

# 扩张与洗牌并存 国产GPU厂商发力AI生态



证券时报记者 王一鸣

前几年广受资本追捧的GPU(图形处理器)这一硬核算赛道正在步入洗牌期,一些原本备受瞩目的通用GPU设计企业先后陷入经营困境。与此同时,有头部企业继续求新求变,已步入AI(人工智能)核心战场。

对于行业现状,北京社科院研究员王鹏向证券时报记者分析,当前国产GPU产业处于快速发展与洗牌并存的阶段。一方面,随着AI大模型的需求激增,国产GPU市场迅速崛起,涌现出如摩尔线程、壁仞科技等一批具有创新能力的企业。这些企业在技术研发、产品创新等方面取得了显著进展,为国产GPU行业注入了新的活力。另一方面,市场竞争也日益激烈,部分初创企业由于资金链断裂、技术瓶颈等问题陷入困境,甚至面临裁员、诉讼等风险。

“长远来看,国产GPU行业仍具有广阔的发展前景。这源于人工智能、大数据、云计算等技术的快速发展,对高性能GPU的需求持续增长;同时,国家政策对半导体产业的支持力度不断加大,为国产GPU行业提供了良好的发展环境。此外,随着国产芯片公司在生态构建、技术创新等方面的不断努力,国产GPU有望逐步缩小与国际领先水平的差距。”王鹏说。

## 几家欢喜几家愁

近日,国产GPU独角兽壁仞科技一则动态引起业内关注。该公司在上海证监局办理IPO辅导备案登记,辅导券商为国泰君安。这是继今年8月26日壁仞科技IPO辅导备案报告公布后,又一家启动IPO的上海AI芯片企业。

壁仞科技创立于2019年,致力于研发原创性的通用计算体系,建立高效的软硬件平台,同时在智能计算领域提供一体化的解决方案。截至目前,公司已完成多轮融资,公开融资总额超过50亿元,投资方包括启明创投、IDG资本、中芯聚源、格力创投等。目前,壁仞科技首款国产高端通用GPU(GPGPU)壁仞系列已量产。在9月举行的2024全球AI芯片峰会上,壁仞科技还首次公布了其自研的异构GPU协同训练方案,该方案突破了大模型异构算力孤岛难题,实现了在本地异构多GPU芯片算力训练技术领域的首次突破。

另一家国内GPU独角兽企业也有新动作。今年7月,摩尔线程宣布对其夸娥(KUAE)智算集群解决方案进行重大升级,从当前千卡级别大幅扩展至万卡规模。摩尔线程创始人兼CEO张建中表示,在多元趋势下,万卡已是AI模型训练主战场的标配。事实上,在国内,万卡集群建设正高歌猛进。去年7月,华为昇腾AI集群规模从最初的4000卡集群扩展至16000卡,是国内首个万卡AI集群。不过,当前国内GPU产业的分化亦十分明显,一些原本备受瞩目的企业接连陷入困境。今年8月底,曾获得2024年度重庆市“独角兽企业”称

号的象帝先被曝约400员工被裁。公司在回应中坦陈:“发展面临一定的市场调整压力……正在进行组织结构和人员配置的优化等……并寻找外部融资机会。”另一方面,象帝先股东还在起诉其创始人唐志敏,因其未能筹集到承诺的5亿元B轮融资。

在象帝先陷入困境之前,同处在GPU赛道的砺算科技也曾预警。于今年8月19日,东芯股份公告拟以不超过2亿元收购砺算科技37.88%股份,砺算科技的资金紧张局面方得以缓解。

## GPU占AI芯片八成份额

在中关村物联网产业联盟副秘书长袁帅看来,当前国产GPU行业处于一个复杂而关键的转型期,短期行业出现降温分化,主要是由于市场竞争加剧、技术门槛提高、资本退潮等因素导致企业资金压力增大。在此背景下,部分实力较弱的初创企业在面对挑战时往往更加脆弱,而头部企业则凭借其技术实力、市场布局 and 资金优势逐渐崭露头角。

按定义和应用看,GPU起初是一种显示芯片,用于加速计算机上的图像处理。随着技术进步和应用范围的扩大,GPU衍生出了GPGPU,后者是通过利用GPU并行处理能力特别强、计算能效比高的特点,处理更为复杂的计算任务(如AI模型训练与推理等)。基于此,如今的GPU已被广泛应用于人工智能、虚拟现实、游戏、科学计算等诸多领域。

