

大咖对话 | 陈天石：AI 芯片需要技术和资本的双重密集支撑

2019-03-08 寒武纪创始人兼 CEO 陈天石

技术领导力300讲

[进入课程 >](#)



讲述：刘飞

时长 11:03 大小 10.12M



你好！

本周大咖对话的嘉宾是寒武纪创始人兼 CEO 陈天石博士，近年来，AI 成为科技领域内最受关注的话题，芯片行业也开始高频率地出现在大众的视野之中。而寒武纪作为国产 AI 芯片创业领域的头部独角兽，如何在巨头涌入芯片赛道后继续保持“独立”？我国芯片产业的前景何在？芯片产业当下的挑战和未来的发展趋势又是什么？在采访中，陈天石博士与我们分享了他的观点。

极客时间：陈博士您好，能先简单介绍一下寒武纪目前的芯片产能和自研情况吗？

陈天石：寒武纪拥有终端智能处理器和云端智能芯片两条产品线。

针对终端应用，我们在 2016 年就推出了世界首款终端人工智能专用处理器——寒武纪 1A

处理器（Cambricon-1A），面向智能手机、智能视觉、可穿戴设备、无人机、机器人和智能驾驶等各类终端设备，早于国外同类型产品两年以上。

2017 年开始，我们先后发布了第二代多核智能处理器寒武纪 1H 及第三代高性能智能处理器寒武纪 1M。1A、1H 分别应用于麒麟 970 与麒麟 980 芯片中，迄今已累计服务了千万台智能手机终端。

2018 年 5 月，我们发布了云端人工智能芯片 MLU100，开始从云到端全面的人工智能算力生态布局，目前 MLU100 已进入大规模量产阶段，与浪潮、联想、曙光、H3C 等服务器厂商的适配机型已开始陆续出货。

对于 2019 年，除了完成现有产品线的升级迭代外，面对更复杂的云端智能训练应用、轻量级的边缘端场景和消费类电子产品，我们还将推出一系列新的芯片和 IP 产品。另外，在与民生紧密相关的 AIoT 领域，寒武纪也会做重点赋能。

极客时间：芯片行业并不是个新兴产业，市面上已有如 CPU、GPU 等诸多类型的处理器长期存在。和其他类型芯片相比，智能处理器诞生与发展的必要性是什么？

陈天石：深度学习是目前 AI 领域机器学习方法中最为有效的算法，深度学习模型需要大量的数据训练，这就要求处理器有极高的运算速度作为支撑。

传统 CPU 基于低延时的设计，拥有复杂的内部结构，优点是拥有针对各种不同类型数据的计算能力以及逻辑判断能力，适合进行复杂逻辑的问题处理和计算设备内的管理调度。而深度学习模型所需的运算能力极高，传统架构处理器无法做到功耗与速度上的平衡——通常需要成百上千条指令才能完成对一个神经元的处理。

目前人工智能行业内主流做法是采用 GPU 并行计算神经网络。GPU 是专门进行图像运算工作的处理器，因为它在浮点运算、并行计算等方面的性能优势与人工智能的需求不谋而合，于是在人工智能爆发的关键期，GPU 被广泛使用。但是作为图像处理器，GPU 在推理阶段无法充分发挥并行优势，另外由于 GPU 的计算方式不是为深度学习算法专门设计，性能峰值无法被完全利用。

所以，人工智能行业需要经过专属设计优化的处理器，来应对计算密集型应用场合——不需要面面俱到、八面玲珑，但要有强大的执行力，速度一马当先；同时，也不会带来过高的功耗，这是 AI 专有的速度与激情。

而寒武纪芯片是专门针对深度神经网络计算而设计的 AI 处理器。针对传统 ASIC（将单个特定算法硬件化）思路无法解决深度学习处理需求这个问题，我们通过多年研究，采用硬件神经元虚拟化，稀疏神经网络处理器架构等技术以及深度学习指令集打造了寒武纪的 AI 芯片，使得寒武纪芯片在执行 AI 计算任务时能效可以实现最大程度的优化，和同类产品比速度可以实现大幅提升。

极客时间：现阶段芯片发展到了什么阶段？目前面临的最大挑战点和难点在哪？

陈天石：从人工神经网络的雏形在上世纪四、五十年代被提出开始，人工智能发展几经起伏。近年来，得益于算力、算法、大数据等各个要素上的全面突破和创新，在现阶段，AI 技术应用发展如火如荼。

作为 AI 核心的底层硬件——芯片，也同样经历了漫长的发展过程，从 CPU 到 GPU 再到专门的 AI 芯片，我们一直在试图通过硬件架构创新，去追求计算效率、性能和能效比等性能上的进一步攀升。

