**ASIC芯片市场迎来新变局**

来源：中国电子报、电子信息产业网 作者：杨鹏岳

全球AI算力市场正在将目光聚焦到定制化的ASIC芯片领域，并开始上演一场没有硝烟的“车轮战”：OpenAI正在初步测试部分谷歌的张量处理器（TPU）；英伟达正式发布NVLink Fusion，与由一众科技巨头组成的UALink联盟展开正面交锋……随着定制化算力竞赛进入深水区，ASIC具备的优势逐渐清晰，并有望在明年迎来超越GPU的临界点。



今年4月，谷歌推出第一款专为推理而设计的定制AI加速器

**“鲶鱼”入局定制化算力竞赛**

在高性能计算芯片从通用向专用的发展趋势下，ASIC芯片凭借其针对特定应用优化的特性，在多个领域崭露头角。其中，人工智能是ASIC芯片的重要应用方向之一，ASIC芯片可针对深度学习算法进行底层优化，在AI模型训练和推理任务中，以高效的计算能力和低能耗表现，支撑起自然语言处理等应用。

近日有消息爆出，OpenAI已开始租用谷歌的TPU 芯片为其ChatGPT及其他AI 产品提供算力支持。对此，摩根士丹利分析指出，对谷歌而言，这是OpenAI对其AI基础设施能力的重要背书，将推动谷歌云业务增长并巩固其在ASIC生态系统中的领先地位。

就在今年4月，谷歌在其一年一度的谷歌云大会上重磅发布第七代TPU芯片——Ironwood。值得关注的是，它是谷歌迄今为止性能最强、可扩展性最高的定制AI加速器，也是首款专为推理设计的加速器，直接叫板英伟达Blackwell B200。据谷歌云员工透露，谷歌虽然向其竞争对手开放TPU芯片，但还是会将更强大的TPU保留给自己的AI团队开发，供自家的Gemini模型使用。

TPU是ASIC芯片中的一个典型架构，而OpenAI将目光转向TPU，被视为AI基础设施市场转向的一个关键节点。作为英伟达GPU的最大采购商之一，OpenAI若是今后大规模采购TPU，将直接削弱英伟达GPU的优势地位。

“随着芯片产业链的不断成熟，相较于通用芯片或者FPGA芯片，ASIC芯片在一些业务逻辑确定且需求量较大的场景下可以提供更高的性能、更低的功耗和更具备竞争力的价格。”中昊芯英解决方案架构师顾立程向《中国电子报》记者表示，不同领域对芯片的性能、功耗、尺寸等要求各异，促使 ASIC 芯片市场呈现多样化需求，市场规模也随之不断扩大。

事实上，面对逐渐崛起的ASIC生态，吃遍GPU红利的英伟达已经急了。5月19日，英伟达正式发布并公开提及NVLink Fusion技术，这是一种半定制 AI 基础设施解决方案，其核心在于将英伟达的高速互连技术NVLink与第三方ASIC 、CPU等异构芯片深度融合。据悉，首批采用 NVLink Fusion的厂商不乏联发科、Marvell、新思科技和Cadence等半导体企业。黄仁勋表示，NVLink Fusion将NVIDIA AI平台和丰富的生态系统对外开放，助力合作伙伴构建专用AI基础设施。

NVLink Fusion的推出是英伟达与UALink联盟展开的正面竞争。面对日益增长的AI算力需求和复杂的业务场景挑战，显然单靠英伟达并不现实，越来越多的云服务科技巨头纷纷朝着自研ASIC的定制化方向布局。而UALink联盟作为AI服务器芯片互连组织，旨在建立开放互连生态，其成员来自云计算、半导体、处理器IP、软件公司、OEM等最顶尖厂商，囊括了亚马逊AWS、阿里巴巴、AMD、苹果、谷歌、Meta、微软等。

在这场战况愈加激烈的AI芯片群雄竞赛中，ASIC生态逐步完善，倒逼已构建GPU护城河的英伟达以技术开放之名，行生态扩张之实。随着入场ASIC的“鲶鱼”增多，未来AI算力市场的权力格局将如何重塑？

**AI算力市场迎来新的临界点**

过去几年，AI领域的热点被训练大模型占据，但随着AI推理模型大量涌现，AI Agent（智能体）在营销、客服、运维等场景的落地进一步放大推理性能需求，推理已成为AI经济的关键驱动力。

在AI算力重心由训练端转向推理端之际，ASIC芯片的高度定制化和能效优势趋势得以逐渐清晰。高盛在其近期发布的报告中引入ASIC AI服务器作为新类别，认为随着AI推理需求上升，具备定制化优势与预算友好特征的ASIC服务器将在2025至2026年全球服务器市场中占比达38%-40%。

在此背景下，定制化算力竞赛进入深水区，以谷歌、Meta、微软、AWS等为代表的全球云服务厂商争相推进自研ASIC的布局。除了已迭代七代TPU芯片的谷歌，Meta 通过与博通合作，最早将于今年第四季度推出其首款 AI ASIC 芯片 MTIA T-V1，规格或将超过英伟达下一代 AI 芯片“Rubin”。AWS已启动不同版本的Trainium v3开发，预计于2026年陆续量产。当然，云巨头的自研ASIC之路并非一帆风顺，微软于近日被爆出在自研AI芯片方面遇阻，原本计划今年量产的AI芯片Braga或被推迟到2026年投产。

尽管如此，ASIC的前进势头依旧不减。从ASIC芯片设计大厂博通和Marvell的最新财报来看，博通2025年二季度AI相关业务收入达44亿美元，同比增长46%。博通表示未来三年AI芯片市场规模有望达到600亿~900亿美元。Marvell在2026财年一季度，数据中心业务收入达14.4亿美元，占总收入的76%，其中AI带动下的定制芯片业务成为核心引擎。Marvell公司CEO Matt Murphy强调，AI定制芯片是公司未来的关键战略方向，并指出其正处于构建AI基础设施转型的核心位置。

在ASIC芯片大厂、云巨头等助推下，AI算力市场正在迎来新的临界点。根据野村证券的最新报告，目前英伟达GPU占 AI 服务器市场 80% 以上，ASIC 仅占 8%-11%。2025 年，预计谷歌和亚马逊 AWS两家的ASIC芯片出货量合计约为英伟达 GPU 出货量（500 万至 600 万）的 40%-60%。到 2026 年，随着 Meta与微软大规模部署，ASIC 出货量有望超越英伟达 GPU。届时，从投资驱动转变为应用拉动，属于ASIC的时代将正式到来。

在这个临界点到来之前，ASIC芯片依旧是机遇与挑战并存。由于ASIC芯片从设计、制造到量产的流程复杂，周期漫长，且一次性投入巨大，相关企业会面临研发周期与成本挑战，还会遇到竞争愈加激烈和技术追赶的现实压力。

从应用场景来看，ASIC行业正从云端向边缘端深度渗透。而放眼全球市场，ASIC领域呈现出“北美主导、中国加速、新兴市场爆发”的格局。

“国际巨头通过技术壁垒与生态闭环垄断市场，同时跨界竞争加剧，科技巨头与初创企业纷纷涉足该领域。”顾立程表示，当下北美AI和半导体圈正在发生AI算力硬件的转向，AI芯片市场也进入了更具竞争性的新阶段。相比之下，国内企业在技术研发水平、高端人才储备、产业链协同等方面仍存在差距。国内企业可以借鉴北美企业在技术创新、市场拓展等方面的经验，加快技术迭代速度，提高产品竞争力，在全球ASIC 芯片市场中占据更有利的地位。