

算力、网力、存力全面自研 阿里亮出“中国芯”新家底

9月22日,2026杭州云栖大会上,阿里巴巴集团CEO吴泳铭在主题演讲中,将AI模型、AI芯片、AI云定义为“机器智能”时代的三大基石,并用一个比喻点出芯片的位置:“如果说Token是AI时代的电,芯片就是发电机。”

同日,这家公司给出了“发电机”清单的最新一页:平头哥发布新一代训推一体AI芯片真武V900,并首次公开倚天服务器CPU的产品路线图。平头哥方面称,其数据中心核心芯片已覆盖AI芯片、服务器CPU、互联、网络和存储等环节。

■新快报记者 郑志辉

单芯片性能提升至三倍

据介绍,真武V900基于自研并行计算架构设计,搭载216GB大容量显存,片间互联带宽达1200GB/s,单芯片性能为上一代真武M890的3倍,原生支持FP8、FP4低精度计算,可覆盖万亿级参数大模型的训练与推理全场景。吴泳铭称,这是目前中国算力性能最强的AI芯片。

按照规划,该芯片将于2027年第一季度量产售卖,并在阿里云数据中心规模化部署。

V900通过平头哥自研的ICN Switch互联芯片实现千卡全带宽高速互联,具备原生内存语义与内存统一编址能力,上千颗芯片可像一颗“超级芯片”协同工作。

与之配套的新一代磐久超节点服务器,集成真武V900、ICN Switch、磐脉智能网卡及镇岳SSD主控芯片,实现算、存、网全栈系统级协同,结合阿里云新一代网络架构,单一集群可扩展至50万卡规模。

上一代产品已经过验证:基于M890的超节点已大规模应用,跑通Qwen3.8、

Kimi K3等超2万亿参数大模型。截至2026年6月,真武系列已服务超过650家企业客户。

吴泳铭表示,由于平头哥芯片产品线的成熟和客户的广泛应用,其AI芯片年出货量将大幅提升。

CPU路线图首次公开

Agent的任务规划、状态管理、工具调用、代码沙箱与多Agent编排高度依赖CPU,海量数据的预处理与吞吐同样离不开它。

平头哥此番首次公布倚天服务器CPU规划:2027年推出倚天720和倚天730两代产品,前者将实现单核性能、核密度与能效的全面提升,后者首次基于全自研CPU微架构设计,单核SPECint2017/GHz性能最高可达倚天710的1.4倍;后续还将推出搭载第二代自研核的倚天750,支持自研ICN片间互联总线协议,使CPU与真武AI芯片直接互联。

加上真武AI芯片、倚天服务器CPU、ICN Switch互联芯片、磐脉智能网卡、镇岳存储主控芯片,平头哥方面称已实现算力、网力、存力的数据中心核心芯片全栈自研。

模型开始反哺芯片设计

芯片之外,阿里首次集中展示“芯片—模型—云”协同飞轮,其中一项是模型参与芯片设计:仅基于一份真实的芯片模块规范,Qwen3.8-Max自主运行超过60小时、调用EDA工具逾万次,完成从功能架构设计、验证到物理



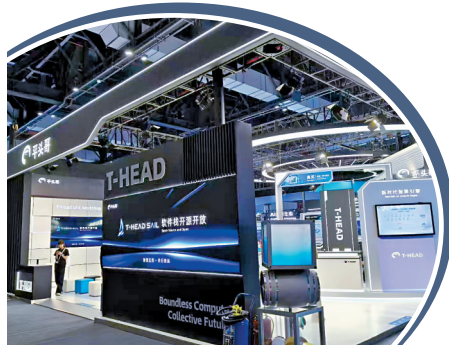
实现的完整流程,将模块物理面积压缩42%。而更好的芯片硬件又将加速模型智能的持续突破,在芯片和模型间形成一个自主迭代的循环。

反向优化同样发生在推理侧。在真武M890上,Qwen3.8-Max“从零适配”全新模型并交付可运行的推理栈,相较首个可用版本,长文本首Token延迟与每输出Token延迟分别下降47%和60%,日常对话场景单实例吞吐提升96%。

来自云上业务的需求,也在反向影响下一代芯片的产品路线。规划中的倚天730面向AI Head Node强化单核性能,倚天720则以192核支持Agentic Rack的大规模并发执行;后续倚天

750还将通过ICN总线与真武AI芯片直接互联,提高协作效率。下一代训推一体AI芯片真武V900,则进一步面向更大模型和更长上下文升级。配合超节点架构与云平台优化,这些硬件能力将为后续模型训练、推理和Agent服务提供支撑。