IDC数据显示,预计到2025年,GPU仍将占据全球AI芯片八成市场份额。Gartner此前预测,2023年全球人工智能芯片(包含GPU、TPU等)的市场规模达到530亿美元,并预计在2024年增长至671亿美元。到2027年,全球人工智能芯片市场规模预计达到1194亿美元。

从行业格局来看,作为GPU的发明者,英伟达占据绝对领先地位。Jon Peddie Research(JPR)数据显示,2024年第一季度英伟达的全球AIB显卡(也称独立GPU)市场份额飙升至88%,而AMD的市场份额下降至12%。JPR预计,2024年第二季度,AI训练器GPU的出货量将继续增加,而游戏AIB市场可能持平或走低。在AI芯片主战场,据TechInsights数据,英伟达2023年的数据中心GPU出货量约为376万片,比2022年多出近100万片,市场占比98%,收入达362亿美元。压倒性击败了AMD和英特尔等竞争对手。

微软公司和Meta等数据中心运营商使用英伟达的芯片来开发和运行AI模型,巨大的需求使其销售额和股价飙升。英伟达市值在今年6月曾登顶全球第一,达到3.33万亿美元。最新财报显示,在英伟达Hopper(GPU构架平台)、GPU计算和网络平台强劲需求的推动下,公司数据中心今年第二季度收入达到262.72亿美元。

英伟达达成如今成绩单的是什

么?袁帅告诉证券时报记者,这主要源于英伟达在GPU设计、制造工艺以及软件优化等方面具有深厚的技术积累,其产品在性能、功耗和稳定性等方面均处于行业领先地位。同时,该公司建立了完善的生态系统,包括CUDA平台、软件开发工具包(SDK)以及丰富的第三方应用支持等,为开发者提供了便捷的开发环境和广泛的应用场景。

“英伟达的芯片产品先进且功能强大,配套软件生态完善渗透率高,并且产品迭代进化速度快速,所以市场地位尚难以撼动。”深度科技研究院院长张孝荣认为。

## 机遇和挑战并存

国内市场方面,自2019年以来,一批国产GPU初创公司先后成立,其中涌现了壁仞科技、摩尔线程、燧原科技等AI芯片设计独角兽。同时,还有一批老牌芯片设计公司也入局GPU领域。

这些公司的机会在于美国出口管制背景下的国产替代。2022年10月,美国政府禁止英伟达向中国出口高性能H100、A100芯片。英伟达随后专门面向中国市场推出了不违反美国出口管制要求的A800和H800芯片,至去年10月,这两项芯片也被限制。2024年2月,英伟达向中国特供的H20芯片接受预订,但该芯片与H100配置相比,GPU核心数量减少41%,性能降低28%。对此,英伟达2023年11月曾透露,数据中心业务两成以上受出口管制影响。

与此同时,国内GPU厂商正迎头赶上。IDC在2023年中发布的一项中国AI加速卡(公开市场)出货量统计显示,2022年上半年到2023年上半年,中国AI加速卡出货约109万张,英伟达市场份额为85%,华为市占率为10%,百度市占率为2%,寒武纪和燧原科技均为1%。

从国内产品成果来看,不仅华为昇腾系列表现突出。2024年5月份,摩尔线程与智谱AI开展了新一轮GPU大模型适配、性能测试。结果表明,在推理方面,摩尔线程自研全功能GPU显著优于基准对比产品RTX 3090和RTX 4090;在训练方面,摩尔线程夸娥千卡智算集群的训练精度与A100集群相比,误差在1%以内,而集群性能扩展系数超过90%,模型算力利用率(MFU)与A100集群相当。

壁仞科技则于2022年8月推出采用7nm制程的GPGPU芯片BR100,该芯片峰值算力达到国际厂商彼时在售旗舰产品3倍以上,创下国内互连带宽纪录。

不过,挑战也随之而来,一些行业领先企业先后被美国政府“盯上”。2022年12月,寒武纪等36家被美国政府列入“实体清单”。2023年10月,壁仞科技、摩尔线程等亦被添加到实体

清单中。

这对相关企业影响几何?寒武纪在2024年半年报中表示,公司及部分子公司已被列入“实体清单”,将对公司供应链的稳定造成一定风险。受“实体清单”等供应链不利因素影响,公司上半年营收6476.53万元,较上年同期下降43.42%。

