

ChatGPT:存算一体，算力的下一极

AIGC行业深度报告(6)

华西计算机团队

2023年4月4日

分析师：刘泽晶

SAC NO: S1120520020002

邮箱：liuzj1@hx168.com.cn

核心逻辑:

- ◆ **ChatGPT开启大模型“军备赛”，存储作为计算机重要组成部分明显受益:** ChatGPT开启算力军备赛，大模型参数呈现指数规模，引爆海量算力需求，模型计算量增长速度远超人工智能硬件算力增长速度，同时也对数据传输速度提出了更高的要求。XPU、内存、硬盘组成完整的冯诺依曼体系，以一台通用服务器为例，芯片组+存储的成本约占70%以上，芯片组、内部存储和外部存储是组成核心部件；存储是计算机的重要组成结构，“内存”实为硬盘与CPU之间的中间人，存储可按照介质分类为ROM和RAM两部分。
- ◆ **存算一体，后摩尔时代的必然发展:** 过去二十年中，算力发展速度远超存储，“存储墙”成为加速学习时代下的一代挑战，原因是在后摩尔时代，存储带宽制约了计算系统的有效带宽，芯片算力增长步履维艰。因此存算一体有望打破冯诺依曼架构，是后摩时代下的必然选择，存算一体即数据存储与计算融合在同一个芯片的同一片区之中，极其适用于大数据量大规模并行的应用场景。存算一体优势显著，被誉为AI芯片的“全能战士”，具有高能耗、低成本、高算力等优势；存算一体按照计算方式分为数字计算和模拟计算，应用场景较为广泛，SRAM、RRAM有望成为云端存算一体主流介质。
- ◆ **存算一体前景广阔、渐入佳境:** 存算一体需求旺盛，有望推动下一阶段的人工智能发展，原因是我们认为现在存算一体主要AI的算力需求、并行计算、神经网络计算等；大模型兴起，存算一体适用于从云至端各类计算，**端测方面**，人工智能更在意及时响应，即“输入”即“输出”，目前存算一体已经可以完成高精度计算；**云端方面**，随着大模型的横空出世，参数方面已经达到上亿级别，存算一体有望成为新一代算力因素；存算一体适用于人工智能各个场景，如穿戴设备、移动终端、智能驾驶、数据中心等。我们认为存算一体为下一代技术趋势并有望广泛应用于人工智能神经网络相关应用、感存算一体，多模态的人工智能计算、类脑计算等场景。
- ◆ **投资建议:** 关注四条投资主线：1)具备算力能力的厂商，受益标的为：**首都在线、中科曙光、拓维信息、神州数码、浪潮信息、寒武纪、海光信息、龙芯中科、景嘉微**；2)具备存储芯片能力的厂商，受益标的为：**东芯股份、兆易创新、北京君正、澜起科技、聚辰股份、普冉股份**；3)具备存储模组能力的厂商，受益标的为：**江波龙、佰维存储、德明利**；4)与存储相关的其他厂商，受益标的为：**中科曙光、易华录、同有科技、恒烁股份**。
- ◆ **风险提示:** 核心技术水平升级不及预期的风险、AI伦理风险、政策推进不及预期的风险、中美贸易摩擦升级的风险。



目录

01 存算一体，开启算力新篇章

02 存算一体，打开海量应用空间

03 投资建议：梳理AIGC相关受益厂商

04 风险提示

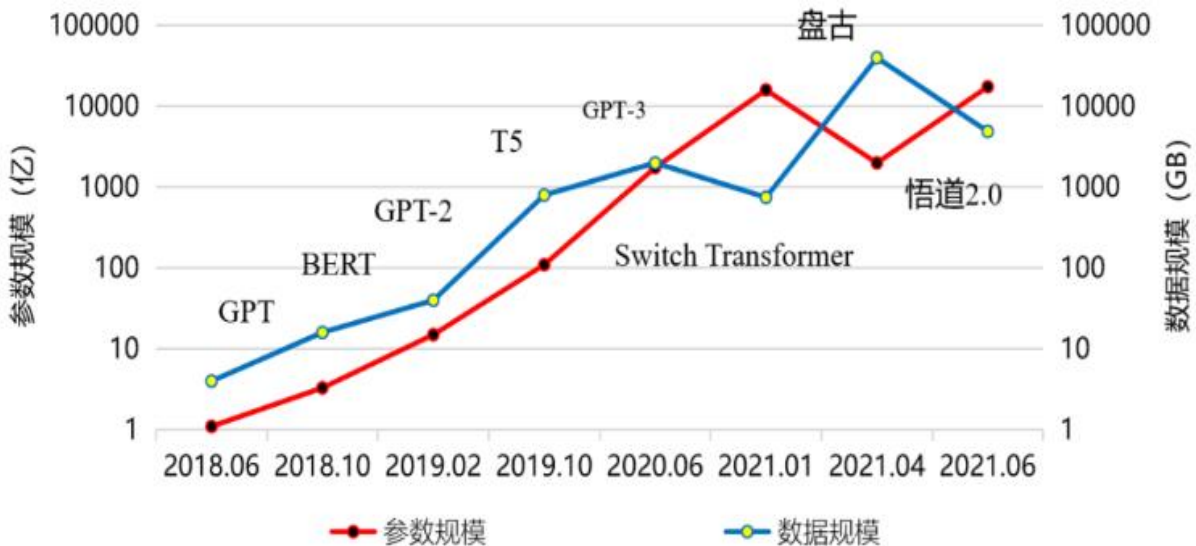


01 存算一体，开启算力新篇章

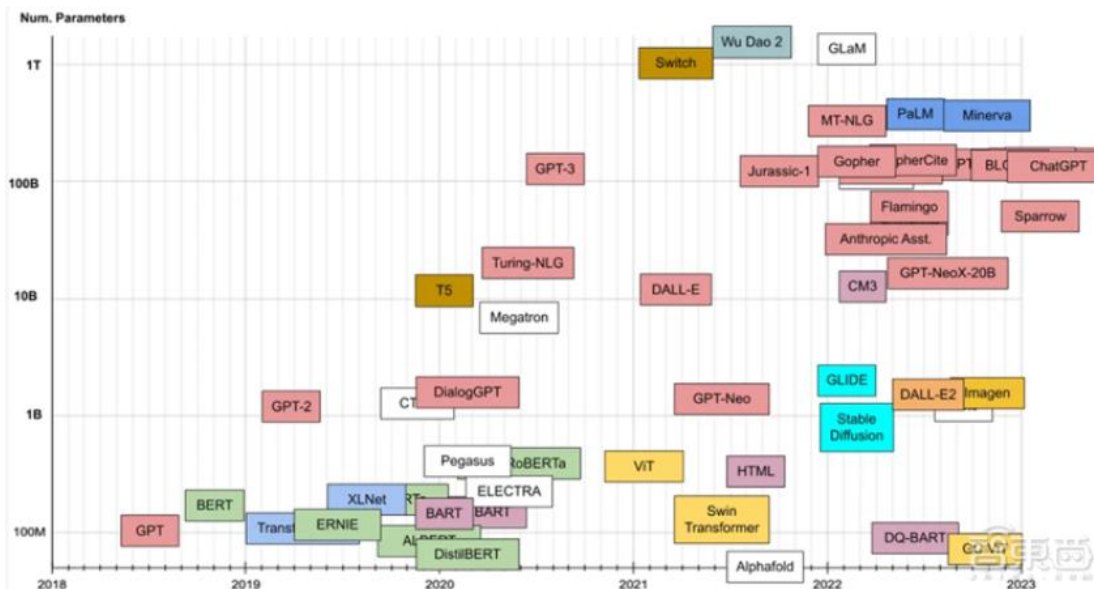
1.1 ChatGPT开启大模型“军备赛”，算力呈现明显缺口

- ◆ **ChatGPT开启算力军备赛**：我们已经在《ChatGPT：百度文心一言畅想》中证明数据、平台、算力是打造大模型生态的必备基础，且算力是训练大模型的底层动力源泉，一个优秀的算力底座在大模型(AI算法)的训练和推理具备效率优势；同时，我们在《ChatGPT打响AI算力“军备战”》中证明算力是AI技术角逐“入场券”，其中AI服务器、AI芯片等为核心产品；此外，我们还在《ChatGPT，英伟达DGX引爆 AI “核聚变”》中证明以英伟达为代表的科技公司正在快速补足全球AI算力需求，为大模型增添必备“燃料”。
- ◆ **大模型参数呈现指数规模，引爆海量算力需求**：根据财联社和OpenAI数据，ChatGPT浪潮下算力缺口巨大，根据OpenAI数据，模型计算量增长速度远超人工智能硬件算力增长速度，存在万倍差距。运算规模的增长，带动了对AI训练芯片单点算力提升的需求，并对数据传输速度提出了更高的要求。根据智东西数据，过去五年，大模型发展呈现指数级别，部分大模型已达万亿级别，因此对算力需求也随之攀升。

