

编者按:当前,存储行业正处于史无前例的“超级上升周期”,不仅是频繁涨价,还面临着技术迭代加速、市场需求结构变化、供应链格局调整等多重机遇与挑战,各大企业需要在产业链供需平衡与技术升级节奏之间进行深度博弈。

为探索新时代下存储产业的发展之道,《中国电子报》推出“对话存储‘芯’力量”栏目,邀请全球存储企业负责人,全方位解读产业发展趋势、企业战略与技术创新方向,共同探讨AI时代下,存储技术发展和产能供给的布局与思考。

紫光闪芯总经理东猛：“看透”并用好AI存储超级周期



对话人：紫光闪芯总经理 东猛

《中国电子报》总编辑 连晓东

时 间：2026年5月27日

地 点：新紫光集团会客厅

本报记者 许子皓

三大核心因素触发 存储超级周期并非泡沫

连晓东:当前由AI驱动的存储超级周期备受行业关注,您认为这是真实刚需还是存在一定泡沫?

东猛:这一轮由人工智能驱动的存储超级周期,本质上是真实的产业刚需,并不是短期供需差带来的炒作,也不是单纯的价格抬升行为,而是面向未来数字经济与人工智能发展的正常市场驱动。

因为,与过去经历的多轮存储周期相比,这一轮有着本质上的三个核心不同:

一是需求的驱动因素彻底改变,过去全球半导体闪存的周期波动,几乎都是由消费电子需求主导,手机、个人电脑等终端产品的换代更新,消耗了市场上80%甚至90%的闪存晶圆,消费电子需求的起伏直接导致原厂产能扩张或收缩,进而带来价格的周期性涨跌;而这一轮需求的核心驱动力是人工智能,尤其是生成式AI与大模型从训练阶段全面进入推理与落地应用阶段,模型参数、快照缓存、KV Cache等海量数据,已经远远超出HBM和DRAM内存的承载能力,同时大量数据还会占用GPU算力形成明显瓶颈,这就让大容量、高性能的高速闪存成为人工智能基础设施不可或缺的核心部分,这种需求瞬间爆发,形成了跨越传统消费电子周期的全新增长动力。

二是供应端存在刚性约束,内存晶圆和闪存晶圆在部分制造工序上是共用的,当前HBM内存供应极度紧缺,直接挤占了闪存的可用产能,海外主流闪存晶圆厂的扩产计划已经排到2028年,高端AI数据中心所需要的高性能闪存芯片,在短期内看不到产能大幅缓解的可能。

三是产业底层逻辑被AI全面重构,人工智能正在重新定义整个信息技术体系,从行业应用到消费电

子再到各类智能终端,都需要更高性能、更大容量、更稳定可靠的存储支撑,这种由技术革命带来的需求具备长期性、确定性和不可替代性,尤其是PCIe 5.0以上能够直接对接大模型的企业级大容量高速固态硬盘,是真正的刚性缺口,也让这一轮周期具备了持续成长的基础。

连晓东:您认为中国本土存储企业面临的机会更多还是挑战更多?

东猛:对于中国本土的存储企业来说,一定是机会远大于挑战。目前全球行业已经形成共识,中国是仅次于美国的全球第二大人工智能需求市场,随着国家经济实力、科技水平的持续提升,未来我们甚至有可能超越北美,成为全球最大的AI需求市场。庞大且持续增长的本土需求,是国内企业最坚实的发展基础。

同时,国内以本土化CPU、GPU、操作系统、数据库为核心的本土生态持续完善,本土存储厂商拥有海外友商不具备的先天优势,我们更理解国内行业场景,在金融、电信、能源、医疗、教育等领域的适配性更强,联合研发的效率更高,能够快速响应客户的定制化需求。对于提前做好技术布局、产品布局和供应链布局的国内厂商来说,这恰恰是抢占市场、建立壁垒的最佳窗口期。

挑战也客观存在,AI数据中心级别的存储产品具备很高的技术门槛,对性能、功耗、稳定性、断电数据保护能力、数据中心兼容性都有极为严苛的要求,传统存储厂商如果没有长期的研发积累和技术投入,很难达到高端AI存储产品的标准;同时从PCIe 5.0到PCIe 6.0的技术迭代速度很快,每一代产品都需要10到12个月的研发周期,国内企业能否紧跟高端技术路线、持续投入研发、完善产业生态,也是我们必须面对和突破的考验。

连晓东:有观点认为,在不断涨价背景下,国内存储企业的机会在端侧而非企业级,您是否认同这种看法?

