

推理需求爆发 国产AI芯片获热捧



旺盛的AI推理需求支撑下,国产AI芯片厂商业绩支撑力大增。

目前国内AI芯片核心公司中,寒武纪连续六年亏损的纪录在2024年四季度终结,并将这种趋势延续到今年一季度;海光信息业绩处在稳定态势,在今年一季度更是实现合同负债大幅增长。

不过相比之下,聚焦在相对专用市场的AI芯片企业面临阶段性难题。景嘉微和龙芯中科在2024年度和今年一季度业绩均面临压力。两家公司不约而同提到在专用市场发展面临瓶颈。

当然这也是目前国内AI芯片产业发展的不同切面,通用市场固然在经历足够积淀后,借力生态兼容能实现较快发展;专有市场的蓄力也必不可少。

近两年来,超节点服务器、智算一体机等多种类型产品出现,也是在聚焦解决目前国产算力效率提升和应用落地难题。在智算产业链的多方聚力之下,国产AI芯片市场在图谋更大成长空间。

GPU巨头六年扭亏

寒武纪终于摆脱连续六年亏损的泥淖,在2024年第四季度扭亏为盈,这一趋势延续到今年一季度。这也是中国AI芯片公司业绩正待释放的一个缩影。

财报显示,2024年度寒武纪实现营业收入11.74亿元,同比大幅增加65.56%,而在2023年公司营收同比还是微跌2.7%;2024年第四季度公司已经实现扭亏为盈,净利润为2.72亿元。

公告指出,收入增长的原因在于,报告期内寒武纪持续拓展市场,助力人工智能应用落地,使得报告期内云端产品线收入同比大幅增长。2024年云端产品线占公司整体收入的99%,报告期内收入同比增长1187.78%,毛利率减少3.94个百分点。

考虑到寒武纪报告期内境内收入占比99.9%,意味着其正承接来自国内庞大的AI计算需求。

这一情况延续到了今年第一季度。财报显示,今年第一季度寒武纪实现营收11.11亿元,同比增长40倍,不过环比增幅有所下降。回看历年寒武纪的业绩不难发现,每年第四季度都是公司营收高点,因此第一季度增幅下滑在情理之中。收入增长原因是公司持续拓展市场,助力人工智能应用落地。

在净利润方面,2024年第一季度寒武纪亏损2.27亿元,今年第一季度寒武纪盈利3.55亿元。

总体来说,多年亏损的寒武纪经过持续研发投入,叠加AI需求爆发和国内对采用自主芯片的需求趋势下,业绩正式迎来爆发。

不过寒武纪多年来都存在单一客户的潜在风险,年报显示,公司近三年前五大客户的销售金额合计占营业收入比例分别为84.94%、92.36%和94.63%。其中第一大客户占公司总销售额79.15%。公告指出,该客户为公司长期合作伙伴,本报告期内其大量采购,主要系公司的智能芯片产品竞争力不断提升,在长期合作过程中得到客户认可。

国海证券认为,寒武纪存货和预付款的高增长,体现出在当前国产芯片供应链不稳定的背景下,公司或具备了较好的供货能力,为未来业务增

长奠定基础。

通用计算需求旺盛

产品线相对丰富的海光信息业绩增长更为稳健。2024年,海光信息实现营业收入91.62亿元,同比增长52.4%;归母净利润19.31亿元,同比增长52.87%。

公告显示,这是源于2024年海光信息围绕通用计算市场,不断实现技术创新、产品性能提升。CPU(通用处理器)产品进一步拓展市场应用领域,扩大市场份额,DCU(协处理器)产品快速迭代发展,得到市场更广泛认可。加之国产化市场占比进一步提升,促进了业绩显著增长。

这种趋势延续到了今年,第一季度海光信息实现营收24亿元,同比增加50.76%;归母净利润5.06亿元,同比增加75.33%。海光信息产品竞争力保持市场领先,市场需求不断增加,带动公司营业收入快速增长;营收增长带动了销售回款增加,已签订合同的预收货款同比增加较多。

海光信息财务数据中有一个显著的变化,合同负债大幅增长,今年一季度为32.37亿元,上一季度末为9.03亿元,同比增长258.47%。一般来说,合同负债意味着客户的预订合同款项,结合公司在此前年报中的表述,这意味着其为客户提供的产品和服务将大增。

