

AI眼镜热潮引发“芯”变革



本报记者 许子皓

从Meta到小米, AI眼镜正形成“百镜大战”之势。作为可穿戴领域的新宠, AI眼镜的功能在不断迭代和丰富, 而在这背后, 少不了各类芯片的加持, 无论是技术层面的革新突破, 还是需求数量的爆发式增长, 都是强劲的“芯”动力在发挥作用。当然, AI眼镜市场的快速扩张, 对各类芯片的需求呈现出爆发式增长, 也为芯片市场注入了新的强心剂。

核心算力引擎： 主控芯片 SoC

AI眼镜的运行依赖多种关键芯片协同工作, 各类芯片的性能也是AI眼镜实现各种强大功能的基石, 直接影响着AI眼镜的运算速度、图像识别能力、音频处理效果以及功耗管理等关键性能。其中, SoC芯片作为核心的主控芯片, 如同AI眼镜的“大脑”, 集成了中央处理器(CPU)、图形处理器(GPU)、神经网络处理器(NPU)以及图像信号处理器(ISP)等多个模块, 负责整体运算、图形处理、AI任务执行以及图像数据处理等重要功能。

当前, 主流的SoC方案主要有三种。

第一种是系统级SoC方案, 该方案是将多个关键的电子组件集成为一个芯片上, 实现高度集成化, 具备较强的计算能力和丰富的功能支持, 能应对复杂的AI运算和多任务处理。以高通的AR1 Gen1芯片为例, 其采用6nm制程, 具备高集成度和低功耗特性, 搭载的多核处理器可支持复杂计算任务和实时图像处理, 集成的第三代Hexagon NPU, 为视觉搜索、实时翻译、定向音频采集等功能提供了支持, 目前被应用于Meta Ray-Ban智能眼镜、雷鸟X3 PRO、小米AI眼镜等企业旗舰产品中。然而, 系统级SoC也存在一些劣势, 其功耗相对较高, 在散热方面也面临挑战, 同时芯片成本较高, 平均约55美元, 占整机硬件成本的31.6%, 限制了其在中低端AI眼镜产品中的应用。

第二种是MCU级SoC+ISP方案, 该方案以MCU作为主控单元, 集成了基础通信与传感模块, 提供基础运算能力, 但由于无法集成ISP, 需要外接ISP芯片用于影像功能。例如, 恒玄2800采用6nm制程, 集成多核CPU/GPU、NPU等, 为满足拍摄需求, 需外挂ISP芯片。这种方案在一定程度上平衡了成本与性能, 成本仅10~15美元, 功耗低至100mW, 但难以支持复杂的AI模型。因此, 通常应用于一些功能相对简单、对成本和功耗要求较高入门级AI眼镜产品中, 提供基本的图像拍摄和简单的智能交互体验。

第三种是SoC+MCU双架构, 该方案让SoC和MCU分工协作, SoC负责处理高计算需求的任务, 如运行分时操作系统、执行人工智能算法以及处理摄影功能相关的高负载任务, 利用其优秀的算力和丰富的功能模块, 保障AI眼镜在复杂任务下的高效运行; MCU则专注于音频等低功耗场景, 通过分时调度实现对功耗的精细管理, 从而优化续航表现。通过这种分工方

式, 该方案能够在保证AI眼镜性能的同时, 实现高效的电源管理, 有效延长设备的续航时间, 非常适合需要长时间佩戴使用的AI眼镜产品。不过, SoC+MCU方案由于采用了两颗芯片, 成本也相对较高。例如, 星辰科技SSC833在AI视觉处理中采用了这一架构, 成本约20~30美元。

随着AI技术在眼镜设备中的深入应用, 对主控芯片算力的要求呈指数级增长。赛迪顾问集成电路中心副总经理杨俊刚表示, 从简单的语音识别、图像分析, 到更为复杂的实时翻译、智能场景感知, 都需要芯片具备强大的计算能力, 以实现快速、准确的处理。然而, 算力的提升往往伴随能耗的急剧增加, 这与AI眼镜对续航能力的要求形成了矛盾。

解决这一矛盾, 当前主流的方法是采用更先进的制程, 如从传统的12nm工艺演进到更先进的6nm、4nm工艺, 能够在更小的芯片面积内集成更多的晶体管, 增强集成度, 减少芯片面积的同时, 提高芯片的运算效率, 同时降低单位运算的能耗。另一方面, 引入动态电压频率调节(DVFS)技术, 根据芯片的工作负载实时调整电压和频率。当芯片执行简单任务时, 降低电压和频率以减少能耗; 而在处理复杂任务时, 提高电压和频率以满足算力需求。此外, 一些芯片还采用了异构核分工技术, 如恒玄BES2800芯片, 通过大核处理突发任务、小核维持待机, 实现了平均功耗低于300mW, 在保证芯片性能的同时, 优化了能耗管理。

