

# 企业研究报告

中科寒武纪

2019年





寒武纪是中科院计算所下孵化的AI芯片研发单位，团队曾参与研发国产“龙芯”芯片，在AI芯片领域发表了多篇领先业界的芯片架构及指令集设计论文，多篇论文获取学界顶级奖项，技术储备雄厚。



公司自2016年成立以来已进行多轮融资，当前估值已达到25亿美元。公司股东除实际控制人及中科院外还有多家带有国资背景的基金，以及以阿里、科大讯飞、联想等为代表的上市公司。



公司自16年成立已经推出了面向终端市场的1A/1H/1M产品及针对云数据中心训练端的MLU100产品，并且开发了针对于NN运算的通用ISA和针对于自家芯片的SDK。1A和1H产品已经被华为麒麟系列芯片SoC所采纳，此外，公司与多家服务器厂商推出针对云端推理场景的智能服务器产品。



目前已有大量科技公司及人才涌入智能芯片领域，其中除近几年成立的大量创业公司外，传统的芯片、科技巨头也在布局该领域。在面对行业竞争及挑战时，相比起终端市场，公司在云计算领域有更强的产品（硬件+软件）及产业竞争优势



推测公司未来有望成为首批科创板挂牌企业，通过上市实现二级市场融资以支持公司进一步发展资金需求。

公司基本情况介绍

1

AI芯片简介

2

公司业务&竞争力分析

3

总结

4

# 公司基本情况介绍

## 公司是中国人工智能芯片领域独角兽企业

寒武纪是全球智能芯片领域的先行者，是全球第一个成功流片并拥有成熟产品的智能芯片公司。公司的使命是打造各类智能云服务器、智能终端以及智能机器人的核心处理器芯片，让机器更好地理解和服务人类。

公司在2016年推出的寒武纪1A处理器是世界首款商用深度学习专用处理器，已应用于数千万智能手机中，入选了第三届世界互联网大会评选的十五项“世界互联网领先科技成果”。公司在2018年推出的MLU100机器学习处理器芯片，运行主流智能算法时性能功耗比全面超越CPU和GPU。目前，寒武纪已与智能产业的各大上下游企业建立了良好的合作关系。

### 公司工商信息概览

项目	内容
公司名称	北京中科寒武纪科技有限公司
成立时间	2016年3月
注册资本	137.65582万元
主营业务	人工智能AI芯片的研发、生产和销售

来源：天眼查；国家企业信用信息公示系统。

# 公司核心团队介绍

团队成员有深厚的中科院计算机所背景，曾参与中国首款通用MIPS架构CPU“龙芯”的研发和设计工作。

## 公司核心团队人员背景介绍

创始人	公司职务	个人履历
陈天石	公司创始人、CEO	2005年毕业于中国科学技术大学少年班获得理学学士学位。2010年获得中国科学技术大学计算机学院工学博士学位，历任中国科学院计算技术研究所助理研究员、副研究员、研究员（正教授）和博士生导师 在处理器架构深耕十余年，曾获得国家自然科学基金委员会“优青”、中国计算机学会科学技术一等奖、CCF-Intel青年学者奖、中国计算机学会优秀博士论文奖
刘少礼博士	公司副总裁	2009年获得南开大学数学学院学士学位，2014年获得中国科学院计算机研究所博士学位，历任中国科学院计算机研究所助理研究员、副研究员（副教授）和硕士生导师。 2016年3月加盟寒武纪，具有8年的芯片研发经验，参与“龙芯”的结构设计，担任世界首个深度学习处理器芯片的主架构师
刘道福博士	公司副总裁	2010年获得中国科学技术大学计算机学院学士学位，2015年获得中国科学院计算技术研究所博士学位，历任中国科学院计算技术研究所助理研究员、高级工程师 2016年加盟寒武纪科技，具有8年芯片研发经验，有负责多款芯片流片的经验。
王在博士	公司副总裁	2006年获得中国科学技术大学计算机学院学士学位，2017年获得中国科学技术大学博士学位，研究方向为人工智能。 2011年至2015年曾任郑州商品交易所任核心交易系统开发工程师，2015年至2016年任中原银行渠道组负责人管理电子银行、手机银行、微信银行及呼叫中心等众多项目，2016年3月加盟寒武纪。
陈云霄	首席科学家	中国科学院计算技术研究所研究员、博士生导师，担任中国科学院脑科学卓越中心特聘研究员、中国科学院大学教授，师从龙芯首席科学家胡伟武，曾任龙芯3号总设计师。曾获得2014年“CCF青年科学家奖”、首届国家自然科学基金“优秀青年基金”、首届国家万人计划“青年拔尖人才”、中国计算机学会青年科学家奖以及中科院青年人才奖，其所带领的科研团队获得全国“青年文明号”和中央国家机关“青年文明号”的称号 曾在国际顶级学术会议及期刊上发表论文60余篇，目前带领实验室研制寒武纪深度学习处理器。

来源：寒武纪官网；网络公开数据。

公司是“中科系”背景芯片研发国家队



# 公司发展历史

## 公司团队有十年的发展史并积攒了丰富的技术储备

公司工商主体虽然成立于2016年，但实际初创团队从2008年即开始了相关技术的研发工作，并先后在世界顶级学术会议上发表了多篇AI芯片硬件架构及指令集论文，其中多篇论文入选了当年的顶级论文。公司自16年成立以来，快速实现了技术的产业化输出，先后推出了用于终端场景的1A/1H/1M系列芯片及云端场景的MLU100芯片及计算卡产品，其中1A/1H应用于华为麒麟系列芯片，MLU系列产品应用于多家云服务器厂商产品，获得了不错的市场反馈。公司自成立以来，已经获得了3轮股权投资，总共融资金额达到了数亿美元，为公司持续研发备足了充分的“弹药”。