大模型参数数量和训练数据规模快速增长



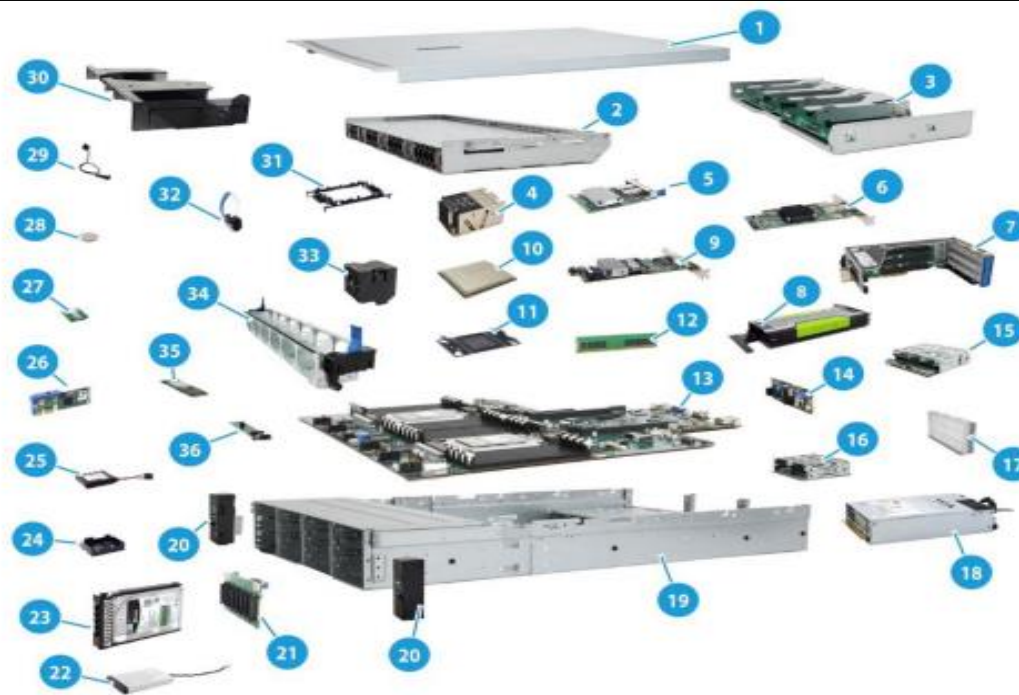
近年大模型的参数规模增长趋势



1.2 深度拆解服务器核心硬件组成部分

- ◆ **服务器的组成:** 我们以一台通用服务器为例，服务器主要由主板、内存、芯片组、磁盘、网卡、显卡、电源、主机箱等硬件设备组成；其中芯片组、内部存储和外部存储是组成核心部件。
- ◆ **GPU服务器优势显著:** GPU服务器超强的计算功能可应用于海量数据处理方面的运算，如搜索、大数据推荐、智能输入法等，相较于通用服务器，在数据量和计算量方面具有成倍的效率优势。此外，GPU可作为深度学习的训练平台，优势在于1、GPU服务器可直接加速计算服务，亦可直接与外界连接通信；2、GPU服务器和云服务器搭配使用，云服务器为主，GPU服务器负责提供计算平台；3、对象存储 COS 可以为 GPU 服务器提供大数据量的云存储服务。
- ◆ **AI服务器芯片组价值成本凸显:** 以一台通用服务器为例，主板或芯片组占比最高，大约占成本50%以上，内存(内部存储+外部存储)占比约为20%。此外，根据Wind及芯语的数据，AI服务器相较于高性能服务器、基础服务器在芯片组(CPU+GPU)的价格往往更高，AI服务器(训练)芯片组的成本占比高达83%、AI服务器(推理)芯片组占比为50%，远远高于通用服务器芯片组的占比。

H3C UniServer R4900 G5服务器硬件结构拆解



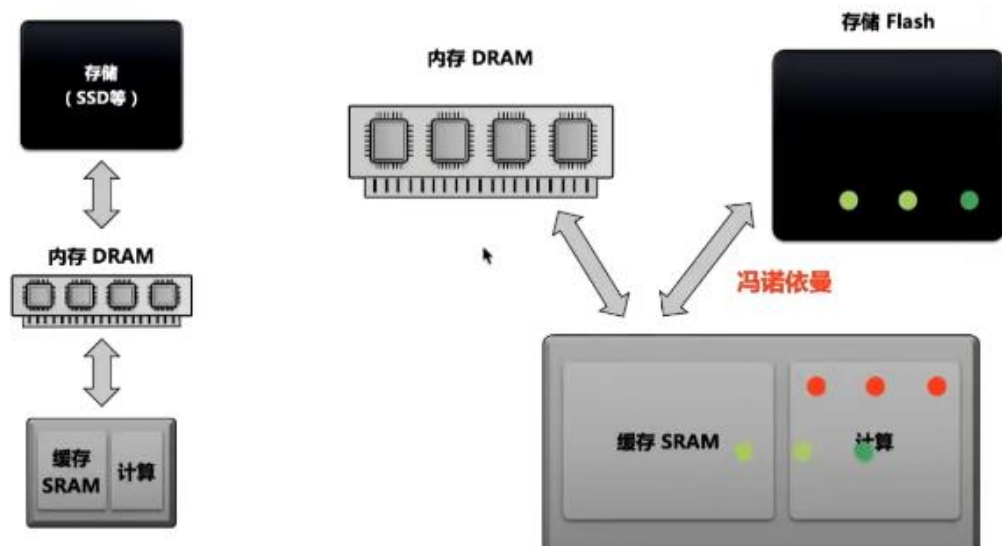
H3C UniServer R4900 G5服务器硬件结构注释

编号	名称	编号	名称
3	中置GPU模块	12	内存
5、6	网卡	13	主板
7	Riser卡	18	电源模块
8	GPU卡	23	硬盘
9	存储控制卡	25	超级电容
10	CPU	27	加密模块
12	内存	28	系统电池

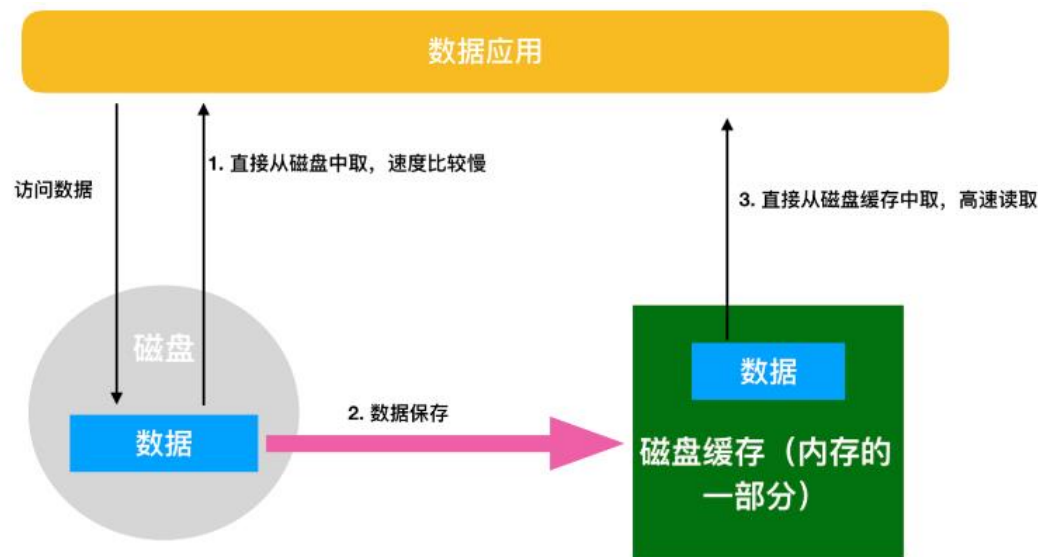
1.3 存储，计算机的重要组成结构

- ◆ **存储是计算机的重要组成结构**：存储器是用来存储程序和数据的部件，对于计算机来说，有了存储器才有记忆功能，才能保证正常工作。存储器按其用途可分为主存储器和辅助存储器，主存储器又称内存储器（简称内存），辅助存储器又称外存储器（简称外存）。
 - ✓ **内存**：主板上的存储结构，与CPU直接沟通，并用其存储数据的部件，存放当前正在使用的（即执行中）的数据和程序，一旦断电，其中的程序和数据就会丢失；
 - ✓ **外存**：磁性介质或光盘，像硬盘，软盘，CD等，能长期保存信息，并且不依赖于电力来保存信息。
 - ✓ **XPU、内存、硬盘组成完整的冯诺依曼体系**：“内存”实为硬盘与CPU之间的中间人，CPU如果直接从硬盘中抓数据，时间会太久。所以“内存”作为中间人，从硬盘里面提取数据，再让CPU直接到内存中拿数据做运算。这样会比直接去硬盘抓数据，快百万倍；CPU里面有一个存储空间Register（寄存器），运算时，CPU会从内存中把数据载入Register，再让Register中存的数字做运算，运算完再将结果存回内存中，因此运算速度Register > 内存 > 硬盘，速度越快，价格越高，容量越低。