东猛:我们完全不认同这种观点,这种判断本质上是对上游供应链的信心不足,以及对高端技术布局的认知不够。紫光闪芯依托新紫光集团的产业背景,拥有稳定的企业级存储颗粒供应能力,同时覆盖本土化颗粒和国际化颗粒两大来源,通过自主研发的芯片和固件开发能力,我们可以让不同代际、不同类型的颗粒达到同等水平的AI SSD性能,同时还在固件中加入专属AI算法,提前将模型参数加载到大容量闪存中,有效提升系统整体性能,这也是我们的产品能够获得客户高度认可的核心原因。

紫光闪芯在2023年筹建之初,就主动摒弃了低毛利、低技术门槛的产品路线,2024年直接立项研发企业级PCIe 5.0高端产品,提前三年布局AI存储赛道,用两年时间完成核心技术研发,今年产品量产迅速跻身国内第一梯队,成功导入多家头部云厂商和AI数据中心。对于中国企业来说,AI数据中心是需求最旺盛、价值最高、成长空间最大的赛道,再叠加国内无比庞大的市场需求,这里的机会远比端侧市场更大,这也是紫光闪芯最核心的差异化优势和战略选择。

依托新紫光集团产业生态 提供云到端全场景闪存产品

连晓东:紫光闪芯在整个存储行业的核心定位是什么?全球化思路是怎样的?

东猛:紫光闪芯是一家专注于大容量高速闪存,集研发、设计、制造、客户端应用于体的全产业链半导体闪存研发企业,能够全面支持国内及全球数据中心、手机、无人机、车载嵌入式、AI PC等全产业链AI存储需求。

紫光闪芯是新紫光集团设立的新建创新型企业,也是集团围绕人工智能存储产业布局的核心产业平台。在不久之前的新紫光集团创新峰会上,新紫光集团发布了“新紫光集团全家桶”,覆盖IC、ICT、AI三大领域,其中IC层面,贯通通信、计算、控制、联接、存储,器件六大芯片品类,紫光闪芯属于IC领域中的存储领域,代表新紫光集团在闪存方面的重要战略布局。

公司的出发点是依托于新紫光集团强大的产业生态、技术积累和平台支持,公司的发展目标绝不仅仅局限于集团体系内部。新紫光集团有限公司董事长李滨曾提出,集团所有新创公司的未来,都面向本土化市场、全国市场乃至全球市场,国内本土化存储是千亿级别的市场,而全球存储市场是万亿级别的市场,紫光闪芯的目标就是立足新紫光生态的孵化与协同,凭借自主研发的产品和技术实力,通过市场化的方式不断打磨产品、锻炼团队,持续追赶全球领先存储企业的技术水平,最终成为AI时代覆盖芯片、固件、半导体设计、封装测试、产能保障的全产业链大容量高速闪存高科技企业,服务国内所有本土化AI服务器、云厂商、互联网企业。同时依托新紫光集团在海外60多个办事处的全球化布局,开拓海外核心客户,参与全球市场竞争。

我们还注册了UMEMORY海外专属品牌,在全球22个国家完成了商标注册和市场布局。特别要指出的是,公司的全球化路线,并不是只面向“一带一路”等新兴市场,而是直接对接全球最前沿、最高端的AI需求市场,与海外大型AI数据中心设备厂商、物联网终端厂商开展产品适配合作,解决海外市场面临的存储缺货、高性能产品不足等痛点。同时我们坚持双向协同,把海外的前沿技术、应用经验、市场需求引入国内,反哺国内产品迭代和生态优化,让国内技术与全球前沿保持同步,真正实现“立足中国、服务全球”的全球化目标。

连晓东:请展开讲讲公司的产

品布局及能力?

东猛:公司将按照人工智能的实际应用场景,把核心产品线划分为三大方向,全面覆盖AI时代的核心存储需求。

第一类是AI数据中心企业级SSD,目前公司PCIe 5.0规格的产品已经实现量产,eTLCSSD最高容量达到30.72TB,性能达到14GB/s,能够匹配当前主流大模型的应用需求;同时我们正在研发128TBcQLC-SSD,单盘大容量闪存产品,预计今年第四季度正式发布,明年将推进PCIe 6.0规格产品的量产,性能将提升至28GB/s,持续跟上大模型参数指数级增长带来的容量与性能需求。

第二类是AI推理与边缘计算嵌入式闪存产品,其中最核心的是车载存储,随着新能源汽车和智能驾驶的发展,汽车已经成为一个小型移动数据中心,对存储的容量、性能、稳定性要求大幅提升,公司提供1TB、2TB高规格车规级存储方案,满足-40℃到105℃的宽温适配要求,和国内整车厂、紫光展锐联合研发,推进本土嵌入式存储在新一代车型上的落地应用;同时覆盖无人机、机器人等端侧场景,满足端侧算力与存力持续提升的需求。

第三类是AI PC SSD,深度适配本土化CPU、操作系统和数据库,目前已经量产出货超过100万片,支撑本土化PC和AI PC的升级换代需求。

紫光闪芯虽然在2024年才注册成立,2025年全年营收就达到上亿元,这体现了我们的产品能力,我们今年的营收目标是实现三倍以上增长。而从营收的构成看,AI数据中心企业级业务已经从过去占比20%的辅助地位,一跃成为市场绝对主力。未来,我们预测公司50%以上的营收由AI数据中心企业级SSD贡献,端侧业务包括车载、无人机、机器人等占比30%;AI PC相关业务占比20%,这一结构也完全贴合全球存储市场向AI高端存储需求集中的大趋势。

连晓东:目前紫光闪芯的产能和供货能力处于什么水平?未来是否能支撑公司上述产品规划?