## 生态建设待完善

王鹏指出,当前国际政治经济环境的变化对GPU行业产生了一定影响。特别是美国对中国AI产业的打压政策限制了国内相关企业的进口渠道和供应链稳定性。

袁帅向证券时报记者表示,在面临外部封锁前提下,国产GPU厂商要站稳脚跟并在高端AI芯片领域争得一席之地并非易事。在硬件层面,国内芯片代工企业在高端工艺和产能方面相对不足,难以完全满足国产GPU设计厂商对高端AI芯片的生产需求;在软件层面,英伟达等国际巨头在软件生态(CUDA等)和IP授权方面拥有强大优势。国产GPU厂商在构建和完善软件生态以及获取关键IP授权领域仍显薄弱。

事实上,CUDA被业内普遍认为是英伟达最大护城河。英伟达CEO黄仁勋在今年也曾强调CUDA软件系统生态对于AI时代的重要性。“目前,CUDA已实现良性的发展循环,全球拥有500万开发者,覆盖医疗保健、金融服务、计算机行业、汽车行业等领域。”黄仁勋说。

中国工程院院士、高性能计算研究所所长郑纬民在今年7月7日的信息百人会上指出:“这些年国产芯片无论是软件硬件都有很大的进展,但是用户不太喜欢用,原因是国产卡的生态系统不太好。”

郑纬民认为,这需要做好系统设计和相关软件优化,具体包括十个方面:编程框架、并行加速、通信库、算子库、AI编译器、编程语音、调度器、内存分配系统、容错系统、存储系统等。在他看来,在国产算力支撑大模型训练时,国产AI芯片只能达到国外芯片60%的性能,但如果把前述十个方面的软件生态做好了,客户也会满意。

国泰君安在今年3月发布的《重视AI芯片配套的软件生态》报告中总结:谁能做出中国版“CUDA”,就能占领AI算力高地。国内各大厂商已在GPU生态上积极布局,包括驱动程序、开发工具、应用程序接口(API)等,以确保其产品能够更好地服务于人工智能、大数据处理、云计算和游戏等多样化的应用场景。例如,华为推出了CANN以及对应的生态体系,力求突破技术瓶颈。此外,像景嘉微、摩尔线程等公司也在自主研发GPU产品,并努力打造与之兼容的软件环境,以减少对外部供应商的依赖,提升国产GPU的市场竞争力。

# 三连板卓朗科技提示风险 称不存在华为折叠屏相关收入

证券时报记者 韩忠楠

华为折叠屏手机上市后,相关概念股备受市场关注,多只个股股价持续飘红。9月23日晚间,已经三连板的卓朗科技(600225)披露股票交易风险提示公告,称经公司核实自查,公司不存在网络安全及华为折叠屏相关收入。

目前,卓朗科技的主营业务为软件和信息技术服务业务,其主要产品及服务分为软件开发及云服务业务、信息系统集成、互联网数据中心建设与托管服务三大类。

据悉,卓朗科技旗下公司的卓朗昆仑云曾在2019年受邀成为华为鲲鹏计算产业联盟首批生态合作伙伴,并与华为鲲鹏完成了产品兼容性互认证,推出了融合双方技术优势的云政务解决方案。

公开信息显示,2024年上半年,卓朗科技信息系统集成收入占营业收入的比例为83.03%,互联网数据中心建设与托管收入占营业收入的比例为12.01%,软件开发及云服务收入占营业收入的比例为0.18%,其他业务占营业收入的比例为4.78%。

值得关注的是,今年3月,卓朗科技收到证监会《立案告知书》,因涉嫌信息披露违法违规,证监会决定对卓朗科技进行立案。截至记者发稿前,证监会的调查尚在进行中,卓朗科技尚未收到就上述立案调查事项的结论性意见或决定。

卓朗科技称,若后续经证监会行政处罚认定的事实,触及《上海证券交易所股票上市规则》规定的重大违法强制退市情形,公司股票将被实施重大违法强制退市。

今年8月28日,卓朗科技发布公告称,公司及相关人员收到了天津证监局出具的警示函。天津证监局在警示函中指出,卓朗科技2023年年度业绩预告与2023年年度报告中披露的相关数据存在较大差异,信息披露不准确。

卓朗科技在更正公告中解释数据存在差异的原因:由于部分客户在产品交付验收后提出新的整改和验收要求,公司决定履行整改和调试工作并重新验收,导致相应收入及成本被冲回。同时,部分软件客户提出退货或变更要求,因此公司冲回相关已确认收入。

