

项目榜单

榜单名称	基于可重构数据流架构的高性能AI芯片研发		
行业领域	人工智能	专业方向	智能芯片
(计划)启动时间	2024年4月1日	计划完成时间	2027年3月31日
榜单提出目的	<p>1.重要性</p> <p>首先，本项目将可重构数据流方法从单芯片扩展到多片多机的集群系统，使可重构数据流架构的高芯片利用率特性可以扩展到大规模应用。使用多个芯片支持单个大型模型是未来的趋势，而大多数现有产品只是将现有的网络栈应用于芯片。本项目将针对大型数据中心的数百/千个互连芯片设计协议和拓扑，这是国内首次尝试解决数据中心规模的大模型问题。</p> <p>其次，本项目将创建一个编译工具链生态来支持各种顶层人工智能框架和更多的第三方独立开发者。项目的解决方案不是融入CUDA的现有生态，而是替换CUDA，同时保持现有算法开发人员的编程习惯和开发环境的灵活性。</p> <p>2.必要性</p> <p>面对国际在产品和软件生态方面已经形成的明显竞争优势，只有通过架构创新才能打破当前头部企业对于人工智能算力芯片的垄断态势，并且国际主要AI优势初创企业也走上了相同的道路。跟随国际企业技术路线，不可避免会陷入国外技术垄断的困境中。数据流架构芯片不再跟随国际企业已有技术路线，通过底层架构创新，不依赖先进制程工艺，能够基于国内可量产工艺，打造性能更高、成本更低的算力芯片产品，支撑我国人工智能产业跨过“算力天堑”，实现“算力自由”。</p>		
榜单任务内容	<p>1.拟解决的问题</p> <p>针对工艺受限条件下如何满足数据中心AI应用以及大模型应用快速增长的算力需求的问题，本项目将设计研发一种基于创新性数据流架构的高性能AI计算加速芯片，并基于该加速芯片开发配套的软件工具链，进而设计支持多芯片并行的计算系统，从而提供完整的算力解决方案。</p> <p>2.技术指标</p> <p>项目总体目标是完成基于数据流架构的高性能AI芯片研发，采用12nm工艺，实现128TOPS的INT8算力和64TOPS的FP16算力，芯片利用率峰值达到90%，并基于该芯片开发与之配套的以及面向多种场景的高性能数据流工具链和适应多卡及多机互联的应用系统。</p> <p>3.产业化指标</p> <p>项目实施期最后一年，将开展基于芯片的AI加速产品的销售工作，考虑产品导入周期及产品单价，预计收入不低于2000万元。预计在项目结束后3年内，基于本项目芯片的产品销售额将超过3亿元。申请专利≥50件，其中发明专利≥40件。集成电路布图设计登记证书1件。</p>		

<p>榜单效益目标</p>	<p>随着人工智能在产业数字化进程中从“单点突破”迈向“泛在智能”，一个以数字化、网络化、智能化为特征的智慧社会正加速到来。AI算力作为人工智能的基石，是算力网络构建多要素融合新型信息基础设施的关键领域，已成为数字经济高质量发展的核心引擎。</p> <p>AI芯片已经在智慧电力、智能制造、智慧城市等诸多领域产生了爆发式的应用。AI芯片的需求量也呈快速增长趋势，亿欧智库统计数据显示，预计2023年中国AI芯片市场规模将达到1038.8亿元，2024年中国AI芯片市场规模将达到1405.9亿元，2025年将达到1780亿元，是一个前景巨大且飞速增长的市场。</p> <p>当前国内AI芯片市场超过千亿规模，以芯片3-5年的产品周期和项目单位的客户积累及整体市场情况，预计在项目结束后3年内，基于本项目芯片的产品销售额将超过5亿元。</p> <p>随着大模型的飞速发展，各个产业对于人工智能愈发重视。在当前形势下，广东省可发挥CAISA芯片优势，“盘活一子，带动全产业链”：解决算力瓶颈，吸引龙头企业的战略新产品、新业态、新兴企业落地广东，共同打造基于创新架构的人工智能产业生态，催生新的独角兽企业，并最终建设成为国内人工智能产业生态高地。</p>
---------------	--