中国平安 金融·科技

平安证券 PING AN SECURITIES

计算机

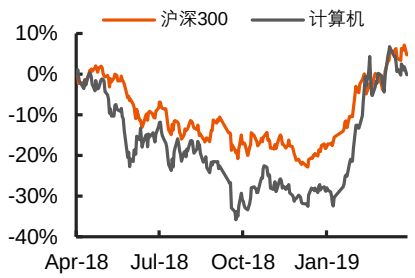
2019年4月24日

人工智能研究系列

AI 芯片市场正快速起航，国内边缘芯片面临更大机遇

强于大市（维持）

行情走势图



证券分析师

闫磊 投资咨询资格编号 S1060517070006
010-56800140
YANLEI511@PINGAN.COM.CN

研究助理

付强 一般从业资格编号 S1060118050035
FUQIANG021@PINGAN.COM.CN

陈苏 一般从业资格编号 S1060117080005
010-56800139
CHENSU109@PINGAN.COM.CN

- **人工智能芯片发展现状及趋势。**当前，在算力、算法和大数据三驾马车的支撑下，全球人工智能进入第三次爆发期。然而，作为引爆点的深度学习算法，对现有的算力尤其是芯片提出了更为苛刻的要求。在 AI 场景中，传统通用 CPU 由于计算效率低难以适应 AI 计算要求，GPU、FPGA 以及 ASIC 等 AI 芯片凭借着自身特点，要么在云端，要么在边缘端，有着优异表现，应用更广。从技术趋势看，短期内 GPU 仍将是 AI 芯片的主导，长期看 GPU、FPGA 以及 ASIC 三大技术路线将呈现并行态势。从市场趋势看，全球 AI 芯片需求将保持较快增长势头，云端、边缘芯片均具备较大增长潜力，预计未来 5 年市场增速将接近 50%；国内虽然芯片技术差距较大，但随着 AI 应用的快速落地，AI 芯片需求增长可能更为迅速。

- **AI 芯片主要应用场景。**一般来讲，只要涉及到 AI 训练和推理的环节，都需要应用 AI 芯片，目前场景主要包括数据中心、自动驾驶、安防、智能家居以及机器人等。数据中心是 AI 训练芯片最主要的客户之一，其中 GPU、ASIC 在该领域均有着大量应用；自动驾驶对算力、时延和可靠性要求近乎苛刻，目前多使用 GPU+FPGA 的解决方案，后续随着算法的稳定，ASIC 有望获得市场空间；安防是视觉芯片最主要的落地场景，未来随着 5G 的部署，云边结合将进一步加快，国内企业在边缘端推理市场机会将增多；智能家居是语音交互芯片重点突破的市场，目前参与企业主要是算法厂商，这些企业将业务延伸到芯片设计，借此提升产品的附加值并降低成本；机器人市场增长较快，其控制系统对 AI 芯片的需求也会增多。

- **国内外芯片企业概览。**近年来，各类势力均在发力 AI 芯片，参与者包括传统芯片设计、IT 厂商、技术公司、互联网以及初创企业等，产品覆盖了 CPU、GPU、FPGA 、ASIC 等。从总体竞争格局看，欧美韩日继续领先该市场，且基本垄断中高端云端芯片，国内企业有进步但主要集中在边缘端，云端差距较为明显。在 AI Chipset Index TOP24 榜单中，前十依然是欧美韩日企业，国内芯片企业如华为海思、联发科、Imagination（2017 年被中国资本收购）、寒武纪、地平线机器人等企业进入该榜单，其中华为海思排 12 位，寒武纪排 23 位，地平线机器人排 24 位。

股票名称	股票代码	股票价格	EPS				P/E				评级
		2019-04-23	2018A	2019E	2020E	2021E	2018A	2019E	2020E	2021E	
科大讯飞	002230	33.91	0.26	0.39	0.65	0.91	130.42	86.95	52.17	37.26	推荐
四维图新	002405	25.00	0.37	0.30	0.37	0.47	67.57	83.33	67.57	53.19	推荐
中科创达	300496	32.35	0.40*	0.56	0.79	-	80.88	57.77	40.95	-	推荐
中科曙光	603019	41.25	0.67	0.93	1.22	1.55	61.57	44.35	33.81	26.61	推荐

注：中科创达年报未发，2018 年 EPS 为预测数据。

- **投资建议：**从当前 AI 芯片市场前景和竞争格局看，我们认为，国内 AI 芯片企业在边缘端的机会多于云端。一方面，在边缘场景下，语音、视觉等领域国内已经形成了一批芯片设计企业队伍，相关芯片产品已经在安防、数据中心推理、智能家居、服务机器人、智能汽车等领域找到落地场景，未来随着 5G、物联网等应用的兴起，相关企业的市场空间将进一步扩大。另一方面，在云端，国内企业也正在加速追赶，未来个别企业有望取得突破。尤其是寒武纪，作为云端芯片重要的技术厂商，有望通过授权等方式，为下游芯片设计、服务器企业赋能。目前，AI 芯片上市公司标的较为稀缺，覆盖标的中，重点推荐**中科曙光、科大讯飞、中科创达以及四维图新**。

- **风险提示：**（1）场景落地不及预期。国内 AI 芯片主要集中在边缘端推理领域，竞争异常激烈，可能影响到企业芯片产品落地效果。（2）技术方向存在不确定性风险。AI 芯片和算法、终端、应用场景等密切相关，但在当前信息技术加速迭代的背景下，AI 技术路线调整的风险非常高，一旦调整，将对企业发展造成重大影响。（3）研发进度不及预期。AI 芯片研发难度大，需要投入大量的资金和人力，国内创业企业资金实力相对薄弱，且在人才竞争中将处于劣势，两方面的因素可能直接导致企业研发进度落后于预期。

正文目录

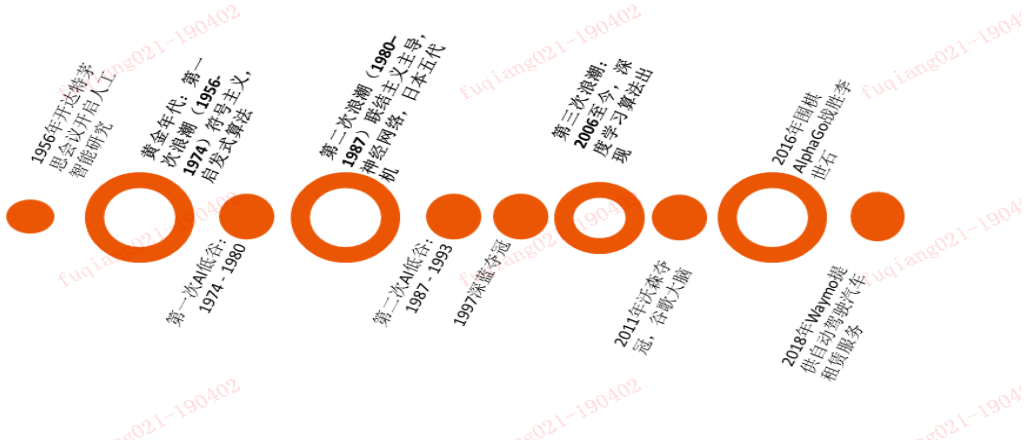
图表目录

一、 人工智能芯片发展现状及趋势

1.1 深度学习算法对芯片要求更为苛刻，通用 CPU 性价比相对较差

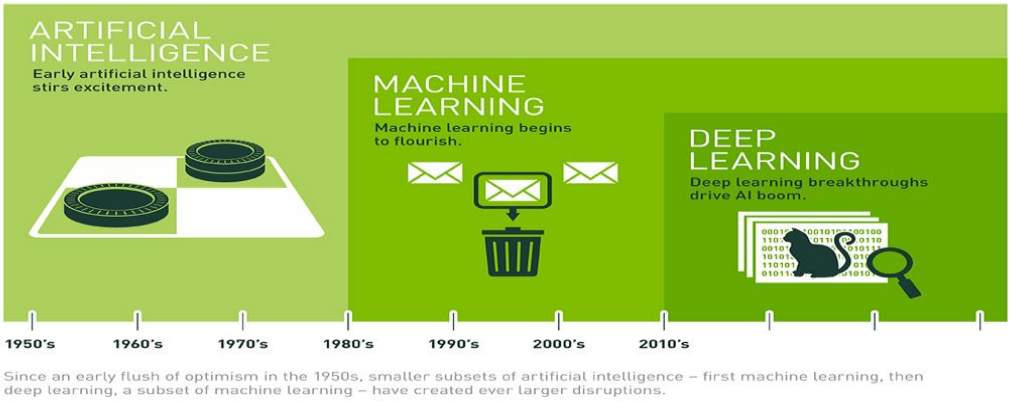
经历了 60 多年的起起伏伏之后，人工智能终于迎来了第三次爆发。第三次爆发的核心引爆点是深度学习算法的出现，但其背后的支撑是数据和算力。对整个 AI 行业来讲，算法、数据和算力三大基本要素中，数据尤其是海量数据的获取和处理难度在下降，算法也在深度学习模型的基础上不断优化，而负责将数据和深度算法统一协调起来的芯片能否获得大的飞跃，成为市场关注的焦点。

图表1 人工智能发展历程



资料来源:wind，平安证券研究所

图表2 深度学习算法发展历程



资料来源:英伟达官网、平安证券研究所

深度学习算法对芯片性能需求主要表现在三个方面：一、海量数据在计算和存储单元之间的高速通信需求。这不但需要芯片具备强大的缓存和片上存储能力，而且还需要计算和存储单元之间有较大的通信带宽。二、专用计算能力需求高。深度学习算法中有大量卷积、残差网络、全连接等特殊计算需要处理，还需要提升运算速度，降低功耗。三、海量数据自身处理同样也对芯片提出了新的要求，尤其是非结构化数据的增多，对传统芯片结构造成了较大的压力。

通用 CPU 在深度学习中可用但效率较低。比如在图像处理领域，主要用到的是 CNN（卷积神经网络），在自然语言识别、语音处理等领域，主要用到的是 RNN（循环神经网络），虽然这两种算法模

型有着较大的区别，但本质上都是向量和矩阵运算，主要是加法和乘法，辅助一些除法和指数运算。传统 CPU 可用于做上述运算，但是 CPU 还有大量的计算逻辑控制单元，这些单元在 AI 计算中是用不上的，造成了 CPU 在 AI 计算中的性价比较低。

图表3　人工智能深度学习算法的算力需求特点

计算类型	算力需求特点
矩阵相乘（Matrix Multiplication）	➢ 应用范围广，几乎所有深度学习算法均需要用到该计算模式； ➢ 运算密集、量大
卷积（Convolution）	常用，主要是进行浮点运算
循环层（Recurrent Layers）	用于深度学习算法的反馈层，主要是矩阵相乘和卷积计算的组合
Allreduce	将多个目标数组减少为单个数组，主要用到加法、比较大小等

