半导体

人工智能芯片——新架构改变世界

人工智能倒逼芯片底层的真正变革

人类精密制造领域（半导体制造是目前为止人类制造领域的最巅峰）遇到硅基极限的挑战，摩尔定律的放缓似乎预示着底层架构上的芯片性能的再提升已经出现瓶颈，而数据量的增长却呈现指数型的爆发，两者之间的不匹配势必会带来技术和产业上的变革升级。**变革从底层架构开始。计算的体系处于碎片化引发架构变革**。数据的扩张远大于处理器性能的扩张，依靠处理器性能在摩尔定律推动下的提升的单极世界已经崩溃，处理器性能提升的速度并不足以满足 AI 所需的应用程序的需求。大量数据消耗的数字运算能力比几年前所有数据中心加起来还要多。基于冯诺伊曼架构的拓扑结构已经持续了很多年并没有本质上的变化。而人工智能带来的，**是在摩尔定律放缓维度下引发芯片底层架构重构的变革**。有可能引发的是一次超越以往任何时代的科技革命

基于摩尔定律的机器时代的架构——从 Wintel 到 AA

冯诺伊曼架构带来了计算体系的建立并通过 Intel 实现了最大化；ARM 通过共享 IP 的商业模式带来了更开放的生态体系，实现了软硬件的结合延伸了人类的触角观察 Intel 和 ARM 的黄金十年，站在现在时点往后看，我们提出以下观点：**过去十年以下游的应用驱动设计公司的成长转换为由设计公司主导应用正在发生**。从需求层面看企业成长空间。类似 90 年代的 PC 和 10 年的智能手机带来的亿级大空间增量市场将很容易推动企业的快速增长。设计企业能够在成长轨迹上实现跨越式突破的可能性来自于赛道的选择。但站在现在时点看，人工智能是确定性的方向，在所有已有领域的人工智能渗透，都将极大的改变人类的生活。**处于最前沿的芯片公司的革新正在以此而发生，重新定义底层架构的芯片，从上游推动行业的变革**。在并没有具体应用场景爆发之前已经给予芯片公司充分的高估值就是认可设计公司的价值

人工智能芯片——新架构的异军突起

观察人工智能系统的搭建，以目前的架构而言，主要是以各种加速器来实现深度学习算法。讨论各种加速器的形式和实现，并探讨加速器变革下引发的行业深层次转变。**认为人工智能芯片将有可能在摩尔定律放缓维度下引发芯片底层架构重构的变革**。

从 2 个维度测算人工智能芯片空间

从两个维度讨论人工智能芯片的市场空间测算。维度一从人工智能总市场规模空间反推芯片，**维度二详细拆分云端/移动端所需人工智能加速器的 BOM** 进而推断人工智能芯片市场空间。二个维度印证到 2020 年人工智能芯片将达到百亿美元市场

重点标的：Intel，台积电，NVIDIA，全志科技，富瀚微，北京君正

风险提示：人工智能芯片发展不达预期

证券研究报告
2017 年 09 月 16 日

投资评级

行业评级	强于大市(维持评级)
上次评级	强于大市

作者

农冰立 分析师
SAC 执业证书编号：S1110516110006
nongbingli@tfzq.com

陈俊杰 分析师
SAC 执业证书编号：S1110517070009
chenjunjie@tfzq.com

行业走势图

资料来源：贝格数据

相关报告

1 《半导体-行业研究周报:一周半导体行业观点：智能手机 8 月挹注台积电营收，4Q 将成半导体产业链今年最强季，关注产业链公司》 2017-09-10

2 《半导体-行业点评:边际改善提升业绩增长，关注半导体族群》2017-09-09

3 《半导体-行业点评:抽丝剥茧——探寻本轮半导体元器件涨价背后的原因》2017-04-01

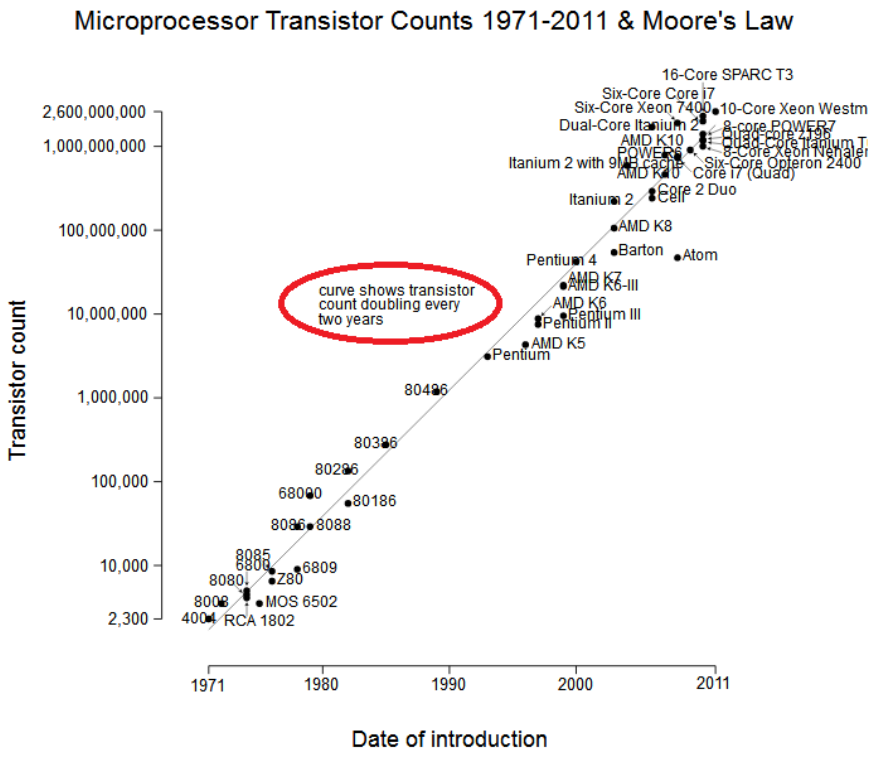
内容目录

图表目录

1. 人工智能倒逼芯片底层的真正变革

研究人类的科技发展史，发现科技的进步速度呈现指数型加速态势。尤其在 1950 年以后进入芯片时代，摩尔定律推动下的每 18 个月“芯片晶体管同比例缩小一半”带来的性能提升以倍数计。每一次加速的过程推动，都引发了产业的深层次变革，带动从底层到系统的阶跃。我们本篇报告将着重从底层芯片角度出发，探讨人工智能芯片带来的深层次变革。

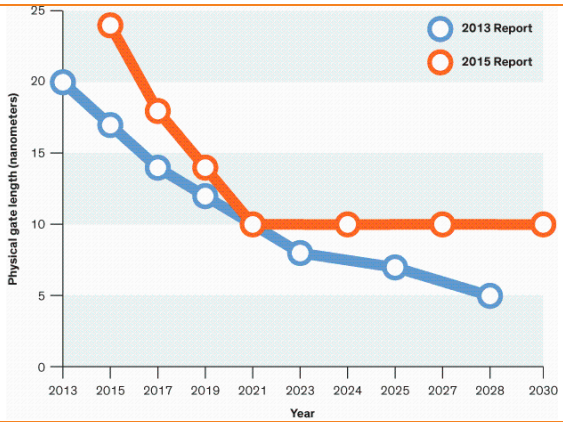
图 1：遵从摩尔定律发展到微处理器发展



资料来源：Gartner，天风证券研究所

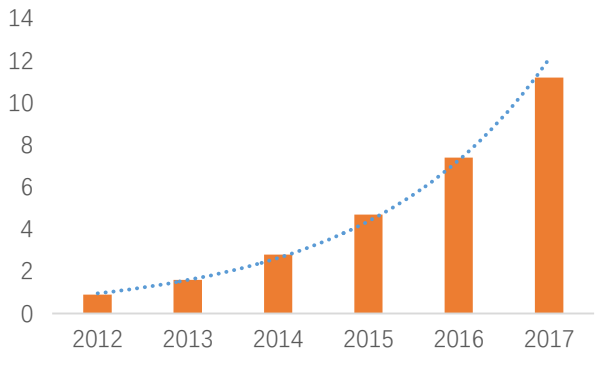
然而时至今日，**人类精密制造领域（半导体制造是目前为止人类制造领域的最巅峰）遇到硅基极限的挑战，摩尔定律的放缓似乎预示着底层架构上的芯片性能的再提升已经出现瓶颈，而数据量的增长却呈现指数型的爆发**，两者之间的不匹配势必会带来技术和产业上的变革升级。

图 2：摩尔定律在放缓



资料来源：IFS，天风证券研究所

图 3：全球智能手机每月产生的数据量（EB）5 年提升了 13X



资料来源：Cisco VNL，天风证券研究所

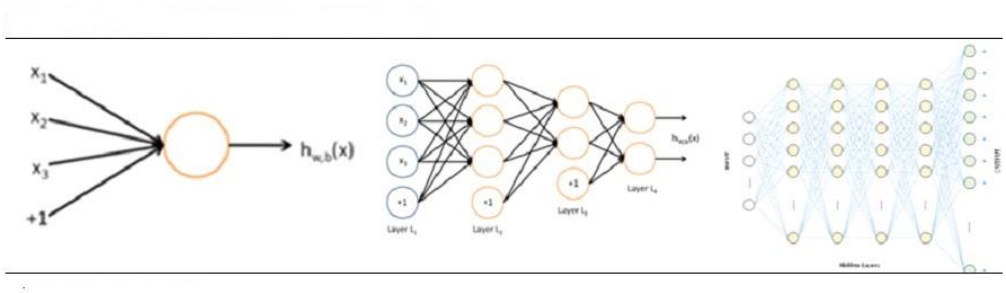
变革从底层架构开始

计算芯片的架构 50 多年来都没有发生过本质上的变化，请注意计算架构的决定是资源的组织形式。而传统的冯诺伊曼是采取控制流架构，采用的是线性的记忆体和布尔函数作为基线计算操作。处理器的架构基于流水线串行处理的机制建立，存储器和处理器分离，流水线的计算过程可以分解为取指令，执行，取数据，数据存储，依次循环。依靠整个串行的过程，逻辑清晰，但性能的提升通过两种方式，**一是摩尔定律下推动下晶体管数量的增多实现性能倍增；二是通过并行多个芯片核来实现。无论何种方式，本质上都是线性的性能扩张。**

人工智能芯片根据数据流的碎片化和分布式而采取神经网络计算范式，特征在于分布式的表示和激活模式。变量由叠加在共享物理资源上的向量表示，并且通过神经元的激活来进行计算。以神经元架构实现深度学习人工智能的临界点实现主要原因在于：**数据量的激增和计算机能力/成本。**

深度学习以神经元为架构。**从单一的神经元,再到简单的神经网络,到一个用于语音识别的深层神经网络。层次间的复杂度呈几何倍数的递增。数据量的激增要求的就是芯片计算能力的提升。**

图 4：单一神经元 VS 复杂神经元



资料来源：NVIDIA，天风证券研究所

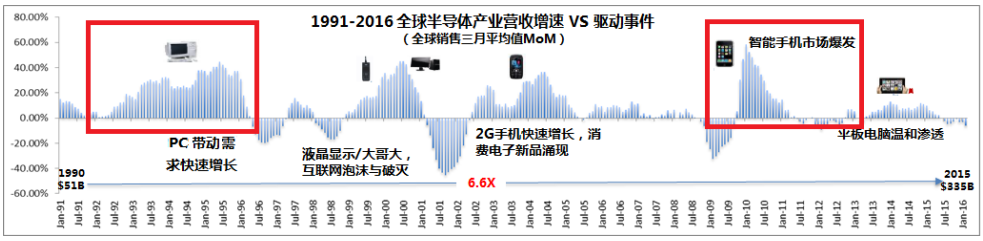
计算的体系处于碎片化引发架构变革。数据的扩张远大于处理器性能的扩张，依靠处理器性能在摩尔定律推动下的提升的单极世界已经崩溃，处理器性能提升的速度并不足以满足 AI 所需的应用程序的需求。大量数据消耗的数字运算能力比几年前所有数据中心加起来还要多。