东猛:产能与供货的稳定性是大客户选择紫光闪芯产品的关键因素,目前我们在供应端具备显著优势。公司第一代高端产品就完成了本土化颗粒的全适配研发,在去年行业供应最紧张的时候,我们又顺利导入国际化颗粒资源,实现了本土化颗粒与国际化颗粒双来源稳定保障,能够充分满足头部云厂商、AI服务器厂商的批量供货需求。

在产能方面,我们严格遵循新紫光集团“大研发、大市场、大制造”的整体战略,集团内部的产业公司可以为我们提供封测、贴片等环节的支持,同时我们与业内成熟、稳定的头部代工厂建立长期合作,并且提前锁定了AI赛道专属的产能资源,保障当前业务的快速增长。

展望2028年至2029年,预计营收目标将达到20亿元到50亿元的量级,产品出货量达到千万级,届时,我们将规划建设自有贴片厂,实现50%产能自主保障,50%市场化灵活配置的供应链结构,最终打造从芯片、固件、半导体设计,到封装测试、产能保障的全产业链AI大容量高速闪存企业,实现对核心行业客户的长期稳定交付。

三大硬核能力赋能 全方位推动AI原生

连晓东:紫光闪芯在技术创新上的核心能力体现在哪些方面?

东猛:紫光闪芯的技术创新能力,围绕大容量高速闪存全产业链构建,核心体现在三个关键维度,也是公司能够在高端AI存储市场站稳脚跟的核心壁垒。

一是全自主固件开发能力。公司拥有所有兼容性验证的软件开发

源代码,针对本土CPU、GPU、操作系统深度定制优化算法,同时结合国内外主流大模型的参数特点,在固件中嵌入AI优化策略,在硬件架构定型的基础上,持续通过软件算法提升存储产品在AI场景下的性能表现,让存储更适配大模型的训练与推理需求。

二是全场景生态适配能力,公司与国内主流CPU、GPU、AI服务器厂商建立深度协同合作,和新华三设立联合实验室,完成了数十种本土及全球CPU硬件平台的全面适配验证,优先覆盖国内外主流AI服务器平台,确保产品研发完成后可以快速导入客户现场,大幅缩短客户的测试与上线周期。

三是颗粒适配与硬件设计能力,公司拥有经验丰富的核心硬件团队,长期专注于存储颗粒筛选、硬件架构设计,针对国内和国际主流供应商的新一代颗粒,开展大量模拟实验与适配调试工作,能够在新一代颗粒上市时快速完成适配,优先推出高性价比的大容量闪存产品,抢占市场先机。这些技术能力已经落地为实际的商业成果,我们的PCIe 5.0高端产品成功导入多家核心云厂商和AI数据中心,顺利入围中国移动采购项目,成为国内第一代AI SSD的标杆应用案例。

一句话,我们的技术创新能力十分符合“以全栈本土化、大容量、高性能的闪存产品,成为AI时代数据存储基石”的定位。

连晓东:紫光闪芯在AI原生方向有怎样的规划?

东猛:AI原生的核心是硬件、算法、架构全面适配人工智能。我们的产品正全面迈向AI原生,所有研发设计都围绕AI原生的需求展开。首先在硬件层面,我们选用的芯片、颗粒、硬件架构等指标,完全按照AI大模型的实际需求设定,从源头满足AI场景的高性能、大容量、低延迟要求;其次在算法层面,我们针对大模型存力瓶颈问题,在固件中定制开发专属AI算法。未来还会推出针对不同大模型的定制化SSD产品,可以将约三分之一的模型参数预装在闪存中,用户开机即可实现性能加速,大幅降低GPU算力负载,提升推理效率;同时我们高度关注行业前沿技术方向,包括HBF高带宽闪存、GPU直连闪存、异构封装、晶圆级集成等技术路线,对标海外新一代GPU架构设计理念,未来将推出真正意义上面向AI原生的闪存产品,这也是公司未来最重要的研发方向之一。

连晓东:未来闪存是否有可能替代一部分DRAM?