资料来源：CSDN、平安证券研究所

1.2 GPU、FPGA 以及 ASIC 各有优劣，成为当前 AI 芯片行业的主流

正因为 CPU 在 AI 计算上的弱点，给了可以实现海量并行计算且能够对进行计算加速的 AI 芯片留下了市场空间。从广义上讲，面向 AI 计算的芯片都可以称为 AI 芯片，包括基于传统架构的 GPU、FPGA 以及 ASIC（专用芯片），也包括正在研究但离商用还有较大差距的类脑芯片、可重构 AI 芯片等。

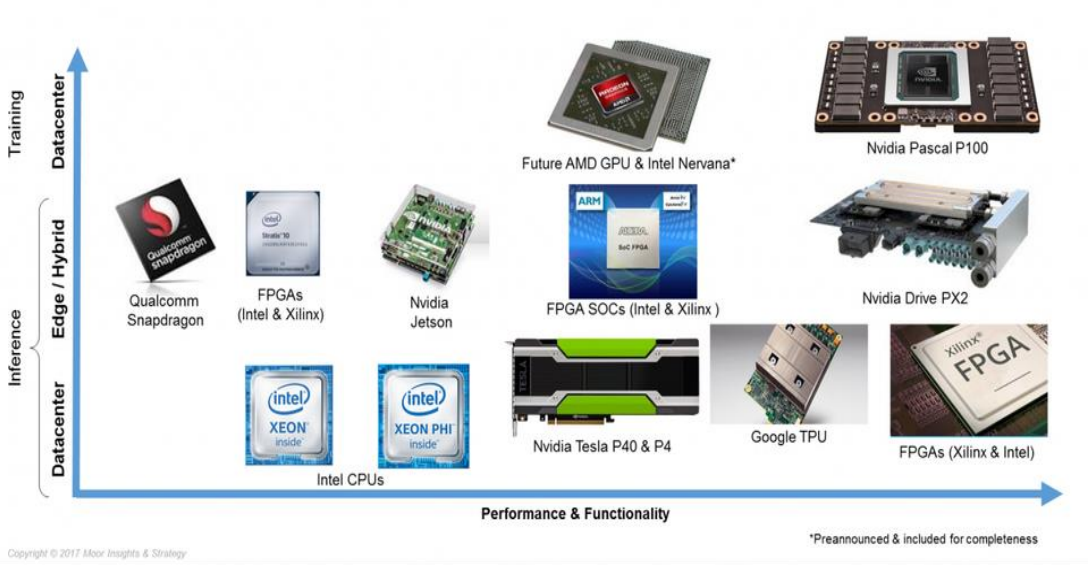
■ **云端训练芯片市场较为集中，而推理市场云、边两端均有大量企业参与**

按照部署位置划分，AI 芯片可以分为云端芯片和边缘端芯片。云端芯片部署位置包括公有云、私有云或者混合云等基础设施，主要用于处理海量数据和大规模计算，而且还要能够支持语音、图片、视频等非结构化应用的计算和传输，一般情况下都是用多个处理器并行完成相关任务；边缘端 AI 芯片主要应用于嵌入式、移动终端等领域，如摄像头、智能手机、边缘服务器、工控设备等，此类芯片一般体积小、耗电低，性能要求略低，一般只需具备一两种 AI 能力。

按照承担的任务分，AI 芯片可以划分为训练芯片和推理芯片。训练是指通过大量标记过的数据在平台上进行“学习”，并形成具备特定功能的神经网络模型；推理则是利用已经训练好的模型输入新数据通过计算得到各种结论。训练芯片对算力、精度要求非常之高，而且还需要具备一定的通用性，以适应多种算法的训练；推理芯片更加注重综合能力，包括算力能耗、时延、成本等因素。

综合来看，训练芯片由于对算力的特殊要求，只适合在云端部署，而且多采用的是“CPU+加速芯片”类似的异构模式，加速芯片可以是 GPU，也可以是 FPGA 或者是 ASIC 专用芯片。AI 训练芯片市场集中度高，英伟达和谷歌领先，英特尔和 AMD 正在积极切入。推理在云端和终端都可进行，市场门槛相对较低，市场参与者较多。云端推理芯片除了传统的英伟达、谷歌、赛灵思等芯片大厂外，Groq 等国际新兴力量也在加入竞争，国内寒武纪、比特大陆也有不错表现；终端推理芯片市场较为分散，场景各异，参与者除了英伟达、英特尔、ARM 和高通之外，国内企业如寒武纪、地平线、云知声、云天励飞等在各自细分领域均有所建树。

图表4　AI 深度学习芯片产品图谱



资料来源: Moor Insights & Strategy、平安证券研究所

图表5　GPU、FPGA、ASIC 性能特点对比

指标	GPU	FPGA	ASIC
定制化程度	通用	半定制化	全定制化
灵活性	高	高	低
成本	高	较高	低
功耗	高	较高	低
主要优点	计算能力强、产品成熟	平均性能较高、功耗低、灵活性强	平均性能强、功耗低、体积小
主要缺点	效率不高、编程难度大	峰值计算能力较弱、编程语言难度大	不可编辑、研发时间长、技术风险较高
主要计算场景	云端训练和推理	云端和终端推理	云端训练和推理，终端推理

资料来源：CSDN、平安证券研究所

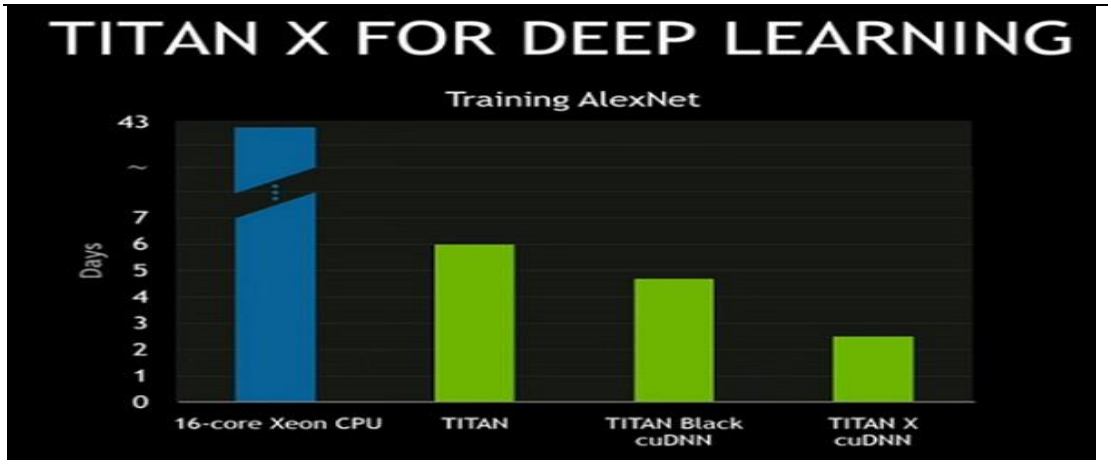
■　**GPU 擅长云端训练，但需与 CPU 异构、功耗高且推理效率一般**

GPU（Graphics Processing Unit）是一种由大量核心组成的大规模并行计算架构，专为同时处理多重任务而设计的芯片。正是由于其具备良好的矩阵计算能力和并行计算优势，最早被用于 AI 计算，并在云端获得大量应用。GPU 中超过 80%部分为运算单元（ALU），而 CPU 仅有 20%，因此 GPU 更擅长于大规模并行运算。以英伟达的 GPU TITAN X 为例，该产品在深度学习中所需训练时间只有 CPU 的 1/10 不到。但 GPU 用于云端训练也有短板，GPU 需要同 CPU 进行异构，通过 CPU 调用才能工作，而且本身功耗非常高。同时，GPU 在推理方面需要对单项输入进行处理时，并行计算的优势未必能够得到很好的发挥，会出现较多的资源浪费。

图表6　CPU 与 GPU 架构对比

资料来源: Elecfans、平安证券研究所

图表7　主流 GPU 处理器与 CPU 在深度学习训练时长比较



资料来源:英伟达官网、平安证券研究所

■　**FPGA 芯片算力强、灵活度高，但技术难度大国内差距较为明显**

FPGA（Field-Programmable Gate Array）即现场可编程门阵列，该芯片集成了大量的基本门电路以及存储器，其灵活性介于 CPU、GPU 等通用处理器和专用集成电路 ASIC 之间，在硬件固定之前，允许使用者灵活使用软件进行编程。FPGA 在出厂时是“万能芯片”，用户可根据自身需求，用硬件描述语言对 FPGA 的硬件电路进行设计；每完成一次烧录，FPGA 内部的硬件电路就有了确定的连接方式，具有了一定的功能；输入的数据只需要依次经过各个门电路，就可以得到输出结果。

FPGA 应用于 AI 有以下优势：

（1）算力强劲。由于 FPGA 可以同时进行数据并行和任务并行计算，在处理特定应用时效果更加明显，对于某一个特定的运算，FPGA 可以通过编辑重组电路，生成专用电路，大幅压缩计算周期。从赛灵思推出的 FPGA 产品看，其吞吐量和时延指标都好于 CPU 和 GPU 产品。

（2）功耗优势明显。FPGA 能耗比是 CPU 的 10 倍以上、GPU 的 3 倍。由于在 FPGA 中没有取指令与指令译码操作，没有这部分功耗；而在复杂指令集（X86）的 CPU 中仅仅译码就占整个芯片能耗的约 50%，在 GPU 里取指与译码也会消耗 10%至 20%的能耗。

（3）灵活性好。使用通用处理器或 ASIC 难以实现的下层硬件控制操作技术，利用 FPGA 可以很方便的实现，从而为算法的功能实现和优化留出了更大空间。

（4）成本相对 ASIC 具备一定优势。FPGA 一次性成本（光刻掩模制作成本）远低于 ASIC，在芯片需求还未成规模、深度学习算法暂未稳定需要不断迭代改进的情况下，利用具备可重构特性的 FPGA 芯片来实现半定制的人工智能芯片是最佳选择。

正因为存在上述优势，FPGA 被广泛用于 AI 云端和终端的推理。国外包括亚马逊、微软都推出了基于 FPGA 的云计算服务，而国内包括腾讯云、阿里云均在 2017 年推出了基于 FPGA 的服务，百度大脑也使用了 FPGA 芯片。

从市场格局上看，全球 FPGA 长期被 Xilinx（赛灵思）、Intel（英特尔）、Lattice（莱迪思）、Microsemi（美高森美）四大巨头垄断。其中，赛灵思和英特尔合计占到市场的 90%左右，赛灵思的市场份额超过 50%，国内厂商刚刚起步，差距较大。