我们在下一章将观察历史上两次重要的电子产业变革，试图证明无论是 PC 时代的“Wintel”还是智能手机时代的“ARM+Android”，都还无法摆脱机器本身的桎梏。换句话说，截止于现阶段的一切技术和应用，基于冯诺伊曼架构的拓扑结构已经持续了很多年并没有本质上的变化。而人工智能带来的，**是在摩尔定律放缓维度下引发芯片底层架构重构的变革。**有可能引发的是一次超越以往任何时代的科技革命。

2. 基于摩尔定律的机器时代的架构——从 Wintel 到 AA

本章我们重点讨论两次芯片架构变化引发的产业变革和应用爆发。Intel 与 Windows 结合构建 PC 生态 ，本质上诞生了软硬件结合的机器时代。而在其基础上的延升， 2010 后苹果带来的智能手机引发的 ARM 与 Android 生态，将机器与人的结合拓展到了移动端。我们回顾历史上的芯片架构历史，**认为冯诺伊曼架构带来了计算体系的建立并通过 Intel 实现了最大化；ARM 通过共享 IP 的商业模式带来了更开放的生态体系，实现了软硬件的结合延伸了人类的触角。**

图 5：2 次应用驱动芯片发展



资料来源：SIA，天风证券研究所

观察 Intel 和 ARM 的黄金十年，站在现在时点往后看，我们提出以下观点：**过去十年以下游的应用驱动设计公司的成长转换为由设计公司主导应用正在发生。**从需求层面看企业成长空间。类似 90 年代的 PC 和 10 年的智能手机带来的亿级大空间增量市场将很容易推动企业的快速增长。设计企业能够在成长轨迹上实现跨越式突破的可能性来自于赛道的选择。但站在现在时点看，人工智能是确定性的方向，在所有已有领域的人工智能渗透，都将极大的改变人类的生活。**处于最前沿的芯片公司的革新正在以此而发生，重新定义底层架构的芯片，从上游推动行业的变革。**在并没有具体应用场景爆发之前已经给予芯片公司充分的高估值就是认可设计公司的价值

2.1. Intel——PC 时代的王者荣耀

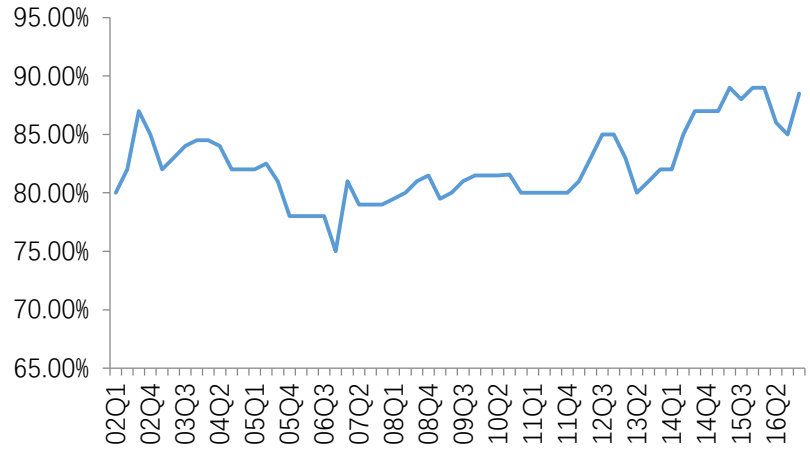
本节重点阐述 Intel 公司在 X86 时代的芯片架构产品以及此架构下公司以及行业的变化。

2.1.1. Intel 公司简介

Intel 是一家成立于 1968 年的半导体制造公司，总部位于美国加州。随着个人电脑的普及和全球计算机工业的日益发展，公司逐渐发展成为全球最大的微处理器及相关零件的供应商。公司在 2016 年实现营业收入 594 亿美元，世界 500 强排名 158。

公司分为 PC 客户端部门、数据中心部门、物联网、移动及通讯部门、软件及服务运营，其他还有笔记本部门、新设备部门及 NVM 解决方案部门。公司主要营业收入来自于 PC 客户部门，其次是数据中心部门。公司的主要产品 X86 处理器占主导地位，接近 90%，包括苹果在 2006 年放弃 PowerPC 改用英特尔的 x86 processors。

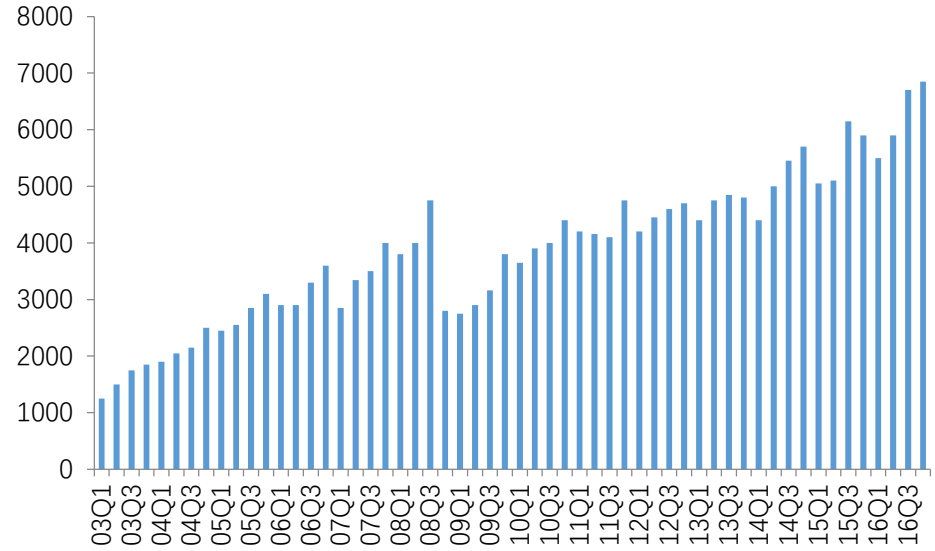
图 6：英特尔 x86 处理器总市场份额



资料来源：Intel，天风证券研究所

Intel 是第一家推出 x86 架构处理器的公司。Intel 从 8086 开始，286、386、486、586、P1、P2、P3、P4 都用的同一种 CPU 架构，统称 X86。大多数英特尔处理器都是基于 x86 指令集，被称为 x86 微处理器。指令集是微处理器可以遵循的基本命令集，它本质上是微处理器的芯片级“语言”。英特尔拥有 x86 架构的知识产权和给 AMD 和 Via 做处理器的许可权。

图 7：使用 X86 架构的单元



资料来源：wind，天风证券研究所

2.1.2. Intel 带来的 PC 行业的市场规模变革和产业变化

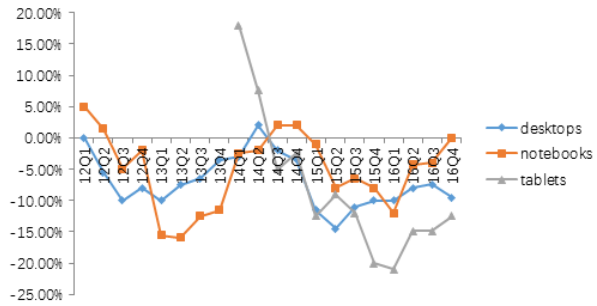
回顾 Intel 90 年代至今发展历程，清晰看到 90 年代是 Intel 发展最快的阶段并在 2000 年前后达到了峰值。**显而易见的原因是个人电脑的快速普及渗透。而遵从摩尔定律的每一代产品的推出，叠加个人电脑快速渗透的乘数效应，持续放大了企业的市值，类似于戴维斯双击，推动股价的一路上扬。**

图 8：摩尔定律下推动下的 Intel 股价上扬

资料来源：Wind，天风证券研究所

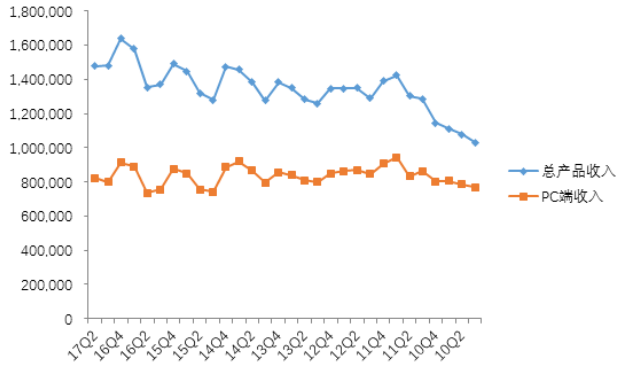
冯诺伊曼架构带来了计算体系的建立并通过 Intel 实现了最大化，但从本质上说，英特尔参与的是机器时代的兴起和计算芯片价值体现。但时至今日，在人口红利消散，PC 渗透率达到稳定阶段，依托于 PC 时代的处理器芯片进入了稳定常态。英特尔在总产品收入提升的情况下，PC 端提供的收入增长机会停滞。处理器依靠摩尔定律不断推经延续生命力，但在应用增长乏力的阶段缺乏爆发式的再增长。**PC 时代的处理器设计遵从了下游应用驱动上游芯片的实质。**

图 9：Intel 2012Q1-2016Q4 各产品线增速



资料来源：Intel，天风证券研究所

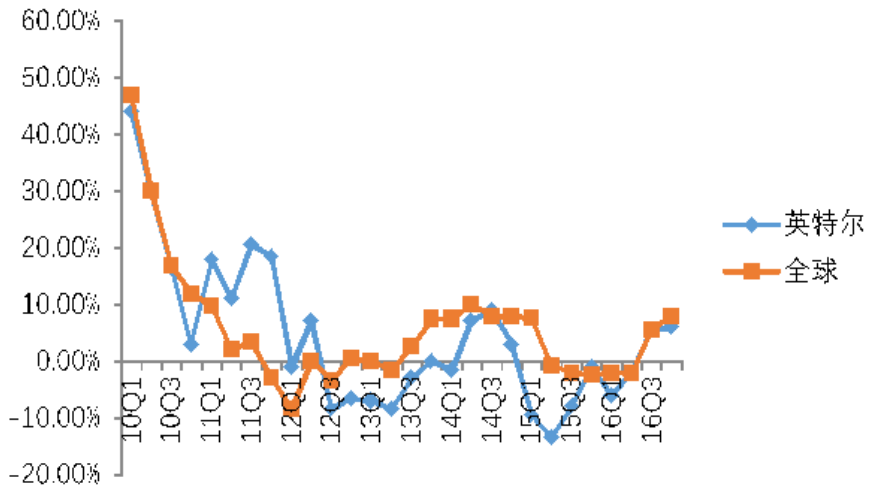
图 10：Intel 总产品收入 VS PC 端收入



资料来源：Intel，天风证券研究所

进入 2010 年后，英特尔的处理器增速同半导体行业基本协同一致，毫无疑问超越行业增速的增长已经需要新的应用拉动。摩尔定律支撑了 10 多年的快速增长再出现边际改善的增长需要重新审视。