数据存储示意图



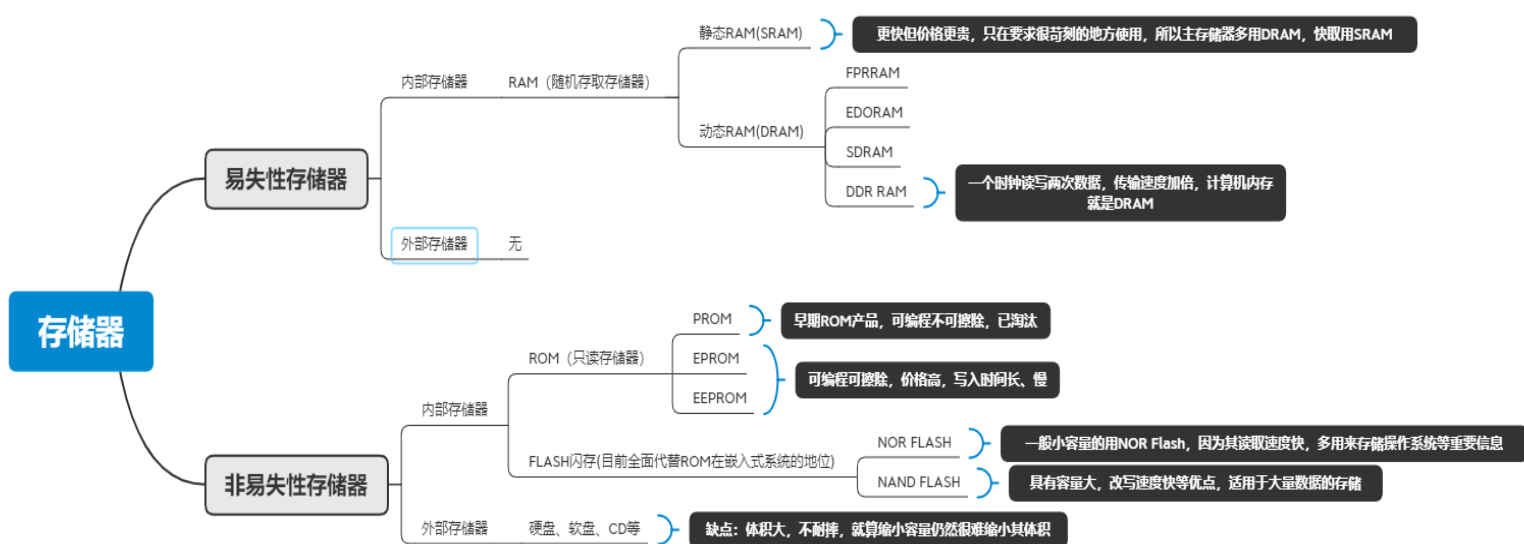
数据存储示意图



1.3 存储，计算机的重要组成结构

- ◆ **存储按照易失性分类**：分别为ROM(只读存储器)是Read Only Memory的缩写，RAM(随机存取存储器)是Random Access Memory的缩写。ROM在系统停止供电的时候仍然可以保持数据，而RAM通常都是在掉电之后就丢失数据，典型的RAM就是计算机的内存。
- ◆ **RAM(随机存取存储器)作为内存架构广泛应用于计算机中**：是与中央处理器直接交换数据的内部存储器。可以随时读写且速度很快，通常作为操作系统或其他正在运行中的程序的临时资料存储介质。RAM可分为静态SRAM与动态DRAM, SRAM速度非常快，是目前读写最快的存储设备了，但是价格昂贵，所以只在要求很苛刻的地方使用，譬如CPU的一级缓冲，二级缓冲；DRAM保留数据的时间很短，速度也比SRAM慢，不过比任何的ROM都要快，但从价格上来说DRAM相比SRAM要便宜，因此计算机内存大部分为DRAM架构；
- ◆ **ROM(只读存储器)作为硬盘介质广泛使用**：Flash内存的存储特性相当于硬盘，它结合了ROM和RAM的长处，不仅具备了电子可擦除可编程的性能，还不会断电丢失数据同时可以快速读取数据，近年来Flash已经全面替代传统ROM在嵌入式系统的定位，目前Flash主要有两种NOR Flash和NAND Flash。Nand-flash存储器具有容量较大，改写速度快等优点，适用于大量数据的存储，因此被广泛应用在各种存储卡，U盘，SSD，eMMC等等大容量设备中；NOR-Flash则由于特点是芯片内执行，因此应用于众多消费电子领域。

存储分类示意图



易失性存储器分类

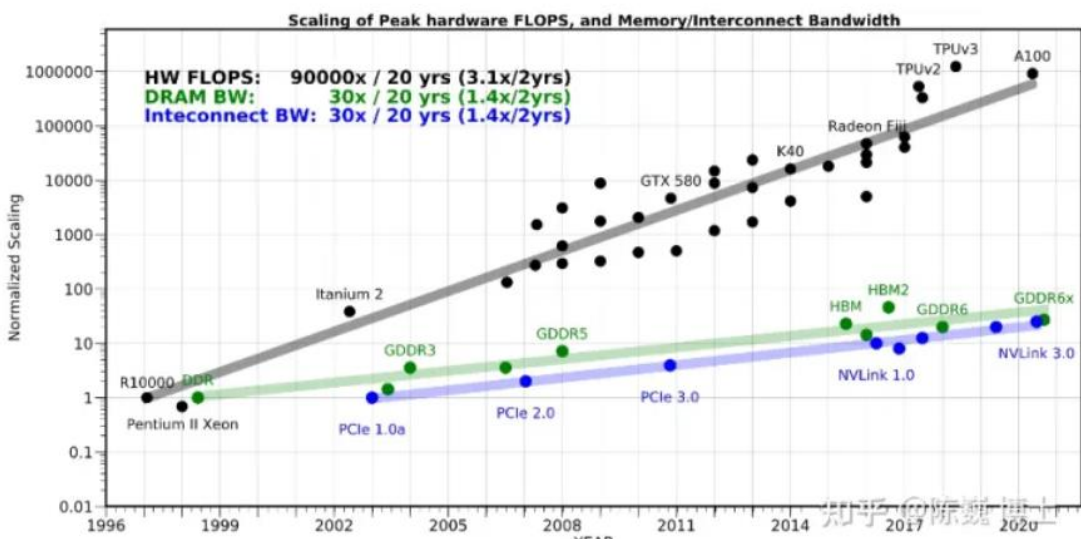
比较	动态RAM (DRAM)	静态RAM (SRAM)
存储原理	电容存储电荷的方式	触发器的方式
集成度	高	低
芯片引脚	少	多
功耗	小	大
价格	低	高
速度	慢	快
刷新	有	无

<https://blog.csdn.net/mj475002864>

1.4 存算一体，后摩尔时代的必然发展

- ◆ **算力发展速度远超存储，存储带宽限制计算系统的速度**：在过去二十年，处理器性能以每年大约55%的速度提升，内存性能的提升速度每年只有10%左右。因此，目前的存储速度严重滞后于处理器的计算速度。能耗方面，从处理单元外的存储器提取所需的时间往往是运算时间的成百上千倍，因此能效非常低；“存储墙”成为加速学习时代下的一代挑战，原因是数据在计算单元和存储单元的频繁移动。
- ◆ **存储墙、带宽墙和功耗墙成为首要限制关键**：在传统计算机架构中，存储与计算分离，存储单元服务于计算单元，因此会考虑两者优先级；如今由于海量数据和AI加速时代来临，不得不考虑以最佳的配合方式为数据采集、传输、处理服务，然而存储墙、带宽墙和功耗墙成为首要挑战，虽然多核并行加速技术也能提升算力，但在后摩尔时代，存储带宽制约了计算系统的有效带宽，芯片算力增长步履维艰。
- ◆ **存算一体有望打破冯诺依曼架构，是后摩时代下的必然选择**：存算一体是在存储器中嵌入计算能力，以新的运算架构进行二维和三维矩阵乘法/加法运算。存内计算和存内逻辑，即存算一体技术优势在于可直接利用存储器进行数据处理或计算，从而把数据存储与计算融合在同一个芯片的同一片区之中，可以彻底消除冯诺依曼计算架构瓶颈，特别适用于深度学习神经网络这种大数据量大规模并行的应用场景。

算力发展速度远超存储器

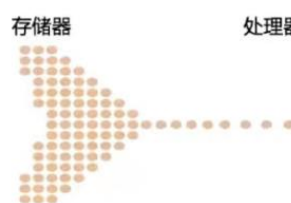


存储墙瓶颈

- 数据搬运速度是运算的瓶颈
- SRAM : 10-100TB/s
- DRAM : 40GB-1TB/s
- AI运算需要1PB/s
- 数据搬运速度还需提高近百倍

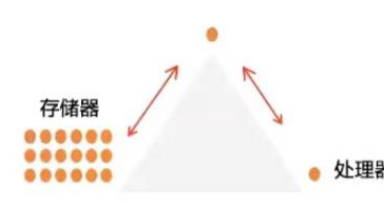
Operation	Energy(pJ)
Computation Energy Cost	
Integer Add (32b)	0.1
Integer Multiply (32b)	3.1
Floating Point Add (32b)	0.9
Floating Point Multiply (32b)	3.7
Memory Access Energy Cost	
8KB SRAM (64b)	10
1MB SRAM (64b)	100
DRAM	2000

~650X



存储器 → 处理器

问题1：数据搬运慢



存储器 → 处理器

问题2：搬运能耗大

1.4 存算一体，后摩尔时代的必然发展

- ◆ **存算一体概念已有50年历史**：早在1969年，斯坦福研究所的Kautz等人提出了存算一体计算机的概念。但受限于当时的芯片制造技术和算力需求的匮乏，那时存算一体仅仅停留在理论研究阶段，并未得到实际应用。然而为了打破冯诺依曼架构，降低“存储-内存-处理单元”过程数据搬移带来的开销，业内广泛采用3D封装技术实现3D堆叠提供更大带宽，但是并没有改变数据存储与数据处理分离的问题；
- ◆ **近年来，存算一体随着人工智能的驱动得到较快发展**：随着半导体制造技术突破，以及AI等算力密集的应用场景的崛起，为存算一体技术提供新的制造平台和产业驱动力。2016年，美国加州大学团队提出使用RRAM构建存算一体架构的深度学习神经网络(PRIME)。相较于传统冯诺伊曼架构的传统方案，PRIME可以实现功耗降低约20倍、速度提升约50倍；此外，2017年，英伟达、微软、三星等提出存算一体原型；同年起，国产存算一体芯片企业开始“扎堆”入场，例如千芯科技、智芯微、亿铸科技、后摩时代、莘芯科技等。