此外, 3D堆叠封装技术有望成为提升芯片性能与集成度的关键路径。通过TSV(硅通孔)技术, 3D堆叠封装能够实现存储、计算单元的三维集成, 将原本在二维平面上布局的芯片组件在垂直方向上进行堆叠, 从而大幅提升芯片的能效比, 为AI眼镜实现更复杂的功能和更长的续航提供有力支持。

视觉之眼： CMOS图像传感器

在AI眼镜的“视界”里, CMOS图像传感器是实现高清拍摄、精准图像识别等功能的核心组件, 其性能直接决定了AI眼镜的成像质量、视觉交互体验以及各项功能的实现效果。

当前, 索尼IMX681在AI眼镜市场中占据着统治地位, Ray-Ban Meta、雷鸟V3、Rokid Glasse、小米AI眼镜等产品均采用该芯片。索尼IMX681之所以能够占据主要市场, 源于以下优势:

首先, 其尺寸小巧, 像素排列为3024×4032, 像素数达到1200万, 像素间距为1微米, CMOS宽边长

度略大于4.032mm, 长边长度略大于3.024mm, 仅约为手机摄像头传感器的25%, 这使得它能轻松适配AI眼镜的紧凑设计。在功耗方面, IMX681采用背照式堆栈工艺, 优化了能耗管理, 相比手机的CMOS传感器更节能, 低功耗特性有效降低了AI眼镜整体的发热问题, 避免影响佩戴体验。在拍摄质量上, 它拥有不错的HDR支持以及全局快门技术, AI眼镜用于AR体验、物体识别时, 经常需要拍摄运动物体, 全局快门能避免传统滚动快门带来的运动畸变问题, 用于实时场景识别(OCR、手势检测、人脸识别等), 也能提供更清晰的图像, 有助于AI算法的处理。

此外, 索尼在智能手机影像领域长期耕耘所构建的成熟开发生态, 也是IMX681的一大优势。众多AI眼镜厂商采用高通AR1平台作为计算平台, 而索尼IMX681与高通AR1平台深度适配, 厂商无须投入大量人力进行影像优化, 就能获得成熟稳定的影像体验, 大大缩短了产品的研发周期, 降低了开发成本。

然而, 随着AI眼镜市场从小众探索走向大众消费, 用户对影像质量、AI计算能力、续航的需求不断升级, “统治”市场的索尼IMX681也面临着挑战。近年来, 以豪威和思特威为代表的中国CMOS厂商在AI眼镜领域取得了显著进展, 以优秀的性价比和丰富的定制能力, 开始打破索尼IMX681的垄断局面, 有望为AI眼镜市场带来新的活力与变革。

豪威作为韦尔股份旗下的CMOS厂商, 是国内手机CMOS领域的佼佼者, 已被小米、华为等旗舰手机采用, 其产品在像素尺寸、动态范围、对焦速度等关键指标上表现出色。豪威已经成功进军亚马逊的AI+AR眼镜项目, 这标志着其与索尼的竞争中取得了重要突破。

思特威同样在手机市场占有一席之地, 2023年全球手机摄像头CMOS出货量排名中紧随索尼、三星和豪威之后。2025年5月8日, 思特威全新推出1200万像素AI眼镜应用CMOS图像传感器——SC1200IOT。在尺寸设计上, SC1200IOT采用1/3.57”靶面设计, 像素尺寸为1μm, 封装尺寸小至5.1毫米×3.7毫米, 能很好地适配主流AI眼镜边框视觉模块的微型摄像头。功耗方面, 得益于内部电路模块的精准化控制设计, SC1200IOT实现了功耗的显著优化, 这不仅能有效减少使用中的设备发热, 还能提升AI眼镜的续航能力。成像效果上, 与行业同规格产品相比, SC1200IOT的感光度提升约29%。目前, SC1200IOT已接受送样, 并于2025年Q2实现量产, 有望在AI眼镜市场中占据一席之地。