图 11：Intel VS 全球半导体 增速



资料来源：Intel，天风证券研究所

2.2. ARM——开放生态下移动时代的新王加冕

本节重点阐述 ARM 在移动时代的芯片架构产品以及此架构下公司以及行业的变化。

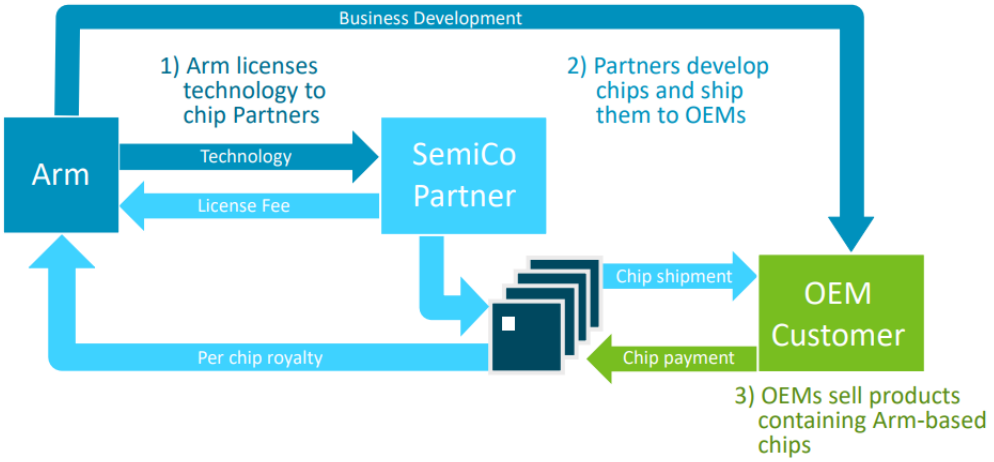
2.2.1. ARM 公司简介

ARM 公司是全球领先的半导体知识产权 (IP) 提供商，专门从事基于 RISC 技术芯片设计开发，并因此在数字电子产品的开发中处于核心地位。公司的前身 Acorn 于 1978 年在伦敦正式成立。1990 年 ARM 从 Acorn 分拆出来。得益于 20 世纪 90 年代手机的快速发展，基于 ARM 技术的芯片出货量飞速增长，**并于 2017 年宣布正式达成 1000 亿芯片出货量的里程碑**。2016 年 7 月，日本软银以 320 亿美元收购了 ARM。

ARM 本身不直接从事芯片生产，只设计 IP，包括指令集架构、微处理器、图形核心和互连架构，依靠转让设计许可由合作公司生产各具特色的芯片，目前它在世界范围有超过 1100 个的合作伙伴。

ARM 的创新型商业模式为公司带来了丰厚的回报率：它既使得 ARM 技术获得更多的第三方工具、制造、软件的支持，又使整个系统成本降低，使产品更容易进入市场被消费者所接受，更具有竞争力。正因为 ARM 的 IP 多种多样以及支持基于 ARM 的解决方案的芯片和软件体系十分庞大，全球领先的原始设备制造商 (OEM) 都在广泛使用 ARM 技术，**因此 ARM 得以在智能手机、平板上一枝独秀，全世界超过 95%的智能手机都采用 ARM 架构。**

图 12：ARM 的商业模式



资料来源：ARM，天风证券研究所

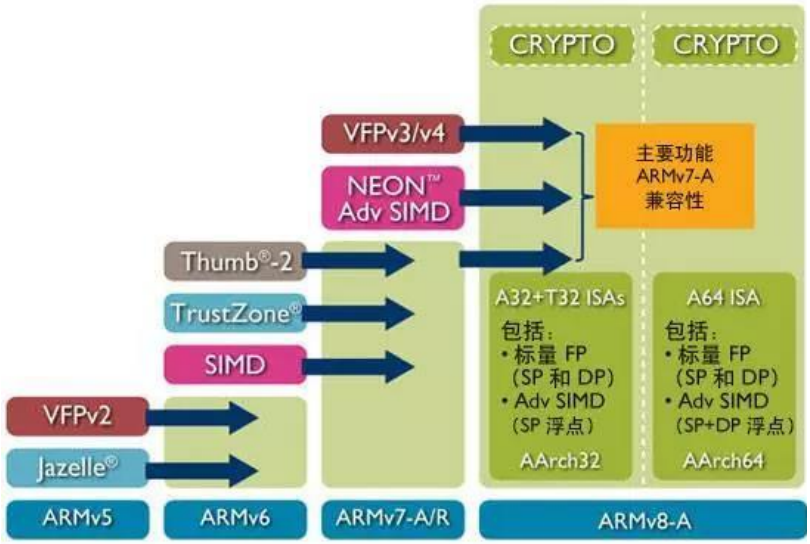
2.2.2. ARM 架构——重新塑造移动智能时代

ARM 沿用了冯诺伊曼架构，在性能和功耗上做到了更加平衡。在底层架构没有发生根本性　　变革的情况下，在架构的横向延伸上寻找到了技术的转换，从而实现了智能手机时代移动端的产品阶跃。

处理器架构在根源上看 ARM 延续了 X86 的底层架构。正如我们在之前讨论架构时指出，处理器一般分为取指令，译码，发射，执行，写回五个步骤。而我们说的访存，指的是访问数据，不是指令抓取。访问数据的指令在前三步没有什么特殊，在第四步，它会被发送到存取单元，等待完成。与 X86 不同的是在指令集方面，ARM 架构过去称作进阶精简指令机器（Advanced RISC Machine），更早时期被称作 Acorn RISC Machine，是 32 位精简指令集（RISC）处理器架构，被广泛地使用在嵌入式系统设计中。在应用场景上有所不同。

ARM 指令集架构的主要特点：一是体积小、低功耗、低成本、高性能，因此 ARM 处理器非常适用于移动通讯领域；二是大量使用寄存器且大多数数据操作都在寄存器中完成，指令执行速度更快；三是寻址方式灵活简单，执行效率高；四是指令长度固定，可通过多流水线方式提高处理效率。

图 13：ARM 架构的发展



资料来源：ARM，天风证券研究所

表 1：ARM 架构汇总

架构	代表处理器	简介
ARM V1	ARM1	
ARM V2	ARM2、ARM3	该版本架构对 V1 进行了扩展，包含了对 32 位乘法指令和协处理器指令的支持。版本 2a 是版本 2 的变种，ARM3 芯片采用了版本 2a，是第一片采用 Cache 的 ARM 处理器。
ARM V3	ARM6、ARM7	ARM 作为独立的公司，在 1990 年设计的第一个微处理器采用的就是版本 3 的 ARM6。它作为 IP 核、独立的处理器、具有片上高速缓存、MMCU 和写缓冲的集成 CPU。变种版本有 3G 和 3M。版本 3G 是不与版本 2a 相兼容的版本 3。版本 3M 引入了有符号和无符号数乘法和乘加指令。
ARM V4	ARM7-TDMI，ARM720-T，ARM9-TDMI，ARM920-T，ARM940-T 等	V4 版架构在 V3 版上作了进一步扩充，V4 版架构是目前应用最广的 ARM 架构。V4 首次增加 Thumb 指令集，不再强制要求与 26 位地址空间兼容，而且还明确了哪些指令会引起未定义指令异常。
ARM V5	ARMv5TE 指令集：ARM9-E-S，ARM1020-E，ARM940-T 等；ARMv5EJ 指令集：ARM926-EJ-S，ARM7-EJ-S，ARM1026-EJ-S 等	V5 版架构是在 V4 版基础上增加了一些新的指令，包括带有链接和交换的转移 BLX 指令；计数前导零 CLZ 指令；BRK 中断指令；增加了数字信号处理指令（V5TE 版）；为协处理器增加更多可选择的指令。
ARM V6	ARM1136-J(F)-S，ARM1156-J(F)-S，ARM1176-J(F)-S，ARM11 MPCore 等	V6 版架构于 2001 年正式发布，首先被应用在 ARM11 处理器。V6 版架构在降低耗电量的同时，还强化了图形处理性能。它还引进了包括单指令多数据(SIMD) 运算在内的一系列新功能。通过追加有效进行多媒体处理的 SIMD（Single Instruction,Multiple Data，单指令多数据）功能，将语音及图像的处理功能提高到了原型机的 4 倍。此外，还引进了作为 ARMv6 体系结构的变体的 Thumb-2 和 TrustZone 技术。
ARM V7	Cortex-A、Cortex-M、Cortex-R 等	全新的 ARMv7 架构是在 ARMv6 架构的基础上诞生的。ARMv7 架构采用了 Thumb-2 技术，它是在 ARM 的 Thumb 代码压缩技术的基础上发展出来的，并且保持了对已存 ARM 解决方案的完整的代码兼容性。此外，ARMv7 还支持改良的运行环境，来迎合不断增加的 JIT 和 DAC 技术的使用。ARMv7

		架构还包括 NEON™ 技术扩展，可将 DSP 和媒体处理吞吐量提升高达 400%，并提供改进的浮点支持以满足下一代 3D 图形和游戏以及传统嵌入式控制应用的需要。
ARM V8	Cortex-A23、Cortex-A57、Cortex-A53、Cortex-R52、Cortex-M23、Cortex-M33 等	ARMv8 是 ARM 公司的首款支持 64 位指令集的处理器架构，可在 32 位 和 64 位 之间切换。由于 ARM 处理器的授权内核被广泛用于手机等诸多电子产品，故 ARMv8 架构作为下一代处理器的核心技术而受到普遍关注。ARMv8 是在 32 位 ARM 架构上进行开发的，主要被用于对扩展虚拟地址和 64 位数据处理技术有更高要求的产品领域。ARMv8 是近 20 年来，ARM 架构变动最大的一次。它引入的 Execution State、Exception Level、Security State 等新特性，已经和我们对旧的 ARM 架构的认知。

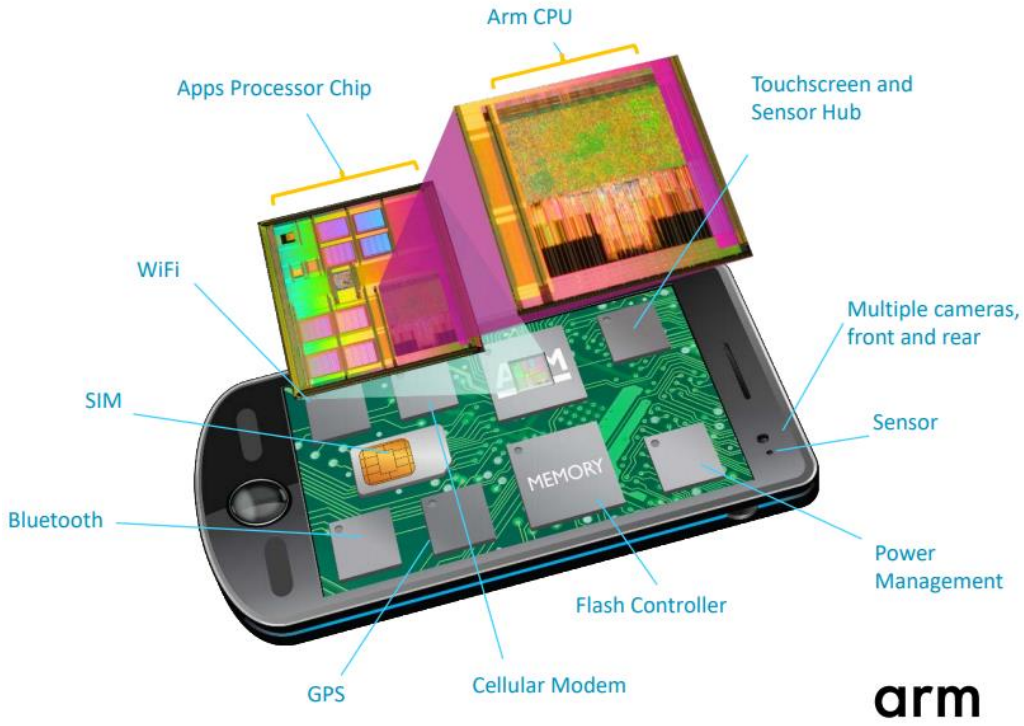
资料来源：ARM、满天芯，天风证券研究所

2.2.3. 生态的建立和商业模式的转变——ARM 重塑了行业

ARM 的商业模式值得真正的关注。ARM 通过授权和版税来赚取收入。使用 ARM 的授权，跟据流片的次数，可以付一次流片的费用，也可以买三年内无限次流片，更可以永久买断。芯片量产后，根据产量，会按百分比收一点版税。**Intel 通过售卖自己的芯片来赢得终端客户和市场，而 ARM 则是通过授权让全世界的芯片制造商使用自家的产品来推广。**ARM 的商业模式之所以在智能手机时代能够推广，是因为移动端的生态更为开放，自上而下的生态建立，不仅是芯片开发者，也包括软件开发者，都被构建在生态的范围内。