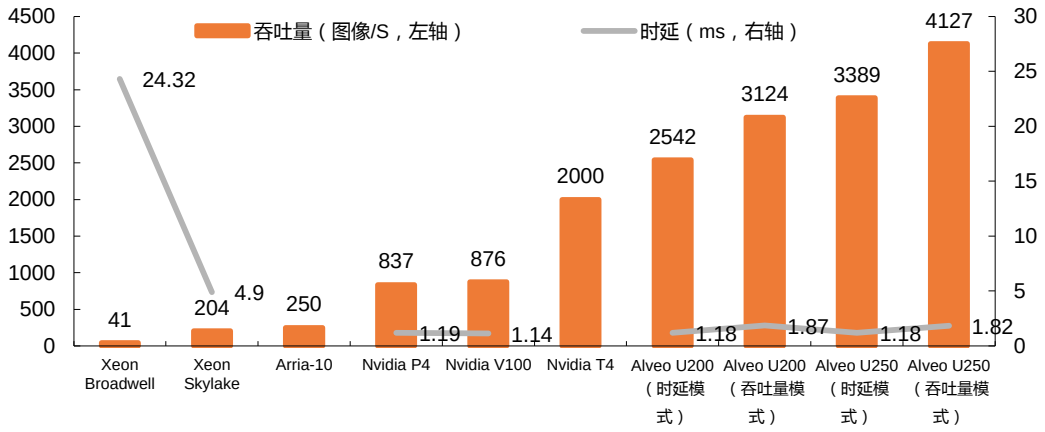
图表8　Alveo U200 数据中心加速器卡

资料来源:赛灵思官网、平安证券研究所

图表9　英特尔® Agilex™ F 系列 FPGA

资料来源:英特尔官网、平安证券研究所

图表10　赛灵思 FPGA 产品图像吞吐量和时延指标表现优异



资料来源:赛灵思官网、平安证券研究所

■　**专用芯片（ASIC）深度学习算法加速应用增多，可提供更高能效表现和计算效率**

ASIC（Application Specific Integrated Circuits），即专用芯片，是一种为特定目的、面向特定用户需求设计的定制芯片，具备性能更强、体积小、功耗低、可靠性更高等优点。在大规模量产的情况下，还具备成本低的特点。

ASIC 与 GPU、FPGA 不同，GPU、FPGA 除了是一种技术路线之外，还是实实在在的确定产品，而 ASIC 只是一种技术路线或者方案，其呈现出的最终形态与功能也是多种多样的。近年来，越来越多的公司开始采用 ASIC 芯片进行深度学习算法加速，其中表现最为突出的 ASIC 就是 Google 的 TPU（张量处理芯片）。

TPU 是谷歌为提升 AI 计算能力同时大幅降低功耗而专门设计的芯片。该芯片正式发布于 2016 年 5 月。TPU 之所以称为 AI 专用芯片，是因为它是专门针对 TensorFlow 等机器学习平台而打造，该芯片可以在相同时间内处理更复杂、更强大的机器学习模型。谷歌通过数据中心测试显示，TPU 平均比当时的 GPU 或 CPU 快 15-30 倍，性能功耗比（TFOPS/Watt）高出约 30-80 倍。

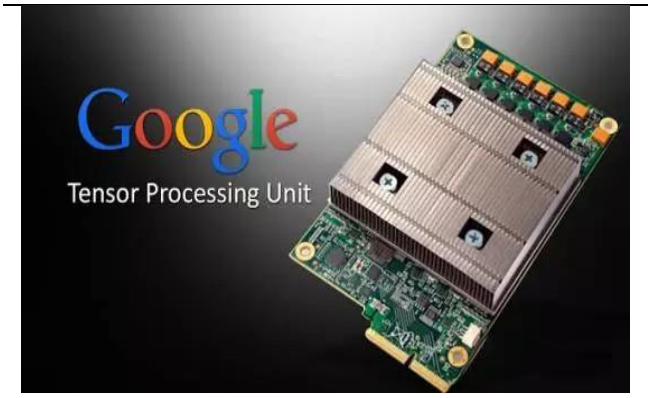
但是，ASIC 一旦制造完成以后就不能修改了，且研发周期较长、商业应用风险较大，目前只有大企业或背靠大企业的团队愿意投入到它的完整开发中。国外主要是谷歌在主导，国内企业寒武纪开发的 Cambricon 系列处理器也广泛受到关注。其中，华为海思的麒麟 980 处理器所搭载的 NPU 就是寒武纪的处理器 IP。

图表11　国内外主要企业专用芯片进展情况

公司	芯片名称	简介
谷歌	TPU	面向机器学习的张量处理加速芯片
IBM	TrueNorth	以分布、并行方式存储和处理信息的芯片，支持 SNN（脉冲神经网络）
高通	Zeroth	按照人类神经网络传输信息的方式而设计的芯片，支持 SNN
英特尔	神经网络芯片	支持片上学习的 SNN 芯片
中星微	-	开发出中国首个嵌入式神经网络芯片 NPU
寒武纪	-	全球首个提出深度学习处理器芯片指令集
云知声	UniOne	面向物联网的全栈芯片
阿里达摩院	Ali-NPU	用于图像视频分析和机器学习等推理计算
地平线机器人	-	专注于人工智能本地化机器学习芯片
灵汐科技	-	类脑处理芯片，支持 DNN/SNN 混合模式

资料来源：清华大学、中国工程院、平安证券研究所

图表12　谷歌 TPU 芯片



资料来源:谷歌官网、平安证券研究所

图表13　寒武纪 MLU100 智能芯片

资料来源:寒武纪科技官网、平安证券研究所

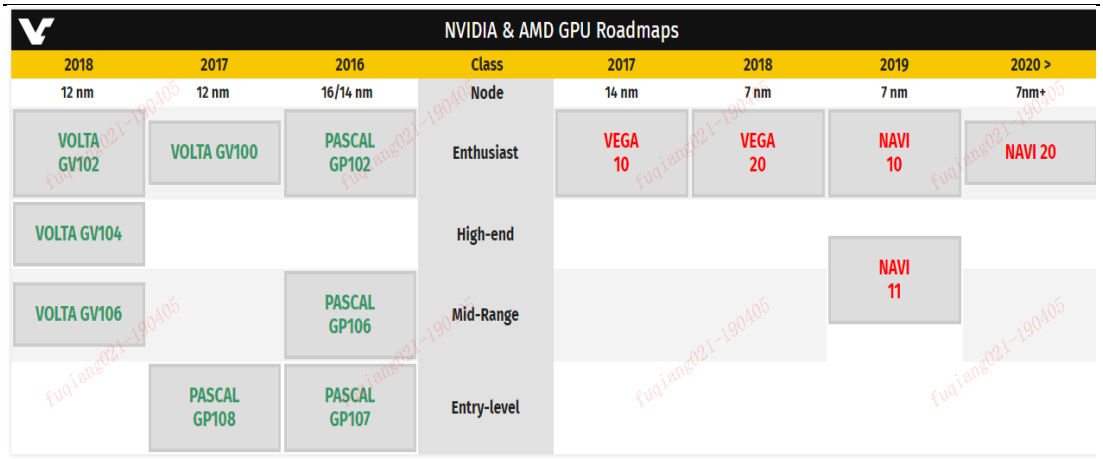
1.3 短期内 GPU 仍将是 AI 芯片主导，长期看三大技术路线将呈现并行态势

■ **短期内 GPU 仍将主导 AI 芯片市场，FPGA 的使用将更为广泛**

GPU 短期将延续 AI 芯片的领导地位。GPU 作为市场上 AI 计算最成熟、应用最广泛的通用型芯片，应用潜力较大。凭借其强大的计算能力、较高的通用性，GPU 将继续占领 AI 芯片的主要市场份额。

当前，两大 GPU 厂商都还在不断升级架构并推出新品，深度学习性能提升明显，未来应用的场景将更为丰富。英伟达凭借着其在矩阵运算上的优势，率先推出了专为深度学习优化的 Pascal GPU，而且针对 GPU 在深度学习上的短板，2018 年推出了 Volta 架构，正在完成加速-运算-AI 构建的闭环；AMD 针对深度学习，2018 年推出 Radeon Instinct 系列，未来将应用于数据中心、超算等 AI 基础设施上。我们预计，在效率和场景应用要求大幅提升之前，作为数据中心和大型计算力支撑的主力军，GPU 仍具有很大的优势。

图表14　英伟达及 AMD GPU 架构演进路线图



资料来源: videocardz、平安证券研究所

FPGA 是短期内 AI 芯片市场上的重要增长点，FPGA 的最大优势在于可编程带来的配置灵活性，在当前技术与运用都在快速更迭的时期，FPGA 具有明显的实用性。企业通过 FPGA 可以有效降低研发调试成本，提高市场响应能力，推出差异化产品。在专业芯片发展得足够完善之前，FPGA 是最好的过渡产品，正因为如此，科技巨头纷纷布局云计算+FPGA 的平台。随着 FPGA 的开发者生态逐渐丰富，适用的编程语言增加，FPGA 运用会更加广泛。因此短期内，FPGA 作为兼顾效率和灵活性的硬件选择仍将是热点所在。

■ **长期来看 GPU、FPGA 以及 ASIC 三大类技术路线将并存**

GPU 主要方向是高级复杂算法和通用型人工智能平台。(1) 高端复杂算法实现方向。由于 GPU 本身就具备高性能计算优势，同时对于指令的逻辑控制上可以做的更复杂，在面向复杂 AI 计算的应用方面具有较大优势。(2) 通用型的人工智能平台方向。GPU 由于通用性强，性能较高，可以应用于大型人工智能平台够高效地完成不同种类的调用需求。

FPGA 未来在垂直行业有着较大的空间。由于在灵活性方面的优势，FPGA 对于部分市场变化迅速的行业最为实用。同时，FPGA 的高端器件中也可以逐渐增加 DSP、ARM 核等高级模块，以实现较为复杂的算法。随着 FPGA 应用生态的逐步成熟，FPGA 的优势也会逐渐为更多用户所认可，并得以广泛应用。