东猛:闪存一定会替代一部分DRAM,这一趋势已经被海外头部厂商验证,并且成为未来计算架构的重要方向。英伟达下一代Rubin GPU架构就明确释放了信号,其对DRAM和HBM内存的需求降低50%,同时对大容量闪存的需求提升50%,核心逻辑就是用大容量、高带宽的闪存作为缓存层,直接卸载GPU的算力压力,打破传统分层存储的性能瓶颈,同时缓解HBM产能极度紧缺的行业痛点。这一减一加的变化,重构了计算与存储的协同架构,为闪存行业打开了全新的增量空间,也让闪存从传统的存储角色,向更靠近算力核心的位置升级。

连晓东:您如何看待“韜(τ)定律”对闪存行业的影响?

东猛:“韜(τ)定律”对整个半导体闪存行业的影响非常深远,也为我们突破制程限制提供了新的思路。

“韜(τ)定律”一方面直指核心芯片的制程卡脖子问题,目前国内晶圆厂在先进制程上与海外存在差距,主要受制于生产设备等关键环节;另一方面提出了颗粒直贴、异构集成等创新路线,通过绕开传统堆叠技术的制程限制,对晶圆颗粒进行深度筛选与优化,直接贴装在主板上,或者通过打孔串联形成更高并发的闪存介质,

提供新一代半导体存储解决方案,我们关注的HBF高带宽闪存技术,也遵循这一核心逻辑。

夯实系统工程能力 中国存储可加速赶超

连晓东:AI原生驱动下,中国存储企业是否有机会实现弯道超车?

东猛:中国存储企业完全有机会实现弯道超车,而且这一过程正在发生。

今年国内以中国移动为代表的云厂商、大型互联网企业、AI服务器厂商,已经全面切换到AI原生应用场景,高端存储产品快速升级到PCIe 5.0以上规格,国内的生态成熟度和市场需求水平,已经媲美国外主流设备厂商;同时以DeepSeek为代表的本土大模型,已经和海外主流大模型形成同台竞争的格局。

超大容量SSD的技术壁垒,并不在于存储颗粒本身,而在于系统性工程能力:跨多客户平台适配,颗粒批次一致性管控、业务QoS服务质量保障、散热和功耗的长期稳态平衡,以及固件算法的持续迭代优化,每一项都需要长期技术深耕与场景打磨,而绝非简单硬件堆叠。紫光闪芯的PCIe 5.0高端闪存产品,在性能、容量、稳定性等核心指标上,已经可以媲美国际一线品牌的同类方案,紫光闪芯还将面向AI大模型厂商、AI科技企业及AI数据中心推出企业级122TB QLC SSD产品。

PCIe 6.0作为下一代核心方向的高端闪存技术,正成为国内存储企业突破发展的关键抓手,从产业需求端看,AI大模型迭代持续抬升存储性能门槛,类似DeepSeek等大模型产品参数成倍扩容,未来新一代大模型运行落地,将高度依赖PCIe 6.0高速闪存支撑。同时海外GPU架构迎来变革,英伟达新一代B300、Rubin架构大幅削减DRAM、HBM使用需求,闪存缓存承载算力卸载的作用持续放大,异构存储架构落地进一步拔高了高速PCIe6.0闪存的长期市场价值。

紫光闪芯目前已全面启动PCIe 6.0产品研发与落地规划,提前卡位PCIe 6.0技术,帮助国内存储打破海外技术壁垒,在全球AI存储市场建立核心竞争力。

连晓东:您如何判断未来3~5年存储行业的发展趋势?

东猛:我坚定看好人工智能为信息技术产业带来的长期红利,即便2028年海外晶圆厂扩产计划落地见效,届时,人工智能对大容量高速闪存的需求也会翻倍增长,高端AI存储产品的供应依然会保持紧张状态,整个行业将出现明显的分化格局:能够提供高端AI数据中心、边缘计算、智能终端完整存储解决方案,拥有核心技术壁垒、生态护城河和稳定供应链的厂商,将占据市场主导地位,持续享受行业红利;而只做低端同质化产品,没有技术和生态壁垒的企业,将面临巨大的生存压力。

紫光闪芯拥有深厚的高科技产业基因,我们始终聚焦人工智能和AI原生存储,完成从高端到中端的全产品线布局,依托新紫光集团生态加持,稳定供应链保障、持续高强度研发投入,未来3~5年,我们不追求做“大而空”的规模型企业,而是要做强而坚韧的专精特新型企业,让更多人工智能应用在我们的存储产品上跑得更快、更稳、更好,成为全球AI存储领域的核心供应商。

连晓东:请您为中国存储产业的未来发展送上一句寄语。

东猛:希望中国半导体存储产业立足人工智能大发展的时代机遇,扎扎实实做好技术研发与产业落地,让中国存储产品服务中国、走向全球,用科技改变生活,让未来因人工智能而更加美好。