智能移动设备上包含多件 ARM 的处理器/技术，每当智能手机上新增一个功能时，就为新的 ARM 知识产权带来了新的机会。2016 年，ARM 在移动应用处理器（包括智能手机、平板电脑和笔记本电脑）上，根据量的测算，其市场份额高达 90%，同时 **ARM 估计移动应用处理器规模将从 2016 年的 200 亿美元增长到 2025 年的 300 亿美元。**

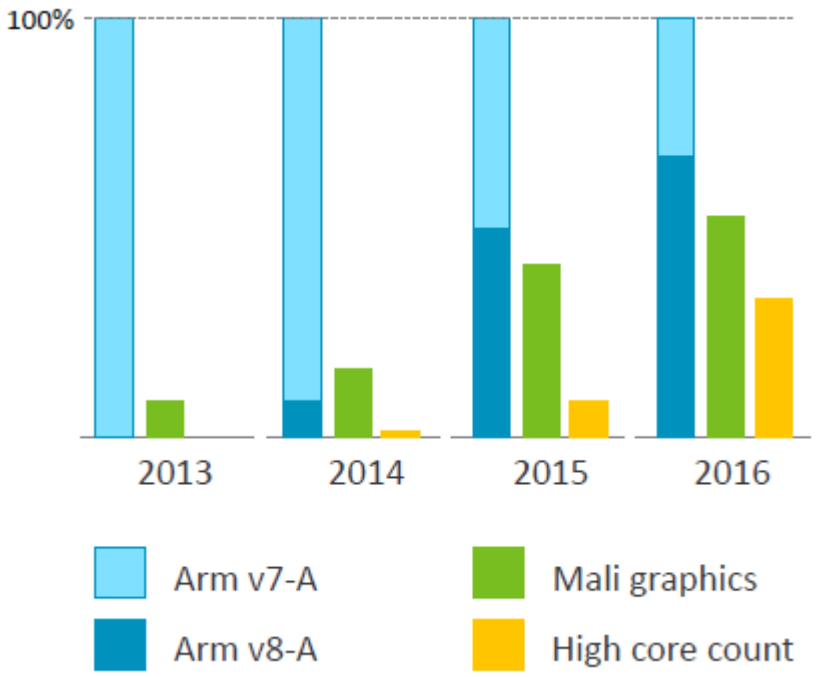
图 14：高级消费电子产品正在结合更多的 ARM 技术



资料来源：ARM，天风证券研究所

2016 年，ARM 各项技术在智能手机领域都有良好的渗透率：ARMv7-A 技术早已完全渗入，ARMv8-A 技术渗透率达到 70%，Mali graphics 达到 50%，高核数技术（high core count）则为 35%.

图 15：ARM 在智能手机中的成分



资料来源：ARM，天风证券研究所

根据 ARM 的预测，到 2025 年为止，智能手机设备的 CAGR 为 3%左右，而 ARM 在这一板块的专利收入将会以大于 5%的 CAGR 上涨。

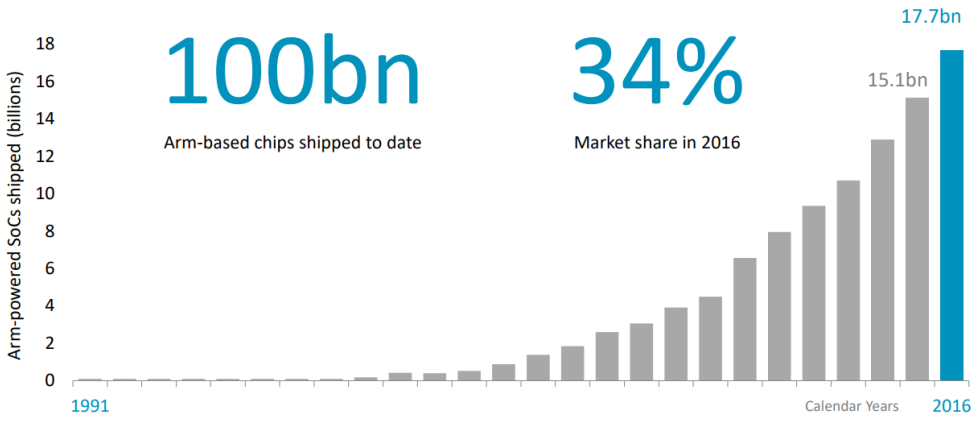
表 2：2020 年 ARM 在各类型智能手机部件中的可获得的单机收入

智能手机类型	部件	ASP
高端智能机	应用处理器	$15-$20
	连接传感器	$5-$10
中端智能机	应用处理器	$5-$15
	连接传感器	$2-$3
低端智能机	应用处理器	<$5
	连接传感器	$1-$2

资料来源：ARM，天风证券研究所

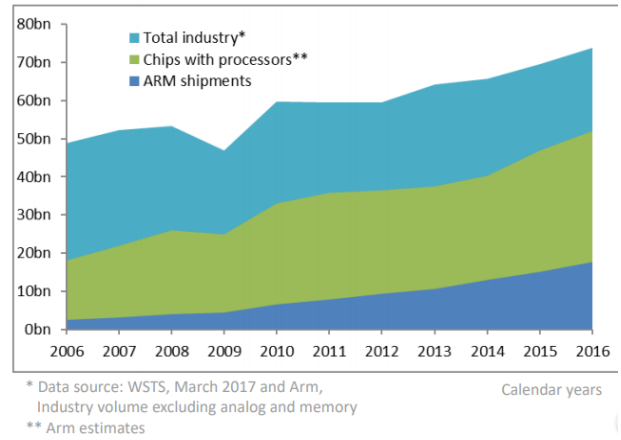
ARM 的累计出货量已经超过 1000 亿支，2016 年全年发出的基于 ARM 技术芯片达到 177 亿，发货量在过去 5 年时间中 CAGR 将近 15%。ARM 的增长完美契合了智能手机的快速增长 10 年。

图 16：基于 ARM 芯片的出货量



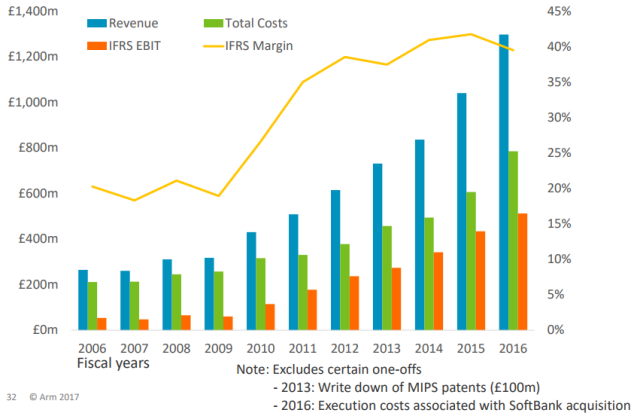
资料来源：ARM，天风证券研究所

图 17：ARM 在载有处理器芯片部门的市场占有率



资料来源：ARM，天风证券研究所

图 18：ARM 收入及利润情况



资料来源：ARM，天风证券研究所

3. 人工智能芯片——新架构的异军突起

观察人工智能系统的搭建，以目前的架构而言，主要是以各种加速器来实现深度学习算法。本章讨论各种加速器的形式和实现，并探讨加速器变革下引发的行业深层次转变，并从 2 个维度给出详细的测算人工智能芯片的潜在空间

首先我们必须描述人工智能对芯片的诉求，深度学习的目标是模仿人类神经网络感知外部世界的方法。深度学习算法的实现是人工智能芯片需要完成的任务。在算法没有发生质变的前提下，追根溯源，所有的加速器芯片都是为了实现算法而设计。

表 3：人工智能系统

架构单元	芯片功能	芯片类型	芯片厂商
处理器	收发指令，逻辑运算	CPU	Intel，ARM，AMD
存储器	数据/指令读写	NAND、DRAM	三星、海力士、美光
加速器	大规模并行计算	GPU、FPGA、ASIC	Nvidia，Google、Movidius
通信接口	信息交换	WiFi、Bluetooth	Avago，Skyworks，CSR

资料来源：Wind，天风证券研究所

我们整理了人工智能芯片相关的类型和产业链公司，**传统的芯片厂商/生态的建立者/新进入者。 传统的芯片制造厂商**：Intel，Nvidia 和 AMD。他们的优势在于在已有架构上对人工智能的延伸，对于硬件的理解会优于竞争对手，但也会困顿于架构的囹圄；2 上层生态的构建者进入芯片设计，比如苹果和 Google，优势在于根据生态灵活开发定制各类 ASIC，专用性强；新进入者，某些全新的架构比如神经网络芯片的寒武纪，因为是全新的市场开拓，具有后发先至的可能。**新进入者的机会，因为是个全新的架构机会，将有机会诞生独角兽。**

图 19：人工智能芯片产业链

资料来源：Wind，半导体行业观察，天风证券研究所

3.1. GPU——旧瓶装新酒

GPU 使用 SIMD（单指令多数据流）来让多个执行单元以同样的步伐来处理不同的数据，原本用于处理图像数据，但其离散化和分布式的特征，以及用矩阵运算替代布尔运算适合处理深度学习所需要的非线性离散数据。作为加速器的使用，可以实现深度学习算法。**但注意的是，GPU 架构依然基于冯诺伊曼。**

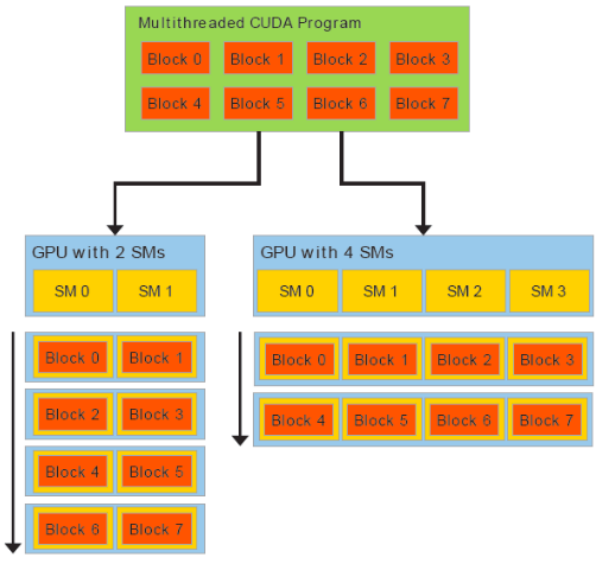
我们以 GPU 和 CPU 的对比来说明 GPU 所具有的架构特点。GPU 由并行计算单元和控制单元以及存储单元构成 GPU 拥有大量的核（多达几千个核）和大量的高速内存，擅长做类似图像处理的并行计算，以矩阵的分布式形式来实现计算。同 CPU 不同的是，GPU 的计算单元明显增多，特别适合大规模并行计算。

图 20：CPU VS GPU 架构

资料来源：NVIDIA，天风证券研究所

注意 GPU 并行计算架构，其中的流处理器组（SMs）类似一个 CPU 核，多个流处理器组可实现数据的同时运算。因此，GPU 主要适用于在数据层呈现很高的并行特性（data-parallelism）的应用。