### 公司发展历史概览



来源：寒武纪官网；网络公开数据。

公司基本情况介绍

1

AI芯片简介

2

公司业务&竞争力分析

3

总结

4

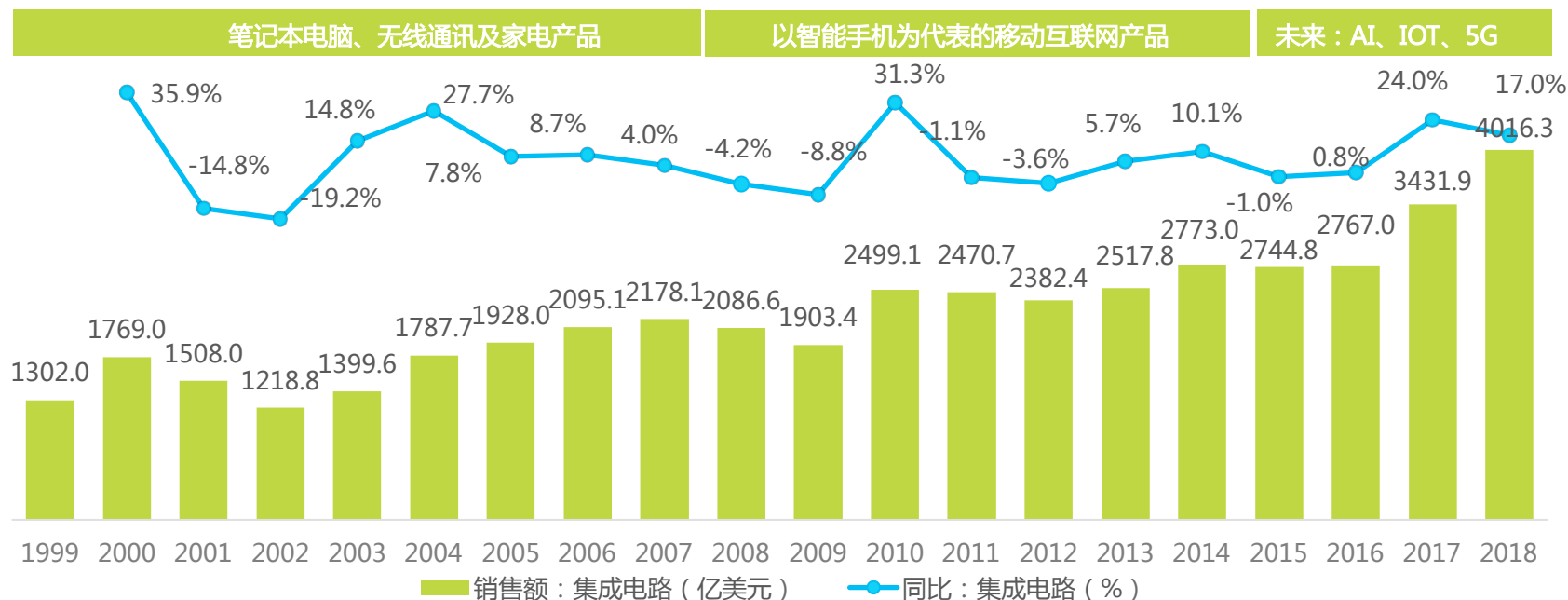


# 半导体行业发展周期

## 人工智能发展推动下一个半导体景气周期的到来

2017年全球半导体销售额达到4086.91亿美元，同比增长20.6%，历史首次突破4000亿美元，全球半导体销售额连续19个月环比增长，景气度创历史新高，而据Garnter预计18年全球半导体市场销售额将达到4510亿美元。半导体产业伴随下游应用市场的发展而不断迎来其上涨周期，21世纪初是通讯、网络以及手机的市场，2004年是MP3、笔记本电脑等3C产品，2010年开始是智能手机市场的快速扩张，2014年开始是云服务器市场以及智能手机共同带动了存储器的发展，而下一个将会是以物联网应用、5G、AI等新兴技术为主带动的半导体景气周期。

### 1999-2018年全球半导体市场销售额



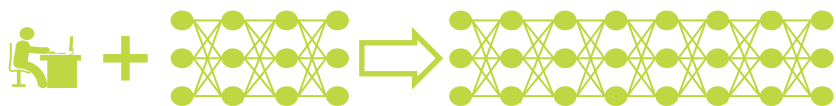
来源：Wind，全球半导体贸易统计；《国之重器、拥抱芯片科技红利！》——中泰证券

# 人工智能行业的第三次崛起

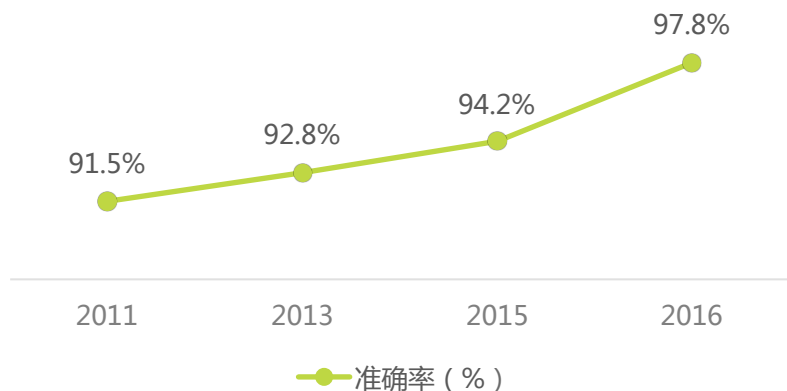
## 数据、算法和算力提升奠定本轮人工智能崛起及商业化应用

1956年，在“达特茅斯会议”上计算机科学家第一次提出了“AI”的概念，但受限于当时的技术水平人工智能仅停留在概念阶段。2016年，以DeepMind开发的Alpha Go围棋机器人为代表，通过引入深度学习技术、并在云端服务器的算力支持下首次击败了人类世界围棋冠军李世石。伴随着大数据的累计、深度学习等新兴算法的提出以及具有高并发量计算能力的芯片的发展使得人工智能不再是停留在文学作品中的臆想，而是在技术上实现了实质性的突破，人工智能在棋牌运动中的表现仅仅是人工智能技术的初步试水，其在安防、教育、医疗、金融、交通、政法法检和人机交互等多领域拥有更广阔的应用前景和市场空间，人们看到了其巨大的商业价值及产业转化的可能性。

算法：由“人工+浅层网络”到“深层神经网络”