吴泳铭表示,当前客户AI需求强劲,行业中长期需求远超供给能力,全球AI数据中心相关供应链的紧缺仍在限制算力增长速度。阿里巴巴将与合作伙伴共同投入AI基础设施建设,目标是到2032年,阿里云运营的全球数据中心规模超过20GW,以满足快速增长的AI需求。



从被断供到“掀桌子”:安世中国的12英寸突围

新快报讯 记者郑志辉报道 9月21日午间,东莞黄江镇安世半导体(中国)有限公司厂区内,工人们惬意漫步,享受难得的秋凉。一年前那场全球罕见的“930”风暴,似乎已不足介怀。在他们身后,封测产线上,12英寸晶圆单片有效芯片产出已达8英寸的2.2至2.5倍。一年之间,安世中国经历了从“被断供”到“自主闭环”的极限突围。

一场被逼出来的国产化重构

2025年9月30日,荷兰经济事务与气候政策部对安世半导体全球30个主体实施为期一年的运营冻结。随后阿姆斯特丹企业法庭裁决暂停时任CEO张学政职务、将闻泰科技所持大部分股权交由第三方托管,安世荷兰与安世中国进入事实“分治”状态。

闻泰科技随即依法反击:2026年5月依据反外国制裁法起诉安世相关主体及管理人员,索赔暂计80亿元;8月底,东莞中院裁定冻结对方21.39亿元财产。7月的中荷经贸混委会上,双方同意推动企业协商化解纠纷。

法律应战的同时,安世中国区业务走向独立运营。但东莞封测厂这座占全球出货量约70%的最大单体工厂,一度

因境外晶圆断供面临“无米下锅”。公司最初引入国内头部晶圆厂资源,验证却“不足以达到车规级标准”,被迫转向全面导入12英寸晶圆工艺平台——单片成本显著下降,但涉及产品结构、架构与制造逻辑的系统性重构。

生产端国产化同步推进:部分购自飞利浦时期的设备已迭代至第三代,核心零部件均可国产替代;投资1.24亿元的新产线设备全部国产化,无锡封测基地以国产装备为主,承接原东南亚工厂品类。如今“国产晶圆+东莞封测”的闭环逐步成形。客户端,公司签署多份MOU、完成审厂与资质迁移;由于车规变更周期长达一至两年,部分海外客户仍自供晶圆、委托东莞封测。

代价写在财报里:闻泰科技2026年上半年营收15.14亿元、同比下降94.02%,归母净亏损4.06亿元;晶体管靠前期库存交付,MOSFET则依托东莞产能爬坡。

用12英寸重新定义行业价格

2026年5月底,闻泰科技董事长杨沐宣布安世中国独立运营体系基本搭建完成。8月23日,公司连发三大产品线新品,其中12英寸双极型二极

管、三极管的研发量产,被称全球首家掌握。

今年以来功率半导体行业均价同比上涨15%至20%,安世中国却反向出牌:8月提出主流产品降价10%至20%、部分高价值产品降幅达30%。公司将此比喻为“掀桌子”——底气来自12英寸规模效应、良率提升与供应链本土化三者叠加:综合成本节约10%至30%,开发周期从24至36个月压缩至6至12个月,交付周期从12至16周缩短至4至6周。

闭环之后:从“活下来”到“过好生活”

按最新规划,2026年四季度进入产能爬坡期,明年二季度目标恢复至断供前约七成,三大产品线实现100%自主可控。目前12英寸晶圆良率约93%,达到风险量产标准;上海合作晶圆厂年底若实现月产能5万片,可覆盖原德国工厂8英寸供应量,满足约1000亿颗年度需求。

需求端同样看好:AI服务器对功率器件需求密度约为传统服务器4倍;汽车单车用量正从燃油车约1000颗向智能电动车3000至5000颗迁移。公司正协同开发车规级40V V-BCD、120V高压BCD等工艺平台。



“先活下来,再过好生活。”这家遭遇泛政治化打压的中国半导体龙头,正试图用一块12英寸晶圆,重新定义自己在全球功率半导体格局中的位置。

风暴另一端,安世荷兰9月17日与印度塔塔电子签约,将利用其300毫米晶圆厂及封测产能。业内人士判断,印度新建产能最快要到2028年才能形成有效供给。



更多优质数智资讯
请关注新快数智视频号及新快网数智频道