存算一体发展简史

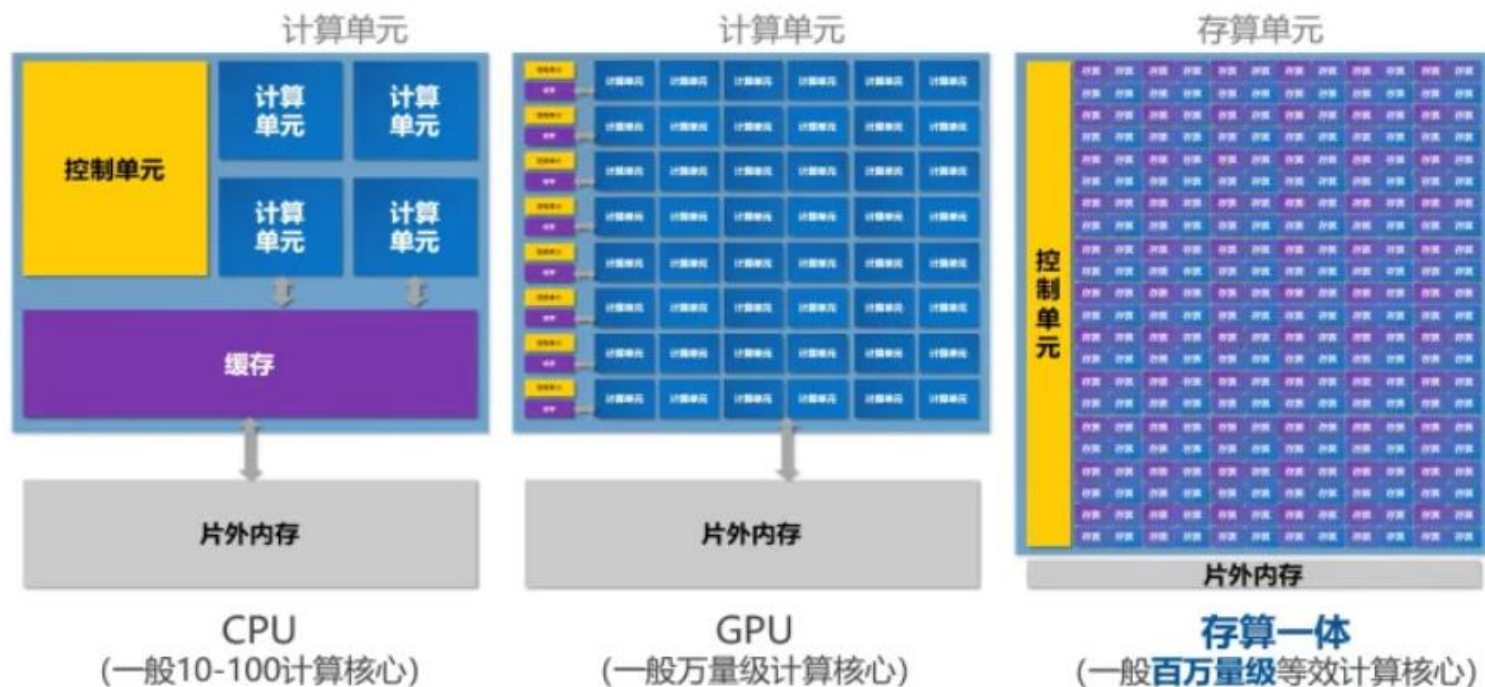


1.5 存算一体: AI芯片的“全能战士”

◆ **存算一体优势显著，被誉为AI芯片的“全能战士”其优势如下:**

- ✓ **1、成百上千倍的提高计算效率，降低成本:**存算一体的优势是打破存储墙，消除不必要的搬运延迟和功耗，使用存储单元提升算力；
- ✓ **2、特定领域提供更高算力与能效:**存算一体架构消除了计算与存储的界限，直接在存储器内完成计算，因此属于非冯诺伊曼架构，在特定领域可以提供更大算力（1000TOPS以上）和更高能效（超过10-100TOPS/W），明显超越现有ASIC算力芯片；
- ✓ **3、存算一体代表了未来AI计算芯片的主流架构:**除AI计算外，存算技术也可用于感存算一体芯片和类脑芯片，可减少不必要的搬运与使用存储单元参与逻辑计算提升算力，原因在于等效于在面积不变的情况下规模化增加计算核心数。

存算一体对比示意图

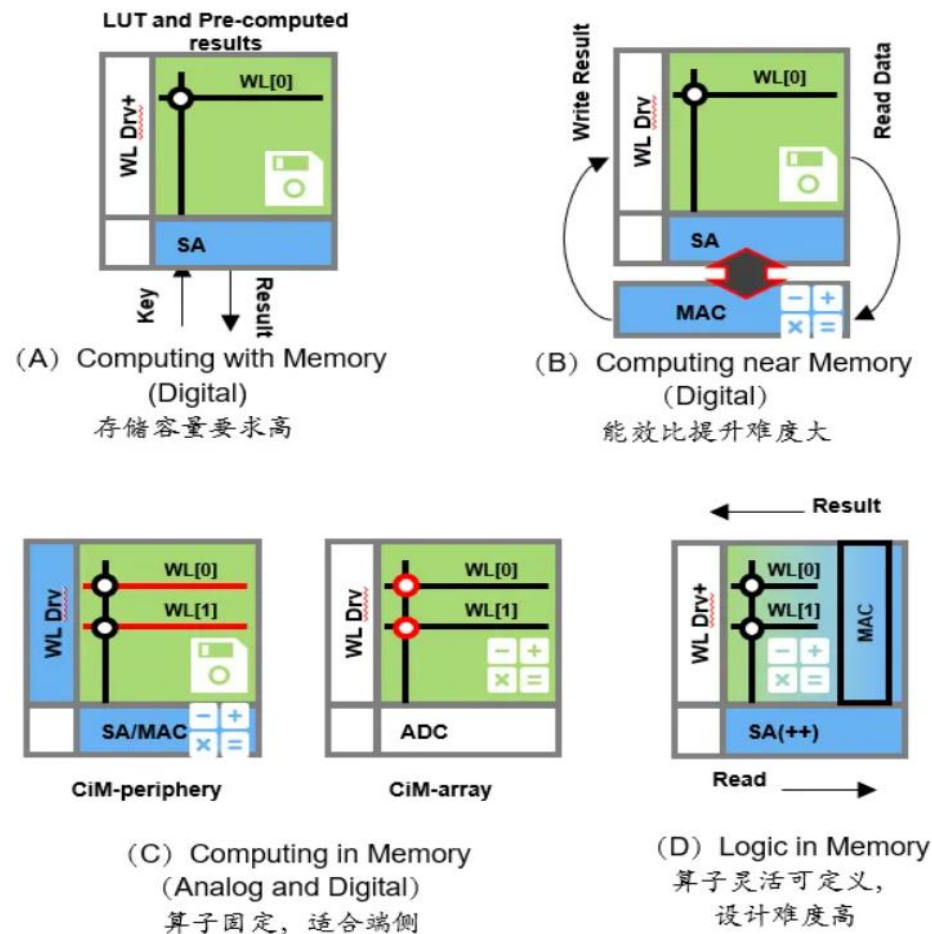


1.5 存算一体: AI芯片的“全能战士”

◆ 目前存算技术按照历史路线顺序演进:

- ✓ **A、查表计算:** GPU中对于复杂函数就采用了这种计算方法, 通过在存储芯片内部查表来完成计算操作, 目前应用较为广阔, 且技术相较成熟;
- ✓ **B、近存计算:** 计算操作由位于存储区域外部的独立计算芯片/模块完成。这种架构设计的代际设计成本较低, 适合传统架构芯片转入。例如AMD的Zen系列CPU、三星的HBM-PIM、特斯拉Dojo(AI训练计算机)、阿里达摩院等, 近存计算技术早已成熟, 被广泛应用在各类CPU和GPU上;
- ✓ **C、存内计算:** 计算操作由位于存储芯片/区域内部的独立计算单元完成, 存储和计算可以是模拟的也可以是数字的。这种路线一般用于算法固定的场景算法计算, 典型代表如Mythic、千芯科技、闪亿、知存、九天睿芯等;
- ✓ **D、存内逻辑:** 这种架构数据传输路径最短, 同时能满足大模型的计算精度要求。通过在内部存储中添加计算逻辑, 直接在内部存储执行数据计算。典型代表为TSMC和千芯科技等。

存算技术按照历史路线顺序演进

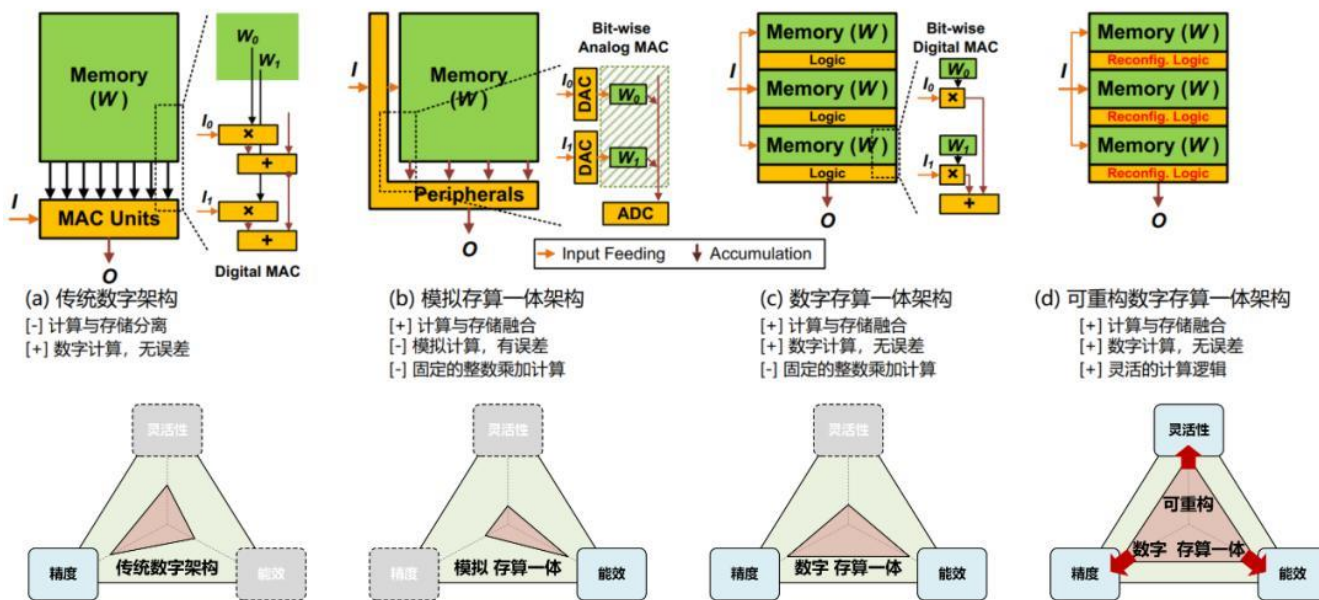


1.5 存算一体: AI芯片的“全能战士”