海光信息更多受益于其通用计算产品的布局。旗下产品包括CPU和DCU,后者就是一种GPGPU(通用图形处理器)产品。

山西证券认为,从技术路线看,海光信息与英伟达同属于GPGPU阵营,前者推出的DCU产品采用该架构有助于降低CUDA(并行计算平台和编程模型)迁移难度并减少性能损失。目前大模型已经深入到AI芯片硬件层面做优化,达到芯片片上缓存大小优化的级别。因此,芯片架构会影响大模型以及上层AI应用的性能,而具有同样架构的芯片,其模型和应用迁移的性能损失相对较小。

国海证券指出,海光DCU具备自主研发的软件栈,是目前国内AI芯片生态中最为完备的生态之一,极大减少了应用迁移难度。

专用市场面临挑战

相比之下,此前多聚焦在相对专用市场的AI芯片公司短期内面临业绩压力。

景嘉微2024年度财报显示,报告期内实现营业收入4.66亿元,同比下降34.62%;归母净利润为亏损1.65亿元,同比下降376.67%。

从整体业务构成来看,对业绩带来较大影响的是“图形显控领域产品”,该部分贡献了公司整体收入的53%,年内实现收入2.45亿元,同比下滑47.27%、毛利率同比下降17.91个百分点。

公告显示,这一类目的产品分为图形显控模块产品和加固类产品,目前主要应用于专用市场,将不断开拓在通用市场的应用,已在针对更为广阔的车载、船舶显控和通用市场等应用领域,持续研发并提供相适应的图形显控模块及其配套产品。

“芯片领域产品”是景嘉微唯一实现收入同比增长的业务,2024年增速33.72%,毛利率同比下滑6.03个百分点,但下滑幅度高于公司整体毛利率承压水平。该业务主要指公司GPU(图形处理器)芯片带来的相关收益。

这两类业务共为景嘉微贡献了八成收入。公告显示,景嘉微全力推进由“专用”到“专用+通用”的发展战略,持续开展高性能GPU、模块及整机等产品的研发。

龙芯中科长期以来推动自主指令集架构LoongArch,因此其生态搭建需要一定时间进行积累,应用落地也呈现逐步拓展行业应用的路径。

不同于目前市场上生态建设已经相对成熟的x86和Arm体系,龙芯的目标是打造第三套生态系统,其逻辑与目前活跃的RISC-V(开源指令集架构)略有接近,但推进情况也有所不同。据称,LoongArch在服务器、桌面、终端、工控等通用领域得到日益广泛的应用。

2024年龙芯中科实现营业收入5.04亿元,同比微降0.28%;归母净利润为-6.25亿元,相比2023年亏损幅度扩大了2.96亿元。

公告显示,2024年是龙芯中科开展生态建设和面向开放市场三年研发转型(2022年至2024年)的决战之年。2024年,公司在具有传统优势的安全应用工控市场停滞,该业务营收大幅下降的情况下,

紧抓电子政务市场复苏的时机,发挥公司新产品的性价比优势,信息化类芯片营收大幅增长,推动下半年公司营收重新进入增长周期。

财报显示,2024年龙芯中科信息化类芯片产品实现收入2.69亿元,同比增长193.70%,带动毛利率回升到26.99%,但固定成本分摊的影响仍在,加之报告期内与桌面CPU配套出货的桥片成本较高,进而影响毛利率仍未恢复至理想水平。工控类芯片产品受安全应用领域需求仍未恢复的影响,实现营业收入0.90亿元,同比下降44.56%,毛利率也因同一因素下降为49.93%、同比减少17.5个百分点。

一种行业观点认为,国内高水平的计算技术体系采取了不同技术路线,其中“高铁模式”,即源头创新来自海外,强调“全兼容”,从应用发扬光大,海光就是其中代表;“北斗模式”则强调“全自主”,这条路线也很重要,但毕竟是从头做起,生态建设将经历较为缓慢的过程,典型案例就是龙芯、寒武纪等。

竞速智算生态

当前国产AI芯片需求旺盛的另一面是,DeepSeek旗下模型产品引起关注一定程度起到了重要作用,这也点燃了AI推理的应用需求。

在今年初,众多国内AI芯片厂商先后宣布实现与DeepSeek旗下不同模型陆续展开适配工作。

海光信息指出,因其DCU芯片采用了GPGPU通用加速计算架构,DeepSeek模型可直接在DCU上运行,并不需要大量适配工作,技术团队的主要工作是进行精度验证和持续的性能优化。

