
# 万卡集群成为大模型训练标配

## ——访摩尔线程创始人兼CEO张建中

本报记者　路轶晨

7月3日，摩尔线程宣布其AI旗舰产品夸娥（KUAE）智算集群解决方案实现重大升级，从当前的千卡级别大幅扩展至万卡规模。摩尔线程夸娥（KUAE）万卡智算集群目标是打造能够承载万卡规模、具备万P级浮点运算能力的国产通用加速计算平台。此外，当天摩尔线程联合中国移动青海公司、中国联通青海公司、北京德道信科集团、中国能源建设股份有限公司总承包公司、桂林华崛大数据科技有限公司分别就三个万卡集群项目进行了战略签约，多方聚力共同构建好用的国产GPU集群。会议期间，《中国电子报》记者就AI大模型发展趋势以及如何突破我国算力瓶颈等问题对摩尔线程创始人兼CEO张建中进行了专访。

### 大模型演进 呼唤高性能算力

关于业界热议的大模型未来走向，张建中认为，有三方面演进趋势值得关注。

一是标度律（Scaling Law）将持续奏效。Scaling Law自2020年被提出以来，已揭示了大模型发展背后的“暴力美学”，即通过算力、算法、数据的深度融合与经验积累，实现模型性能的飞跃，这也成为业界公认的将持续影响未来大模型的发展趋势。Scaling Law将持续奏效，需要单点规模够大并且通用的算力才能快速跟上技术演进。

二是Transformer架构不能实现大一统，会和其他架构持续演进并共存，形成多元化的技术生态。生成式人工智能的进化并非仅依赖于规模的简单膨胀，技术架构的革新同样至关重要。Transformer架构虽然是当前主流，但新兴架构如Mamba、RWKV和RetNet等不断刷新计算效率，加快创新速度。随着技术迭代与演进，Transformer架构并不能实现大一统，从稠密到稀疏模型，再到多模态模型的融合，技术的进步都展现了对更高性能计算资源的渴望。

三是AI、3D和HPC跨技术与跨领域融合不断加速，推动着空间智能、物理AI和AI4Science、世界模型等领域的边界拓展，使得大模型的训练和应用环境更加复杂多元，市场对于能够支持AI+3D、AI+物理仿真、AI+科学计算等多元计算融合发展的通用加速计算平台的需求日益迫切。

### 万卡已是模型训练主战场的标配

“多元趋势下，AI模型训练的主战场，万卡已是标配。”张建中强调，随着计算量不断攀升，大模型训练亟须超级工厂，即一个“大且通用”的加速计算平台，以缩短训练时间，实现模型能力的快速迭代。当前，国际科技大厂都在通过积极部署千卡乃至超万卡规模的计算集群，以确保大模型产品的竞争力。随着模型参数量从千亿迈向万亿，模型能力更加泛化，大模型对底层算力的诉求进一步升级，万卡甚至超万卡集群成为这一轮大模型竞赛的入场券。

然而，构建万卡集群并非一万张GPU卡的简单堆叠，而是一项高度复杂的超级系统工程。它涉及超大规模的组网互联、高效率的集群计算、长期稳定性和高可用性等诸多技术难题。这是难而正确的事情，摩尔线程希望能够建设一个规模超万卡、场景够通用的加速计算平台，并优先解决大模型训练的难题。

### 中国如何突破算力瓶颈？

“当前，我们正处在生成式人工智能的黄金时代，技术交织催动智能涌现，GPU成为加速新技术浪潮来临的创新引擎。”张建中表示，“中国的人工智能落地场景相比国外来说更加广泛，因为中国在人工智能应用领域里面开发者数量很多，落地速度也更快。”

张建中认为，目前国内大模型行业发展面临的主要问题，不是中国公司的研发人员技术水平比国外差，归根结底还是缺少充足的算力。而这个问题不是光靠堆积GPU的数量就能解决的。

