Arm物理AI事业部执行副总裁Drew Henry：

物理AI计算平台需要专属架构设计

本报记者 姬晓婷

近日，在Arm组织的以物理AI为主题的媒体交流会上，Arm物理AI事业部执行副总裁Drew Henry介绍称：物理AI计算平台与云计算平台所需应对的技术难题截然不同，该平台需要专属的架构设计，而“时延”是理解物理AI最核心的指标。

“时延”是理解物理AI最核心的指标

物理AI是将AI深度嵌入各类智能设备并实现实体化落地的过程，而时延是物理AI最核心的指标。

2025年7月，英伟达创始人兼首席执行官黄仁勋与之江实验室主任、阿里云创始人王坚对话时首次明确提出，“人工智能的下个浪潮是‘物理AI’”。这引发了产业界对物理AI的追捧与讨论。在Drew Henry看来，物理AI是将AI深度嵌入各类智能设备并实现实体化落地的过程。换言之，把AI嵌入执行器（actuators）以及机器人平台、自动驾驶汽车平台等可自主运动的各类设备，就是物理AI。

而“时延”，Drew Henry认为是物理AI最核心的指标。所谓时延，指的是电子系统内部从感知信号到实际动作执行之间的时间。在汽车中，时延可以指代汽车从感知到前方障碍物到执行刹车的时间；在机器人中，时延可以指代观察到目标物体到机械臂运动、机器人移动的时间。在具体的场景中，“从感知信号到执行控制”之间的计算需要在微秒或毫秒级时间内完成。“这是一个与数据中心AI完全不同、复杂度极高的计算挑战。”Drew Henry表示。

在Arm看来，物理AI的实现需要深入理解四大计算层级。

第一个计算层级为感知驱动，聚焦于自主运行（Autonomous Operation）。它关乎感知系统，即赋予机器人或汽车“看见”周围环境的能力，并据此做出快速的实时决策。这一层级的核心要求是在极短时间内完成实时运算。其重要的考察指标，便是从传感器感知信号到执行器启动运作的时延表现。

第二个计算层级是交互驱动层。当乘客乘坐自动驾驶汽车前往目的地时，依然需要和车辆进行交互：乘客可能想查看路线与导航信息、确认行程，也可能因为路途枯燥，想要观看影片。要让车内乘客、或是与人形机器人互动的用户获得流畅体验，交互系统就必须提供相应的算力支撑，因此需要专门设计交互计算层。这一层不需要感知层那样的强实时性，所采用的计算体系也有所区别。

第三个计算层级是驱动执行层。它负责精准控制机器人手中的各类微型执行器，也负责自动驾驶汽车中制动系统与转向系统的控制和执行。这一系统由大量微型器件组成，需要上层系统统一协调调度，这也让整体设计变得极为复杂。

第四个计算层级是云端层，目的主要是实现人形机器人、自动驾驶、机器人系统与云端环境的交互。一方面，用户可以在云端完成新模型训练，再下载到这些终端设备；另一方面，所有设备可以通过云端整合成一个集群，以集群方式协同作业。

此外，Drew Henry提到，这些系统必须做到功能安全、信息安全。

物理AI计算平台与云计算平台截然不同

未来十年，模型与需求都会持续升级，每一代产品都会对性能、能效与成本效益提出更高要求。

具身智能行业与IC企业的技术变革之间，谁将是更强的驱动力？

Drew Henry表示，未来十年，具身智能及其所需的模型势必持续迭代演进，人形机器人平台与自动驾驶平台的应用需求也将不断变化，模型与需求都会一代又一代持续升级，而每一代产品都会对性能、能效与成本效益提出更高要求。随着行业对具身智能技术实现路径的探索不断深入，相关工作负载与模型也将持续优化调整。这是一项需要深耕十年乃至更久的计算领域难题，这也正是该领域有望成为有史以来规模最大的市场之一的关键所在。

在物理AI领域，由于感知驱动型智能体系，即传感器从采集输入数据，到转化为设备的实际执行动作的过程必须在微秒乃至毫秒级完成，这就意味着，我们所设计的系统并非以极致性能和超高内存带宽为核心设计目标。该系统的设计核心，是在极短时间内实现最快速、最高效的指令执行，达成数据输入、动作输出的即时闭环。

而这与面向云端设计的计算平台所应对的技术难题截然不同，二者属于不同的计算平台，物理AI计算平台也因此需要专属的架构设计。这是一类截然不同的计算领域难题，这一难题也将推动未来十年的系统架构迎来变革。

三星电子与AMD扩大AI基础设施内存芯片合作

本报讯 近日，三星电子宣布，已与半导体芯片大厂AMD（超威）在三星平泽厂签署了谅解备忘录（MOU），以扩大在人工智能（AI）基础设施内存芯片供应方面的战略合作。并且，双方还将探讨代工合作的机会，即三星可能为AMD下一代产品提供芯片代工制造服务。

根据该谅解备忘录，三星和AMD将在下一代AMD AI加速器AMD Instinct MI455X GPU的HBM4主要供应方面，以及代号为“Venice”的第六代AMD EPYC CPU的先进DRAM解决方案方面达成合作共识。这些技术将支持结合AMD Instinct GPU、AMD EPYC CPU和机架式架构（例如AMD Helios平台）的下一代AI系统。