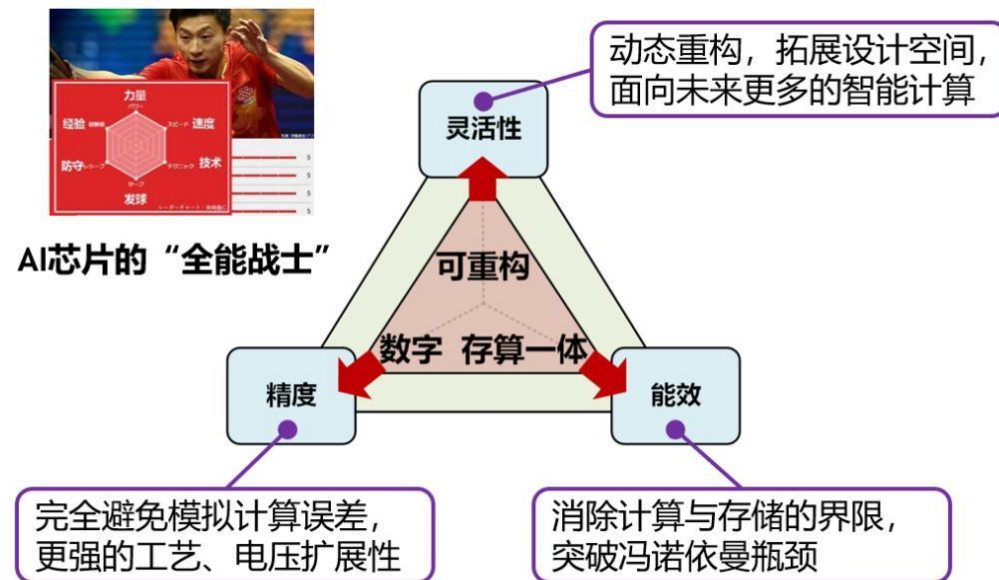
◆ 存算一体按照计算方式分为数字计算和模拟计算:

- ✓ **模拟计算:** 模拟存算一体通常使用FLASH、RRAM、PRAM等非易失性介质作为存储器件, 存储密度大, 并行度高, 但是对环境噪声和温度非常敏感。模拟存算一体模型权重保持在存储器中, 输入数据流入存储器内部基于电流或电压实现模拟乘加计算, 并由外设电路对输出数据实现模数转换。由于模拟存算一体架构能够实现低功耗低位宽的整数乘加计算, 因此非常适合边缘端AI场景。
- ✓ **数字计算:** 随着AI任务的复杂性和应用范围增加, 高精度的大规模AI模型不断涌现。这些模型需要在数据中心等云端AI场景完成训练和推理, 产生巨大的算力需求, 相比于边缘端AI场景, 云端AI场景具有更多样的任务需求, 因此云端AI芯片必须兼顾能效、精度、灵活性等方面以保证各种大规模AI推理和训练; 数字存算一体主要以SRAM和RRAM作为存储器件, 采用先进逻辑工艺, 具有高性能高精度的优势, 且具备很好的抗噪声能力和可靠性, 因此较为适合在云端大算力高能效的商用场景。

AI芯片架构对比



高算力AI芯片新范式



1.6 存算一体存储介质百花齐放

- ◆ **存算一体应用场景极其广阔**：端侧小算力场景，算力范围约从16TOPS至100TOPS，例如智能可穿戴设备、智能安防、移动终端、AR\VR等。大算力场景，算力序曲在1000TOPS以上，例如云计算数据中心、自动驾驶、GPT-4等大模型等。我们认为云和边缘大算力场景，是存算一体芯片的优势领域，存算一体更具备核心影响力和竞争力；
- ◆ **SRAM、RRAM有望成为云端存算一体主流介质**：目前可用于存算一体的成熟存储器有NOR FLASH、SRAM、DRAM、RRAM、MRAM等；Nor Flash属于非易失性存储介质，在工艺制成上有明显瓶颈；DRAM成本较低，容量大，但是速度慢；SRAM在速度方面和能效比方面具有优势，特别是在存内逻辑技术发展起来之后具有明显的高能效和高精度特点；此外，PRAM有望成为下一代主流研究方向，虽然目前来看材料依旧不稳定，但是具有高速、结构简单的优点，有希望成为未来发展最快的新型存储器。

存算一体终端应用场景


场景分类	算力范围	存算一体优势
智能可穿戴设备	~100GOPS	可穿戴设备等消费类电子通常采用电池供电要求设备具有极低的功耗以延长使用时间。存算技术能够提高近一百倍不等的效率的同时降低10-20倍的功耗。
智能安防	~16TOPS	存算一体自带的高并行计算能力使得计算的实时性更高
移动终端	~32TOPS	在视觉信号处理上可以达到端侧产品低功耗要求
AR/VR	~64TOPS	在电池技术没有突破的当下，低功耗存算一体芯片是实现轻薄化的重要路径；同时在语音、手势识别中，存算一体在计算效率和实时性的优势能够得到充分发挥。

存算一体大算力应用场景

场景分类	算力范围	存算一体优势
数据中心	单板卡 1000TOPS 及以上	在较低成本、低功耗、低延迟下，初创公司能够借着存算一体芯片，用较小的代价实现高算力，突破芯片大厂的生态壁垒
自动驾驶		
GPT -4等大模型		

存储器类型及适合场景示意图

存储器类型	优势	不足	适合场景
RRAM	与CMOS工艺兼容性高、可并行处理大量数据	严重器件变化性	云端AI推理、智能汽车、AI数据中心、大模型等算力场景
SRAM(数字模式)	能效比高，高速高精度，对噪声不敏感，工艺成熟先进，适合IP化	存储密度略低	大算力、云计算、边缘计算
SRAM(模拟模式)	能效比高，工艺成熟先进	对PVT变化敏感，对信噪比敏感，存储密度略低	小算力、端侧、不要求待机功耗
DRAM	高存储密度，整合方案成熟	只能做近存计算，速度略低，工艺迭代慢、拓展性差	适合现有冯氏架构向存算过渡
MRAM	耐久性接近无线、功耗低	目前制造和设计比较困难、价格昂贵	智能电表、端侧、待机时间长、需要处理庞大数据量应用场景
Nor Flash	期间工艺成熟、存储阵列大、能够实现大规模运算	容量密度肖、写入速度慢	车载摄像头、激光雷达等端侧、待机时间长的场景



02 存算一体，打开海量应用空间

2.1 存算一体前景广阔、渐入佳境

- ◆ **存算一体需求旺盛，有望推动下一阶段的人工智能发展：**我们认为现在存算一体主要AI的算力需求、并行计算、神经网络计算等，因此存算一体需求旺盛；以数据中心为例，百亿亿次（E级）的超级计算机成为各国比拼算力的关键点，为此美国能源部启动了“百亿亿次计算项目”，我国则联合国防科大、中科曙光等机构推出首台E级超算，而E级超算面临的主要问题为功耗过高、现有技术超算功率高达千兆瓦，**需要一个专门的核电站来给它供电**，而其中50%以上的功耗都来源于数据的“搬运”，本质原因是计算与存储分离所致。
- ◆ **大模型兴起，存算一体适用于从云至端各类计算：**ChatGPT等“大模型”兴起，本质即为神经网络、深度学习等计算，因此，我们认为对算力需求旺盛；**端测方面**，人工智能更在意及时响应，即“输入”即“输出”，同时，随着存算一体发展，存内计算和存内逻辑，已经可以完成高精度计算；**云端方面**，随着大模型的横空出世，参数方面已经达到上亿级别，因此对算力的能耗方面考核更加严格，随着SRAM和PRAM等技术进一步成熟，存算一体有望成为新一代算力因素，从而推动人工智能产业的发展。

存算一体重点需求和相对优势

场景	重点需求	存算一体优势
端侧	低延时、低功耗、低成本、隐私性	当前存内计算产品已成功在端侧初步商用，提供语音、视频等 AI 处理能力，并获得十倍以上的能效提升，有效降低了端侧成本。
边侧	低延时、低功耗、低成本、通用性	存算一体在深度学习等领域有独特优势，可以提供比传统设备高几十倍的算效比，此外存内计算芯片通过架构创新可以提供综合性能全面兼顾的芯片及板卡，预计将在边侧推理场景中有着广泛的应用，为广泛的边缘 AI 业务提供服务。
云侧	大算力、高宽带、低功耗	存内计算可通过多核协同集成大算力芯片，结合可重构设计打造通用计算架构，存内计算作为智算中心下一代关键 AI 芯片技术，正面向大算力、通用性、高计算精度等方面持续演进，有望为智算中心提供绿色节能的大规模 AI 算力。

存算一体典型应用场景



2.1 存算一体前景广阔、渐入佳境

- ◆ 存算一体适用于人工智能各个场景，如穿戴设备、移动终端、智能驾驶、数据中心等。按照算力大小划分：
 - ✓ 端侧：针对端侧的可穿戴等小设备，对算力的要求远低于智能驾驶和云计算设备，但对成本、功耗、时延、开发难度很敏感。
 - ✓ **云测与边缘测**：针对云计算和边缘计算的大算力设备，是存算一体芯片的优势领域。我们认为存算一体本身在云和边缘测具有极高的适用性，原因是其具有高算力、低功耗、高性价比等优势；尤其是针对智能驾驶、数据中心等大算力应用场景，在可靠性、算力方面有较高要求，此外云计算市场玩家相对集中，因此我们认为存算一体在云计算市场有望先于端侧市场落地。
- ◆ **我们认为存算一体明确为下一代技术趋势**：目前存算一体国内外已就位起步阶段，相应差距并不大，原因是芯片设计层面皆为创新；存算一体是计算系统和存储系统的整合设计，比模拟IP和存储IP更复杂，依赖于多次存储器流片而积累的经验，因此我们认为在存储方面有优势和经验的企业具备先发优势；
 - ✓ **目前行业主要分为两种路径**：分别是小算力场景，例如音频类、终端应用场景；和大算力场景，例如云计算、智能驾驶、机器人等领域；