“集中力量办大事，打造好用的万卡级别的算力集群，才能让用户真正地使用好大模型。”张建中说道。

张建中强调，在技术层面，中国企业完全有信心有能力去追赶全球顶级GPU企业，做出更先进、性价比更高的芯片。但光有技术还不够，更重要的是生态环境的完善，这需要政府及产业链上下游企业共同努力。

# 数字补偿技术改善射频系统性能

本报记者　许子皓

近日，在2024世界移动通信大会（MWC上海）上，各种通信创新技术不断涌现，5G-A、AI、低空经济等新兴领域的创新产品更是层出不穷，让人目不暇接。通信领域如此高速的发展，对于芯片的要求也越来越高，各大通信芯片企业的创新突破也成为本次展会最受关注的焦点之一。得翼通信首次公开了其在射频领域的突破性技术——射频补偿芯片（RPU），得翼通信创始人兼CEO王子明表示：“我们希望通过数字补偿技术改善射频系统性能，重新定义射频器件的性能标准，满足用户对无感知网络连接的需求。”

当前，AI的爆发式增长可谓是大势所趋，各种AI终端和数据中心之间频繁交互产生了大量的实时数据，给无线通信连接的承载能力带来了很大的挑战。射频器件和射频系统的性能该如何跟上生成式AI等新业态的发展步伐，成为各大厂商必须优先解决的关键问题。

无线连接作为AI时代的电力线，是连接物理世界和数字世界的桥梁，而终端和网络设备的射频系统性能对于无线连接的传输速率和覆盖范围有着显著影响。与数字系统追求极致性能相比，射频系统一直在追求平衡，这是因为射频系统会同时面临着发射功率、信号的线性度和系统效率三者之间两两互斥的矛盾。增大发射功率就会导致信号失真，而追求信号线性不失真又会降低系统效率，增加散热负担。再加上传统的射频器件通过不同的化合物材料和工艺提升性能的方式趋于物理极限，难以在高带宽、高功率和低功耗之间找到平衡点，因此发展速度比较缓慢，即使不考虑成本和代价，也很难拥有逻辑IC遵循的摩尔定律的发展速度。

王子明在接受《中国电子报》记者专访时表示：“传统的射频器件是用化合物来堆积的。比如说一个射频器件能够覆盖的面积不够，可以用两个射频器件组合在一起输出，使用更多的天线，从而获得更大的功率输出，但它的成本是一个指数级增加。如果想提升50%的射频性能，可能需要2倍以上成本才能达到35%的性能提升，因为它已经到达极限，需要更多冗余设计才能让它稳定工作。因此，现在射频器件性能落后用户需求10倍，甚至更多。”

而得翼通信作为新入者，决定不再追随传统的迭代路线，而是另辟蹊径，发布了RPU射频补偿芯片，选择采用数字补偿技术，通过数字预失真算法，对射频器件的输出信号进行实时补偿和优化，从而在不增加额外功率的情况下，显著提升信号覆盖范围和传输能力。这一技术有助于突破传统射频器件的性能瓶颈，并降低设备的综合成本（TCO）。

据王子明介绍，如果仅仅采用数字芯片进行补偿，让射频器件负责输出很高功率，会导致信号失真。但增加一个RPU射频补偿芯片，使射频器件和射频补偿芯片协作，能让信号的输出干净、提升功率并降低功耗，将失真部分补偿回来。“其提升性能的原理类似于我们生活中佩戴的眼镜和降噪耳机，能够校正视力，消除噪音，让我们获得的信息更干净纯粹。该方案不仅在系统架构方面实现创新，还能做到比传统高端元器件高10倍的性能，让用户获得对连接的无感知体验。”

其实，数字预失真算法很早就广泛应用在宏基站领域，是宏基站提供商的通用做法，而得翼通信将这项技术适配到了路由器等消费电子产品中，以应对成本、算力和体积等方面的挑战。据了解，得翼通信最新推出的面向Wi-Fi、无人机和基站的RPU系列产品，不仅适用于Wi-Fi路由器这样的小功率但高调制1024/4096QAM的场景，也适用于100W、400M带宽的宏基站的场景。