三星电子HBM4采用最先进的第六代10纳米（nm）级DRAM工艺（1c）和4nm逻辑基础裸片，处理速度高达每秒13Gbps，最大带宽高达3.3TB/s，超过了行业现有标准。

三星电子副董事长兼首席执行官Young Hyun Jun表示，三星和AMD共同致力于推进人工智能计算的发展，这项协议体现了双方合作范围的不断扩大。从HBM4和下一代内存架构，到尖端的晶圆代工和先进封装技术，三星能够提供无与伦比的一站式解决方案，以支持AMD不断发展的人工智能路线图。

当前，三星和AMD正在紧密合作，共同研发用于人工智能和数据中心工作负载的先进内存技术，此次合作将有助于为客户提供更优化的AI基础设施。

AMD董事长兼首席执行官苏姿丰表示，构建下一代人工智能基础设施需要整个行业的深度合作。从芯片到系统再到机架，贯穿整个计算堆叠的集成对于加速人工智能创新至关重要，而这种创新最终将转化为大规模的实际应用。（文 轩）

佰维存储2025年净利润同比增长429%

本报讯 近日，国内存储芯片厂商佰维存储发布2025年年度报告。报告显示，公司在2025年实现营业收入113.02亿元，同比增长68.82%；归属于上市公司股东的净利润为8.53亿元，同比增长429.07%。

2025年，存储行业呈现显著的“前低后高”反转态势。佰维存储紧紧把握行业上行机遇及AI技术革命带来的增长契机，大力拓展全球头部客户，实现了市场与业务的成长突破，产品销量同比大幅提升。

公司在智能穿戴等AI新兴端侧领域深耕多年，构建了差异化竞争优势。公司ePOP等代表性存储产品具有低功耗、快响应、轻薄小巧等优势，已被Meta、Google、阿里、小米、小天才、Rokid、雷鸟创新等国内外知名企业应用于其AI/AR眼镜、智能手表等智能穿戴设备上。2025年，公司AI新兴端侧存储产品收入约17.51亿元，同比大幅增长。

报告期内，佰维存储凭借高性能存储解决方案，产品已广泛应用于多个主流领域。在智能移动领域，公司产品进入OPPO、vivo、荣耀、传音控股、摩托罗拉等知名客户；在PC领域，SSD产品进入联想、小米、Acer、HP、华硕等国内外知名PC厂商；在企业级（服务器）领域，产品成功进入头部OEM厂商、AI服务器厂商及头部互联网企业的核心供应链；在智能汽车领域，产品已进入20余家国内主流主机厂及核心Tier1供应商的供应链。

2025年，公司研发费用为6.32亿元，同比增长41.34%。在芯片设计领域，公司第一款国产自研主控eMMC（SP1800）已成功量产，已批量交付客户，在智能穿戴、手机应用及车规应用领域均实现出货。（宣 闻）

中科曙光以“算存传一体化”架构赋能智算基建

本报讯 记者**许子皓**报道：近日，中科曙光就此前发布的新一代分布式存储解决方案和全栈自研400G无损高速网络scaleFabric举办了媒体沟通会，中科曙光将存储与计算、网络深度融合，构建“算存传一体化”紧耦合架构，打破传统I/O瓶颈，为智能时代算力基础设施提供高效、绿色的核心支撑。

曙光信息产业（北京）有限公司总裁助理、分布式存储产品部总经理石静告诉记者，单纯拼算力、拼存储的时代已经结束，存算传强协同才是智算基建的核心竞争力。中科曙光提出的存算传一体化紧耦合架构，并非物理上把设备集成在一起，而是物理分离、逻辑上实现一体化强协同，既适配中大型算力中心，又能打通数据全链路。

当前AI大模型训练、智算中心建设等场景对数据吞吐与传输效率提出极高要求，传统存储与计算、网络分离架构易出现网络拥塞、资源竞争等问题，严重制约算力利用率。

中科曙光分布式存储创新性的将存储、计算、网络深度融合，打造协同一体的新型架构，让国内智算实现“算得快、算得高效”。性能层面，中科曙光自研超级隧道技术与国产原生RDMA网络深度结合，构建高效数据通道。硬件上为每个数据域配置独享RDMA网络连接与PCIe通道，基于NUMA亲和性优化资源分配，杜绝通道竞争；软件层面采用数据域感知设计，实现线程调度、内存分配、缓存管理的资源绑定与隔离，彻底解决RDMA网络拥塞、PCIe通道竞争、CPU与内存带宽饱和等核心痛点，保障数据高速稳定流通。

曙光信息产业（北京）有限公司scaleFabric产品经理纵瑞博表示，现在行业正从100G、200G向400G逐步过渡，中科曙光400G scaleFabric刚好处于换代关键节点。他用通俗比喻区分技术差异，原生RDMA像“预约制高铁”，有座位再发车，全程不丢包；RoCE则像开放式高速，数据包先发后控，在“超时重返”机制影响下，极易拥堵丢包。根据有关研究数据，0.1%丢包就会让算力衰减50%。