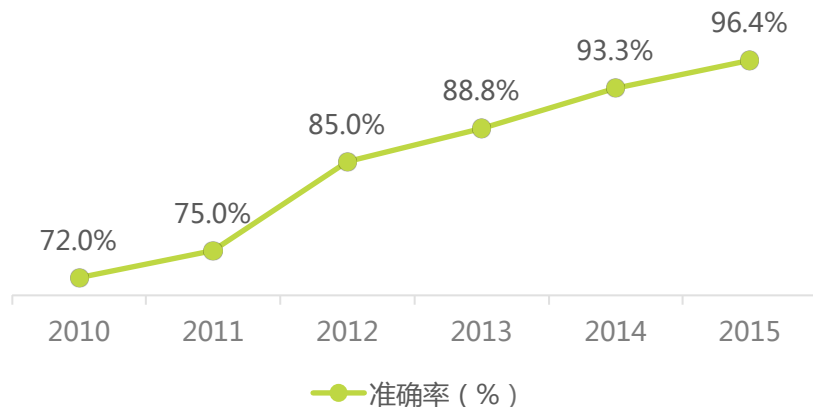
CHiME语音识别竞赛历届冠军准确率



硬件：由CPU到“CPU+GPU/FPGA/ASIC架构”



ImageNet图像识别竞赛历届冠军准确率

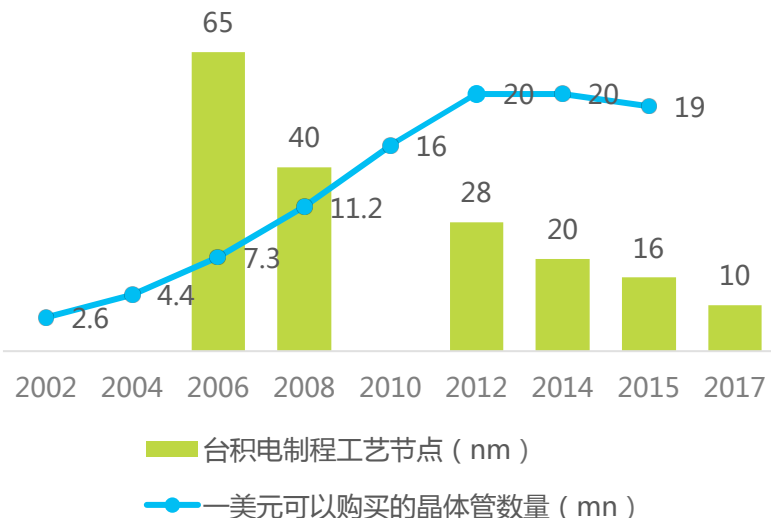


# 人工智能硬件技术突破

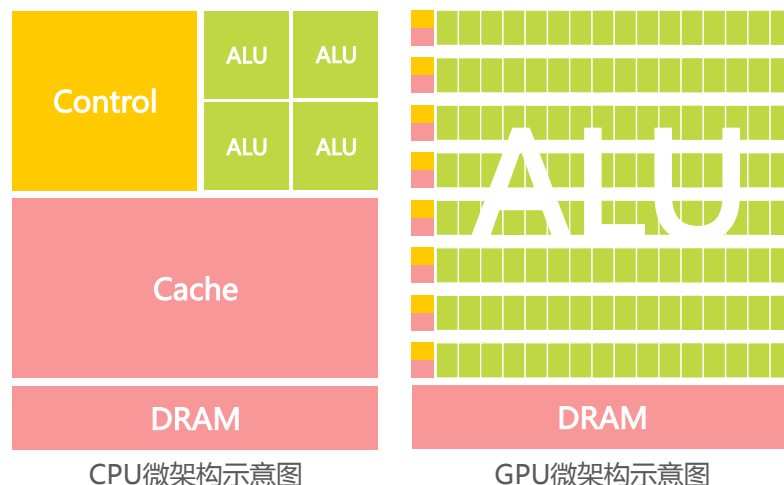
## 新架构芯片超越摩尔定律为人工智能提供充分的算力支持

深度学习需要芯片能够提供极高的并发计算能力对大量训练数据进行处理。传统的逻辑芯片架构，如：ARM、X86等CPU，由于其计算单元并不多（4、8、16核等），无法满足深度学习对于算力的需求。而摩尔定律随着制程工艺接近极限（10nm以下时由量子隧穿效应造成的晶体管漏电以及光刻机精度限制）而逐渐失效，意味着在传统CPU设计逻辑上通过进一步提升制程工艺和主频很难再获得更高的算力表现。英伟达在2011年将其GPU产品首先应用于深度学习后发现当时12颗GPU的计算能力相当于2000颗传统CPU在深度学习应用中的运算能力表现，以GPU为代表的异构（非CPU）芯片通过对结构的定向化设计为深度学习的开发和应用的部署提供了足够的算力支撑，极大的推动了人工智能技术的发展和商业化应用进程。