据悉，得翼通信第一代产品已向头部客户进行导入测试，预计今年年底将实现量产。未来迭代方向将包括更小型化、更便宜的芯片设计以及支持更多天线和频段的版本。

# 韩国6月芯片出口额创单月纪录

**本报讯**　记者**姬晓婷**报道：记者从韩国产业通商资源部官网了解到，7月1日，韩国产业通商资源部公布2024年上半年及6月进出口数据。数据显示，韩国6月半导体出口创下历史最高值，达到134亿美元，继续成为拉动韩国贸易顺差的重要驱动力。

该公告显示，6月份，韩国15大主力出口产品种类中有6种实现出口增长，其中半导体居于榜首。包括半导体、显示器、电脑、无线通信设备在内的所有类型IT产品出口量连续4个月增加，合计出口额连续8个月增加，这带动了韩国出口的正增长。其中半导体出口额创下历史最高纪录，达到134.2亿美元，同比增长50.9%，实现连续8个月正增长。2024年1—6月份，半导体作为韩国最大出口产品类型，出口额同比增长52.2%，达到657亿美元，实现大幅提升的原因是存储器价格上涨和服务器等产业需求扩大。

从存储器价格上涨情况来看，自2023年第四季度以来，韩国两大存储器领军企业三星、SK海力士存储器价格已连续3个季度上涨，单季涨幅在10%~25%。

从服务器产业需求来看，芯谋研究企业服务部总监王笑龙在接受《中国电子报》记者采访时表示，用于服务器的HBM是拉动出口增长的重要推动力。SK海力士曾在2023年财年报告中表示，公司主力产品DDR5 DRAM和HBM3的收入同比分别增长4倍和5倍以上。在2024年第一季度财报中，SK海力士再次表示，市场对HBM的强劲需求助推了DRAM价格。

从出口地区来看，韩国对美出口规模最大，创下6月份最高纪录110.2亿美元，同比增长14.7%。自去年8月转为正值以来，连续11个月刷新单月最高出口业绩。

面向对美出口需求，韩国财政部7月1日表示，从当天开始，韩元对美元外汇市场交易时间将延长至凌晨2点，这是韩国为提高投资者便利和改善市场准入所做努力的一部分。韩国财政部表示，韩元对美元交易市场的营业时间将从目前的每天上午9点至下午3点30分，改为上午9点至次日凌晨2点。韩元对非美货币的交易时间保持不变。此举旨在提升国内外投资者外币兑换的便利性，降低外汇交易成本。

当前，三星、海力士均在加码HBM生产。摩根大通在研报中表示，截至2023年年底，三星已完成与三大GPU客户的HBM3资格认证，并进入大规模生产阶段。HBM3E 12-Hi的样品测试正在进行中，预计2024年第二季度开始晶圆贴片，第三季度起将大规模增产。摩根大通估计，到2024年年底，三星的HBM月产能将从2023年年底的6万片提升至13万片。

韩国SK海力士母公司SK集团近期表示，到2028年SK海力士将投资103万亿韩元（约合746亿美元），以加强其芯片业务，专注于人工智能。SK集团还表示，计划到2026年确保80万亿韩元的资金，用于投资人工智能和半导体领域，以及为股东回报提供资金，并对超过175家的子公司进行精简。

# 三星电子第二季度利润同比增长1452%

**本报讯**　7月5日，三星电子公布了2024年第二季度业绩指引，预计其第二季度营业利润约为10.4万亿韩元，同比增长1452.2%，环比增长57.3%，远超之前市场预测的8.8万亿韩元；预计营收约为74万亿韩元，同比增长23.3%，环比增长2.89%。完整的三星电子第二季度财报预计将在本月底发布。

今年以来，三星电子迎来大幅增长。2024年第一季度营收利润达到6.61万亿韩元，相比于2023年第四季度增长932.8%。若今年第二季度的数据如三星电子预测，则其营收利润同比增幅将达到惊人的1452.2%。