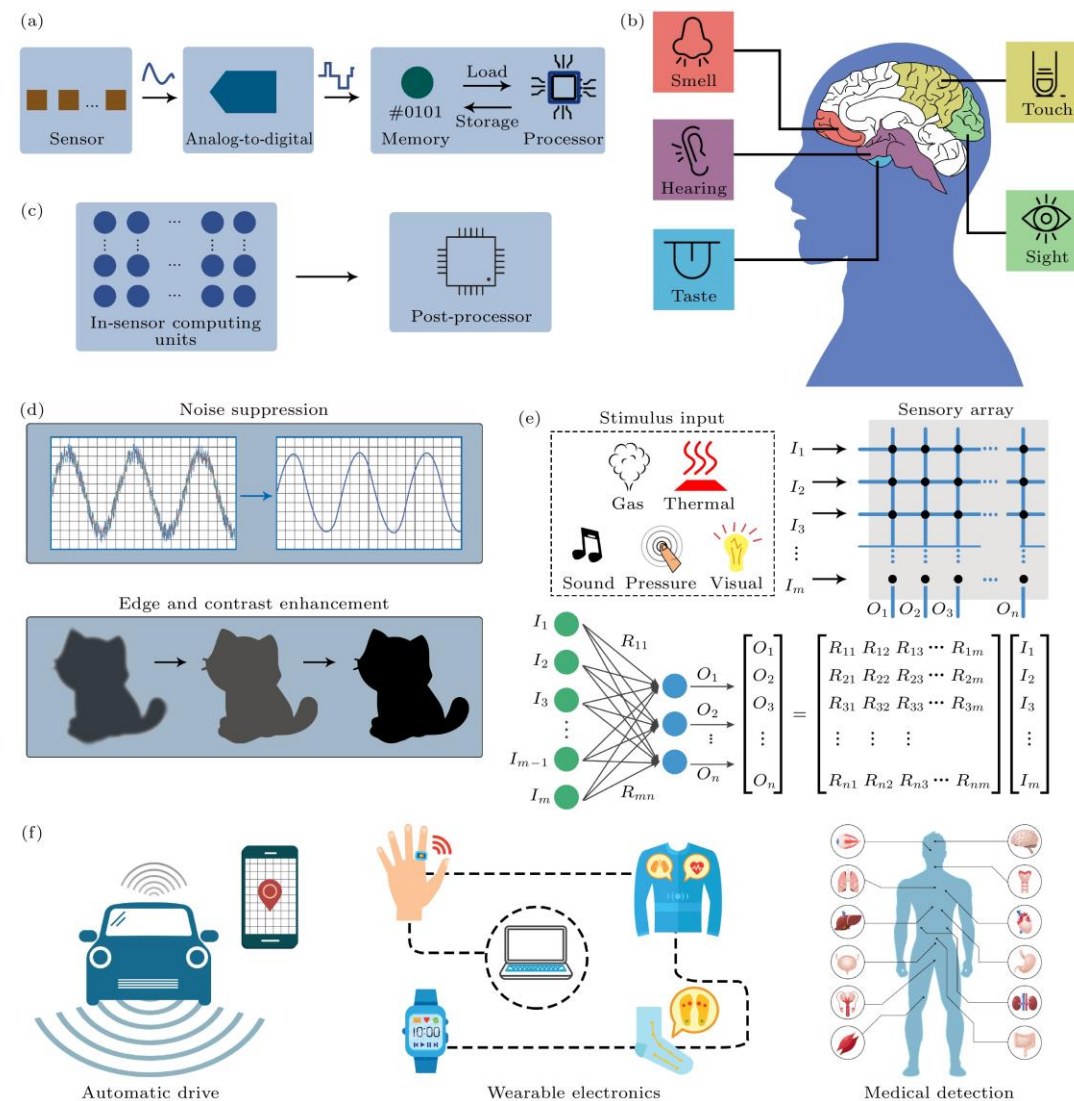
存算一体具体应用场景示意图



2.2 存算一体为人工智能发展的必然选择

- ◆ 我们认为存算一体为下一代技术趋势，原因及创新应用科技场景如下：
 - ✓ **人工智能神经网络相关应用：**存算一体具有能耗优势，其使用成本能大幅降低，可将带AI计算的中大量乘加计算的权重部分存在存储单元中从而在读取的同时进行数据输入和计算处理，从而完成卷积运算。因此存内计算和存内逻辑非常适合人工智能的深度神经网络应用和基于AI的大数据技术；
 - ✓ **感存算一体，多模态的人工智能计算：**集传感、储存和运算为一体构建感存算一体架构，我们认为其为多模态时代下的应用场景，通过传感器AI存算一体芯片，来实现零延时和超低功耗的智能视觉处理能力。我们认为其融入了CV(计算机视觉)、PRA(过程自动画)、神经网络计算等多种计算方式，存算一体芯片具有时效、能耗等多重优势；
 - ✓ **存算一体是类脑计算的关键技术基石：**类脑计算又称神经形态计算，是借鉴生物神经系统信息处理模式和结构的计算理论、体系结构、芯片设计等算法总称；类脑计算属于大算力高能耗反冲，存算一体极其适合大算力领域，因此天然适合应用在类脑计算领域，是类脑计算的关键技术基石。

感存算一体多模态示意图





03 投资建议: 梳理AIGC相关受益厂商

3.1 投资建议: 梳理AIGC的受益厂商

- ◆ 我们认为存算一体的出现是人工智能发展的必然选择，率先布局存储的厂商在介质上和技术上兼具备先发优势，因此积极地推荐以下四条投资路线：
 - ✓ 1)具备算力能力的厂商，受益标的为：**首都在线、中科曙光、拓维信息、神州数码、浪潮信息、寒武纪、海光信息、龙芯中科、景嘉微**；
 - ✓ 2)具备存储芯片能力的厂商，受益标的为：**东芯股份、兆易创新、北京君正、澜起科技、聚辰股份、普冉股份**；
 - ✓ 3)具备存储模组能力的厂商，受益标的为：**江波龙、佰维存储、德明利**；
 - ✓ 4)与存储相关的其他厂商，受益标的为：**中科曙光、易华录、同有科技、恒烁股份**。

3.1 投资建议: 梳理AIGC的受益厂商

AIGC的A股受益标的

公司名称	股票代码	收盘价	市值(亿元)	EPS(元)			PE(倍)		
		2023/4/4	2023/4/4	2021	2022E	2023E	2021	2022E	2023E
首都在线	300846.SZ	18.76	87.58	0.05	-	-	352.6	-	-
寒武纪	688256.SH	215.51	863.80	-2.06	-2.79	-1.89	-	-	-
拓维信息	002261.SZ	15.67	196.79	0.07	-0.04	0.15	223.9	-	103.4
神州数码	000034.SZ	31.60	211.40	0.37	1.56	1.88	85.7	20.2	16.8
龙芯中科	688047.SH	157.50	631.58	0.66	0.43	0.65	238.6	365.9	241.6
浪潮信息	000977.SZ	36.51	534.40	1.38	1.67	2.03	26.5	21.9	18.0
景嘉微	300474.SZ	130.85	595.54	0.97	0.64	0.93	134.9	205.9	140.4
海光信息	688041.SH	85.88	1996.14	0.16	0.40	0.61	531.4	214.1	141.8
东芯股份	688110.SH	39.59	175.09	0.77	0.79	1.03	51.4	50.0	38.3
兆易创新	603986.SH	130.56	870.87	3.54	4.00	3.95	36.9	32.6	33.1
北京君正	300223.SZ	110.24	530.88	1.97	2.10	2.34	55.9	52.4	47.1
澜起科技	688008.SH	74.00	840.70	0.73	1.17	1.63	101.4	63.1	45.4
聚辰股份	688123.SH	100.42	121.41	0.90	3.10	4.68	111.6	32.4	21.5
普冉股份	688766.SH	188.00	95.35	9.64	5.93	-	19.5	31.7	-
江波龙	301308.SZ	103.89	428.92	2.73	0.49	0.96	38.1	210.2	108.3
佰维存储	688041.SH	85.88	1996.14	0.16	0.40	0.61	531.4	214.1	141.8
德明利	001309.SZ	78.97	63.32	1.64	-	-	48.2	-	-
中科曙光	603019.SH	41.16	602.59	0.80	1.05	1.37	51.5	39.0	30.1
易华录	300212.SZ	46.50	309.60	-	0.20	0.32	-	238.2	143.2
同有科技	300033.SZ	226.97	1220.19	3.56	3.41	4.11	63.8	66.5	55.2
恒烁股份	688416.SH	57.99	47.92	2.43	0.29	1.00	23.9	199.7	57.7

注: 均来自wind一致预测

3.2.1 浪潮信息：中国服务器/AI服务器市占率稳居榜首

- ◆ **浪潮信息是全球领先的新型IT基础架构产品、方案及服务提供商：**公司是全球领先的 AI 基础设施供应商，拥有业内最全的人工智能计算全堆栈解决方案，涉及训练、推理、边缘等全栈 AI 场景，构建起领先的 AI 算法模型、AI 框架优化、AI 开发管理和应用优化等全栈 AI 能力，为智慧时代提供坚实的基础设施支撑。
- ◆ **公司算力技术壁垒浓厚：**生产算力方面，公司拥有业内最强最全的 AI 计算产品阵列，业界性能最好的Transformer 训练服务器 NF5488、全球首个 AI 开放加速计算系统 MX1、自研 AI 大模型计算框架 LMS。聚合算力层面，公司针对高并发训练推理集群进行架构优化，构建了高性能的NVMe 存储池，深度优化了软件栈，性能提升 3.5 倍以上。调度算力层面，浪潮信息 Aistation 计算资源平台可支持 AI 训练和推理，是业界功能最全的 AI 管理平台；同时，浪潮信息还有自动机器学习平台 AutoML Suite，可实现自动建模，加速产业化应用。

