术驱动创新，以芯力量拥抱数智未来的研发理念，主要打造面向汽车、家电、工业等领域的智联网芯片平台。

AI 芯片：以 CPU 为主，兼顾部分 ASIC 芯片。平头哥目前产品分为四大类：1）玄铁系列的 CPU 芯片，此类芯片包含 8、9 以及无剑三大系列，基于 RISC-V 架构进行设计，由于 RISC-V 本身的架构特性，适用范围较广，既能用于智能监控、机器视觉、人工智能、5G、边缘服务器等对处理器性能要求很高的应用领域，又能用在对功耗和成本极其敏感的 IoT、MCU 等领域。2）倚天系列的服务器芯片，倚天 710 采用 2.5D 封装，分为两个 DIE，总计 600 亿晶体管。包含 128 个 Armv9 高性能 CPU 核，用于服务器。3）含光 AI 芯片，含光 800 基于 12nm 工艺， 集成 170 亿晶体管，性能峰值算力达 820 TOPS（INT 8），支持 Tensorflow、MXNet、Caffe、ONNX 等主流深度学习框架。4）羽阵 RFID 芯片，羽阵 600 是一颗低功耗、高性能超高频 RFID 电子标签芯片，用于智慧物流、智慧仓储、智慧零售、资产管理等应用场景。

图 41：阿里平头哥产品矩阵

资料来源：阿里平头哥官网，中信证券研究部

图 42：阿里平头哥 AI 芯片含光 800 架构及参数示意图

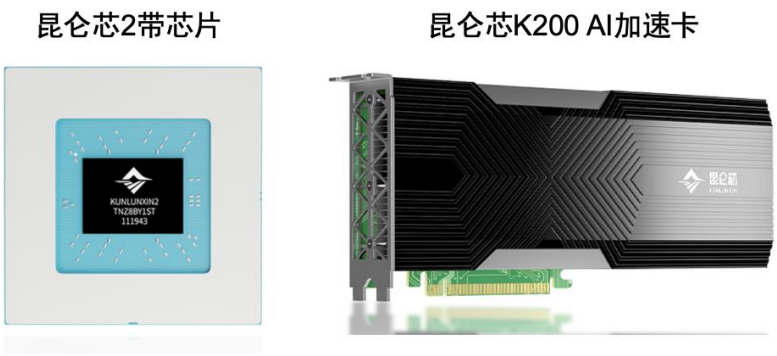
资料来源：阿里平头哥官网，中信证券研究部

商业化：服务阿里巴巴自身业务的同时，对外进行输出。作为阿里巴巴旗下的芯片平台，平头哥高性能产品直接用于阿里云相关产品，例如含光 800 已经广泛用于阿里云、阿里电商搜索与营销等领域。但在 AI 之外，平头哥 RISC-V 架构芯片由于适用范围大，亦广泛用于其他场景，根据阿里平头哥公开披露，截至 2020 年，玄铁系列 CPU 已经出货 20 亿颗，自研嵌入式 CPU IP 核授权客户超 100 家。根据纳思达在 2021 年 6 月公开披露，公司是阿里平头哥国产玄铁系列 CPU 的最大客户，基于玄铁系列 CPU 的芯片出货量累计已超过了 5 亿颗。2022 年阿里平头哥与国内 MCU 厂商爱普特达成合作。在服务阿里自身业务的同时，对外进行多维度拓展。

昆仑芯：产品聚焦 AI 加速芯片，自研 XPU 架构赋能智慧应用

专注 AI 加速，打造全链路服务体系。昆仑芯科技是一家 AI 芯片公司，于 2021 年 4 月正式从百度独立出来，当前已完成 130 亿人民币和 20 亿美元两轮融资。昆仑芯前身是百度智能芯片及架构部，于 2011 年 6 月设立，期间在实际业务场景中持续深耕 AI 加速领域，是一家在体系结构、芯片实现、软件系统和场景应用均有深厚积累的 AI 芯片企业。

图 43：昆仑芯产品示意图



资料来源：昆仑芯科技官网，中信证券研究部

自研 XPU 产品架构，赋能智慧应用场景。昆仑芯科技研发实力雄厚，CEO 欧阳剑是原百度首席架构师（T11），智能芯片业务总经理，基础技术体系联席技术委员会主席，百度无人驾驶初始团队成员。团队成员拥有全球顶尖学术背景，多数成员来自百度、高通、Marvell、Tesla 等行业头部公司，并提出了 100%自研的、面向通用人工智能计算的核心架构 XPU。目前，昆仑芯科技已与智能产业的上下游企业建立了良好的合作生态，通过向不同行业提供以人工智能芯片为基础的算力产品，辐射互联网、智慧城市、智算中心、智慧工业、智慧应急、智慧交通、智慧金融等“智慧+”产业。

表 14：昆仑芯产品简介

产品系列	主要产品	2022E	2023E	2024E
K 系列	昆仑芯 1 代芯片	256 TOPS@INT8，64 TFLOPS@FP16，14nm	云数据中心 智慧城市	百度搜索 微亿智造