**工艺提升：制程工艺接近极限导致通过增加单位硅片面积内的晶体管数量以提升算力的逻辑失效**



**结构设计：GPU等异构芯片通过构建更多的计算核心以及重构计算和存储单元关系提升计算效率**



# 人工智能芯片分类

## 人工智能芯片由通用化向定制化发展

不同种类芯片应用于人工智能领域的优缺点横向对比

CPU	GPU	FPGA	ASIC
<p><b>概念：</b>中央处理器，计算机的运算和控制中心，其包括了中央运算器和中央控制器；</p> <p><b>优点：</b>通用性好，适用于传统的编程计算模式，串行运算能力强；</p> <p><b>缺点：</b>开发难度极大，大量的晶体管用于构建控制电路和高速缓冲存储器，而实际用于执行数据计算的ALU模块并不多，难以通过提升主频来处理深度学习的数据运算需求；</p> <p><b>应用场景：</b>个人电脑、移动终端、服务器；</p> <p><b>代表厂商：</b>Intel、AMD、ARM</p>	<p><b>概念：</b>显示核心，主要由大量的流处理器和显存控制器组成；</p> <p><b>优点：</b>相比CPU，由于其多线程结构，拥有较强的并行运算能力，与FPGA和ASIC相比通用性更强；</p> <p><b>缺点：</b>价格、功耗等不如FPGA和ASIC，并行运算能力在推理端无法完全发挥；</p> <p><b>应用场景：</b>个人电脑、游戏机、工作站、数据中心人工智能训练端</p> <p><b>代表厂商：</b>Nvidia、AMD（收购ATI），在2011年推出了CPU+GPU的APU异构计算方案</p>	<p><b>概念：</b>可编程逻辑门阵列，采用了逻辑单元阵列(LCA)，包括可配置逻辑模块(CLB)，输出/入模块(IOB)和内部链接，可通过编程将计算逻辑映射到硬件上，通过调整将不同的逻辑部件联通在一起完成计算任务；</p> <p><b>优点：</b>可针对特定算法修改电路，灵活、计算效率高，且功耗远远小于CPU和GPU</p> <p><b>缺点：</b>单个单元的计算能力比较弱，硬件成本较高；</p> <p><b>应用场景：</b>数据中心人工智能训练端/推断端；</p> <p><b>代表厂商：</b>Xilinx、Altera</p>	<p><b>概念：</b>专用集成电路，在人工智能领域被称作神经网络处理器，从硬件角度对神经网络结构进行模拟</p> <p><b>优点：</b>由于是针对专门的任务进行设计，因此体积更小、功耗更低、可靠性高、性能高、成本低；</p> <p><b>缺点：</b>开发难度高、初始投入大，一旦完成芯片设计，难以再对结构进行更改；</p> <p><b>应用场景：</b>数据中心、边缘计算，如移动设备、自动驾驶等；</p> <p><b>代表厂商：</b>Google、寒武纪、地平线、比特大陆</p>

通用性强

运算效率更高、能耗更低、器件成本更低

# AI芯片应用场景及代表企业

## AI芯片可分为云端/边缘两大应用场景并实现训练/推理功能

AI芯片是一种实现ML计算（主要是矩阵加乘）加速的特殊芯片，其主要的应用场景分为云计算数据中心和边缘计算，边缘计算典型场景如：摄像头IPC SoC、自动驾驶车辆DCU以及手机SoC等。AI芯片主要功能包括：训练——基于训练集实现深度学习模型的构建；推断——基于训练好的模型通过代入新的观察值导出模型推断结果。

### 人工智能芯片产业图谱



来源：根据公开网络数据整理自主绘制。

公司基本情况介绍

1

AI芯片简介

2

公司业务&竞争力分析

3

总结

4

公司基本情况介绍	1
AI芯片简介	2
公司业务&竞争力分析	3
当前公司业务发展分析	3.1
在不同场景下公司竞争力分析	3.2
总结	4



# 公司团队是全球AI芯片理论的先行者

## 全球最早提出神经网络专用芯片架构及通用指令集设计理论

传统的基于GPU/FPGA的加速以及直接硬件化算法的ASIC并没有真正解决内存访问和带宽限制，公司创始团队提出

“DianNao”系列微架构实现“小芯片”解决大规模机器学习算法问题，“DianNao”系列可以在损失一小部分计算精度（半精度计算）下更高效（降低芯片面积和功耗）的完成神经网络计算任务。“PuDianNao”实现了对多种深度学习算法的兼容，“DianNaoYu”（Cambricon）提出了神经网络计算芯片指令集架构（ISA），实现对十种代表性神经网络（NN）的兼容，提升了神经网络芯片指令效率，奠定了设计通用型神经网络计算芯片（ASIC）的基础。Google TPU主架构师曾与寒武纪共同开发DianNao架构，其TPU论文大量引用“DianNao”系列论文成果。

### DianNao (2014) ASPLOS2014最佳论文奖 (亚洲第一次)

原型神经网络处理器结构，针对CNN\DNN\RNN神经网络计算设计，相比起传统的硬件加速器方案拥有更灵活的可拓展性。其包含一个处理器核，主频为0.98GHz，运算峰值达到0.45TOPs，65nm工艺下能耗仅为0.485W。性能超过主流CPU100倍，但是面积和功耗仅为其1/10，平均性能与GPGPU相当，但功耗仅为其百分之一量级。

### DaDianNao (2014) MICR2014最佳论文奖 (美国以外国家第一次)

是DianNao的多核并行架构版本，揭示神经网络的可分特性使加速器具备可扩展性。包含16个NFU核和更大的片上存储，并支持多芯片间直接高速互连，避免高昂的内存访问开销。主频为606MHz，28nm工艺下功耗约16W。性能超过主流GPU21倍，而功耗仅为其1/330。该架构可实现对深度学习training过程的支持。

### ShiDianNao (2015) ISCA2015

由于DRAM的读写会有相当大的功耗并产生延迟。“ShiDianNao”提出通过加速器与传感器直连而绕过内存，从而降低芯片运算对内存访问的依赖，而CNN算法共享权值存储于SRAM中，避免了对于DRAM的使用。

### PuDianNao (2015) ASPLOS2015

实现了包括k-最近邻、k-均值、朴素贝叶斯、线性回归等7种机器学习算法的兼容。主频为1GHz，峰值性能达每秒1.06TOPs，面积3.51mm<sup>2</sup>，功耗为0.596W（65nm工艺下）。PuDianNao运行上述机器学习算法时的平均性能与主流GPGPU相当，但面积和功耗仅为主流GPGPU百分之一量级。

### DianNaoYu (2016) ISCA2016 (评分排名位列第一)

全球首个神经网络通用指令集架构，兼容十种代表性的神经网络。针对大规模的神经元计算，单条指令即可完成一次向量或矩阵运算。Cambricon架构下的代码长度分别比GPU\X86\MIPS短6.41、9.86、13.38倍，性能是X86和GPU的91.72倍和3.09倍，而GPU的功耗是Cambricon的130.53倍。