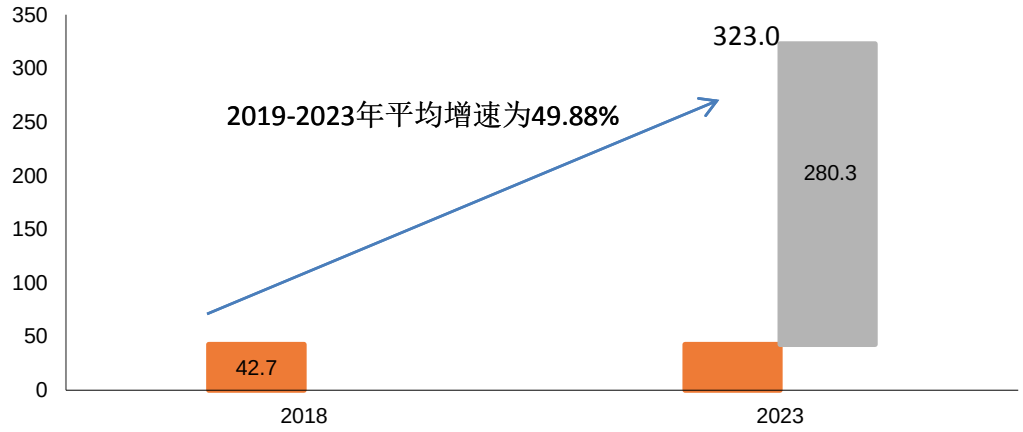
ASIC 长远来看非常适用于人工智能，尤其是应对未来爆发的面向应用场景的定制化芯片需求。ASIC 的潜力体现在，AI 算法厂商有望通过算法嵌入切入该领域，以进入如安防、智能驾驶等场景。由于

其具备高性能低消耗的特点，可以基于多个人工智能算法进行定制，以应对不同的场景，未来在训练和推理市场上都有较大空间。

1.4 国内外 AI 芯片市场需求将保持较快增长势头，云端、边缘均具备潜力

近年来，伴随着全球 AI 产业的快速增长，AI 芯片需求大幅上升。按照 Gartner 最新数据，2018 年全球 AI 芯片市场规模达到 42.7 亿美元。未来几年，全球各大芯片企业、互联网巨头、初创企业都将在该市场上进行角逐，预计到 2023 年全球市场规模将达到 323 亿美元。未来五年（2019-2023 年）平均增速约为 50%，其中数据中心、个人终端、物联网芯片均是增长的重点。

图表15　2018-2023 年全球 AI 芯片市场规模（亿美元）



资料来源:Gartner、平安证券研究所

国内人工智能芯片行业发展仍处在起步阶段。长期以来，我国在 CPU、GPU 和 DSP 设计上一直处于追赶状态，绝大多数芯片依靠国外的 IP 核进行设计，自主创新能力不足。但我们也看到，国内人工智能产业的快速发展，也为国内芯片产业实现换道超车创造了机会。由于国内外在芯片生态上并未形成垄断，国内芯片设计厂商尤其是专用芯片设计厂商，同国外竞争对手还处在同一起跑线上。

目前国内人工智能芯片市场呈现出百花齐放的态势。AI 芯片的应用领域广泛分布在金融证券、商品推荐、安防、消费机器人、智能驾驶、智能家居等众多领域，催生了大量的人工智能创业企业，如地平线、深鉴科技、寒武纪、云知声、云天励飞等。我们认为，未来随着国内人工智能市场的快速发展，生态建设的完善，国内 AI 芯片企业将有着更大的发展空间，未来 5 年的市场规模增速将超过全球平均水平。

二、 AI 芯片主要应用场景

2.1 数据中心（云端）

数据中心是 AI 训练芯片应用的最主要场景，主要涉及芯片是 GPU 和专用芯片（ASIC）。如前所述，GPU 在云端训练过程中得到广泛应用。目前，全球主流的硬件平台都在使用英伟达的 GPU 进行加速，AMD 也在积极参与。亚马逊网络服务 AWS EC2、Google Cloud Engine（GCE）、IBM Softlayer、Hetzner、Paperspace 、LeaderGPU、阿里云、平安云等计算平台都使用了英伟达的 GPU 产品提供深度学习算法训练服务。

在云端推理市场上，由于芯片更加贴近应用，市场更多关注的是响应时间，需求也更加的细分。除了主流的 CPU+GPU 异构之外，还可通过 CPU+FPGA/ASIC 进行异构。目前，英伟达在该市场依然保持着领军位置，但是 FPGA 的低延迟、低功耗、可编程性优势（适用于传感器数据预处理工作以及小型开发试错升级迭代阶段）和 ASIC 的特定优化和效能优势（适用于在确定性执行模型）也正在凸显，赛灵思、谷歌、Wave Computing、Groq、寒武纪、比特大陆等企业市场空间也在扩大。

图表16　全球人工智能硬件平台 AI 芯片配置情况

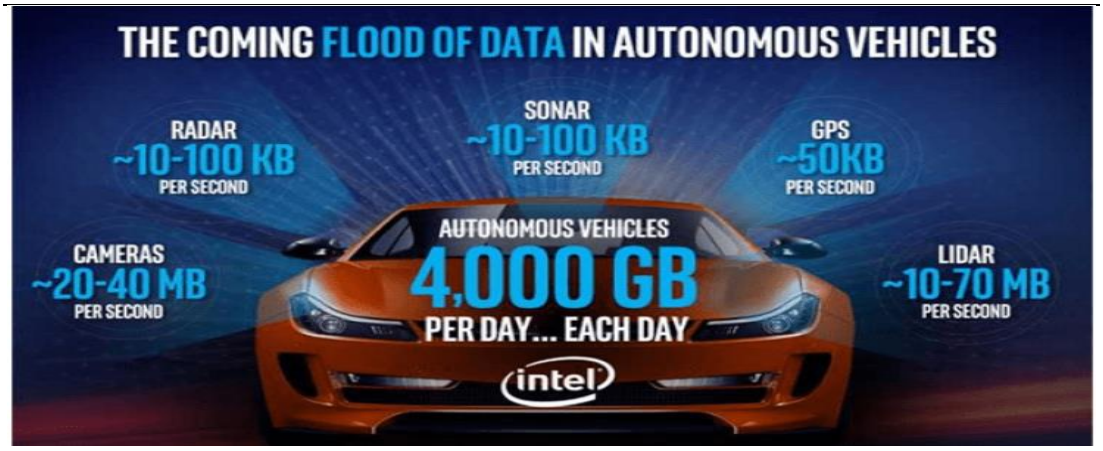
硬件平台	档次	GPU 产品型号
亚马逊 AWS	低端	英伟达特斯拉 K80
	高端	英伟达特斯拉 V100
Google Compute Engine	低端	英伟达特斯拉 K80
	高端	英伟达特斯拉 V100
PaperSpace	低端	英伟达 Quadro P6000
	高端	英伟达特斯拉 V100
IBM softlayer	低端	2*英伟达特斯拉 M60
	高端	4*英伟达特斯拉 K80
LeaderGPU	低端	2*英伟达 GTX 1080
	高端	2*英伟达特斯拉 P100
腾讯云	-	N*英伟达特斯拉 V100
平安云	-	N*英伟达特斯拉 V100/P100
百度云	-	N*英伟达特斯拉 V100
阿里云	-	AMD S7150/英伟达特斯拉 V100

资料来源：公司网站，平安证券研究所

2.2 自动驾驶

自动驾驶汽车装备了大量的传感器、摄像头、雷达、激光雷达等车辆自主运行需要的部件，每秒都会产生大量的数据，对芯片算力有很高的要求， 但受限于时延及可靠性，有关车辆控制的计算不能再依托云端进行，高算力、快速响应的车辆端人工智能推理芯片必不可少。

图表17　自动驾驶数据量预测



资料来源:英特尔、平安证券研究所

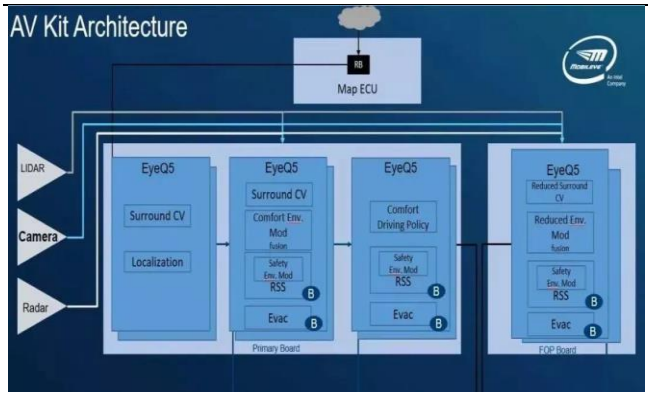
目前，自动驾驶所使用的芯片主要基于 GPU、FPGA 和 ASIC 三条技术路线。但由于自动驾驶算法仍在快速更迭和进化，因此大多自动驾驶芯片使用 GPU+FPGA 的解决方案。未来算法稳定后，ASIC 将成为主流。按照 SAE International 的自动驾驶等级标准，目前已商用的自动驾驶芯片基本处于高级驾驶辅助系统（ADAS）阶段，可实现 L1-L2 等级的辅助驾驶和半自动驾驶（部分宣称可实现 L3 的功能）；而面向 L4-L5 超高度自动驾驶及全自动驾驶的 AI 芯片离规模化商用仍有距离。

根据丰田公司的统计数据，实现 L5 级完全自动驾驶，至少需要 12TOPS 的推理算力，按照 Nvidia PX2 自动驾驶平台测算，差不多需要 15 块 PX2 车载计算机，才能满足完全自动驾驶的需求。AI 芯片用于自动驾驶之后，对传统的汽车电子市场冲击较大，传统的汽车电子巨头（恩智浦、英飞凌、意法半导体、瑞萨）虽然在自动驾驶芯片市场有所斩获，但风头远不及英特尔、英伟达、高通甚至是特斯拉。国内初创企业如地平线、眼擎科技、寒武纪也都在积极参与。在自动驾驶芯片领域进展最快以及竞争力最强的是英特尔和英伟达，英特尔强在能耗，英伟达则在算力和算法平台方面优势明显。

英特尔进入自动驾驶芯片市场虽然较晚，但通过一系列大手笔收购确立了其在自动驾驶市场上的龙头地位。2016 年，公司出资 167 亿美元收购了 FPGA 龙头 Altera；2017 年 3 月以 153 亿美元天价收购以色列 ADAS 公司 Mobileye，该公司凭借着 EyeQ 系列芯片占据了全球 ADAS 70%左右的市场，为英特尔切入自动驾驶市场创造了条件。收购完成之后，英特尔形成了完整的自动驾驶云到端的算力方案——英特尔凌动/至强+Mobileye EyeQ+Altera FPGA。英特尔收购 Mobileye 之后，后者也直接推出了 EyeQ5，支持 L4-L5 自动驾驶，预计在 2020 年量产。