浪潮信息智算中心



浪潮信息智算中心



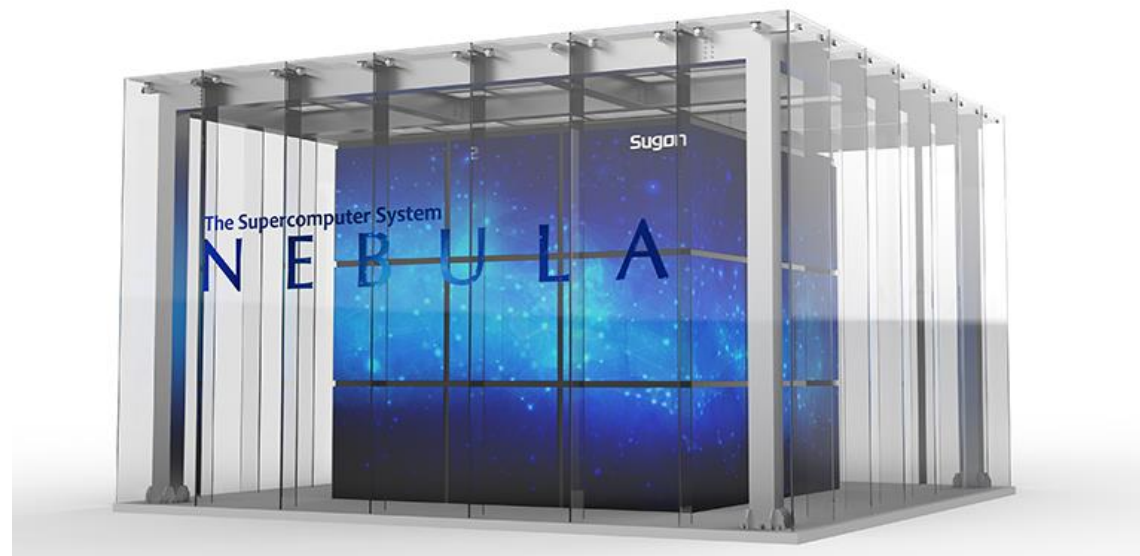
3.2.2 中科曙光：我国高性能计算、智能计算领军企业

- ◆ **中科曙光作我国核心信息基础设施领军企业：**在高端计算、存储、安全、数据中心等领域拥有深厚的技术积淀和领先的市场份额，并充分发挥高端计算优势，布局智能计算、云计算、大数据等领域的技术研发，打造计算产业生态，为科研探索创新、行业信息化建设、产业转型升级、数字经济发展提供了坚实可信的支撑。
- ◆ **依托先进计算领域的先发优势和技术细节，中科曙光全面布局智能计算：**完成了包括AI核心组件、人工智能服务器、人工智能管理平台、软件等多项创新，构建了完整的AI计算服务体系。并积极响应时代需求，在智能计算中心建设浪潮下，形成了5A级智能计算中心整体方案。目前，曙光5A智能计算中心已在广东、安徽、浙江等地建成，江苏、湖北、湖南等地已进入建设阶段，其他地区也在紧张筹备和规划中。

中科曙光主要产品

<p>通用服务器</p> <ul style="list-style-type: none"> 机架式服务器 高密度服务器 刀片服务器 核心应用服务器 	<p>智能计算服务器</p> <ul style="list-style-type: none"> 深度学习训练 智能应用推理 	<p>终端&工作站</p> <ul style="list-style-type: none"> 微型计算机 工作站 	<p>高性能计算机</p> <ul style="list-style-type: none"> 通用高性能计算机 高性能计算机系统组件 高性能计算机的服务支撑
<p>机房冷却设施</p> <ul style="list-style-type: none"> 微模块产品 液冷基础设施产品 	<p>存储产品</p> <ul style="list-style-type: none"> 分布式统一存储 多控统一存储 高密度存储服务器 备份一体机 	<p>网络安全产品</p> <ul style="list-style-type: none"> 数据中心安全产品 汇聚分流设备 智能加速卡 网络内容识别分析系统 网络态势感知系统 	<p>大数据平台软件</p> <ul style="list-style-type: none"> 大数据智能引擎系列 数据工程服务系列 视频智能分析系列 大数据与人工智能实训平台
<p>云计算平台软件</p> <ul style="list-style-type: none"> 云计算操作系统 超融合一体机 云桌面 云容灾 	<p>计算服务</p> <ul style="list-style-type: none"> 弹性计算服务 混合计算服务 专有计算服务 API 托管、运营 	<p>云计算服务</p> <ul style="list-style-type: none"> 云服务器 ECS 裸金属 BMS 对象存储 OSS 云容器实例 CCI 人工智能服务 数据开发 DDS 数据治理中心 DGS 数据服务 DSS 数据可视化 DAV 数据集成 Data Integration 	<p>城市云</p> <ul style="list-style-type: none"> 智慧城市 国资云 交通云 医疗云
<p>5A级智算中心</p>			

中科曙光硅立方液体相变冷却计算机



3.2.3 神州数码: 华为生态核心践行者

- ◆ **神州数码领先的数字化转型:** 神州数码围绕企业数字化转型的关键要素, 开创性的提出“数云融合”战略和技术体系框架, 着力在云原生、数字原生、数云融合关键技术和信创产业上架构产品和服务能力, 为处在不同数字化转型阶段的快消零售、汽车、金融、医疗、政企、教育、运营商等行业客户提供泛在的敏捷IT能力和融合的数据驱动能力。
- ◆ **神州数码为华为生态核心践行者:** 公司旗下的神州鲲泰基于华为鲲鹏处理器多款不同种类的服务器产品, 包括1、单路服务器: R222、R224; 2、双路服务器: R522、R524、R722、R724、R2240、R2260、R2280。3、四路服务器: R822。此外, 公司基于华为鲲鹏920处理器与昇腾Atlas AI加速卡, 神州数码开发了采用ARM架构的一系列AI服务器。

神州数码服务器及相关参数

名称	示意图	形态	处理器	内存支持	AI加速卡/AI处理器	AI算力
KunTai A222		2U单路边缘机架式服务器	1*鲲鹏920处理器, 24核, 主频2.6GHz	4个DDR4 RDIMM, 最高速率3200MT/s内存保护支持ECC、SEC/DED、SDDC、Patrol scrubbing功能; 单根内存条容量支持16GB/32GB/64GB/128GB	最大支持3张Atlas 300V 视频解析卡或Atlas 300I Pro 推理卡或Atlas 300V Pro 视频解析卡	最大420 TOPS INT8
KunTai A722		2U 双路推理型 AI 机架式服务器	2*鲲鹏920处理器, 支持32、48、64核可选, 主频2.6GHz	16个或32个DDR4 RDIMM, 最高速率2933MT/s内存保护支持ECC、SEC/DED、SDDC、Patrol scrubbing功能; 单根内存条容量支持16GB/32GB/64GB/128GB	最大支持8张, Atlas 300V 视频解析卡或Atlas 300I Pro 推理卡或Atlas 300V Pro 视频解析卡	最大1120 TOPS INT8
KunTai A924		4U四路训练型AI机架式服务器	4*鲲鹏920处理器, 支持48核, 主频 2.6GHz	支持32个DDR4内存插槽, 速率最高2933MT/s内存保护支持ECC、SEC/DED、SDDC、Patrol scrubbing功能; 单根内存条容量支持32GB/64GB/128GB	8*昇腾910, 支持直出100G RoCE网络接口	最大512Tops Int8或256Tops FP16

3.2.4 拓维信息: 华为生态重要参与者

- ◆ **拓维信息是领先的软硬一体化解决方案提供商:** 公司1996年成立, 业务涵盖政企数字化、智能计算、鸿蒙生态, 覆盖全国31个省级行政区、海外10+国家, 聚焦数字政府、运营商、考试、交通、制造、教育等重点领域和行业, 服务超过1500家政企客户, 为其提供全栈国产数字化解决方案和一站式全生命周期的综合服务。
- ◆ **拓维信息为华为生态重要参与者:** “兆瀚”系列通用服务器是基于ARM架构, 搭载鲲鹏920处理器设计开发的机架式服务器, 拥有高的性能、可靠性、高效环保、兼容性强等特点; “兆瀚”系列AI服务器能够满足当前各类主流AI场景与AI大模型的训练需求, 已经在国内多个区域人工智能计算中心、城市人工智能中枢、通用AI服务器场景中得到了应用, 已经在国内多家头部互联网企业开展适配测试。

拓维信息旗下“兆瀚”系列服务器产品介绍

种类	名称	示意图	形态	处理器	内存支持	AI加速卡/AI处理器	AI算力
通用服务器	兆瀚RH220系列		2U双路机架	支持两颗华为鲲鹏920处理器, CPU主频2.6GHz。单CPU最多64个内核, 最大功率180w。	最多支持32个DDR4内存DIMM插槽, 最高速率2933MT/s	/	/
	兆瀚RH520系列		4U机架服务器	支持两颗华为鲲鹏920处理器, CPU主频2.6GHz。单CPU最多64个内核, 最大功率180w。	最多支持32个DDR4内存DIMM插槽, 最高速率2933MT/s	/	/
AI服务器	兆瀚RA2300-A		2U推理服务器	支持两颗华为鲲鹏920处理器, CPU主频2.6GHz。单CPU最多64个内核, 最大功率180w。	最多支持32个DDR4内存DIMM插槽, 最高速率2933MT/s	支持Atlas 300I Pro推理卡和Atlas 300V Pro视频解析卡	最大1.12 POPS INT8; 最大560 TFLOPS PF16
	兆瀚SA300		2U智能边缘服务器	支持一颗华为鲲鹏920处理器, CPU主频2.6GHz。单CPU最多64个内核, 最大功率181w。	最多支持4个DDR4内存DIMM插槽, 最高速率2934MT/s	支持Atlas 300I Pro推理卡/Atlas 300V Pro视频解析卡	最大420 TOPS INT8 或 384路1080P 30 FPS视频解析(硬件解码能力)
	兆瀚RA5900-A		4U训练服务器	支持四颗华为鲲鹏920处理器, CPU主频2.6GHz。单CPU最多64个内核, 最大功率182w。	最多32个DDR4内存插槽, 支持RDIMM。单根内存条容量支持32 GB/64GB	8*昇腾910	/
	兆瀚RA2302-B		2U AI 服务器	2*64核青松处理器	32个DDR4内存插槽, 最高3200 MT/s, 支持ECC	最大支持4个Atlas 300I/V Pro	最大560 TPOS INT8

3.2.5 海光信息：支持全精度，GPU实现规模量产

- ◆ **海光信息主要从事高端处理器、加速器等计算芯片产品和系统的研究、开发，主要产品包括海光CPU和海光DCU:**2018年10月，公司启动深算一号DCU产品设计，海光8100采用先进的FinFET工艺，典型应用场景下性能指标可以达到国际同类型高端产品的同期水平。2020年1月，公司启动DCU深算二号的产品研发。
- ◆ **海光DCU性能强大:**海光DCU基于大规模并行计算微结构进行设计，不但具备强大的双精度浮点计算能力，同时在单精度、半精度、整型计算方面表现同样优异，是一款计算性能强大、能效比较高的通用协处理器。海光DCU集成片上高带宽内存芯片，可以在大规模数据计算过程中提供优异的数据处理能力。