半导体资深人士李国强认为，三星电子第二季度营收利润涨幅如此夸张的主要原因是去年存储器价格处于低谷。当时三星电子通过控制产能利用率，采取了“保利润、降营收”的经营策略。现在市场需求恢复，三星电子又在今年多次提高了其存储产品的价格，两者共同推动了三星电子在近期营收利润上的连续大幅增长。集邦咨询的数据显示，DRAM芯片价格在今年第二季度增长了13%~18%，NAND Flash芯片价格增长了15%~20%。同时，AI热潮对HBM价格的推动作用明显。

根据三星电子2024年第一季度的财报说明，存储器和移动应用处理器的设备解决方案部门（DS Division）营收达到23.14万亿韩元，同比增长68%，其中存储业务营收17.49万亿韩元，涨幅高达96%。然而，2024年第一季度三星电子整体的营收增幅仅为13%。在营收利润方面，设备解决方案部门同比增长6.50万亿韩元，同样高于三星电子整体5.97万亿韩元的增长。这样的涨幅差额说明存储业务是三星电子近期最大的增长点，并在其他业务有亏损的情况下带动了三星电子整体的增长。多家机构预测，三星电子设备解决方案部门的第二季度营业利润将达到4.6万亿~5.1万亿韩元，相较于第一季度的1.91万亿韩元有141%~167%的增长。

此外，记者在采访中了解到，在手机去库存周期结束后，手机业务也为三星电子带来了不少利润。“三星电子的手机出货量是三星电子营收增长的主要动力之一。”李国强表示。市场分析机构Canalys报告显示，2024年第一季度三星电子手机出货量达到6000万部，重回全球第一位。

7月5日收盘时，三星电子每股87100韩元，涨幅2.96%，该股价为2021年以来新高。　（吴修齐）

# 信越化学将于2028年量产封装基板制造新设备

**本报讯**　记者**许子皓**报道：近日，日本半导体材料供应商信越化学宣布，计划于2028年正式量产其最新研发的封装基板制造设备。据了解，这款设备可以简化“后工序”中把半导体芯片连接到基板的工序，消除对Chiplet（芯粒）中介层的需求，基板制造初期投资将减少一半以上，从而提升封装基板的生产效率和质量。

封装基板是半导体器件中不可或缺的一部分，它不仅为芯片提供了物理支撑，还承担着电信号传输和热量散发的关键作用。随着5G、人工智能、物联网等技术的快速发展，市场对高性能封装基板的需求日益迫切。

据悉，信越化学开发的这款全新的半导体封装基板制造设备，不再使用光刻设备形成布线的传统方法，而是使用激光在基板上蚀刻布线的全新系统。该系统是一种高性能准分子激光加工系统，通过将半导体前段工艺中使用的“双镶嵌”方法应用于后段工艺的封装基板生产，可以将中介层的功能直接集成到封装基板，在消除对中介层需求的同时，还可以实现微纳加工。

中介层是一种高性能的基板，它位于各个Chiplet之间，负责将这些小芯片相互连接，实现高速的数据传输和通信。但随着Chiplet技术的不断发展，中介层的设计和制造也变得越来越复杂，成为半导体制造商需要突破的难题。

信越化学通过“双镶嵌”方法消除了对中介层的需求，从而缩短了先进半导体制造工艺周期，降低了成本。信越化学表示，计划在未来4年内投入约100亿日元用于新设备的研发和生产线建设，预计新设备将在2028年实现商业化生产，并逐步扩大产能以满足市场需求。

社长：张立 社址：北京市海淀区紫竹院路66号赛迪大厦18层 邮编：100048 每周二、五出版 周二8版 周五8版 零售4.50元 全年定价420元 广告部：010-88558848/8808 发行部：010-88558777 广告许可证：京石工商广登字20170003号 发行单位：中国电子报社 印刷单位及地址：经济日报印刷厂 北京市西城区白纸坊东街2号