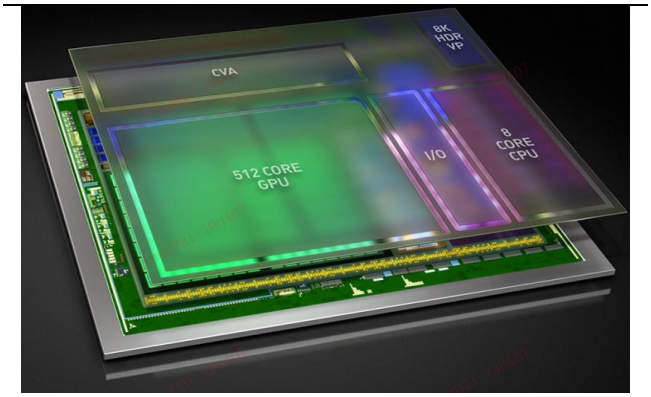
英伟达在汽车 AI 芯片的竞争中不落下风。英伟达在 2015 年推出了世界首款车载超级计算机 Drive PX，紧接着 2016 年推出 Drive PX2，2018 年推出新一代超级计算机 Drive Xavier，同年，基于双 Drive Xavier 芯片针对自动驾驶出租车业务的 Drive PX Pegasu 计算平台面世。2019 CES 上，英伟达推出了全球首款商用 L2+自动驾驶系统 NVIDIA DRIVE AutoPilot。DRIVE AutoPilot 的核心是 Xavier 系统级芯片。该芯片处理器算力高达每秒 30 万亿次，已经投产。

图表18　英特尔自动驾驶完整硬件系统



资料来源:电子工程世界、平安证券研究所

图表19　英伟达 Xavier 芯片



资料来源:电子工程世界、平安证券研究所

图表20　英特尔和英伟达主要自动驾驶芯片性能指标对比

指标	英特尔		英伟达	
AI 芯片或者平台	EyeQ4 芯片	EyeQ5 芯片	Xavier 芯片	Pegasus 平台
量产时间	2018	2020	2018	2018
自动驾驶级别	L3	L4-5	L3-4	L5
配置	MIPS i-class 核心处理器*4；MIPS m-class*1	CPU*2；多线程 CPU*8；下一代视觉处理器*18	Xavier*1；512 GPU*1；CPU*1	Xavier 核心芯片*2；下一代 GPU*2
算力	2.5Tops	24Tops	30Tops	320Tops
能耗	3W	10W	30W	500W
制程	28nm	7nm	12nm	12nm
合作对象	通用、奥迪、宝马、本田、菲亚特、上汽、蔚来、威马等	-	百度、采埃孚等	博世、戴姆勒等

资料来源：搜狐汽车、平安证券研究所

2.3 安防

安防市场是全球及国内 AI 最为确定以及最大的市场，尤其是 AI 中的图像识别和视频处理技术正在全面影响安防产业。其中，在安防产品中，摄像头、交换机、IPC（网络摄像机）、硬盘刻录机、各类服务器等设备都需要芯片，这些芯片也决定了整个安防系统的整体功能、技术指标、能耗以及成本。在安防芯片中，最为关注的还是四类与监控相关的芯片（ISP 芯片、DVR SoC 芯片、IPC SoC 芯片、NVR SoC 芯片）。

ISP 芯片(Image Signal Processing，图像信号处理)主要负责对前端摄像头所采集的原始图像信号进行处理；DVR（DigitalVideoRecorder，数字硬盘录像机）SoC 芯片主要用于模拟音视频的数字化、编码压缩与存储；IPC （IP Camera，IP 摄像机）SoC 芯片通常集成了嵌入式处理器（CPU）、图像信号处理（ISP）模块、视音频编码模块、网络接口模块等，具备入侵探测、人数统计、车辆逆行、丢包检测等一些简单的视频分析功能；NVR (Network Video Recorder，网络硬盘录像机) SoC 芯片主要用于视频数据的分析与存储，功能相对单一，但由于多与 IPC 联合使用，市场增长也较快。

通常情况下，安防视频监控模拟摄像机的核心部件包括一颗图像传感器和一颗 ISP 芯片，安防视频监控网络摄像机的核心部件包括一颗图像传感器和一颗 IPC SoC 芯片。单从国内来看，未来国内视频监控行业增速仍将保持 12%-15%左右的水平增长，其中网络监控设备增长更为迅速，相关芯片产品需求十分旺盛。

安防 AI 芯片市场上，除了传统芯片以及安防厂商，还有大量的创业企业在涌入。国外芯片厂商主要有英伟达、英特尔、安霸、TI、索尼、特威、三星、谷歌等；国内厂商主要有海思（华为）、国科微、中星微、北京君正、富瀚微、景嘉微、寒武纪、深鉴科技、云天励飞、中科曙光等。英伟达、英特尔等企业凭借着通用处理器以及物联网解决方案的优势，长期与安防巨头如海康、大华、博世等保持紧密联系；国内寒武纪、地平线、云天励飞等企业，都有 AI 芯片产品面世，海思本身就有安防摄像机 SoC 芯片，在新加入 AI 模块之后，竞争力进一步提升。

从安防行业发展的趋势来看，随着 5G 和物联网的快速落地，“云边结合”将是行业最大的趋势，云端芯片国内企业预计很难有所突破，但是边缘侧尤其是视频处理相关 AI 芯片还是有较大潜力，国产化替代将加速。但也看到，AI 芯片离在安防领域实现大规模快速落地仍有距离。除了功耗和算力约束外，工程化难度大也是困扰行业的重要因素，尤其是在安防这种产业链长而高度碎片化的产业，新技术落地需要长时间的积累与打磨，以及人力资源的不断投入。

图表21　国内面向安防 AI 芯片的企业及主要产品

企业	产品	产品说明
国科微	国科 GK7102	面向高清网络摄像机的 IPC 芯片，内置优秀的图像处理算法和丰富的智能视频分析算法
景嘉微	图形图像处理芯片	国产 GPU
富瀚微	IPC SoC 和 ISP	视频编解码和图像信号处理芯片
深鉴科技	听涛系列	产品主要面向无人机、安防、数据中心，融入自身算法
华为海思	Hi3516CV500 等	ARM　A7 架构，具备神经网络加速能力
	云端 AI 芯片“昇腾”系列	910 可用于云端，支持 128 通道全高清解码；310 主要用于边缘端，可支持 16 通道全高清解码
云天励飞	NNP100 已经完成流片，基于 FPGA 实现	用于 DeepEye100 智能盒子和 DeepEye200 服务器加速卡
地平线	“旭日 1.0”	嵌入式人工智能视觉芯片
阿里巴巴	Ali-NPU	面向图像和视频处理需求

资料来源：公司网站，平安证券研究所

2.4 智能家居

智能家居近年来也成为人工智能重要的落地场景。从技术应用上讲，人类 90%的信息输出是通过语音，80%的是通过视觉，智能家居领域应用最多的就是智能语音交互技术。近年来，正是看到语音

交互技术与智能家居深度融合的潜力，谷歌、苹果、微软均将其作为进入智能家居领域的重要切入口，发布了多款软硬件平台，如亚马逊推出的智能音箱设备。国内智能语音龙头企业科大讯飞较早就切入了该领域，联合地产商推出了硬件平台魔飞（MORFEI）平台，电视、咖啡机、电灯、空调、热水器等产品都能通过融入相关平台实现智能化。

当前，无论是智能音箱还是其他智能家居设备，智能功能都是在云端来实现，但云端存在着语音交互时延的问题，对网络的需求限制了设备的使用空间，而且由此还带来了数据与隐私危机。为了让设备使用场景不受局限，用户体验更好，端侧智能已成为一种趋势，语音 AI 芯片也随之切入端侧市场。国内主要语音技术公司凭借自身在语音识别、自然语言处理、语音交互设计等技术上的积累，开始转型做 AI 语音芯片集成及提供语音交互解决方案，包括云知声、出门问问、思必驰以及 Rokid。

市场上主流的 AI 语音芯片，一般都内置了为语音识别而优化的深度神经网络加速方案，以实现语音离线识别。随着算法的精进，部分企业的语音识别能力得到了较快提升，尤其是在远场识别、语音分析和语义理解等方面都取得了重要进展。云知声在 2018 年 5 月，推出语音 AI 芯片雨燕，并在研发多模态芯片，以适应物联网场景，目前公司芯片产品已经广泛用于智能家电如空调之中；出门问问也在 2018 年推出了 AI 语音芯片模组“问芯”MobvoiA1；Rokid 也发在 2018 年发布了 AI 语音芯片 KAMINO18；思必驰利用其声纹识别等技术优势，2019 年初推出基于双 DSP 架构的语音处理专用芯片 TH1520，具有完整语音交互功能，能实现语音处理、语音识别、语音播报等功能。

由于语音芯片市场过于细碎，需要企业根据场景和商业模式需要设计出芯片产品，这对传统的通用芯片企业的商业模式是一种颠覆，以致于在 2018 年以前都很少有芯片巨头进入该领域，这也给了国内语音芯片企业较大的施展空间。而对算法公司来说，通过进入芯片市场，进而通过解决方案直接面向客户和应用场景，通过实战数据来训练和优化算法。

图表22　国内主要语音芯片厂商及产品情况

厂商	发布时间	产品	简介
杭州国芯	2017.10	语音 AI 芯片 GX8010	搭载 NPU、DSP 等技术，具备低功耗、可离线等特点
声智科技	2018.3	低功耗麦克风阵列芯片 SAI101C	支持智能电视及机顶盒、智能家居网关等产品
云知声	2018.5	雨燕 UniOne	面向 IoT 的 AI 芯片
出门问问	2018.5	芯片模组 Mobvoi A1	与国芯合作研发，出门问问提供语音交互能力
Rokid	2018.6	AI 语音专用 SoC 芯片 KAMINO18	集成了 ARM、NPU、DSP、DDR、DAC 等多个核心元件，结合其算法优势，能耗较低
思必驰	2019.1	思必驰-深聪 TAIHANG 芯片	与中芯国际合作开发，基于双 DSP 架构设计
全志科技	-	R 系列芯片	集成语音识别应用，用于智能音箱和智能家居

资料来源：各公司官网、平安证券研究所

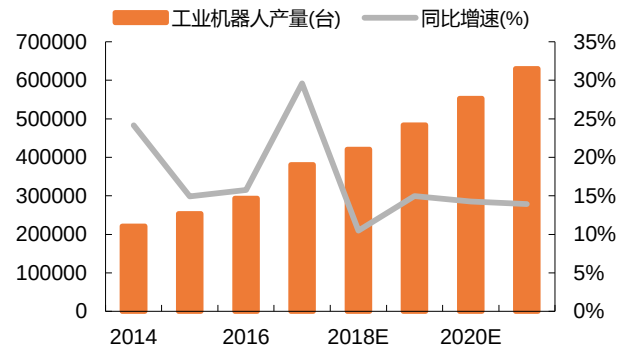
2.5 机器人

机器人是人工智能行业最早的落地形态，也是现在和将来重要的应用方向。机器人主要包括两类——制造环境下的工业机器人和非制造环境下的服务机器人。工业机器人主要是面向工业领域的多关节机械手或多自由度机器人。服务机器人则是除工业机器人之外的、用于非制造业并服务于人类的各种先进机器人。