灼识咨询高级咨询顾问张笑璐表示,众多国产AI芯片厂商与DeepSeek的快速适配潮是中国芯片实现国际化突破的第一步。DeepSeek对于中国AI芯片厂商的利好非常确定,通过合作,中国芯片厂商加速了深度学习框架和分布式训练适配,推动了中国的“国产算力+国产大模型”闭环生态。

“从性能表现角度看,DeepSeek推理任务表现与OpenAI媲美,甚至在部分任务有所领先。这使得国产AI芯片在推理端可以快速实现商业化,同时说明了国产AI芯片替换英伟达GPU非常可行。”她续称,“当然,训练端国产芯片与英伟达的差异依然存在,在更复杂的模型训练场景下,国产芯片仍需提升。此外中国AI芯片生态仍然面临挑战。”

伴随AI推理需求日益活跃,更适配私有数据边缘端运行的智算一体机产品也先后被推出,据统计,目前市场上已经有超过一百款一体机产品,但也面临应用良莠不齐的情况。

业内人士表示,目前国产一体机产品已经实现了“可用”,但做到“好用”还需要持续推进性能优化工作,并丰富产品形态。

根据中国信通院披露,今年4月,相关团队已经启动训推一体机测评工作,除了对通用模块进行测试,同步还推出能力子域专项测评。

不止于此,智算产业链还在寻找能更好提升计算效率和传输速率的路径。

前不久,华为发布了基于新型高速总线架构的CloudMatrix 384超节点集群。海通证券指出,该架构采用全对等互联总线和共享以太网技术,将资源互联带宽提升了10倍以上。

“英伟达此前提出了这一理念,我们也认为,超节点是一个重要方向。”一名通信行业人士表示,其核心要点在于,将数据传输尽可能放在芯片内部,不行再到机内、机间去解决。“但最终还是要看什么场景下如何解决问题,最终让计算能力多好省地释放。”

供稿:《21世纪经济报道》作者:骆轶琪

混凝土生产检测施工告别“灰头土脸”

► 科工前沿

全球最大规模 真空环保高压开关启用

日前,在武汉舵落口220千伏变电站,由特变电工高压开关(GIS)智能工厂自主研发的126千伏绝缘真空开断型环保GIS成套设备,共20组批量投入商业运行。据悉,该套设备全寿命周期内预计可减少等量二氧化碳约10万吨,让电站告别温室效应污染。

原来,传统高压开关设备依赖六氟化硫实现绝缘和灭弧,每公斤六氟化硫的温室效应相当于25200公斤二氧化碳。而这套新制造的环保型高压开关设备,用洁净空气替代沿用60年的温室气体六氟化硫。

“我们实现了126千伏绝缘真空环保GIS全部核心技术及装备的国产化。”特变电工高压开关相关负责人介绍,该套设备可直接对大气排放,不会产生含氟气体及毒性分解物,使全生命周期维护成本大幅降低。同时,在不安装加热装置的情况下,设备能适应零下40摄氏度严寒环境;在电网故障状态下能承受40千安短路电流(相当于雷击强度的2倍),有效保障电网全天候可靠运行。

该负责人表示,该套设备的投用意味着武汉舵落口220千伏变电站成为我国首个大规模应用洁净空气绝缘真空环保GIS的变电站,填补国内GIS规模化商业运行的行业空白,更创下环保高压开关设备全球最大规模商用纪录。

据《湖南日报》作者:唐曦 首安倩

混凝土也称砼,是最基础的建筑材料之一。从高耸入云的摩天大楼,到纵横交错的交通路网,都有混凝土的身影。

如今,随着科技的飞速发展,传统“灰头土脸”的混凝土生产施工场景正在向着数字化、智能化生产施工场景转变——生产效率日益提高、检测水平不断提升、施工过程逐渐精细化。

华南材料从实际应用场景出发,打造出“华南智砼”数智化解决方案。这一解决方案包括“华南智砼”数智化平台、混凝土抗压强度检测设备、混凝土施工设备等,打通了混凝土生产施工上下游,为行业实现全场景智慧赋能。