产品系列	主要产品	2022E	2023E	2024E
	昆仑芯 AI 加速卡 K200	256 TOPS@INT8	智慧工业 智算中心 智能交通 生物计算	江苏银行 宜昌市点军区政府 百度智能云 小度科技 重庆市高级人民法院
	昆仑芯 AI 加速卡 K100	128 TOPS@INT8 算力		
R 系列	昆仑芯 2 代芯片	已量产，XPU-R 架构 256 TOPS@INT8，128 TFLOPS@FP16，7nm		
	昆仑芯 AI 加速卡 R200	256 TOPS@INT8，128 TFLOPS@FP16		
	昆仑芯 R480-X8AI	1 Peta FLOPS@FP16 算力和 256G 显存		

资料来源：昆仑芯科技官网，中信证券研究部

分析师声明

主要负责撰写本研究报告全部或部分内容的分析师在此声明：（i）本研究报告所表述的任何观点均精准地反映了上述每位分析师个人对标的证券和发行人的看法；（ii）该分析师所得报酬的任何组成部分无论是在过去、现在及将来均不会直接或间接地与研究报告所表述的具体建议或观点相联系。

一般性声明

本研究报告由中信证券股份有限公司或其附属机构制作。中信证券股份有限公司及其全球的附属机构、分支机构及联营机构（仅就本研究报告免责条款而言，不含 CLSA group of companies），统称为“中信证券”。

本研究报告对于收件人而言属高度机密，只有收件人才能使用。本研究报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。本研究报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。中信证券并不因收件人收到本报告而视其为中信证券的客户。本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断并自行承担投资风险。

本报告所载资料的来源被认为是可靠的，但中信证券不保证其准确性或完整性。中信证券并不对使用本报告或其所包含的内容产生的任何直接或间接损失或与此有关的其他损失承担任何责任。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可跌可升。过往的业绩并不能代表未来的表现。

本报告所载的资料、观点及预测均反映了中信证券在最初发布该报告日期当日分析师的判断，可以在不发出通知的情况下做出更改，亦可因使用不同假设和标准、采用不同观点和分析方法而与中信证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。中信证券并不承担提示本报告的收件人注意该等材料的责任。中信证券通过信息隔离墙控制中信证券内部一个或多个领域的信息向中信证券其他领域、单位、集团及其他附属机构的流动。负责撰写本报告的分析师的薪酬由研究部门管理层和中信证券高级管理层全权决定。分析师的薪酬不是基于中信证券投资银行收入而定，但是，分析师的薪酬可能与投行整体收入有关，其中包括投资银行、销售与交易业务。

若中信证券以外的金融机构发送本报告，则由该金融机构为此发送行为承担全部责任。该机构的客户应联系该机构以交易本报告中提及的证券或要求获悉更详细信息。本报告不构成中信证券向发送本报告金融机构之客户提供的投资建议，中信证券以及中信证券的各个高级职员、董事和员工亦不为（前述金融机构之客户）因使用本报告或报告载明的内容产生的直接或间接损失承担任何责任。

评级说明

投资建议的评级标准		评级	说明
报告中投资建议所涉及的评级分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后 6 到 12 个月内的相对市场表现，也即：以报告发布日后的 6 到 12 个月内的公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。其中：A 股市场以沪深 300 指数为基准，新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以摩根士丹利中国指数为基准；美国市场以纳斯达克综合指数或标普 500 指数为基准；韩国市场以科斯达克指数或韩国综合股价指数为基准。	股票评级	买入	相对同期相关证券市场代表性指数涨幅 20%以上
		增持	相对同期相关证券市场代表性指数涨幅介于 5%～20%之间
		持有	相对同期相关证券市场代表性指数涨幅介于-10%～5%之间
		卖出	相对同期相关证券市场代表性指数跌幅 10%以上
	行业评级	强于大市	相对同期相关证券市场代表性指数涨幅 10%以上
		中性	相对同期相关证券市场代表性指数涨幅介于-10%～10%之间
		弱于大市	相对同期相关证券市场代表性指数跌幅 10%以上

特别声明

在法律许可的情况下，中信证券可能（1）与本研究报告所提到的公司建立或保持顾问、投资银行或证券服务关系，（2）参与或投资本报告所提到的公司的金融交易，及/或持有其证券或其衍生品或进行证券或其衍生品交易。本研究报告涉及具体公司的披露信息，请访问 https://research.citicsinfo.com/disclosure。