随着云物移大智等信息及智能化技术的发展，机器人在某些领域的工作效率高于人类，并在工业和服务场景中得到了大量应用。据国际机器人联盟统计，2017 年，全球工业机器人产量达到 38.1 万台，同比增长 30%，预计 2018-2021 年全球工业机器人产量将保持 10%以上增速增长，2021 年产量预计将达到 63.0 万台。中国是全球最大的工业机器人生产国，2017 年产量达到 13.79 万台，同比大幅增长 60%。服务机器人主要用于物流、防务、公共服务、医疗等领域，虽然规模不大，但是增长迅速。2017 年全球产量为 10.95 万台，同比大幅增长 85%。预计 2018 年全球专业服务机器人产量将达到 16.53 万台，同比增长 32%，2019-2021 年平均增速将保持在 21%左右。

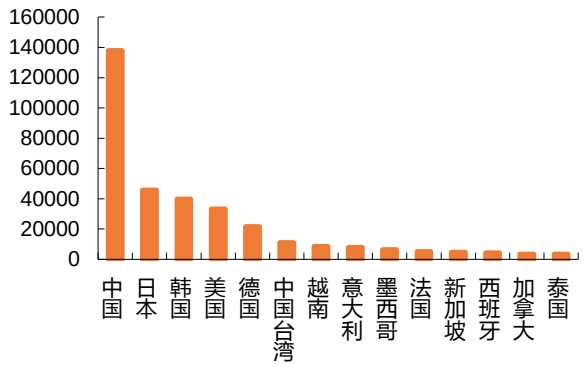
机器人尤其是国内产业规模的快速扩大，将大幅带动国内机器人相关智能芯片产业的发展。机器人由控制、传感、驱动和电源四大装置构成，其中控制装置是机器人的“大脑”，核心是 AI 芯片。机器人芯片需要具备强大的数据计算、自主判断思考和执行能力，国外厂商如高通、英特尔、英伟达都在积极部署该领域，国内企业目前处于追赶状态，相关企业包括**瑞芯微、珠海全志、炬力**等。

图表23　2014-2021 年全球工业机器人产量及增速



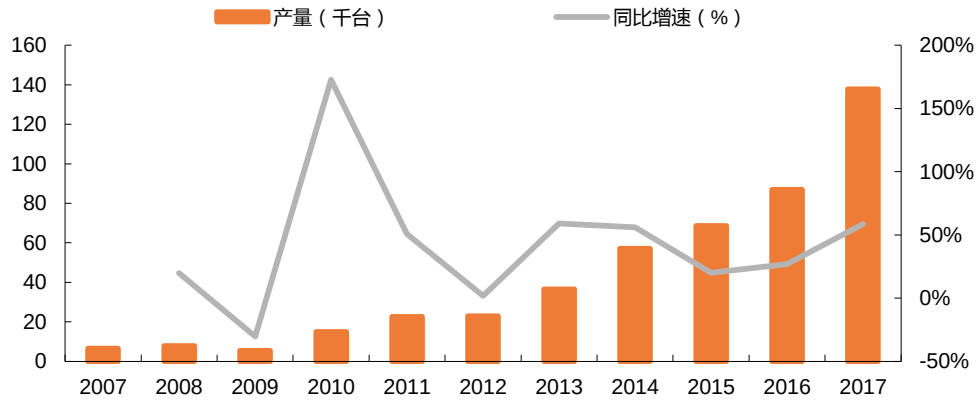
资料来源:国际机器人联盟、平安证券研究所

图表24　2017 年主要国家和地区工业机器人产量（台）



资料来源:国际机器人联盟、平安证券研究所

图表25　2007-2017 年中国工业机器人产量及同比增速



资料来源:国际机器人联盟、平安证券研究所

图表26　国内机器人芯片企业及产品

企业	芯片型号及功能	典型客户
瑞芯微	RK3399：AI+图像识别+定位导航 RV1108：图像视觉定位导航	科沃斯扫地机器人
炬芯	ATS 系列	巴巴腾、智伴
珠海全志	R58	小鱼在家
山景	蓝牙音频方案	小腾陪护机器人

资料来源：公司网站，平安证券研究所

三、 国内外 AI 芯片厂商概览

3.1 整体排名

近年来，各类势力均在发力 AI 芯片，参与者包括传统芯片设计、IT 厂商、技术公司、互联网以及初创企业等，产品覆盖了 CPU、GPU、FPGA 、ASIC 等。在市场调研机构 Compass Intelligence 2018 年发布的 AI Chipset Index TOP24 榜单中，前十依然是欧美韩日企业，国内芯片企业如华为海思、联发科、Imagination（2017 年被中国资本收购）、寒武纪、地平线机器人等企业进入该榜单，其中华为海思排 12 位，寒武纪排 23 位，地平线机器人排 24 位。

图表27　全球 AI 芯片企业排名（TOP24）

排名	企业	指数	排名	企业	指数
1	英伟达	85.3	13	Synopsys	61
2	Intel（mobileye、nervana、Movidus）	82.9	14	联发科	59.5
3	IBM	80.2	15	Imagination	59
4	Google	78	16	Marvell	58.5
5	Apple	75.3	17	赛灵思	58
6	AMD	74.7	18	CEVA	54
7	ARM	73	19	Cadence	51.5
8	高通	73	20	Rockchiip	48
9	三星	72.1	21	Verisilcon	47
10	恩智浦	70.2	22	General Vision	46
11	博通	68.2	23	寒武纪	44.5
12	华为海思	64.5	24	地平线机器人	38.5

资料来源：Compass Intelligence、平安证券研究所

图表28　主要 AI 芯片类型及企业

部署位置	芯片类型	训练	推理
云端	GPU	英伟达、AMD	英伟达
	FPGA	英特尔、赛灵思	英特尔、赛灵思、亚马逊、微软、百度、阿里、腾讯
	ASIC	谷歌	谷歌、寒武纪、比特大陆、Groq、Wave Computing
终端	GPU	-	英伟达、ARM
	FPGA		深鉴科技
	ASIC		寒武纪、地平线、华为海思、高通、ARM

资料来源：CSDN、平安证券研究所

3.2 芯片企业

芯片设计企业依然是当前 AI 芯片市场的主要力量，包括英伟达、英特尔、AMD、高通、三星、恩智浦、博通、华为海思、联发科、Marvell（美满）、赛灵思等，另外，还包括不直接参与芯片设计，只做芯片 IP 授权的 ARM 公司。其中，英伟达、英特尔竞争力最为强劲。

■　**英伟达：AI 芯片市场的领导者，计算加速平台广泛用于数据中心、自动驾驶等场景**

英伟达创立于 1993 年，最初的主业为显卡和主板芯片组。其主板芯片组主要客户以前是 AMD，但是在 AMD 收购 ATI 推出自研芯片组之后，英伟达在该领域的优势就荡然无存。于是，公司全面转向到 GPU 技术研发，同时进入人工智能领域。2012 年，公司神经网络技术在其 GPU 产品的支持下取得重大进展，并在计算机视觉、语音识别、自然语言处理等方面得到广泛应用。

2016 年，全球人工智能发展加速，英伟达迅速推出了第一个专为深度学习优化的 Pascal GPU。2017 年，英伟达又推出了性能相比 Pascal 提升 5 倍的新 GPU 架构 Volta，同时推出神经网络推理加速器 TensorRT 3。至此，英伟达完成了算力、AI 构建平台的部署，也理所当然成为这一波人工智能热潮的最大受益者和领导者。公司的战略方向包括人工智能和自动驾驶。

人工智能方面。英伟达面向人工智能的产品有两个，Tesla 系列 GPU 芯片以及 DGX 训练服务器。Tesla 系列是专门针对 AI 深度学习算法加速设计 GPU 芯片，DGX 则主要是面向 AI 研究开发人员设计的工作站或者超算系统。2018 年，公司包含这两款产品的数据中心业务收入大幅增长 52%，其中 Tesla V100 的强劲销售是其收入的主要来源。

自动驾驶方面。英伟达针对自动驾驶等场景，推出了 Tegra 处理器，并提供了自动驾驶相关的工具包。2018 年，基于 Tegra 处理器，英伟达推出了 NVIDIA DRIVE AutoPilot Level 2+，并赢得了丰田、戴姆勒等车企的自动驾驶订单。同时，2018 年，公司也正在积极推动 Xavier 自动驾驶芯片的量产。

值得关注的是，英伟达还正在通过投资和并购方式继续加强在超算或者数据中心方面的业务能力。2019 年 3 月，英伟达宣称将斥资 69 亿美元收购 Mellanox。Mellanox 是超算互联技术的早期研发和参与者。通过与 Mellanox 的结合，英伟达将具备优化数据中心网络负载能力的能力，其 GPU 加速解决方案在超算或者数据中心领域的竞争力也将得到显著提升。

■ **英特尔加速向数字公司转型，通过并购+生态优势发力人工智能**

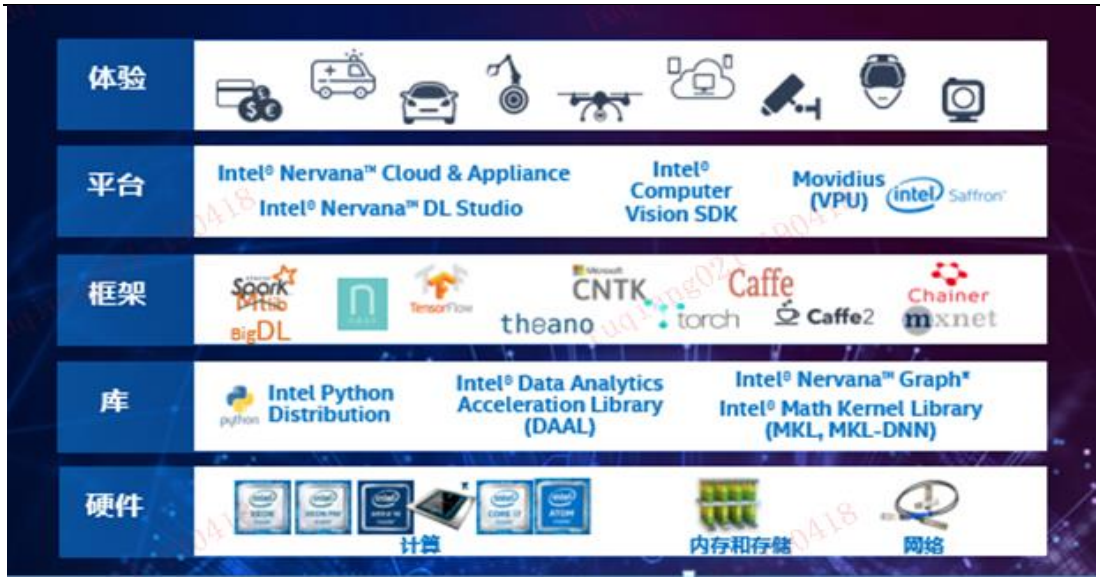
英特尔作为传统的 CPU 设计制造企业，在传统 PC、服务器市场有着绝对的统治力。随着互联网时代的到来以及个人电脑市场的饱和，公司也在开始加快向数字公司转型。尤其在人工智能兴起之后，英特尔凭借着技术和生态优势，打造算力平台，形成全栈式解决方案。