数据流转中心： 存储芯片

存储芯片在AI眼镜中同样关键, 肩负着存储和检索数据的重任, 在AI眼镜的硬件成本中占据着相当高的比例。以Meta的Ray-Ban智能眼镜为例, 佰维存储提供的存储芯片(ROM+RAM)在其BOM成本中占比约7%, 仅次于主控SoC芯片, 单机价值约11美元。

ePOP、eMCP是当前市场主流的AI眼镜存储集成方案。ePOP(Embedded Package on Package)即嵌入式堆叠封装技术, 通过将NAND Flash(非易失性闪存存储器)和LPDDR(低功耗双倍数据速率内存)垂直堆叠在SoC上方, 实现了高度的集成化。这种技术最大的优势在于能够节省约60%的PCB空间, 让AI眼镜的内部布局更加紧凑, 有助于设备实现轻薄化设计。同时, 由于减少了芯片数量和连接线路, 降低了功耗, 从而提升了设备的续航能力。该技术方案凭借小尺寸、低功耗、高性能等特点, 已被Meta、Google、Facebook等企业应用于AI智能眼镜、智能手表等产品, 并通过了高通等平台认证, 有效优化了设备响应速度与多线程处理能力。

eMCP(Embedded Multi-Chip Package)即嵌入式多芯片封装技术, 则是采用多芯片封装技术, 将eMMC(嵌入式多媒体卡)与LPDDR集成在一个封装中。它内置NAND Flash控制芯片, 能够减少主芯片的运算负担, 简化大容量闪存的管理, 同时也能节省主板空间。在一些对存储容量和性能有一定要求, 但又需要控制成本的AI眼镜中, eMCP技术得到了广泛应用。它能够在保证一定性能的前提下, 为设备提供较为经济的存储解决方案, 满足了市场对于中低端AI眼镜的需求。

随着AI眼镜功能的日益强大和丰富, AI眼镜需要处理和存储更多的数据, 如高清视频、高分辨率图像以及不断升级的AI模型等。存储芯片的容量将不断增大, 从目前常见的32GB向更高容量发展, 以确保设备能够流畅运行各种复杂的应用程序。

除了以上芯片种类之外, AI眼镜中还包括音频处理芯片、电源管理芯片、通信芯片、蓝牙芯片等。AI眼镜市场的快速增长, 直接拉动了对各类芯片的需求, 为芯片厂商提供了更广阔的市场空间, 但也对芯片性能、功耗、尺寸等方面提出严格要求, 促使芯片厂商不断探索新技术、新工艺。这些技术创新不仅适用于AI眼镜芯片, 还有望应用于其他领域的芯片产品, 拓展芯片产业的应用边界。

北大团队发布 全球首创芯片设计存算一体排序架构

本报讯 记者许子皓报道:近日, 由北京大学集成电路学院教授杨玉超、人工智能研究院研究员陶耀宇组成的团队, 在国际上首次实现了基于存算一体技术的高效排序硬件架构, 解决了传统计算架构面对复杂非线性排序问题时计算效率低下的瓶颈问题, 将为具身智能、大语言模型、智能驾驶、智慧交通、智慧城市等人工智能(AI)应用提供更高效的算力支持。

陶耀宇在接受《中国电子报》记者采访时表示:“排序, 作为人工智能系统中最常用、最耗时的基础操作之一, 广泛存在于自然语言处理、信息检索、图神经网络、智能决策等人工智能相关领域中。在传统的冯·诺依曼计算架构中, 处理器与存储器分离的特点导致数据被频繁搬运, 严重制约了计算速度和系统能效, 因此将存储与计算融合的‘存算一体’架构是突破硬件性能的必由之路。”

例如, 在智能驾驶场景中, 车辆在高速公路等简单环境中, 只需对周边几台车的行进路线进行排序就能判断是否会对自身安全产生影响, 整个耗时可控制在毫秒级别, 但在复杂的城市环境中风险来自周边的数百个乃至更多的各类交通参与者, 排序的复杂程度和硬件延迟大幅提高; 再比如, 在抖音、小红书等平台大的规模检索与推荐系统中, 存在用户点击、评分机制, 系统经过相似度计算后, 常需要对海量的候选内容进行快速排序, 找出最相关项, 排序性能直接影响系统的响应时间和可扩展性; 此外, 在大语言模型训练、机器人路径规划、强化学习搜索等场景中, 快速评估多个决策或行动的优劣并进行排序, 也是必不可少而又极为费时的步骤。然而, 在传统计算架构下, 大规模的非线性排序难以在端侧或边缘设备高效完成, 这一过程消耗大量时间与功耗, 极大制约了具身智能、智能驾

驶等新兴技术的发展与普及。

近年来, “存算一体”被认为是突破传统硬件架构计算瓶颈的关键技术, 已经在矩阵计算等规则性强的数值计算环节取得显著成果。然而, 由于排序过程仍存在逻辑复杂、操作非线性、数据访问不规则、高度依赖复杂比较器网络, 以及缺乏通用、高效的硬件排序原语等诸多障碍, 目前国际主流的存算一体架构均无法解决大数据排序问题, 上述难题成为了制约下一代人工智能计算硬件发展的前沿焦点与核心卡点问题。