图 21：GPU 架构流程



资料来源：NVIDIA，天风证券研究所

CPU 和 GPU 本身架构方式和运算目的不同导致了 CPU 和 GPU 之间的不同，主要不同点列举如下

图 22：CPU VS GPU

	CPU	GPU
架构区别	70%晶体管用来构建 Cache 还有一部分控制单元，负责逻辑算数的部分并不多	整个就是一个庞大的计算阵列（包括 alu 和 shader 填充
	非常依赖 Cache	不依赖 Cache
	逻辑核心复杂	逻辑核心简单
计算目的	适合串行	适合大规模并行
	运算复杂度高	运算复杂度低

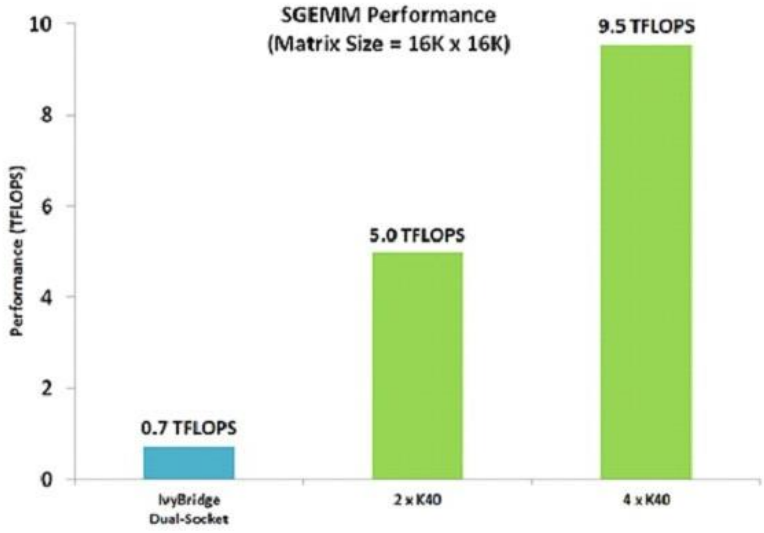
资料来源：Intel，天风证券研究所

深度学习是利用复杂的多级「深度」神经网络来打造一些系统，这些系统能够从海量的未标记训练数据中进行特征检测。因为 GPU 可以平行处理大量琐碎信息。深度学习所依赖的是神经系统网络——与人类大脑神经高度相似的网络——而这种网络出现的目的，就是要在高速的状态下分析海量的数据。GPU 擅长的是海量数据的快速处理

GPU 的特征决定了其特别适合做训练。机器学习的广泛应用: 海量训练数据的出现以及 GPU 计算所提供的强大而高效的并行计算。 人们利用 GPU 来训练这些深度神经网络，所使用的训练集大得多，所耗费的时间大幅缩短，占用的数据中心基础设施也少得多。 GPU 还被用于运行这些机器学习训练模型，以便在云端进行分类和预测，从而在耗费功率更低、占用基础设施更少的情况下能够支持远比从前更大的数据量和吞吐量。

与单纯使用 CPU 的做法相比，GPU 具有数以千计的计算核心、可实现 10-100 倍应用吞吐量，因此 GPU 已经成为数据科学家处理大数据的处理器。

图 23：GPU 性能



资料来源：NVIDIA，天风证券研究所

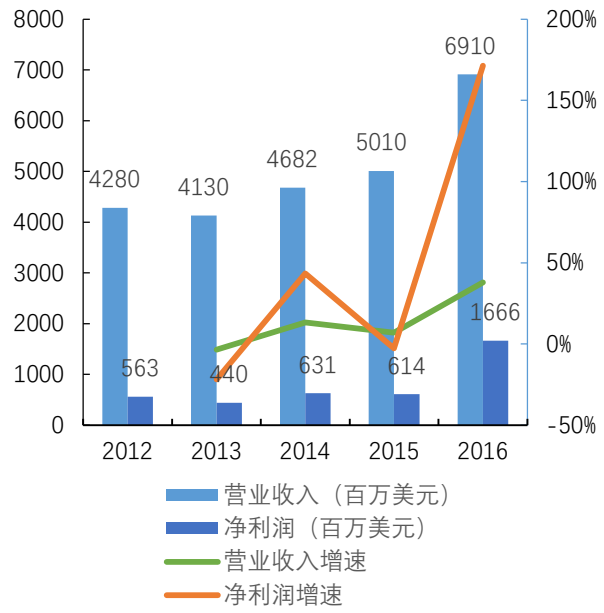
3.1.1. GPU 芯片王者——NVIDIA

NVIDIA 是一家以设计 GPU 芯片为主业的半导体公司，其主要产品从应用领域划分，包括 GPU（如游戏图形处理器 GeForce GPU，深度学习处理器 Tesla，图形处理器 GRID 等）和 Tegra Processor（用于车载，包括 DRIVE PX 和 SHIELD）等。GPU 芯片构成公司最主要收入来源，2017 年上半年，GPU 贡献收入 34.59 亿美元，占公司总收入的 83%；Tegra Processor

贡献收入 6.65 亿美元，占比 16%，其他部分贡献收入 1%。

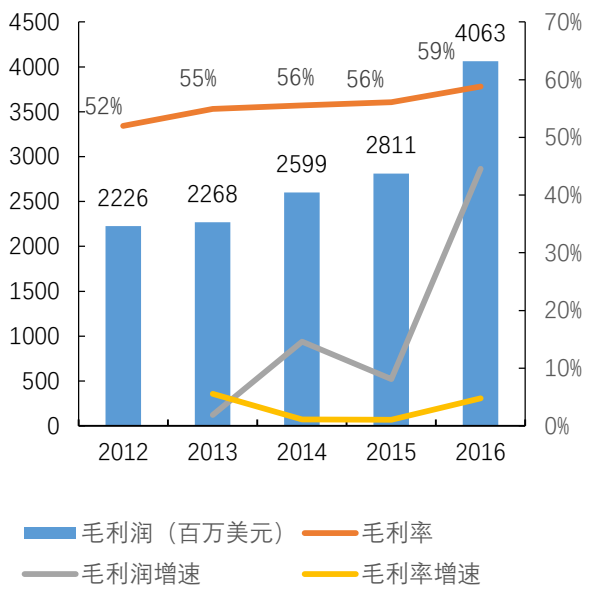
公司业绩稳定，营业收入除 2013 年略有下降外，2012-2016 年均实现稳步增长，从 42.80 亿美元增至 69.10 亿美元，CAGR 为 10.05%；2016 年公司实现净利 16.66 亿美元，相较于 2012 年的 5.63 亿美元，CAGR 达 24.23%。毛利润方面，公司毛利润从 2012 年的 22.26 亿美元增至 2016 年的 40.63 亿美元，实现稳步增长，毛利率维持在 50%以上。

图 24：2012-2016 年 NVIDIA 营收情况



资料来源：公司年报，天风证券研究所

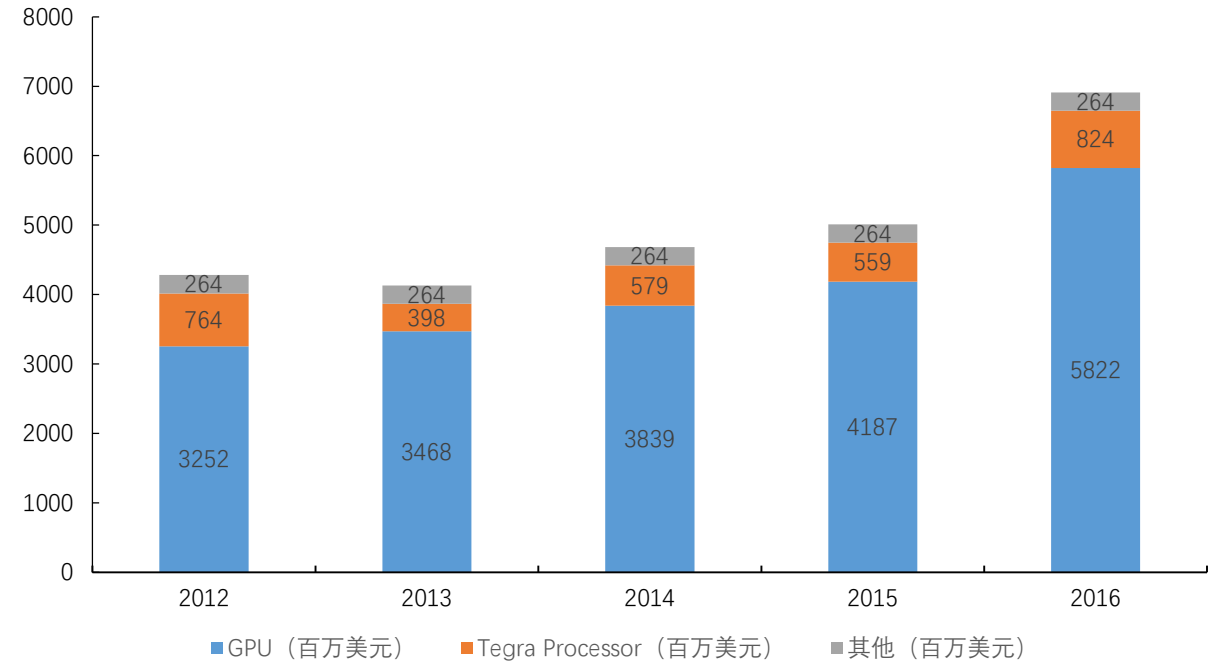
图 25：2012-2016 年 NVIDIA 毛利情况



资料来源：公司年报，天风证券研究所

从收入构成来看，公司 GPU 芯片业务从 2012 年的 32.52 亿美元增至 2016 年的 58.22 亿美元，实现稳步增长，GPU 业务在收入结构中占比稳定在 76%以上。

图 26：NVIDIA2017 年上半年收入构成



资料来源：公司年报，天风证券研究所

在高性能计算机、深度学习、人工智能等领域，NVIDIA 的 GPU 芯片有十分关键的作用。NVIDIA 的 CUBA 技术，大幅度提高了纯 CPU 构成的超级计算机的性能。人工智能和深度学习需要大量的浮点计算，在高性能计算领域，GPU 需求在不断增强。目前 NVIDIA 的高性能显卡已经占有 84%的市场份额。亚马逊的 AWS，Facebook，Google 等世界一级数据中心都需要用 NVIDIA 的 Tesla 芯片，随着云计算和人工智能的不断发展，

我们认为 NVIDIA 的 GPU 芯片业务在未来将继续维持增长态势，我们分拆每个领域的出货量，预计将从 2016 年的 3602 万颗增至 2018 年的 4175 万颗。

表 4：NVIDIA 出货芯片预测（单位：百万颗）

	2016	2017	2018
游戏显卡	30	31	32
高性能计算处理器	3	3.2	3.5
云端加速器	0.05	0.11	0.66
中端汽车芯片	0.63	0.65	0.67
高端汽车芯片	2.34	3.87	4.92
总计	36.02	38.84	41.75

资料来源：Wind，天风证券研究所

3.2. FPGA——紧追 GPU 的步伐

FPGA 是用于解决专用集成电路的一种方案。专用集成电路是为特定用户或特定电子系统制作的集成电路。人工智能算法所需要的复杂并行电路的设计思路适合用 FPGA 实现。FPGA 计算芯片布满“逻辑单元阵列”，内部包括可配置逻辑模块，输入输出模块和内部连线三个部分，相互之间既可实现组合逻辑功能又可实现时序逻辑功能的独立基本逻辑单元。注意 FPGA 与传统冯诺伊曼架构的最大不同之处在于内存的访问。**FPGA 在本质上是用硬件来实现软件的算法，因此在实现复杂算法方面有一些难度。**