# 稳健医疗1.2亿美元海外并购 布局全球发展战略

证券时报记者 张一帆

稳健医疗(300888)计划通过收购加速其全球化战略。公司9月23日晚间公告,以现金形式收购美国Global Resources International,Inc.(简称“GRI公司”)75.2%股权,对应交易对价约为1.2亿美元,交易不构成重大资产重组。截至目前,股权交割已经完成。

据介绍,GRI公司是一家总部位于美国佐治亚州的全球性医疗耗材和工业防护企业,主要产品包括手术包、洞巾、铺单、容器、手术衣、工业防护服。其生产基地遍布全球,包括中国、美国、越南、多米尼加等国家,总生产面积超8万平方米,并在美国与欧洲拥有销售与物流布局,总仓储面积超7万平方米。值得注意的是,GRI公司的管理团队中有多名中美复合背景的成员。

财务报表显示,截至2023年底,GRI公司资产总额1.64亿美元,净资产6559.14万美元。2023年度整体营收1.52亿美元,净利润893.76万

美元。整体来看,GRI公司的收入主要来自欧美市场,其收入按地区分类美国约占70%,欧洲20%,亚洲10%。按业务分类医疗耗材销售约占55%,工业防护产品销售20%,其他合同加工25%。

稳健医疗表示,本次收购契合了稳健医疗打造“全球医疗耗材一站式解决方案”的发展战略,加强海外产能、销售渠道及本地化运营能力是公司坚定迈向全球化战略方向的重要举措,具有里程碑式的意义。同时,双方将形成显著的规模与协同效应,提高公司的国际核心竞争力。

据称,美国是全球最大的医疗耗材销售市场之一,稳健医疗认为提高美国市场的本地化运营能力将助力公司北美业务规模的扩大。据悉,GRI公司在美国、越南、多米尼加等海外国家拥有多个医疗耗材与工业防护产品生产、物流基地,直接填补了公司海外产能的空白,也满足了海外客户对于分散供应链风险的诉求。与自建产能相比,极大地缩短了建设周期,减少了相关建设、运营风险。

# 中国汽研拟投资不超9亿元 设立全资子公司

证券时报记者 韩忠楠

9月23日晚间,中国汽研(601965)公告称,为把握新能源汽车后市场技术服务领域发展机遇,拓展公司发展赛道,公司拟以不超过9亿元的自有资金、银行贷款或其他符合法律法规规定的自筹资金设立全资子公司,主要从事汽车后市场技术服务、新能源特色检测装备产品及行业服务等业务。

中国汽研公告表示,近年来,我国新能源汽车在全球范围内形成了先发优势,产业规模和设计制造技术水平全球领先,已成为引领全球汽车电动化转型发展的重要一极。但随着新能源汽车渗透率及保有量的快速增长,由新能源汽车三电系统故障带来的安全隐患日渐凸显,已成为当下安全监管和消防救援的重点关注事项。目前我国新能源汽车在用车的安全监管和汽车后市场发展亟须新标准、新技术、新模式推动传统产业链改造升级。

基于中国汽研在汽车后市场领域进行的技术储备,中国汽研希望成立子公司来抓住行业发展机遇,探索新能源汽车后市场技术服务领域的业务机会,打造覆盖在用车检

测、维保、交易、回收的全生命周期健康管理新生态。据悉,新成立的子公司业务包括:机动车检验检测技术服务,智能新能源汽车检测系统及设备研制、生产和销售,软件及数据服务,在用车鉴定评估,新能源汽车废旧动力电池回收及梯次利用。

中国汽研方面表示,本次投资将有利于公司开展汽车后市场技术服务、新能源特色检测装备产品及行业服务等业务,通过标准引领、行业资源整合、新技术研发等措施构建核心竞争力,寻求公司新的业务增长点,促进公司业务布局的纵向延伸与完善,进一步丰富公司业务收入结构,增强核心竞争力。

公开信息显示,2024年上半年,中国汽研营业收入为20.00亿元,较去年同期增长20.38%,其中汽车技术服务业务实现收入16.26亿元,同比增幅17.45%;净利润为4.01亿元,较去年同期增长16.88%,销售毛利率为40.59%。

中国汽研表示,业绩增长主要由于公司坚持以市场和客户需求为导向,聚焦战略优势主业和新兴业务,抢抓市场机遇,在夯实经营基本盘的同时提升区域属地化服务能力,推进新赛道新业务布局。