通用性



# 本地到云端的全套软硬件方案（1/3）

## 终端智能处理器：基于DianNao系列论文开发ASIC专用计算芯片，产品性能持续迭代提升

### 寒武纪智能终端产品性能介绍

智能处理器IP (终端)	性能参数	总结
Cambricon-1A	采用台积电10nm工艺生产， 512GFlops@1GHz	<b>技术特点：</b> 1A处理器是全球第一款商用深度学习处理器IP，其稀疏化技术优异，可实现四核通用CPU 25倍以上的性能和50倍以上的能效。兼容Caffe、TensorFlow、Mxnet等主流AI开发平台； <b>性能表现：</b> 寒武纪-1A可识别2500张照片/分钟，优于苹果A11的识别899张/分钟的成绩； <b>商业化应用：</b> 1A IP植入华为Sirin 970 SoC芯片并用于Mate10手机。
Cambricon-1H	采用台积电7nm工艺，提供 1TOPs、2TOPs、4TOPs、 8TOPs@1GHz	<b>技术特点：</b> H系列相比上一代产品拥有更低的功耗和面积，其中H8是针对低功耗、视觉应用设计，而H16拥有更广泛的通用性，可广泛应用于计算机视觉、语音识别、自然语言处理等智能处理领域； <b>性能表现：</b> 能耗比整体达到了上一代产品的2.3倍； <b>商业化应用：</b> 可应用于手机、安防、智能音箱、智能机器人等领域。 <b>1H8 IP</b> 植入华为Sirin 980 SoC芯片中（双核NPU），应用于华为Mate20 和 荣耀Magic2 手机，可实现每分钟识别照片4500张。
Cambricon-1M	采用台积电7nm工艺生产，8 位 运算达到5TOPs/Watt 寒武纪提供2TOPs、4TOPs、 8TOPs三种尺寸的处理器内核	<b>技术特点：</b> 寒武纪1M处理器除了与上两代芯片一样支持CNN、RNN、SOM等多种深度学习模型，还增加了SVM、k-NN、k-Means、决策树等经典机器学习算法的加速； <b>性能表现：</b> 性能比1A提升了10倍； <b>商业化应用：</b> 全球首款支持本地机器学习训练的智能处理器产品，可为视觉、语音、自然语言处理等任务提供高效计算平台，同时可以应用于智能手机、智能音箱、摄像头、自动驾驶等不同领域

来源：中科寒武纪公司官网；网络公开数据。

# 本地到云端的全套软硬件方案（2/3）

## 云端智能处理器：寒武纪首推我国第一款用于云端的人工智能芯片/板卡产品（ASIC），实现了从终端到云端的全线布局

寒武纪智能云端产品性能介绍

寒武纪MLU100 (云端)	性能参数		总结
核心架构	Cambricon MLUv01		<b>性能表现：</b> MLU-100是寒武纪首推的国内第一款云端智能处理芯片，相比之前的产品，其性能得到了极大的提升，主要为视觉、语音等复杂场景的云端智能处理（推断）提供了重要支撑。
半精度浮点运算速度 (FP16)	平衡模式（峰值）	16TFLOIPS（关闭稀疏模式）	
		64TFLOPS（打开稀疏模式）	
	高性能模式（峰值）	20.48TFLOIPS（关闭稀疏模式）	
		83.2TFLOPS（打开稀疏模式）	<b>商业化应用：</b> 中科曙光推出了基于Cambricon MLU100智能处理卡的服务产品系列“PHANERON”。联想推出了基于Cambricon MLU100智能处理卡的ThinkSystem SR650，该产品采用了2块MLU100智能处理卡，打破了37项服务器基准测试的世界纪录，可以满足客户对于机器学习、VDI、虚拟云化、云、数据库、分析、SAP等方向上的业务需求。其他达成产品合作的云计算服务器厂商还包括浪潮、新华三、堃乾智能等。
定点运算速度（INT8）	平衡模式（峰值）	32TFLOIPS（关闭稀疏模式）	
		128TFLOPS（打开稀疏模式）	
	高性能模式（峰值）	41.6TFLOIPS（关闭稀疏模式）	
		166.4TFLOPS（打开稀疏模式）	
内存容量:16GB/32GB	内存位宽：245-bit	内存接口：102.4GB/s	核心频率：1GHz（平衡模式）/ 1.3GHz（高性能模式）
典型最大功耗：80W/110W	系统接口：PCI Express 3.0*16	外形：267mm*112mm, 双槽位	制程工艺：台积电16nm

来源：中科寒武纪公司官网；网络公开数据。

# 本地到云端的全套软硬件方案（3/3）

## 软件：指令集和软件开发包构建寒武纪芯片的生态体系

公司除了拥有从端到云的硬件产品外，还开发了配套的软件开发包（SDK）和专用的指令集（ISA），构建了从硬件到软件到指令集的完善的生态体系，同时公司产品具备支持多种经典的深度学习开发框架的能力，能够很好地适应当前的深度学习开发生态环境。公司充分借鉴了方舟、龙芯等国产芯片的开发经验，在注重硬件性能的同时更注重下游应用生态环境的建立，通用指令集及软件开发站为客户提供了友好的开发环境，客户能够更好的发挥出寒武纪芯片的性能，寒武纪在为客户进行AI赋能的同时也拓展寒武纪的技术架构应用场景，实现“寒武纪+X”生态体系护城河构建。