智能生产兼顾质量效率

商品混凝土也称预拌混凝土(以下简称“混凝土”),其在交付前,要根据用户的不同需求,进行个性化生产。源头生产水平直接影响着产品质量乃至工程安全。

混凝土的生产流程包括原材料准备、配合比设计、计量配料、搅拌等环节,整个过程十分依赖人工经验。华南材料副总裁、安全总监徐宗望介绍,随着建筑行业对精细化施工、数字化管理要求的提升,传统的混凝土生产场景正面临着诸多挑战,如原材料计量精准度不足、生产设备监测滞后、物流管理粗放等。如今,借助“华南智砼”数智化平台,华南材料打造了智慧混凝土生产场景。

华南材料生产管理部副总经理尹征龙表示,智慧混凝土生产场景采用了先进的智能视觉识别、红外精准定位、数据采集分析等技术,可以实现物料自动化称重、磅单自动生成、异常自动报警等功能。

“装载原材料的车辆到达过磅区域后,不需要人工操作,平台可以自动对车辆及原材料信息进行

识别,并自动生成磅单数据上传。”尹征龙介绍,过磅环节的自动化升级显著提高了效率,并且避免了人为因素对原材料管控的影响。

原材料经过称重、配比设计环节,进入搅拌后,设备运行状态同样影响着混凝土产品的生产效率和最终质量。

为了能够实时掌握设备运行状况,华南材料团队在设备关键部位安装先进传感器,可实时收集设备振动、温度、压力等运行参数,并对数据进行深度分析。一旦发现设备存在异常或潜在故障,便会立即报警。

自动检测保证品质可控

混凝土检测是建筑工程质量管控体系中的关键环节。

如今混凝土检测场景正在从依赖人工判断走向自动化智慧检测。例如,混凝土抗压强度检测是判定混凝土质量是否合格的重要途径。以往,从试块取样制作到抗压强度检测,整个过程都需要检测人员严格按照操作流程进行手工操作,不仅工作量大,还容易造成数据不准确等问题。

在智慧检测场景下,试块自动化检测设备可以将新拌混凝土样品按照预设的比例输送到成型系统。进入成型系统后,混凝土会被自动分配到各个模具中,由压制机构对模具中的试样进行压制,使其成为标准的试块。同时,控制系统也会实时监控样品的输送量和比例,确保取样具有代表性、一致性,并且会精确控制压力和时间,以确保试块的密度和强度达到预设的标准。

试块成型后则自动进入养护区域,按照系统预设养护环境的温度和湿度进行养护,确保试块的养护质量。养护完成后,试块会自动脱模,并进行包装和存储。随后,检测设备会自动抓取试块,进行

抗压试验等一系列检测。“以前,整个流程都需要人工操作,工期紧张时大家要轮流倒班做试验,工序多、任务重,一次检查就要持续几个小时。如今,在智慧检测场景下,机器可以24小时不停运转,大大提升了效率。”尹征龙说。

不仅检测流程实现自动化,整个检测场景也在向着灵活化、透明化方向发展。

由华南材料自主研发的厢式卡车搭载了移动智能生产试验中控室,设有试验区和数智化平台区两大功能区。

尹征龙介绍,试验区可实现原材料和混凝土全过程质量检测,如进场原材料快速检测、结构实体的强度以及裂缝检测等,确保混凝土生产全过程品质可控;数智化平台区的两块数智大屏实时联动,可通过信息化系统实现对生产控制、物流运输、交货验收、现场浇筑、工程实体等数字化全过程质量监管。

机器设备提升施工效能

进入施工环节后,混凝土施工流程的规范性、稳定性、精细度,直接影响建筑工程品质。

为了最大限度避免人工操作的局限性,混凝土施工场景正引入越来越多的机器帮手。例如,混凝土激光摊铺机可以实现人工施工难以达到的超平整度和施工效率。该设备借助激光发射器产生一个精准的水平基准面。该基准面如同一个无形的“水准仪”,为混凝土摊铺提供精确的标高参照。

此外,在传统的混凝土施工场景中,刮板整平和喷淋养护通常需要分别使用不同设备、分阶段进行,导致施工效率不高。如今借助华南材料研发的刮板整平和喷淋养护一体化设备,可以将两个工序合二为一,减少了设备来回倒运和工序衔接等待时间,大幅缩短了施工周期。 据《科技日报》作者:都瓦