目前，随着深度神经网络的层数日益增多，规模日益庞大，AI 芯片发展始终在面对严峻挑战，并需要持续完善以下 3 个方面：

1. 算力资源储备 —— 网络规模变大之后，如何用足够的计算能力来支持如此庞大复杂的深度学习模型；
2. 功耗与成本 —— 如何在超大规模并行计算中更好地兼顾性能与功耗，同时成本可接受；
3. 性能与灵活度 —— 如何能在拥有更适合 AI 计算特点的丰富算力的同时，让算力适用于尽量多的行业与应用场景，满足云端计算中的应用多样化需求。

曾有人开玩笑说，现阶段 AI 的最主要矛盾是 PPT 中体系健全 / 理论完备的全行业解决方案与现实中难落地难实现的实际应用之间的矛盾。作为最底层基础，AI 芯片提供的计算力是连接算法和海量数据的桥梁，是 AI 解决方案得以真正在传统应用中顺利落地的钥匙。

极客时间：在您看来，芯片产业未来的发展趋势将会是什么？

陈天石：AI、大数据、云计算等技术的兴起对处理器提出了不加上限的性能需求，而芯片产业经过 50 余年来的发展，面临着通过架构创新来实现突破性进步的瓶颈。

为了解决海量数据的处理效率问题，无论是在大型数据中心的服务器中还是在芯片内部，异构计算已成为目前越来越普遍的架构模式。CPU 担任调度管理角色，其他多种类型协处理器来负责各种类型的数据处理和计算加速，协处理器的性能在某些程度上还会超越 CPU。

对于 AI 芯片来说，如何通过模型压缩以及设计更合理的数据流等技术手段来实现性能和灵活性的进一步提升、功耗与成本的进一步降低，也是我们持续探索的方向。

另外，在未来时代，物联网和 AI 相辅相成，密不可分。打一个比方：物联网让各类的设备和终端有了生命，而 AI 让它们有了智商。如今物联网已经渗透到我们生活的各个领域，它所具备的万物互联和海量数据的特性也推动了云计算、大数据、AI、边缘计算等技术的发展。即将到来的 5G 时代，也会进一步加快物联网和 AI 等技术的爆破式进步。

物联网连接的物理对象多样且应用场景丰富，将来更需要通过云计算、边缘计算、智能终端计算的协同发展、有机部署来实现万物互联的智能世界，而全部场景中的这些智能数据处理，都需要 AI 芯片参与其中。

极客时间：现在大家都在关注芯片动态，有一种观点说中国芯片产业落后国外几十年，很多时候都是自嗨，您如何看待这一观点？

陈天石：AI 芯片这个领域和其他的芯片领域不太一样的是，中国和其他国家相比不存在太多历史积累上的差距，相反中国还在开始设计研发 AI 芯片的时间上还领先一步。大家从同一起跑线出发，AI 芯片市场前景广阔，人工智能产业发展和应用推进也是迅猛的，希望和国内外同行一道共同努力，推进人工智能领域发展进步。

极客时间：当下越来越多的企业，包括巨头互联网公司都在布局 AI，您如何看待这一现象？接下来寒武纪会有什么差异化的打法吗？

陈天石：基于 AI 技术的突飞猛进和对日常生产生活的深入影响，布局 AI 已经成为一种潮流和趋势，包括现在许多巨头通过研发 AI 芯片来实现自身 AI 生态圈的闭环，形成全栈 AI 平台，其实是希望围绕自身的主营业务形成完整或相对完整的行业解决方案，提前部署万物互联时代的战略，以期待减少对芯片供应商的依赖，保护自身核心技术秘密与知识产权。

但客观来说，芯片从设计、制造到封装是一条几乎没有捷径的漫长链条，尤其是芯片的设计，需要技术和资本双重密集支撑。所以，能对计算基础架构进行布局的 AI 公司不多，能真正打造出高性能、可商用的量产 AI 芯片又是一件更不容易的事。

寒武纪的定位一直以来很明确——做独立的芯片设计公司。为下游厂商提供不同尺寸、面向不同应用场景的终端 AI 处理器 IP，提供从前端训练到后端推理的多品类云端 AI 芯片，帮助大家解决“AI 闭环路上最后一公里”的事情。

技术领导力 300讲

每个技术人都应该知道的管理心经

梁宁 / 著名产品人

张雪峰 / 饿了么CTO

陈皓 左耳朵耗子 / 知名创业者

许式伟 / 七牛云创始人兼CEO

李大学 / 前京东CTO

汤峥嵘 / turtorABC COO

右军 / 蚂蚁金服

程浩 / 迅雷创始人

郭炜 / 易观 CTO

新版升级：点击「👤请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 第183讲 | 薛文植：技术管理的本质—要做尊重人性的管理

下一篇 第184讲 | 狼叔：2019年前端和Node的未来—大前端篇（上）

精选留言

写留言

由作者筛选后的优质留言将会公开显示，欢迎踊跃留言。