### 寒武纪智能软硬件产品生态



支持深度学习  
开发框架

Caffe

Caffe2



mxnet



下游应用



HUAWEI

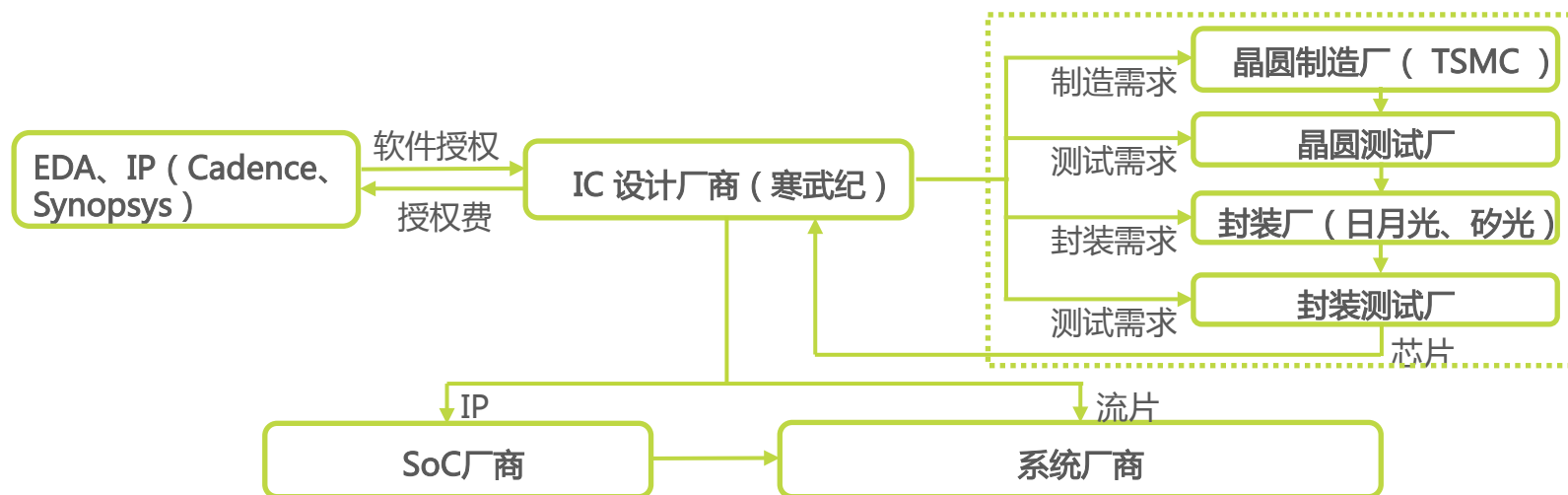


# 产业链及商业模式分析（1/3）

## 芯片设计公司对制造环节上下游EDA/Foundry厂议价能力弱

半导体行业主要有两种生产制造模式，1）IDM模式（英特尔、三星等超大型集成电路公司）：一家公司从设计、到晶圆、芯片制造、封装到产品销售的全流程集成模式；2）Fabless & Foundry模式：随着芯片制造工艺不断提升，研发及产线升级成本也不断攀升，当今越来越多的设计公司舍弃自己生产，而是将设计好的“光罩”交由专业的Foundry厂商负责流片。Foundry厂掌握了最核心的芯片制程工艺及生产能力，顶级Foundry厂（台积电、Intel、三星等）对上游厂商有较强的话语权（华为曾因为苹果订单挤兑了台积电的生产能力而导致麒麟950交付大幅延后）。EDA工具是IC设计的上游，是发展超大型集成电路的基石。目前该领域被国外厂商高度垄断（Cadence、Synopsys、Mentor Graphics三家常年占据该领域全球市场份额的70%以上），IC设计公司对上游议价能力较弱。EDA厂商是采取售卖License的形式获取盈利，一般一套License的3年授权费用在100万美金左右，根据调研，一般一个20人的研发团队，设计一款芯片所需要采购EDA工具的费用大体在100万美元/年（包括EDA和LPDDR等IP购买成本）。

芯片设计制造产业链流程图



来源：根据公开网络数据整理自主绘制。

# 产业链及商业模式分析（2/3）

## 芯片前期投入大风险高，不低于百万级出货量确保盈亏平衡

传统芯片商业模式可分为流片和IP授权。IP授权是指设计厂商仅负责技术开发而并不实际生产芯片，其向客户提供芯片的IP核授权并收取授权费用。由于不涉及芯片制造，资金占用和整体风险较低，但仍然需要支付前期研发费用；另一种模式由于涉及到流片，除前期的研发投入以外，还需要向Foundry（晶圆、晶片、封装、测试）支付千万美元的制造费用，而只有当芯片产量达到一定量级时才能够分摊掉前期的巨额投入。若产品出现一次流片不成功或者市场推广失利等情况，对芯片设计厂商（尤其是初创企业）将会造成毁灭性的打击。以寒武纪16nm制程工艺芯片为例，假设芯片面积分别为60mm<sup>2</sup>、6mm<sup>2</sup>、3mm<sup>2</sup>（数值参考“DianNao”系列论文），随着芯片面积下降和产量上升，芯片成本呈现骤减趋势。由于芯片定价需要大于（晶片总成本+人员+EDA/IP）之和，所以通过以下测算，只有当产量达到千万量级时，芯片定价才能达到市场可接受范围，而这需要公司除了能够成功设计出客户愿意买单的产品外，还需要挑选合适的细分市场。

### 寒武纪芯片流片成本分析（情景分析）

情景分析		假设：12寸晶圆（ 300mm ），4500美元/片；100人研发团队，人均工资50万元/年；研发周期：1.5年；良率：50%										
单位：美元		人员	EDA/IP	掩膜 ( 16nm )	晶片成本			封装	测试	芯片成本		
					60	6	3			60	6	3
					mm <sup>2</sup>	mm <sup>2</sup>	mm <sup>2</sup>			mm <sup>2</sup>	mm <sup>2</sup>	mm <sup>2</sup>
					↑	↑	↑					
					2	2	2			2	2	2
总成本		11,000,000	2,000,000	4,800,000								
产量	10万	110.0	20.0	48.0	8.2	0.8	0.4	6.00	2.00	120.2	112.8	112.4
	100万	11.0	2.0	4.8	8.2	0.8	0.4	0.60	0.20	19.4	12.0	11.6

\*注：芯片成本 = （掩膜 + 封装 + 测试）↓单片成本 / 良品率 + 晶片成本；晶片成本 = 晶圆成本 / （晶片数 / 晶圆）× 良品率；晶圆面积 / 晶片面积 - 晶圆周长 / √0.2 × 晶片面积