表 5：冯诺伊曼架构 VS FPGA 架构

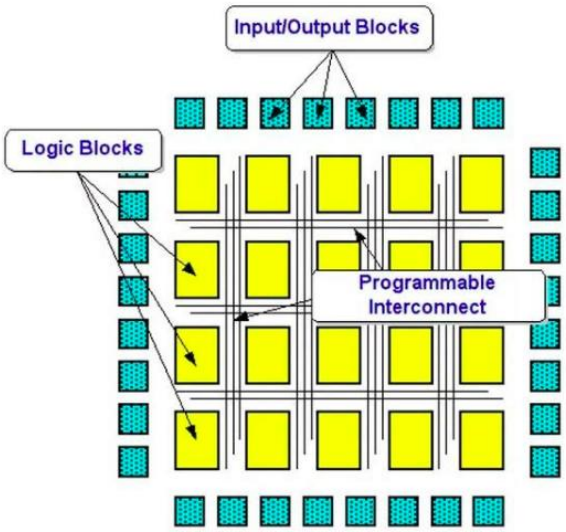
	冯诺伊曼架构	FPGA 架构
内存	共享	专用
访问顺序	访问仲裁，依次调用	无须仲裁和缓存
处理流程	串行	并行
计算效率	低	高

资料来源：Wind，天风证券研究所

架构方面，FPGA 拥有大量的可编程逻辑单元，可以根据客户定制来做针对性的算法设计。除此以外，在处理海量数据的时候，FPGA 相比于 CPU 和 GPU，**独到的优势在于：FPGA**

更接近 IO。换句话说，FPGA 是硬件底层的架构。比如，数据采用 GPU 计算，它先要进入内存，并在 CPU 指令下拷入 GPU 内存，在那边执行结束后再拷到内存被 CPU 继续处理，这过程并没有时间优势;而使用 FPGA 的话，数据 I/O 接口进入 FPGA，在里面解帧后进行数据处理或预处理，然后通过 PCIE 接口送入内存让 CPU 处理，一些很底层的工作已经被 FPGA 处理完毕了（FPGA 扮演协处理器的角色），且积累到一定数量后以 DMA 形式传输到内存，以中断通知 CPU 来处理，这样效率就高得多。

图 27：FPGA 架构



资料来源：人工智能实验室（AiLab），天风证券研究所

性能方面，虽然 FPGA 的频率一般比 CPU 低，但 CPU 是通用处理器，做某个特定运算(如信号处理，图像处理)可能需要很多个时钟周期，而 FPGA 可以通过编程重组电路，直接生成专用电路，加上电路并行性，可能做这个特定运算只需要一个时钟周期。比如一般 CPU 每次只能处理 4 到 8 个指令，在 FPGA 上使用数据并行的方法可以每次处理 256 个或者更多的指令，让 FPGA 可以处理比 CPU 多很多的数据量。举个例子，CPU 主频 3GHz，FPGA 主频 200MHz，若做某个特定运算 CPU 需要 30 个时钟周期，FPGA 只需一个，则耗时情况：CPU：30/3GHz ＝10ns；FPGA：1/200MHz ＝5ns。可以看到，FPGA 做这个特定运算速度比 CPU 块，能帮助加速。

FPGA 相对于 CPU 与 GPU 有明显的能耗优势，主要有两个原因。首先，在 FPGA 中没有取指令与指令译码操作，在 Intel 的 CPU 里面，由于使用的是 CISC 架构，仅仅译码就占整个芯片能耗的 50%；在 GPU 里面，取指令与译码也消耗了 10%～20%的能耗。其次，FPGA 的主频比 CPU 与 GPU 低很多，通常 CPU 与 GPU 都在 1GHz 到 3GHz 之间，而 FPGA 的主频一般在 500MHz 以下。如此大的频率差使得 FPGA 消耗的能耗远低于 CPU 与 GPU。

图 28：FPGA VS CPU 性能

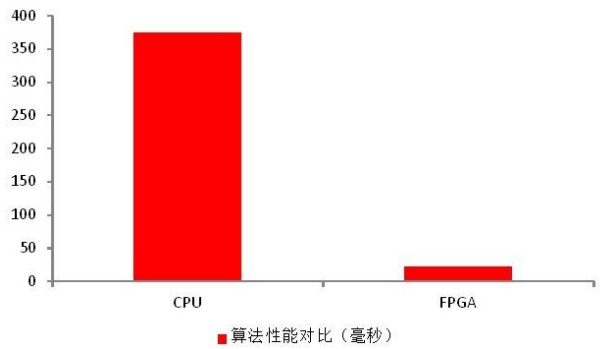
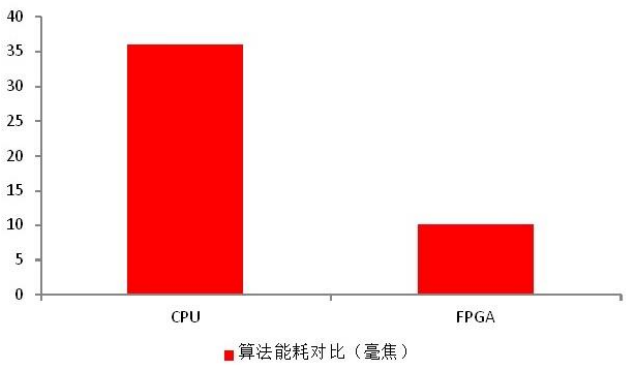


图 29：FPGA VS CPU 功耗



资料来源：Altera，天风证券研究所

资料来源：Altera，天风证券研究所

Intel167 亿美元收购 Altera，IBM 与 Xilinx 的合作，都昭示着 FPGA 领域的变革，未来也将很快看到 FPGA 与个人应用和数据中心应用的整合

根据 Altera 内部文件显示，Altera 很早就在研发使用 FPGA 针对深度学习算法的应用，并在 2015 年 Intel 的论坛上展示了产品的性能。结论是在功耗和性能上相对同等级的 CPU,有较大的优势。CPU+FPGA 在人工智能深度学习领域，将会是未来的一个重要发展方向

图 30：FPGA 性能

Projected AlexNet Performance and Power

CNN Classification Platform	Power (W)	Performance (Image/s)	Efficiency (Images/Sec/W)
E52699 Dual Xeon CPU (18 core per Xeon)	321	1320	4.11
PCIe w/Dual Arria 10 1150	130*	1200	9.27

资料来源：Altera，天风证券研究所

3.3. ASIC——定制化的专用人工智能芯片

ASIC（专用定制芯片）是为实现特定要求而定制的芯片，具有功耗低、可靠性高、性能高、体积小等优点，但不可编程，可扩展性不及 FPGA，**尤其适合适合高性能/低功耗的移动端。**目前，VPU 和 TPU 都是基于 ASIC 架构的设计。

我们梳理针对图像和语音这两方面的人工智能定制芯片，目前主要有专用于图像处理的 VPU，以及针对语音识别的 FAGA 和 TPU 芯片。

表 6：图像应用和语音应用人工智能定制芯片

	应用举例	定制芯片
图像应用	自动驾驶、人脸识别、无人机等	VPU、GPU
声音应用	语音识别、自然语言处理、实时翻译等	TPU

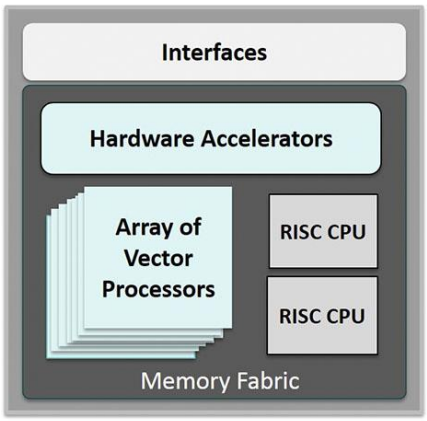
资料来源：Google，天风证券研究所

3.3.1. VPU——你是我的眼

VPU 是专门为图像处理和视觉处理设计的定制芯片。根据特定算法来实现定制化的芯片架构，实现特定的图像处理能力，提高效率，是 VPU 的基础理念。集成在摄像头中的 VPU，直接对输入图像进行识别理解，消除了存储器的读写操作。相较主流的移动处理芯片（集成 GPU 的 SoC），VPU 的尺寸更小，视觉处理运算的效能更高。

以 Movidiu 公司产品 Myriad2 为例，VPU 芯片包括接口电路（Interfaces）、硬件加速器(Hardware Accelerators)，矢量处理器阵列(Array of Vector Processors)，精简指令集的 CPU(RISC CPU)等部分。接口电路支持多路摄像头传感器等外部设备，硬件加速器可以迅速的提高运算处理速度，矢量处理器阵列专门针对机器视觉，精简指令集的 CPU(RISC CPU)主要进行任务分配。

图 31：VPU 架构



Myriad 2 Vision Processor Unit (VPU)

资料来源：Movidius，天风证券研究所

图 32：VPU 模组

资料来源：Movidius，天风证券研究所

表 7：实现原理

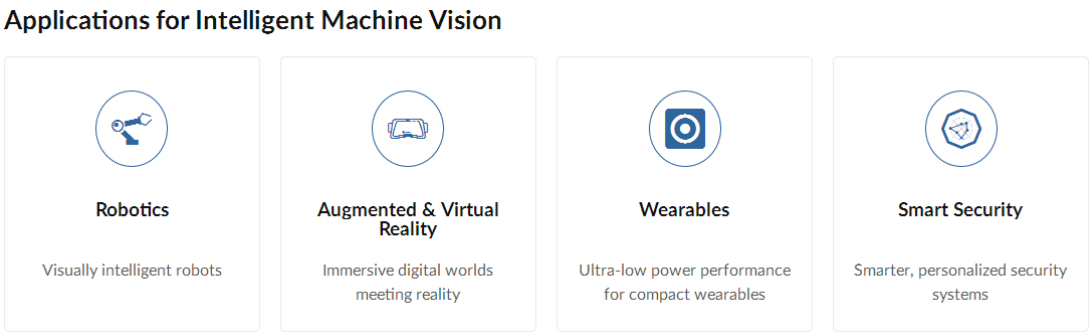
	功能
接口电路	支持多路摄像头传感器，WIFI 设备，SD 卡读写，惯性测量单元等
硬件加速器	支持图像信号和视觉信后的流水线信号处理，而不需再绕回内存进行处理，可以迅速的提高运算处理速度。
矢量处理器阵列	多个 128 位的具有超长指令集的矢量处理器，专门针对机器视觉。
精简指令集的 CPU	32 位的 CPU，进行任务分配

资料来源：Movidius，天风证券研究所

VPU 能够处理各种不同的任务：利用立体摄像机的数据处理深度信息，还有来自声纳传感器的近距离、空间定位，以及用于识别和跟随人的先进光流；它也可以成为虚拟现实、现实增强技术的核心部分，让智能手机以及更便宜的头戴产品达成现如今较为昂贵的系统才能完成的目标。如 HTC Vive，这台设备需要比较诡异的头戴式护目镜，还需要两个激光盒子绘制整个空间，并追踪用户的运动。而装备 VPU 通过移动设备或者耳机就能做到这一点；此外，**具备深度学习能力的 VPU，能够在设备本地就能利用强悍的图像识别计算，设备能够看见和理解周围的世界，不需要检索云端就能做到，避免了延迟的问题**。