英特尔主要产品为 CPU、FPGA 以及相关的芯片模组。虽然 CPU 产品在训练端的应用效率不及英伟达，但推理端优势较为明显。英特尔认为，未来 AI 工作周期中，推理的时长将是训练时长的 5 倍甚至 10 倍，推理端的芯片需求也会放量。同时，即使是云端训练，GPU 也需要同 CPU 进行异构。

目前，英特尔在人工智能芯片领域主要通过三条路径：1）通过并购快速积累人工智能芯片相关的技术和人才，并迅速完成整合。英特尔在收购了 Altera 后，还先后收购了 Nervana、Movidius 与 Mobileye 等初创企业。在完成上述一系列并购之后，英特尔设立了 AI 事业群，整合了 Xeon、Xeon Phi、Nervana、Altera、Movidius 等业务和产品，同时将原有的自动驾驶业务板块并入 Mobileye。2）建立多元的产品线。目前，英特尔正建构满足高性能、低功耗、低延迟等差异化芯片解决方案，除了 Xeon 外，包括可支持云端服务 Azure 的 Movidius VPU 与 FPGA。3）通过计算平台等产品，提供强大的整合能力，优化 AI 计算系统的负载，提供整体解决方案。

在英特尔收购的这些企业中，除了前面已经提到的 Altera、Mobileye 之外，Nervana 也非常值得关注。2016 年 8 月，英特尔斥资超过 3.5 亿美元收购这家员工人数不超过 50 人的创业公司，但是经过不到三年的成长，这家公司已经成为英特尔 AI 事业部的主体。依托 Nervana，英特尔成功在 2017 年 10 月推出了专门针对机器学习的神经网络系列芯片，目前该芯片已经升级至第二代，预计 2019 年下半年将正式量产上市，该芯片在云端上预计能和英伟达的 GPU 产品一较高下。

图表29　英特尔 AI 全栈解决方案框架



资料来源:Intel、平安证券研究所

3.3 IT 及互联网企业

AI 兴起之后，互联网及 IT 企业凭借着在各大应用场景上技术和生态积累，也在积极拓展 AI 相关市场，其中 AI 芯片是部署重点之一。相较而言，互联网企业凭借着数据和场景先天优势，在 AI 算法和芯片领域优势更为明显，如美国谷歌、国内的 BAT。IT 企业如 IBM，在人工智能领域较早开始研究，2018 年年中曾经推出专门针对深度学习算法的原型芯片。

■ **谷歌：TPU 芯片已经实现从云到端，物联网 TPU Edge 是当前布局重点**

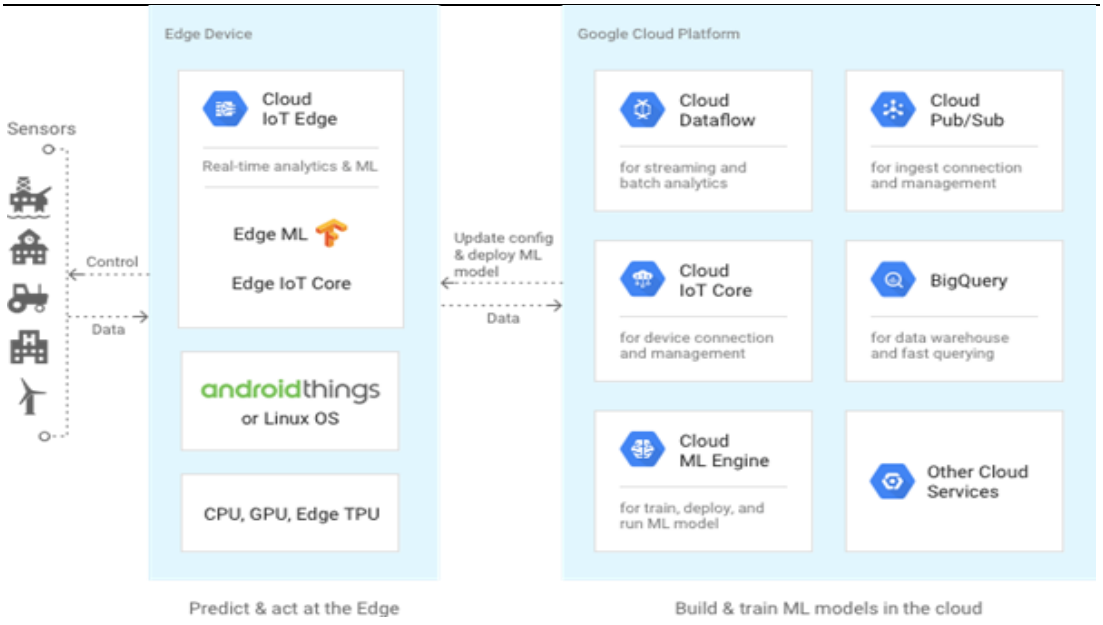
谷歌可谓是 AI 芯片行业的一匹黑马，但是竞争力强劲。谷歌拥有大规模的数据中心，起初同其他厂商的数据中心一样，都采用 CPU+GPU 等异构架构进行计算加速，用来完成图像识别、语音搜索等计算服务。但是，随着业务量的快速增长，传统的异构模式也很难支撑庞大的算力需求，需要探索新的高效计算架构。同时，谷歌也需要通过研发芯片来拓展 AI 平台 TensorFlow 的生态。因此，2016 年，Google 正式发布了 TPU 芯片。

从谷歌 TPU 的本质来看，它是一款 ASIC（定制芯片），针对 TensorFlow 进行了特殊优化，因此该产品在其他平台上无法使用。第一代 Cloud TPU 仅用于自家云端机房，且已对多种 Google 官方云端服务带来加速效果，例如 Google 街景图服务的文字处理、Google 相簿的照片分析、甚至 Google 搜寻引擎服务等。Cloud TPU 也在快速改版，2017 年推出第二代，2018 年推出第三代芯片 TPU 3.0。同时，谷歌对 TPU 的态度也更为开放，之前主要是自用，目前也在对用户开放租赁业务，但没有提供给系统商。

除了云端，谷歌针对边缘端推理需求快速增长的趋势，也在开发边缘 TPU 芯片。2017 年 11 月，Google 推出轻量版的 TensorFlow Lite（某种程度取代此前的 TensorFlow Mobile），使得能耗有限的移动设备也能支持 TensorFlow，2018 年推出的 Edge TPU 芯片即是以执行 TensorFlow Lite 为主，而非 TensorFlow。Edge TPU 性能虽然远不如 TPU，但功耗及体积大幅缩小，适合物联网设备采用。Edge TPU 可以自己运行计算，不需要与多台强大计算机相连，可在传感器或网关设备中与标准芯片或微控制器共同处理 AI 工作。

按照谷歌的规划，Edge TPU 将提供给系统商，开放程度将进一步提升。如果 Edge TPU 推广顺利，支持的系统伙伴将进一步增多，谷歌将尽快推出下一代 Edge TPU 产品。即使推广不顺利，Google 也可能自行推出 Edge 网关、Edge 设备等产品。

图表30　谷歌云端、边缘端 AI 芯片应用情况



资料来源:谷歌官网、平安证券研究所

■ **阿里巴巴：推出自研神经网络处理芯片，同时加速对 AI 企业投资布局**

阿里巴巴作为国内 AI 领域的领军企业，在底层算力、算法技术以及应用平台方面都有较强积累。同 Google 类似原因，阿里巴巴也在近年来开始开发 AI 芯片，同时加大对相关领域的投资布局。

2017 年，阿里巴巴成立阿里达摩院，研究领域之一就是 AI 芯片技术。2018 年 4 月，阿里达摩院对外宣布正研发一款 Ali-NPU 神经网络芯片，预计将在 2019 年下半年问世。这款芯片将主要应用于图像视频分析、机器学习等 AI 推理计算。

阿里巴巴在自研 AI 芯片之前，主要在通过投资的方式布局 AI 芯片领域。目前，寒武纪、深鉴科技、杭州中天微等都有阿里巴巴的入股，其中 2016 年 1 月份还成为了 AI 芯片设计企业杭州中天微的第一大股东。

图表31　阿里巴巴参与投资的 AI 芯片相关创投公司

公司名称	成立时间	业务类型	投资时间	轮次、金额	投资方
Kenron(耐能)	2014	AI 芯片，主打轻量级的 NPU（神经网络处理单元）芯片，主要布局智能家居和物联网	2017.11	A 轮，过千万美金	阿里创业者基金、高通、中科创达、红杉资本等
寒武纪	2016	AI 芯片，拥有终端 AI 处理器 IP 和云端高性能 AI 芯片两条产品线	2017.8	A 轮，1 亿美金	阿里巴巴、联想创投等
深鉴科技	2016	AI 芯片，聚焦于安防	2017.1	A+轮，4000 万美金	蚂蚁金服，联发科，金沙江创投等
Barefoot Networks	2014	AI 芯片，专注于交换系统的超快芯片	2016.11	C 轮，2300 万美金	阿里，腾讯领投
			2016.6	C 轮，5700 万美金	谷歌，丹华资本等
杭州中天微	2001	AI 芯片，大规模量产自主嵌入式 CPU IP Core，面向多媒体、安防、家庭、交通、智慧城市等 IoT 领域	2016.1	未透露	阿里为第一大股东
翱捷科技	2015	AI 芯片，专注于移动智能通讯终端、物联网、导航及其他消费类电子芯片	2017.4、2017.7	超过 1 亿美元	深创投、万容红土基金和阿里巴巴

数据来源：PEdaily 等、平安证券研究所

■ **百度：通过自研、合作以及投资等多种方式部署 AI 芯片**

百度作为搜索企业，其对 AI 芯片的需求更为明确。早在 2011 年，百度就在 FPGA 和 GPU 进行了大规模部署，也开始在 FPGA 的基础上研发 AI 加速器来满足深度学习运算的需要。此后，百度就不断通过合作、投资和自研的方式来推进该业务。