# 产业链及商业模式分析 ( 3/3 )

## 寒武纪为华为提供AI芯片IP，为寒武纪贡献主要收入

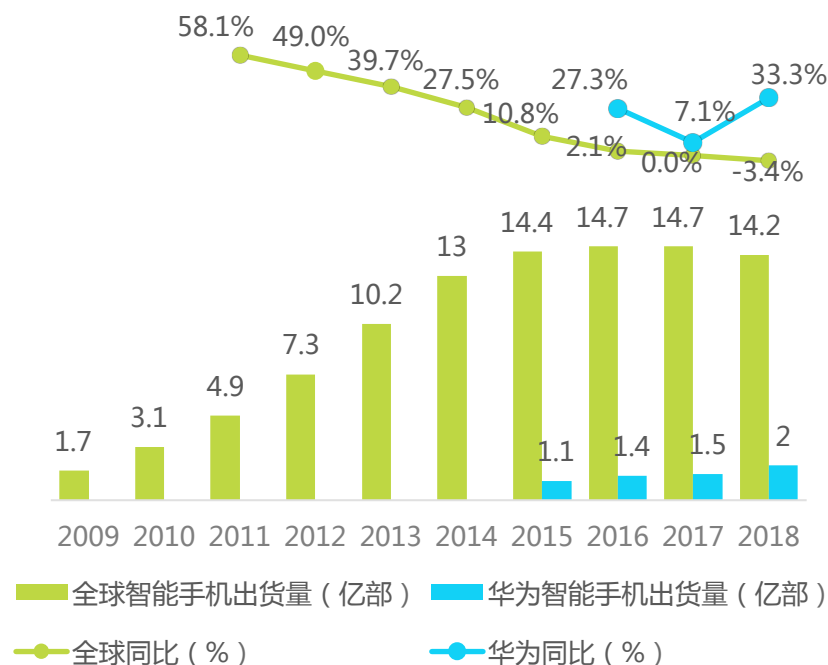
在全球手机市场整体衰退的背景下（渗透率趋于饱和、经济增速放缓），华为18年四季度出货6050万部，同比增43.9%，市场份额达到全球第三，全年出货超过2亿部。寒武纪为华为麒麟970SoC提供了单核1A IP，为麒麟980SoC提供双核1H IP。华为为Mate10、P20、荣耀V10等12款手机配备了麒麟970芯片，为Mate20、荣耀Magic2等6款手机配备麒麟980芯片。从搭载麒麟970手机出货数据来看，若1A核授权费为5美元/片，则超过4000万台手机出货量将为寒武纪带来2亿美元收入。搭载麒麟980的手机集中在18年Q4发售，但Mate20出货也已经超过500万台，对应1H核需求在1000万颗以上。

### 华为手机出货量不完全统计

芯片	手机产品	推出时间	出货量
Kirin970单核1A	Mate10系列： (Mate10、Pro)	2017年10月20日	累计出货超过1000万部
	华为P20系列	2018年3月27日	累计出货超过1600万部
	荣耀10、V10	V10：2017年11月28日 10：2018年4月19日	18年上半年累计出货414万部
	nova3、nova4	nova3：2018年7月10日 nova4：2018年12月27日	Nova3/3i发布一个月出货200万部，推测全年出货累计可达千万部
Kirin980双核1H	Mate20系列	2018年10月16日	到18年底累计发货超过500万部
	荣耀Magic2、V20	Magic2：2018年10月31日； V20：2019年1月	

来源：网络公开数据。

### 2009-2018年全球智能手机出货量



来源：IDC。

公司基本情况介绍	1
AI芯片简介	2
公司业务&竞争力分析	3
当前公司业务发展分析	3.1
在不同场景下公司竞争力分析	3.2
总结	4

公司基本情况介绍

1

AI芯片简介

2

公司业务&竞争力分析

3

总结

4



# 分析总结 ( 1/3 )

## 公司拥有强大的股东背景及深厚的技术研发实力和专利储备

就我们通过市场搜集到的公开信息和数据以及前文分析判断，我们比较看好寒武纪的未来发展前景，原因如下：

角度	分析	
行业	<p>看好整体AI芯片行业发展前景：</p> <ul style="list-style-type: none"><li>- 互联网和实体经济产生大量数据累积并无法得到有效利用。以深度学习为代表的机器学习技术（人工智能）对大数据内在价值进行有效的挖掘和利用，如：智能摄像头在公安政法领域的应用成效；产线、机房自动预警及维护。</li><li>- 人口老龄化问题使中国正在迅速丧失廉价劳动力优势，更多重复性高的工作需要由人工智能技术对人力实现替代；</li><li>- AI芯片是承载人工智能技术的基石，将会被大量部署在云端和边缘市场。到2022年，预计边缘市场AI芯片市场规模预计将会达到231亿美元，云端AI芯片市场规模将会达到256.4亿美元，整体市场规模将会超过3000亿元人民币。</li></ul>	
公司自身实力	团队背景：	公司创始团队是以中科院计算所的陈氏兄弟和刘少礼教授为代表的芯片研发团队，有丰富的芯片研发经验，如：MIPS架构的国产“龙芯”CPU，其设计难度远超AI芯片，“龙芯”芯片的设计开发为团队积攒了宝贵的芯片设计开发及市场推广经验。
	股东背景：	公司为中科院孵化企业，目前中科院及实际控制人共同对公司股权控制比例超过68%，此外公司股东列表中还有大量的带有国资背景的基金、阿里、科大讯飞、联想为代表的上市公司以及疑似由券商系资本主导的投资主体。
	技术实力：	公司创始团队是全球最早提出专用NN计算芯片架构及通用指令集设计思想的团队，其论文多次获得世界顶级半导体学术论坛最佳论文奖，其研究成果多次被谷歌TPU架构团队引用。DianNao系列论文阐释了如何针对NN计算特点进行芯片设计，以及通过对片上存储、芯片连接扩展等优化实现芯片运算效能的提升，并且针对当今具有代表性的十种NN模型设计了通用ISA，使开发通用ASIC AI芯片成为可能。根据天眼查显示，公司共有71项专利和6项软件著作权。