绿色算力方面，中科曙光以全栈液冷技术构建低碳智算中心。将液冷存储、液冷计算子系统与自研液冷IB交换机深度融合，实现计算、存储、网络全组件液冷覆盖与全链路协同优化，有效缓解超大规模集群散热压力，降低能耗，保障高密度算力长期稳定运行。

2025年第四季度全球前十大晶圆代工企业产值增长2.6%

本报讯 TrendForce集邦咨询发布的最新晶圆代工产业研究报告显示，2025年第四季度先进制程得益于AI Server GPU、Google TPU供不应求，加上智能手机新品驱动手机主芯片投片，出货表现亮眼。成熟制程部分，Server、Edge AI的电源管理订单维持八英寸高产能利用率，甚至酝酿涨价，加上十二英寸产能利用率大致持平，推升该季度全球前十大晶圆代工厂合计产值季增2.6%，达到约463亿美元。

报告显示，2025全年前十大晶圆代工企业合计产值为1695亿美元左右，年增26.3%，创下新高。展望2026年，即便上半年有部分消费性产品提前备货，将稳定产能利用率，不过因存储器价格高涨导致主流终端出货承压、需求下降，下半年订单与产能利用率仍有变数。

2025年第四季度台积电晶圆出货量虽略减，但以iPhone 17为主的手机旗舰新品出货量推升3nm晶圆出货，整体平均销售价格（ASP）提高，季度营收因此增长2%至337亿美元，帮助台积电以70.4%的市占率维持第一。

三星方面，则是因2nm新品出货贡献营收，且自家HBM4使用的logic die晶圆也开始产出，淡化整体产能利用率略降的不利因素，2025年第四季度营收季增6.7%，近34亿美元，不仅正式转亏为盈，市占率也从6.8%微幅升至7.1%，位居第二名。

营收第三名为中芯国际，其2025年第四季度营收季增4.5%，上升至近24.9亿美元，动能来自总晶圆出货增加、ASP略增，以及当年年底的光罩出货增量；联电2025年第四季度八英寸、十二英寸皆有大客户维持稳定订单动能，产能利用率持平前一季，营收季增0.9%，约为20亿美元，维持第四名位置。

第五名的格芯因数据中心周边零部件需求增加而晶圆出货、ASP皆成长，2025年第四季度营收季增8.4%，为18亿美元；第六名为华虹集团，旗下华虹宏力2025年第四季度营收由MCU、PMIC需求驱动，季增3.9%，合并上海华力营收后，华虹集团营收近12.2亿美元，季增0.1%。

2025年第四季度硅光子、硅锗等Server相关利基新型应用出货稳健成长，助力高塔半导体营收季增11.1%、上升至4.4亿美元，市占排名前进至第七名，超越世界先进与合肥晶合。世界先进、合肥晶合分列第八名和第九名；力积电在2025年第四季度因存储器代工需求强劲、ASP提升，且逻辑晶圆代工业务大致平稳，营收季增2%，约3.7亿美元，排名第十。（邢 文）

瑞萨子公司推出全新电子研发协同平台Altium Develop

本报讯 记者**许子皓**报道：近日，瑞萨旗下的电子设计与生命周期管理软件公司Altium宣布，其新一代电子研发协同平台Altium Develop已正式推出，其目的是通过构建统一的协同研发环境，使工程、采购与制造团队能够在产品生命周期早期实现协同，帮助企业用户在从概念设计到制造准备的全流程中实现更加一致和更高效的决策。

随着电子系统复杂度持续攀升，传统研发模式中设计、供应链与制造之间的信息割裂与流程断层正逐渐成为制约效率的重要因素。分散工具与数据孤岛使跨团队协作变得复杂，也难以满足智能制造时代对研发效率和协同能力的要求。

为了解决这一瓶颈，Altium发布了Altium Develop，该平台通过构建统一的协同研发环境，使工程、采购与制造团队能够在产品生命周期早期实现协同，并在跨角色协作中显著提升效率。平台通过在统一的云端环境中连接设计数据、研发流程与协作体系，致力于帮助企业能够在从概念设计到制造准备的全流程中实现更一致、更高效的决策。

据了解，Altium Develop致力于帮助中国企业实现传统许可证合规模式转型，让更多工程团队能够更轻松更加便捷地使用平台、无障碍开展协作协同研发。通过降低平台使用门槛、扩大用户群体参与范围，该平台将帮助更多工程团队在产品开发早期实现更高效的协同合作，同时也为中国制造业持续迈向数字化与智能制造提供重要支撑。

瑞萨电子高级副总裁兼软件与数字化事业部总经理、Altium联合创始人Aram Mirkazemi在接受记者采访时表示，中国生态更需可编程平台而非仅可配置工具，以支撑垂直集成与流程自动化，而硬件敏捷开发的核心是创意与实现、硬件与软件同平台，以统一工程上下文让AI真正发挥价值。针对AI是否会取代工程师或者研发平台，他指出：“AI并非要替代工程师与现有研发平台，而是通过深度融合实现能力增强，以人机协同方式提升研发效率与创新速度。”