北大科研团队本次发布的基于存算一体技术的高效排序硬件架构, 围绕“让数据就地排序”的第一性原理目标, 在存算一体架构上攻克了多个核心技术难题, 实现了排序速度与能效的数量级提升。实现了多项突破: 一是开发了一套基于新型存内阵列结构的高并行位读取机制; 二是开创性地引入了忆阻器阵列, 提出了按数、按位、多值三种并行策略, 实现了低延迟、多通路的硬件级并行排序电路设计; 三是在算子层面, 优化了面向人工智能任务的算法—架构协同路径, 同时兼容现有矩阵计算; 四是完全自主设计的器件—电路—系统级技术栈整合。

实测结果显示, 该硬件方案在典型排序任务中实现运算速度超过15倍的提升, 但功耗仅为传统CPU或GPU处理器的1/10。在人工智能推理场景中, 支持动态稀疏度下的推理响应速度可提升70%以上, 特别适用于要求极高实时性的任务环境。

陶耀宇表示, 该成果未来有望广泛应用于国产智能芯片、边缘AI设备、智能制造终端、智慧城市系统等关键领域, 为我国在下一代AI技术与智能硬件领域的竞争中提供了坚实底座, 也为构建安全、高效、自主可控的新一代智能算力体系提供了“加速引擎”。

美撤销芯片设计软件 对华出口限制

本报讯 记者姬晓婷报道:近日, 《中国电子报》记者从新思科技获悉, 新思科技于美国加利福尼亚时间7月2日收到美国商务部工业与安全局来函, 通知基于2025年5月29日收到的限制令所实施的对外出口管制措施现予以撤销, 即时生效。公司正恢复近期受限产品在中国市场的供应和全面客户支持。

此外, 据央视新闻7月3日报道, 德国西门子股份公司收到美国政府的通知称, 美国已取消对中国芯片设计软件的出口限制。根据公司声明, 这家德国供应商已恢复中国客户对其软件和技术的全部访问。

据外国媒体报道, 同为全球EDA产业前三的楷登电子也发布了类似声明。

DDR4价格涨势预计 今年第四季度触顶回落

本报讯 近日, TrendForce最新发布的存储芯片市场观察显示, 自2025年第二季度以来, DDR4的市场价格呈现快速上涨态势, 但这一涨势预计在第四季度将触顶回落。而NAND市场虽受到AI需求的拉动, 但由于供需尚未形成紧张局面, 价格预期保持持平。

自今年第二季度以来, DDR4价格一路飙升。主要原因在于头部芯片制造厂商为追求更高利润, 逐步削减DDR4产能, 加速向DDR5及HBM等高附加值、高性能产品领域转移。以三星为例, 为集中资源全力推进高频宽记忆体的研发进程, 目前已基本停止向市场供应DDR4新品。另一方面, 当前全球供应链稳定性面临挑战, 下游客户出于对供应链中断风险的担忧, 纷纷加大DDR4芯片的采购量, 提前进行大规模囤货。这一系列操作进一步加剧了DDR4市场的供需失衡状况, 从而推动价格持续攀升。

不过, 南亚科等部分厂商基于对市场多元化需求的判断, 逆势扩充DDR4产能。随着这些新增产能在第四季度逐步释放, 市场供应紧张的局面将得到缓解, 预计DDR4价格涨势也将随之触顶回落。

在NAND市场方面, AI技术的快速发展为其带来了显著的需求增长。特别是北美地区的CSPs(云服

务提供商), 已广泛在推理AI服务器中采用大容量QLC企业级固态硬盘, 这对NAND市场起到了有力的支撑作用。

然而, 从供应端来看, 全球市占率排名前五的NAND Flash制造商, 即三星、SK海力士、美光、铠侠与西部数据, 在2025年上半年实施了10%~15%的减产幅度。但这一减产力度尚未使市场供需关系达到紧张状态, 整体供应仍相对充足。

因此, TrendForce预计在未來一段时间内, NAND市场价格将维持持平态势。尽管AI需求在持续增长, 但由于供应端的稳定表现, 短期内难以形成推动价格大幅上涨的合力。DDR4价格走势的变化以及NAND市场价格的持平状态, 将对整个记忆体产业链产生广泛影响。

对于模组厂而言, 需要密切关注DDR4价格波动, 合理安排原材料库存, 避免因价格回落导致库存价值缩水。

对于电子产品制造商而言, 无论是消费电子还是工业电子领域, 都需深入研究市场动态, 加强与供应商的合作与沟通, 共同应对市场变化。在产品研发上, 可探索新的存储解决方案, 以降低对单一存储产品价格波动的依赖, 确保企业生产经营的稳定性和可持续性。

(炫文)