# 分析总结 ( 2/3 )

## 相比终端市场寒武纪更应结合自身优势及政策优势（国产替代）发展云端市场形成差异化竞争优势

角度	分析	
产品/服务竞争力分析	终端市场	<p>终端市场对公司的产品竞争力及市场推广能力提出较大挑战：</p> <ul style="list-style-type: none"><li>- 公司在成立的短短两年时间内即推出了面向终端市场的1A/1H/1M系列产品，其中公司已经为华为麒麟970/980SoC芯片提供1A/1H系列产品的IP，而最新推出的1M芯片实现了本地模型训练能力，摆脱了模型在云端训练后再回传终端加载的过程，降低了终端对于云端数据中心及网络传输的依赖。</li><li>- 公司在AI芯片领域有深厚的技术积累，但仅有技术并不意味着产品能够在终端市场实现成功推广，细分终端场景对AI芯片厂商所提供产品/解决方案提出了不同的要求——如汽车电子领域的车规电子设计标准、智能手机的芯片功耗等，而如何将公司的技术成功适配到下游客户整体产品方案中，或者为下游客户提供完整的硬件+软件的综合解决方案需要极强的行业开发经验和市场推广能力，公司在缺少这两项能力时面对细分行业内存在的市场竞争时欠缺竞争力。</li></ul>
	云端市场	<p>公司在云端市场具有产品和产业链竞争优势：</p> <ul style="list-style-type: none"><li>- 公司推出了针对于服务器的MLU100计算板卡，MLU100主要针对云端推理市场，未来公司还计划推出针对于云端训练场景的MLU200产品</li><li>- 公司撰写的DaDianNao、PuDianNao以及DianNaoYu论文分别解决了：AI芯片针对巨大神经网络问题时的可扩展性；芯片对7种机器学习模型的通用性；以及针对当今最具代表的10种NN模型的ISA，实现了针对NN运算的编程语句优化。此外，公司还为其芯片开发了软件站SDK，使开发者能够更快、更便捷的上手使用公司芯片产品。公司是国内少有的具有开发通用型AI芯片能力的公司。</li><li>- 联想、中科曙光等云端服务器厂商与公司同属于中科院计算所孵化企业，阿里巴巴以及科大讯飞等云厂商及应用层企业亦是公司的股东，公司在云端市场具有较强的产业链优势。因此，基于以上的因素考量，我们认为相比起终端市场，公司在云端市场竞争优势更加显著。</li></ul>

来源：艾瑞研究院。

# 分析总结 ( 3/3 )

## 寒武纪有望成为科创板首批上市企业

角度	分析
未来发展 战略	<b>自身</b> <ul style="list-style-type: none"><li>- 硬件产品：国内目前在云计算AI芯片市场缺少以英伟达为代表的海外科技巨头对标企业，公司应就该领域持续发力，为我国云厂商尽快提供国产产品替代方案；</li><li>- 软件产品：进一步完善软件开发站、开发平台,建立并完善围绕寒武纪芯片的开发生态和应用产品；</li><li>- 团队组建：公司应该加快市场团队组建，推广产品、沟通市场，避免闭门造车的情况发生。</li></ul>
	<b>资本</b> <ul style="list-style-type: none"><li>- 场外融资：公司自16年成立以来已经进行了3轮融资，估值从天使轮1亿美元、A轮10亿美元增长至B轮的25亿美元，复合增长率400%，累计吸收融资数美元，为公司的发展积累了充足的资金储备。公司当前估值达到25亿美元，绝对值较高，但相对于同行业竞品企业估值还是偏低（参考下一页）；公司是中科院计算所孵化的国家队性质的AI芯片研发团队，且有包括上市公司、国资背景基金、券商资本等股东的强大信用背书，公司融资渠道畅通；</li><li>- 场内市场：公司早前在媒体上提到计划境内独立上市，参照主板、创业板上市标准，公司当前发展阶段难以达到证监会审核标准。但就在本篇报告发布前夕，随着国家加快设立科创板的消息不断流出，公司短期内在境内实现上市似乎成为可能，依据如下：<ul style="list-style-type: none"><li>1) 证监会《实施意见》：科创板主要服务国家战略、突破关键核心技术、市场认可度高的科技创新企业。重点支持：新一代信息技术、高端装备、新材料、新能源、节能环保、生物医药等高新技术产业和战略新兴产业，推动互联网、大数据、云计算、人工智能和制造业深度融合。公司业务与科创板支持板块高度重合，中科院背景、研发实力强且掌握高端核心技术，为包括华为在内的客户提供过产品和服务，市场认可度较高；</li><li>2) 比较上交所公布的5套上市标准：a) 科创板主要面向成长型科创企业；b) 实行注册制，财务要求更加灵活非盈利优质科创企业也可实现上市；c) 考察标准差异化，包括：市值、收入、净利润、现金流、核心技术等多重指标。</li><li>3) 12月4日，经济观察网披露：中关村示范区向上交所提交了一份独角兽企业名单作为潜在科创板上市企业。从网上获取信息，不到30家企业中疑似包含寒武纪。</li><li>4) 公司近一轮融资中引入中金、中信两家券商（保荐机构）旗下资本，疑似为企业进一步二级市场资本运作提前做好准备。</li></ul></li></ul> <p>上交所有关负责人表示：预计19年一季度可以完成科创板受理系统技术上线，若届时相关规则完成正式发布，则可在一季度正式受理企业申请。东北证券研究总监付立春评论称：预计科创板今年上半年即可形成雏形，全年上市100家左右</p>

# 关于艾瑞

在艾瑞 我们相信数据的力量，专注驱动大数据洞察为企业赋能。

在艾瑞 我们提供专业的数据、信息和咨询服务，让您更容易、更快捷的洞察市场、预见未来。

在艾瑞 我们重视人才培养，Keep Learning，坚信只有专业的团队，才能更好的为您服务。

在艾瑞 我们专注创新和变革，打破行业边界，探索更多可能。

在艾瑞 我们秉承汇聚智慧、成就价值理念为您赋能。

● 我们是艾瑞，我们致敬匠心 始终坚信“工匠精神，持之以恒”，致力于成为您专属的商业决策智囊。



扫描二维码  
读懂全行业

海量的数据 专业的报告



400-026-2099



ask@iresearch.com.cn

# 法律声明

## 版权声明

本报告为艾瑞咨询制作，报告中所有的文字、图片、表格均受有关商标和著作权的法律保护，部分文字和数据采集于公开信息，所有权为原著者所有。没有经过本公司书面许可，任何组织和个人不得以任何形式复制或传递。任何未经授权使用本报告的相关商业行为都将违反《中华人民共和国著作权法》和其他法律法规以及有关国际公约的规定。

## 免责条款

本报告中行业数据及相关市场预测主要为公司研究员采用桌面研究、行业访谈、市场调查及其他研究方法，并且结合艾瑞监测产品数据，通过艾瑞统计预测模型估算获得；企业数据主要为访谈获得，仅供参考。本报告中发布的调研数据采用样本调研方法，其数据结果受到样本的影响。由于调研方法及样本的限制，调查资料收集范围的限制，该数据仅代表调研时间和人群的基本状况，仅服务于当前的调研目的，为市场和客户提供基本参考。受研究方法和数据获取资源的限制，本报告只提供给用户作为市场参考资料，本公司对该报告的数据和观点不承担法律责任。

为商业决策赋能

EMPOWER BUSINESS DECISIONS



艾 瑞 咨 询