1）加强同芯片设计及 IP 企业合作。2017 年 3 月，百度发布了 DuerOS 智慧芯片，并与紫光展锐、ARM、上海汉枫达成战略合作。这款芯片搭载了对话式人工智能操作系统，可以赋予设备可对话的能力，能广泛用于智能玩具、蓝牙音箱、智能家居等多种设备。2017 年 8 月，百度又与赛思灵（Xilinx）发布了 XPU，这是一款 256 核、基于 FPGA 的云计算加速芯片。同在 2017 年，百度同华为达成合作，推动终端 AI 芯片的落地。

2）参与 AI 芯片企业投资。2018 年 2 月 5 日，美国初创公司 Lightelligence 宣布获得了 1000 万美元种子轮融资，由百度风投和美国半导体高管财团领投。Lightelligence 主要利用基于光学的新技术，来加速人工智能的工作负载，通过光子电路的新兴技术来加速信息处理。

3）自研芯片也正在加速部署。2018 年 7 月，百度正式发布了自研的 AI 芯片“昆仑”，这是当时国内第一款云端全功能 AI 芯片，其中包含训练芯片昆仑 818-300，推理芯片昆仑 818-100。昆仑 AI 芯片是基于百度 CPU、GPU、FPGA 的 AI 加速器研发，能够在 100W 左右的功耗下，提供高达 260 万亿次/秒的运算速度，算力处于业界领先水平。

图表32　百度 AI 芯片合作、合资及自研情况

	公司名称	产品	时间
合作企业	紫光展锐、ARM、上海汉枫	发布 DuerOS 智慧芯片，搭载了对话式人工智能操作系统，可以赋予设备可对话的能力，能广泛用于智能玩具、蓝牙音箱、智能家居等多种设备	2017.3
	赛思灵（Xilinx）	发布 XPU，是一款 256 核、基于 FPGA 的云计算加速芯片	2017.8
	华为	战略合作，弥补自身在硬件方面和端芯片的缺失	2017
投资公司	Lightelligence	利用基于光学的新技术，来加速人工智能的工作负载，通过光子电路的新兴技术来加速信息处理，进行计算的不是电子而是光子	2018.2
自研产品	AI 芯片“昆仑”	中国第一款云端全功能 AI 芯片	2018.7

资料来源：搜狐、Elecfans 等，平安证券研究所

3.4 创业企业

■ **寒武纪：公司同时发力终端和云端芯片，技术综合实力较强**

寒武纪发源于中科院，是目前全球领先的智能芯片公司，由陈天石、陈云霁兄弟联合创办，团队成员主要人员构成也来自于中科院，其中还有部分参与龙芯项目的成员。2018 年 6 月公司，公司获得数亿美元投资，此轮融资之后，寒武纪科技估值从上年的 10 亿美金大幅上升至 25 亿美元。公司是目前国内为数不多的同时具备云端和终端 AI 芯片设计能力的企业。

公司最早发力的是终端芯片，主要为 1A 系列，包括 1A、1H8 和 1H16，公司通过 IP 授权的模式赋能终端或者芯片设计企业，目前主要合作伙伴包括华为，其中麒麟 970 就采用其 1A 处理器。另外，公司还推出了面向低功耗场景视觉应用的寒武纪 1H8，高性能且拥有广泛通用性的寒武纪 1H16，以及用于终端人工智能产品的寒武纪 1M。2018 年 9 月，华为发布的麒麟 980 依然集成了优化版的寒武纪 1H 新一代智能处理器。

公司云端芯片也取得较大突破。云端芯片一直是英特尔、英伟达等公司的领地，国内企业很难进入。2018 年 5 月，寒武纪推出算力达到 128Tops 的 MLU 100 云端智能芯片，可用于训练和推理。MLU100

相比传统的 GPU 和 CPU 芯片，MLU 芯片拥有显著的性能功耗比和性能价格比优势，适用范围覆盖了图像识别、安防监控、智能驾驶等多个重点应用领域。

综合来看，公司在 AI 芯片方面竞争力较强。公司拥有自己的处理器架构和指令集，而且通过硬件神经元虚拟化、开发通用指令集、运用稀疏化处理器架构解决了 ASIC 用于深度学习时存在的三大问题。这三大问题是：云端算力的挑战、能效瓶颈、手机端和云端超大规模计算场景应用问题。

■ **地平线机器人：公司芯片和计算平台在嵌入式及智能驾驶领域具备优势**

地平线成立于 2015 年，主要从事边缘人工智能芯片和计算平台业务，场景聚焦于智能驾驶和 AIoT 边缘计算。2018 年起，公司逐渐实现产品化落地。2019 年 2 月，公司官方宣布已获得 6 亿美元 B 轮融资，SK 中国、SK Hynix 以及数家中国一线汽车集团（与旗下基金）联合领投。B 轮融资后，地平线估值达 30 亿美元。

2017 年 12 月，地平线发布中国首款全球领先的嵌入式人工智能视觉芯片征程（Journey）系列和旭日（Sunrise）系列。旭日 1.0 处理器面向智能摄像头等应用场景，具备在前端实现大规模人脸检测跟踪、视频结构化的处理能力，可广泛用于智慧城市、智慧零售等场景。征程 1.0 处理器面向智能驾驶，具备同时对行人、机动车、非机动车、车道线、交通标志牌、红绿灯等多类目标进行精准的实时检测与识别的处理能力，同时满足车载严苛的环境要求以及不同环境下的视觉感知需求，可用于高性能 L2 级别的高级驾驶辅助系统 ADAS 。

2018 年 2 月，地平线自主研发的高清智能人脸识别网络摄像机，搭载地平线旭日人工智能芯片，提供基于深度学习算法的人脸抓拍、特征抽取、人脸特征值比对等功能。可以在摄像机端实现人脸库最大规模为 5 万的高性能人脸识别功能，适用于智慧城市、智慧零售等多种行业。

2018 年 4 月，公司发布地平线 Matrix1.0 自动驾驶计算平台。目前已经更新到性能更强的升级版本，地平线 Matrix 自动驾驶计算平台结合深度学习感知技术，具备强大的感知计算能力，能够为 L3 和 L4 级别自动驾驶提供高性能的感知系统。地平线 Matrix 自动驾驶计算平台已向世界顶级 Robotaxi 厂商大规模供货，成功开创了中国自动驾驶芯片产品出海和商业化的先河。

四、 投资建议

从当前 AI 芯片竞争格局和市场前景看，我们认为，国内企业在边缘端的机会多于云端。一方面，在边缘场景，国内在语音、视觉等领域已经形成了一批芯片设计企业队伍，相关芯片产品已经在安防、数据中心推理、智能家居、服务机器人、智能汽车等领域找到落地场景，未来随着 5G、物联网等应用的兴起，相关企业的市场空间将进一步扩大。另一方面，在云端，国内企业也正在加速追赶，未来也有望取得突破。尤其是寒武纪，作为云端芯片重要的技术厂商，有望通过授权等方式为下游芯片设计、服务器企业赋能。

AI 芯片上市公司标的较为稀缺，覆盖标的中，重点推荐中科曙光、科大讯飞、中科创达以及四维图新。中科曙光作为“芯-服务器-云”一体化企业，将直接成为国内 AI 芯片发展的受益者。除了 AMD 授权的海光 X86 处理器之外，公司也正在和同为中科体系的寒武纪合作，预计将在 AI 服务器、智能芯片等方面获得突破；科大讯飞作为语音交互领域的龙头，不但持有寒武纪的股份，而且还在同外部合作研发 AI 芯片 Castor（北河二），目前该芯片已经完成测试工作，未来可用于智能家居等语音交互场景；中科创达在嵌入式人工智能领域有着较强的积累，主要为手机及安防终端提供软件解决方案，近年来开始向底层芯片发力，2017 年 11 月跟投了国内神经网络处理器厂商——耐能；四维图新作为自动驾驶领域的重点标的，其收购的杰发科技，车规级 MCU 已经实现量产，为后续进军自动驾驶，实施“汽车大脑”战略打下了良好的基础。

五、 风险提示

（一）场景落地不及预期。国内人工智能芯片主要集中在边缘端推理领域，应用主要集中在安防、智能家居、消费机器人等场景，但是上述场景中 AI 芯片竞争十分激烈，可能影响到芯片企业产品落地效果。

（二）技术方向的不确定性风险。人工智能芯片和算法、终端、应用场景等密切相关，但在当前信息技术加速迭代的背景下，AI 技术调整的风险非常高，相关调整一旦发生，将对企业发展造成重大影响。

（三）研发进度不及预期。AI 芯片研发难度大，需要投入大量的资金和人力。国内创业企业资金实力相对薄弱，在人才竞争中将处于劣势。上述因素，可能直接导致企业研发进度落后于预期。

平安证券研究所投资评级：

股票投资评级:

强烈推荐 （预计 6 个月内，股价表现强于沪深 300 指数 20%以上）

推　　荐 （预计 6 个月内，股价表现强于沪深 300 指数 10%至 20%之间）

中　　性 （预计 6 个月内，股价表现相对沪深 300 指数在±10%之间）

回　　避 （预计 6 个月内，股价表现弱于沪深 300 指数 10%以上）

行业投资评级:

强于大市 （预计 6 个月内，行业指数表现强于沪深 300 指数 5%以上）

中　　性 （预计 6 个月内，行业指数表现相对沪深 300 指数在±5%之间）

弱于大市 （预计 6 个月内，行业指数表现弱于沪深 300 指数 5%以上）

公司声明及风险提示：

负责撰写此报告的分析师(一人或多人)就本研究报告确认：本人具有中国证券业协会授予的证券投资咨询执业资格。

平安证券股份有限公司具备证券投资咨询业务资格。本公司研究报告是针对与公司签署服务协议的签约客户的专属研究产品，为该类客户进行投资决策时提供辅助和参考，双方对权利与义务均有严格约定。本公司研究报告仅提供给上述特定客户，并不面向公众发布。未经书面授权刊载或者转发的，本公司将采取维权措施追究其侵权责任。

证券市场是一个风险无时不在的市场。您在进行证券交易时存在赢利的可能，也存在亏损的风险。请您务必对此有清醒的认识，认真考虑是否进行证券交易。

市场有风险，投资需谨慎。

免责条款：

中国平安 金融 · 科技

平安证券 PING AN SECURITIES

平安证券研究所

电话：4008866338

深圳	上海	北京
深圳市福田区益田路 5033 号平安金融中心 62 楼 邮编：518033	上海市陆家嘴环路 1333 号平安金融大厦 25 楼 邮编：200120 传真：（021）33830395	北京市西城区金融大街甲 9 号金融街中心北楼 15 层 邮编：100033