法律主体声明

本研究报告在中华人民共和国（香港、澳门、台湾除外）由中信证券股份有限公司（受中国证券监督管理委员会监管，经营证券业务许可证编号：Z20374000）分发。本研究报告由下列机构代表中信证券在相应地区分发：在中国香港由 CLSA Limited（于中国香港注册成立的有限公司）分发；在中国台湾由 CL Securities Taiwan Co., Ltd.分发；在澳大利亚由 CLSA Australia Pty Ltd.（商业编号：53 139 992 331/金融服务牌照编号：350159）分发；在美国由 CLSA（CLSA Americas, LLC 除外）分发；在新加坡由 CLSA Singapore Pte Ltd.（公司注册编号：198703750W）分发；在欧洲经济区由 CLSA Europe BV 分发；在英国由 CLSA （UK）分发；在印度由 CLSA India Private Limited 分发（地址：8/F, Dalamal House, Nariman Point, Mumbai 400021；电话：+91-22-66505050；传真：+91-22-22840271；公司识别号：U67120MH1994PLC083118）；在印度尼西亚由 PT CLSA Sekuritas Indonesia 分发；在日本由 CLSA Securities Japan Co., Ltd.分发；在韩国由 CLSA Securities Korea Ltd.分发；在马来西亚由 CLSA Securities Malaysia Sdn Bhd 分发；在菲律宾由 CLSA Philippines Inc.（菲律宾证券交易所及证券投资者保护基金会员）分发；在泰国由 CLSA Securities (Thailand) Limited 分发。

针对不同司法管辖区的声明

中国大陆：根据中国证券监督管理委员会核发的经营证券业务许可，中信证券股份有限公司的经营范围包括证券投资咨询业务。

中国香港：本研究报告由 CLSA Limited 分发。 本研究报告在香港仅分发给专业投资者（《证券及期货条例》（香港法例第 571 章）及其下颁布的任何规则界定的），不得分发给零售投资者。就分析或报告引起的或与分析或报告有关的任何事宜，CLSA 客户应联系 CLSA Limited 的罗鼎，电话：+852 2600 7233。

美国：本研究报告由中信证券制作。本研究报告在美国由 CLSA（CLSA Americas, LLC 除外）仅向符合美国《1934 年证券交易法》下 15a-6 规则界定且 CLSA Americas, LLC 提供服务的“主要美国机构投资者”分发。对身在美国的任何人士发送本研究报告将不被视为对本报告中所评论的证券进行交易的建议或对本报告中所述任何观点的背书。任何从中信证券与 CLSA 获得本研究报告的接收者如果希望在美国交易本报告中提及的任何证券应当联系 CLSA Americas, LLC（在美国证券交易委员会注册的经纪交易商），以及 CLSA 的附属公司。

新加坡：本研究报告在新加坡由 CLSA Singapore Pte Ltd.，仅向（新加坡《财务顾问规例》界定的）“机构投资者、认可投资者及专业投资者”分发。就分析或报告引起的或与分析或报告有关的任何事宜，新加坡的报告收件人应联系 CLSA Singapore Pte Ltd，地址：80 Raffles Place, #18-01, UOB Plaza 1, Singapore 048624，电话：+65 6416 7888。因您作为机构投资者、认可投资者或专业投资者的身份，就 CLSA Singapore Pte Ltd.可能向您提供的任何财务顾问服务，CLSA Singapore Pte Ltd 豁免遵守《财务顾问法》（第 110 章）、《财务顾问规例》以及其下的相关通知和指引（CLSA 业务条款的新加坡附件中证券交易服务 C 部分所披露）的某些要求。MCI（P）085/11/2021。

加拿大：本研究报告由中信证券制作。对身在加拿大的任何人士发送本研究报告将不被视为对本报告中所评论的证券进行交易的建议或对本报告中所载任何观点的背书。

英国：本研究报告归属于营销文件，其不是按照旨在提升研究报告独立性的法律要件而撰写，亦不受任何禁止在投资研究报告发布前进行交易的限制。本研究报告在英国由 CLSA （UK）分发，且针对由相应本地监管规定所界定的在投资方面具有专业经验的人士。涉及到的任何投资活动仅针对此类人士。若您不具备投资的专业经验，请勿依赖本研究报告。

欧洲经济区：本研究报告由荷兰金融市场管理局授权并管理的 CLSA Europe BV 分发。

澳大利亚：CLSA Australia Pty Ltd（“CAPL”）(商业编号：53 139 992 331/金融服务牌照编号：350159) 受澳大利亚证券与投资委员会监管，且为澳大利亚证券交易所及 CHI-X 的市场参与主体。本研究报告在澳大利亚由 CAPL 仅向“批发客户”发布及分发。本研究报告未考虑收件人的具体投资目标、财务状况或特定需求。未经 CAPL 事先书面同意，本研究报告的收件人不得将其分发给任何第三方。本段所称的“批发客户”适用于《公司法（2001）》第 761G 条的规定。CAPL 研究覆盖范围包括研究部门管理层不时认为与投资者相关的 ASX All Ordinaries 指数成分股、离岸市场上市证券、未上市发行人及投资产品。CAPL 寻求覆盖各个行业中与其国内及国际投资者相关的公司。

印度：CLSA India Private Limited，成立于 1994 年 11 月，为全球机构投资者、养老基金和企业提供股票经纪服务（印度证券交易委员会注册编号：INZ000001735）、研究服务（印度证券交易委员会注册编号：INH000001113）和商人银行服务（印度证券交易委员会注册编号：INM000010619）。CLSA 及其关联方可能持有标的公司的债务。此外，CLSA 及其关联方在过去 12 个月内可能已从标的公司收取了非投资银行服务和/或非证券相关服务的报酬。如需了解 CLSA India“关联方”的更多详情，请联系 Compliance-India@clsa.com。

单击此处输入文字。