目前，VPU 的应用市场有机器人、物联网、智能穿戴设备、智能手机、无人驾驶、无人机等。

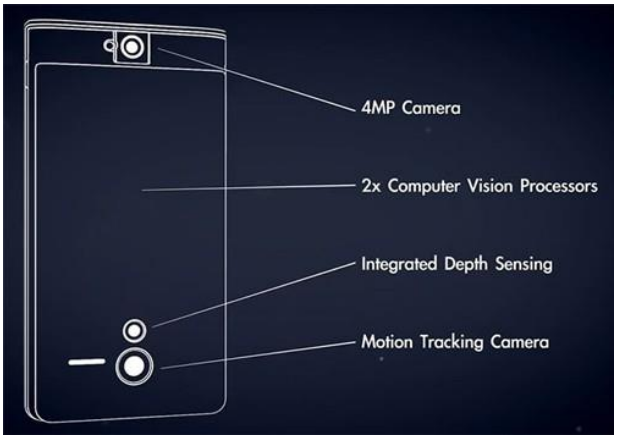
图 33：VPU 应用



资料来源：Movidius，天风证券研究所

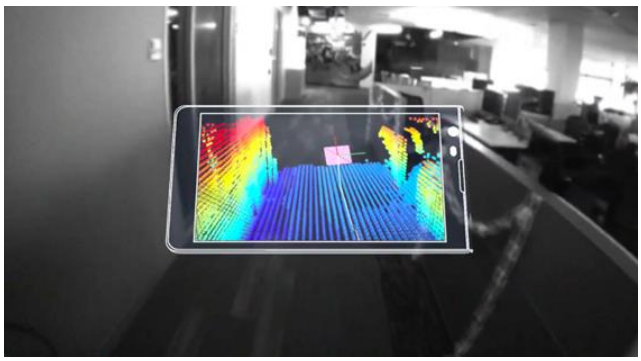
结合光学，在前端实现智能处理识别运算的芯片，正在移动端不断渗透提升。在苹果推出带 3D 感应功能的结构光方案之后，我们预计会深度推动市场在向具有人工智能功能的特定芯片端迈进。VPU 实现了在移动设备端具备 PC 级别的图像处理能力。通常来说这类图像处理芯片能耗非常高，而且也需要电脑支持，但通过 VPU，成功将高级的图像处理方案移植到移动设备中。在前端设备中引入带有 AI 功能的新架构芯片将带来移动端价值量的提升和潜在的变革。

图 34：3D 景深结构



资料来源：Movidius，天风证券研究所

图 35：3D 成像

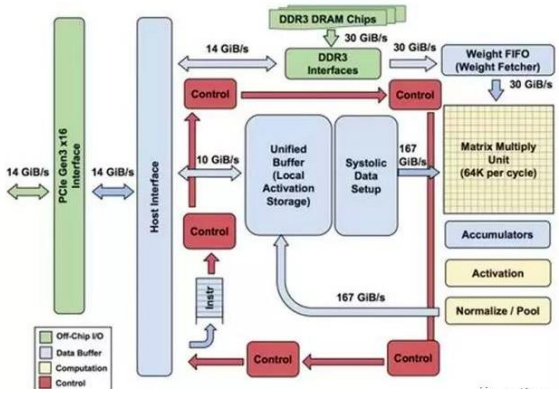


资料来源：Movidius，天风证券研究所

3.3.1. TPU——Google 的野心

TPU (Tensor Processing Unit) 是谷歌的张量处理器，它是一款为机器学习而定制的芯片，经过了专门深度机器学习方面的训练，它有更高效能。

图 3637：Google 公司 TPU 架构



资料来源：Google，天风证券研究所

Google 对 GPU，Intel Xeon E5 v3 CPU 和 TPU 进行了性能对比。在 Google 的测试中，使用 64 位浮点数学运算器的 18 核心运行在 2．3 GHz 的 Haswell Xeon E5－2699 v3 处理器能够处理每秒 1．3 TOPS 的运算，并提供 51GB／秒的内存带宽；Haswell 芯片功耗为 145 瓦，其系统（拥有 256 GB 内存）满载时消耗 455 瓦特。相比之下，TPU 使用 8 位整数数学运算器，拥有 256GB 的主机内存以及 32GB 的内存，能够实现 34GB／秒的内存带宽，处理速度高达 92 TOPS ，这比 Haswell 提升了 71 倍，此外，TPU 服务器的热功率只有 384 瓦。但 TPU 是专为 Google 深度学习语言 Tensor Flow 开发的一种芯片，不具有可扩展性。

图 3839：Google 公司 TPU 性能

Model	Die									Benchmarked Servers				
	mm²	nm	MHz	TDP	Measured		TOPS/s		GB/s	On-Chip Memory	Dies	DRAM Size	TDP	Measured	
					Idle	Busy	8b	FP						Idle	Busy
Haswell E5-2699 v3	662	22	2300	145W	41W	145W	2.6	1.3	51	51 MiB	2	256 GiB	504W	159W	455W
NVIDIA K80 (2 dies/card)	561	28	560	150W	25W	98W	--	2.8	160	8 MiB	8	256 GiB (host) + 12 GiB x 8	1838W	357W	991W
TPU	NA*	28	700	75W	28W	40W	92	--	34	28 MiB	4	256 GiB (host) + 8 GiB x 4	861W	290W	384W

Table 2. Benchmarked servers use Haswell CPUs, K80 GPUs, and TPUs. Haswell has 18 cores, and the K80 has 13 SMX processors. Figure 10 has measured power. The low-power TPU allows for better rack-level density than the high-power GPU. The 8 GiB DRAM per TPU is Weight Memory. GPU Boost mode is not used (Sec. 8). SECDEC and no Boost mode reduce K80 bandwidth from 240 to 160. No Boost mode and single die vs. dual die performance reduces K80 peak TOPS from 8.7 to 2.8. (*The TPU die is ≤ half the Haswell die size.)

资料来源：Google，天风证券研究所

3.4. 人工神经网络芯片

从底层架构的变革角度看，最前沿的革新以深度学习原理打造的人工神经网络芯片。人工神经网络是模仿生物神经网络的计算架构的总称，由若干人工神经元节点互连而成，神经元之间通过突触连接。每个神经元其实是一个激励函数，突触则是记录神经元间联系的强弱权值。

神经网络是多层的，一个神经元函数的输入由与其相连的上一个神经元的输出以及连接突触权重共同决定。所谓训练神经网络，就是通过不断自动调整神经元之间突触权重的过程，直到输出结果稳定正确。然后在输入新数据时，能够根据当前的突触权重计算出输出结果。以此来实现神经网络对已有知识的“学习”。神经网络中存储和处理是一体化的，中间计算结果化身为突触的权重。

冯诺伊曼架构的传统处理器处理神经网络任务时效率低下，是由其本身的架构限制决定的。冯诺伊曼架构存储和处理分离，基本运算为算术和逻辑操作，这两点决定了一个神经元的处理需要成百上千条指令才能完成。以 AlphaGo 为例，总共需要 1202 个 CPU+176 个 CPU。

表 8：冯诺伊曼架构 VS 神经网络芯片架构

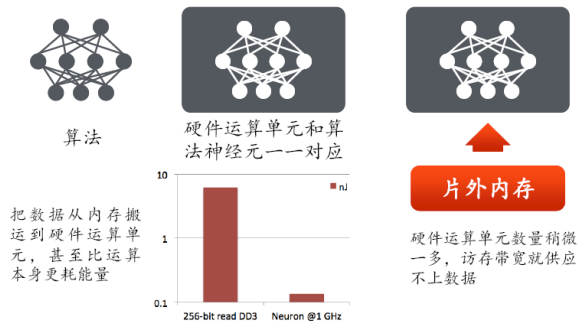
	冯诺伊曼架构	神经网络芯片架构
基本架构	存储/处理分离	存储/处理一体化
运算规则	算术和逻辑操作	激励函数和权重
神经元计算复杂度	成百上千条指令/神经元	一条指令/神经元
计算效率	低	高

资料来源：Wind，天风证券研究所

3.4.1. 寒武纪——真正的不同

真正打造的类脑芯片，寒武纪试图将通过低功耗高性能的架构重塑，颠覆已有的冯诺伊曼架构，实现在移动端/云端的加速器实现。

图 4041：传统硬件处理方式



资料来源：寒武纪资料，天风证券研究所

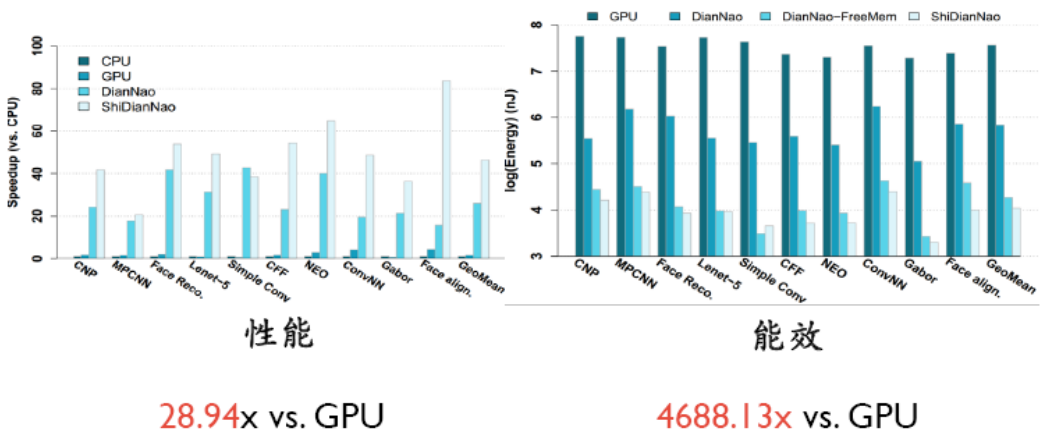
图 4243：寒武纪处理方式

- 小尺度但支持大规模神经网络
- 速度：把访存带宽用起来，尽可能提高性能
- 能耗：通过优化片上存储层次**尽量减少访存次数**

资料来源：寒武纪资料，天风证券研究所

从寒武纪披露的数据来看，其性能远超 GPU 和 CPU。

图 44：寒武纪芯片性能/能效



资料来源：寒武纪资料，天风证券研究所

寒武纪试图将代表性智能算法的处理速度和性能功耗比提升一万倍，在移动端实时完成图像语音和文本的理解和识别，更为重要的是通过实时训练，还能不断进化提升能力，真正实现超越。

图 45：终端和移动端

资料来源：寒武纪资料，天风证券研究所

4. 从 2 个维度测算人工智能芯片空间

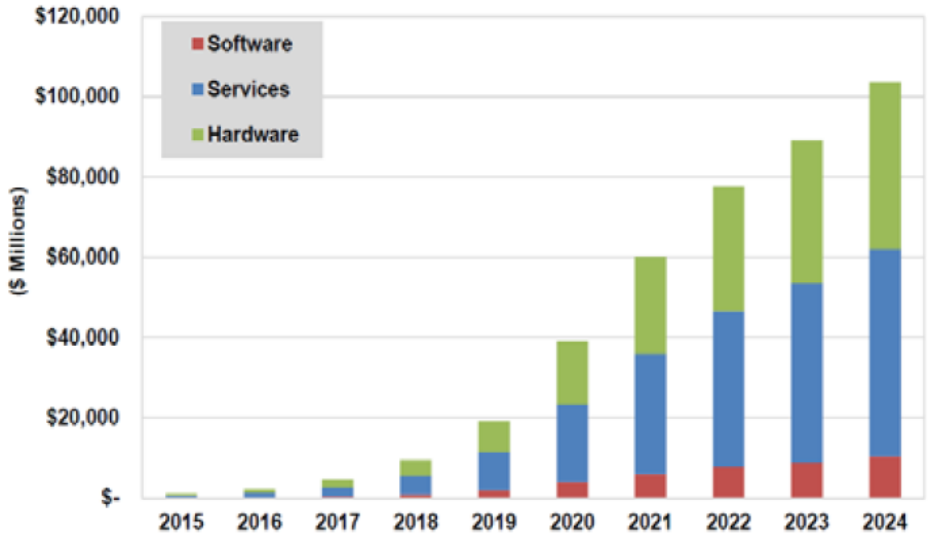
我们在前二章重点讨论了 Intel 和 ARM 的历史发展，认为冯诺伊曼架构带来了计算体系的建立并通过 Intel 实现了最大化；ARM 通过共享 IP 的商业模式带来了更开放的生态体系，实现了软硬件的结合延伸了人类的触角。同时我们认为人工智能芯片将有可能在**摩尔定律放缓维度下引发芯片底层架构重构的变革。**

本章我们重点讨论人工智能芯片的市场空间测算，我们从两个维度来进行估算，给出详细的拆解。

维度一：市场规模反推芯片空间

根据 Nvidia 官方给出的资料统计，到 2020 年，由软件、硬件、服务三者组成的人工智能市场将达到 400 亿美元，其中硬件占到 1/3 强，为 160 亿美元。而硬件的核心是芯片。我们估算硬件的 BOM，芯片会占到 60%，芯片空间将达到 96 亿美元。

图 46：人工智能市场规模

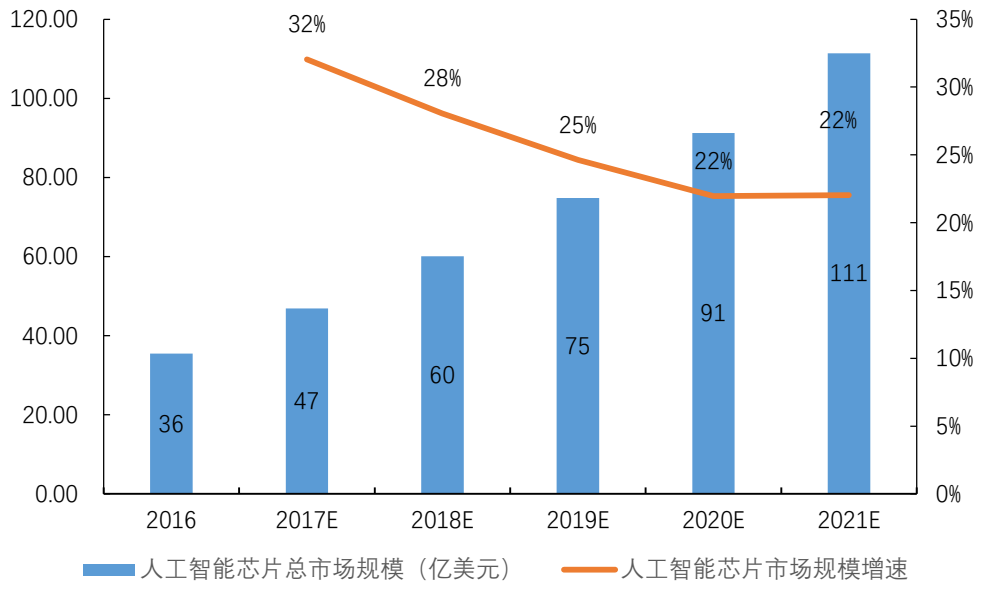


资料来源：NVIDIA，天风证券研究所

维度 2：详细拆分云端/移动端所需人工智能加速器的 BOM

人工智能芯片从用途来看，分为云端加速器芯片和终端（包括智能手机、无人驾驶汽车、）智能芯片。我们基于这两个场景，给出结论，预测至 2021 年，人工智能芯片市场有望达到 111 亿美元，CAGR 达 20.99%。

图 47：人工智能芯片总市场规模



资料来源：Gartner，天风证券研究所

云端加速器详细拆解

具体来看云端方面，根据 Gartner 的统计，到 2020 年，全球云计算市场规模将达到 3834 亿美元，其中，云基础设施服务市场规模达 863.5 亿美元。

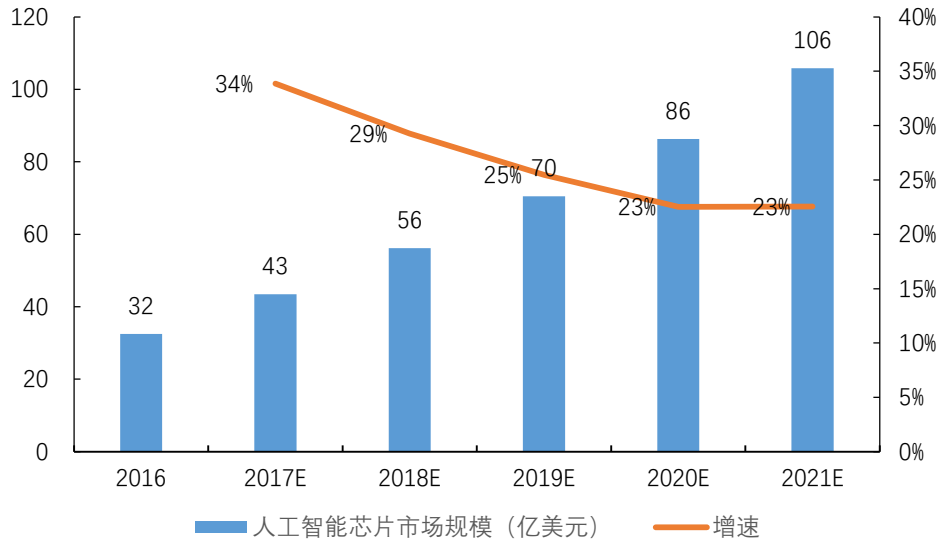
表 48：云端市场规模（单位：百万美元）

	2016	2017E	2018E	2019E	2020E	2021E
云业务流程服务(BPaaS)	40812	43772	47556	51652	56176	61096
云应用程序服务（SaaS）	38567	46331	55143	64870	75734	88417
云管理和安全服务	7150	8768	10427	12159	14004	16129
云广告	90257	104516	118520	133566	151091	170915
云应用基础设施服务	7169	8851	10616	12580	14798	17407
云系统基础设施服务	25290	34603	45559	57897	71552	88428
总计	209245	246841	287821	332724	383355	442393
云基础设施服务市场规模小计	32459	43454	56175	70477	86350	105835

资料来源：Gartner，天风证券研究所

我们假设深度学习相关基础设施占云基础设施的 20%，而其中人工智能芯片占深度学习相关硬件 BOM 的 50%，据此，我们测算云端方面人工智能芯片市场规模将从 2016 年的 32 亿美元增至 2021 年的 106 亿美元，CAGR 达 21.77%。

图 49：云端领域人工智能芯片规模预测



资料来源：Gartner，天风证券研究所

终端加速器市场详细拆解

终端方面，目前人工智能芯片主要应用领域是智能手机、无人驾驶汽车和无人机。我们假设：

1）智能手机全球出货量年均增速 3.3%，主处理器平均价格 15 美元，带人工智能芯片模块占智能手机主处理器 BOM 的 10%

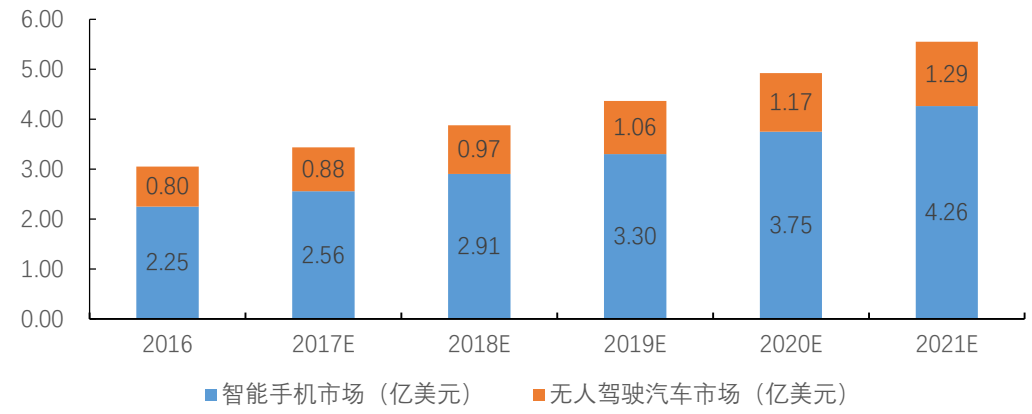
2）带人工智能功能的智能手机渗透率从 2018 的 10%提升到 2020 年的 40%。；

3）无人驾驶汽车市场规模年均增速 10%。因无人驾驶汽车以及其芯片市场均尚未成型，目前成本较高，我们假设芯片成本占总成本的 20%，人工智能芯片占处理器成本的 10%。据此预测终端领域人工智能芯片的市场规模。

据此我们预测，在终端领域，至 2021 年，全球人工智能芯片市场规模由 2016 年的 3.05 亿美元增至 5.55 亿美元，CAGR 为 10.49%。其中，智能手机市场中，人工智能芯片由 2016 年的 2.25 亿美元增至 2021 年的 4.26 亿美元，CAGR 为 11.24%；无人驾驶汽车市场中，人

工智能芯片由 2016 年的 0.80 亿美元增至 2021 年的 1.29 亿美元，CAGR 为 8.27%。

图 50：终端领域人工智能芯片市场规模预测



资料来源：Gartner，天风证券研究所

5. 重点标的

台积电：无论是何种架构的人工智能芯片，都是依赖于台积电最先进制程的代工工艺，在全球只有台积电能够提供 HPC（高性能计算芯片）的工艺平台上，行业的卡位优势已然确立，确定性受益标的。

Intel：收购 Altera，收购 Movidius，CPU+FPGA 方案，Intel 在人工智能领域的布局长远，而通过我们的测算，服务器端将是人工智能芯片未来行业渗透和消耗的重点，而 Intel 在服务器端已经有深厚不可撼动的优势。

NVIDIA：目前人工智能芯片领域的领跑者，深度学习**训练**领域的唯一方案选择。有完整的生态布局，针对云端+汽车自动驾驶，百亿美元新增市场的竞争者。

寒武纪：寒武纪试图将代表性智能算法的处理速度和性能功耗比提升一万倍，在移动端实时完成图像语音和文本的理解和识别，更为重要的是通过实时训练，还能不断进化提升能力，真正实现超越。

富瀚微：国内上市公司智能视频监控领域的前端芯片方案解决商，在前端芯片集成一定的智能算法功能处理。

北京君正：积极进入视频监控领域的芯片方案解决商，曾经的 MIPS 方案芯片设计商，有芯片架构层基因，对标 Movidius。

全志科技：SoC 芯片方案解决商，未来能将 AI 算法模块嵌入 SoC 之中。

分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告所表述的所有观点均准确地反映了我们对标的证券和发行人的个人看法。我们所得报酬的任何部分不曾与，不与，也将不会与本报告中的具体投资建议或观点有直接或间接联系。

一般声明

特别声明

投资评级声明

类别	说明	评级	体系
股票投资评级	自报告日后的 6 个月内，相对同期沪深 300 指数的涨跌幅	买入	预期股价相对收益 20%以上
		增持	预期股价相对收益 10%-20%
		持有	预期股价相对收益-10%-10%
		卖出	预期股价相对收益-10%以下
行业投资评级	自报告日后的 6 个月内，相对同期沪深 300 指数的涨跌幅	强于大市	预期行业指数涨幅 5%以上
		中性	预期行业指数涨幅-5%-5%
		弱于大市	预期行业指数涨幅-5%以下

天风证券研究

北京	武汉	上海	深圳
北京市西城区佟麟阁路 36 号	湖北武汉市武昌区中南路 99	上海市浦东新区兰花路 333	深圳市福田区益田路 4068 号
邮编：100031	号保利广场 A 座 37 楼	号 333 世纪大厦 20 楼	卓越时代广场 36 楼
邮箱：research@tfzq.com	邮编：430071	邮编：201204	邮编：518017
	电话：(8627)-87618889	电话：(8621)-68815388	电话：(86755)-82566970
	传真：(8627)-87618863	传真：(8621)-68812910	传真：(86755)-23913441
	邮箱：research@tfzq.com	邮箱：research@tfzq.com	邮箱：research@tfzq.com