海光信息主要产品



系列	7000系列CPU	5000系列CPU	3000系列CPU	系列	8000系列DCU
核心规格	最大32个物理核心	最大16个物理核心	最大8个物理核心	核心规格	60-64个深度计算单元
应用领域	高端通用服务器、先进计算系统	通用服务器	个人工作站、工控设备等终端产品	应用领域	先进计算系统、人工智能

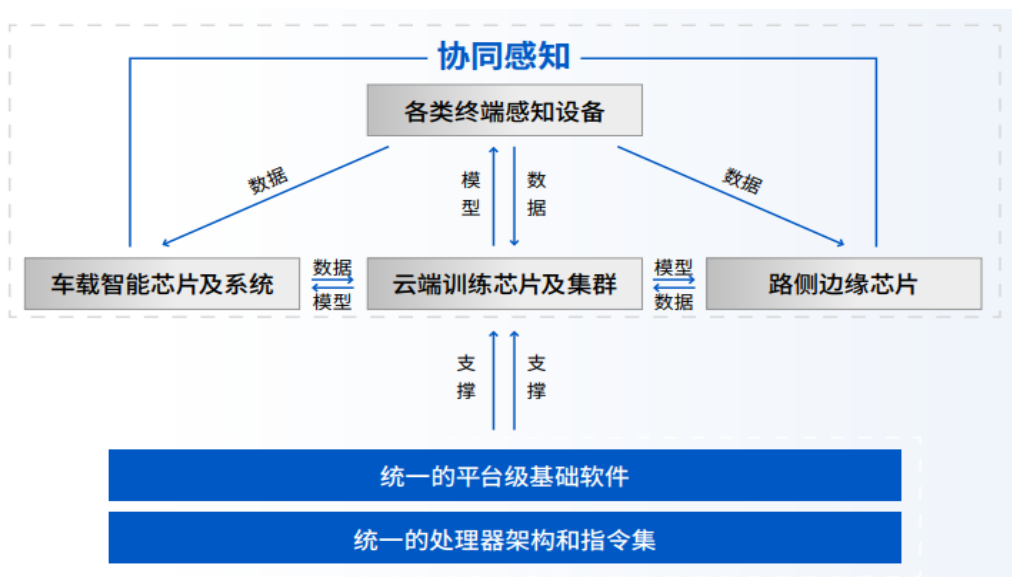
海光深算一号性能达到国际同类产品水平

项目	海光	NVIDIA	AMD
品牌	深算一号	Ampere 100	M1100
生产工艺	7nm FinFET	7nm FinFET	7nm FinFET
核心数量	4096 (64CUs)	2560 CUDA processors 640 Tensor processors	120CUs
内核频率	Up to 1.5GHz (FP64) Up to 1.7Ghz (FP32)	Up to 1.53Ghz	Up to 1.5GHz (FP64) Up to 1.7Ghz (FP32)
显存容量	32GB HBM2	80GB HBM2e	32GB HBM2
显存位宽	4096 bit	5120 bit	4096bit
显存频率	2.0 GHz	3.2 GHz	2.4 GHz
显存带宽	1024 GB/s	2039 GB/s	1228 GB/s
TDP	350 W	400 W	300 W
CPU to GPU 互联	PCIe Gen4 x 16	PCIe Gen4 x 16	PCIe GEN4 x 16
GPU to GPU 互联	xGMI x 2, Up to 184 GB/s	NVLink up to 600 GB/s	Infinity Fabric x 3, up to 276 GB/s

3.2.6 寒武纪：少数全面掌握AI芯片技术的企业之一

- ◆ **寒武纪是目前国际上少数几家全面系统掌握了通用型智能芯片及其基础系统软件研发和产品化核心技术的企业之一**：寒武纪主营业务是应用于各类云服务器、边缘计算设备、终端设备中人工智能核心芯片的研发和销售。公司的主要产品包括终端智能处理器IP、云端智能芯片及加速卡、边缘智能芯片及加速卡以及与上述产品配套的基础系统软件平台。
- ◆ **公司AI技术积累浓厚**：能提供云边端一体、软硬件协同、训练推理融合、具备统一生态的系列化智能芯片产品和平台化基础系统软件。2022年3月，寒武纪正式发布了新款训练加速卡“MLU370-X8”，搭载双芯片四芯粒封装的思元370，集成寒武纪MLU-Link多芯互联技术，主要面向AI训练任务。

寒武纪“云边端车”协同



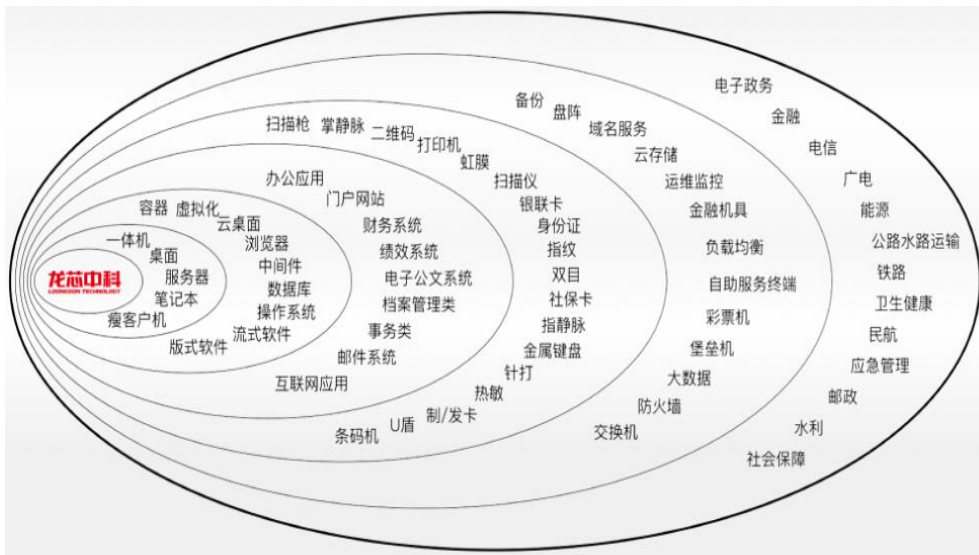
寒武纪产品技术图谱



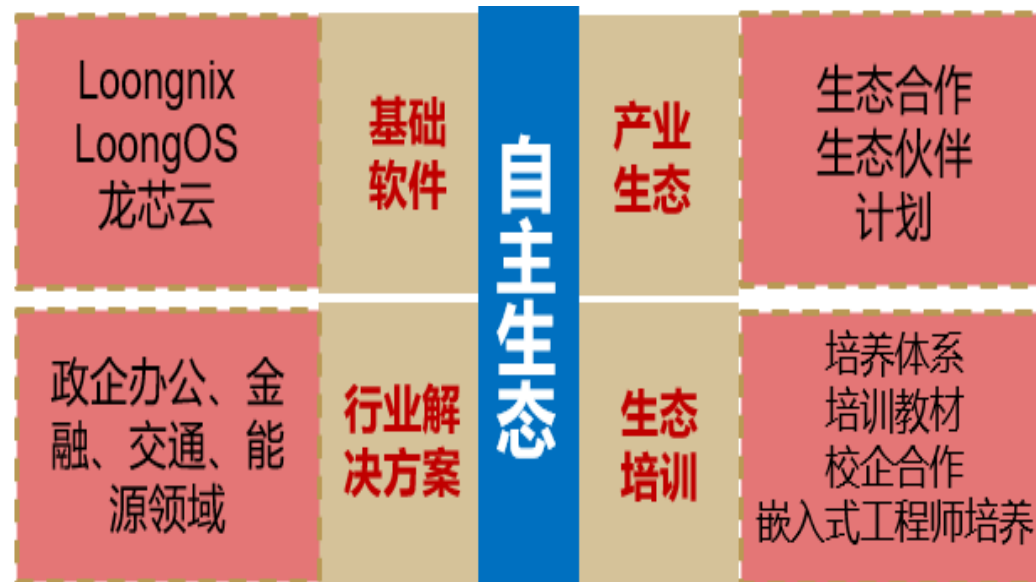
3.2.7 龙芯中科：2K2000系列集成自主GPU

- ◆ **龙芯中科主营业务为处理器及配套芯片的研制、销售及服务：**主要产品与服务包括处理器及配套芯片产品与基础软硬件解决方案业务。公司基于信息系统和工控系统两条主线开展产业生态建设，面向网络安全、办公与业务信息化、工控及物联网等领域与合作伙伴保持全面合作，产品在电子政务、能源、交通、金融、电信、教育等行业领域已获得广泛应用。
- ◆ **公司自主研发2K200系列GPU：**2022年12月，龙芯2K2000完成了初步功能调试及性能测试，达到其设计目标，2023年将推出试用。龙芯2K2000集成了两个LA364处理器核，典型工作频率为1.5GHz，共享2MB的L2缓存，SPEC2006INT (base) 单核定/浮点分值达到13.5/14.9分。龙芯2K2000芯片集成了龙芯自主研发的GPU，并优化了图形算法和性能。

龙芯中科生态合作示意图



龙芯中科自主生态



3.2.8 景嘉微：新一代JM9系列有望打开商用市场

- ◆ **国产GPU龙头企业:** 公司成立于2006年，主要从事军用电子产品的研发、生产、销售，目前形成了三大业务板块分别是图形线控模块、小型专用雷达和芯片业务。GPU方面，2014年首推JM5400实现了军用GPU的国产替代；第二款芯片JM7200于2018年研发成功，具备了PC端的功能；日前，公司9系列芯片研发成功，具备高性能计算能力。
- ◆ **新一代JM9系列有望打开商用市场:** 日前，公司JM9系列图形处理芯片已顺利发布，应用领域涵盖地理信息系统、媒体处理、CAD辅助设计、游戏、虚拟化等高性能显示和人工智能计算领域。目前，信创市场为公司提供了新的业务增长点，JM9系列图形处理芯片的成功发布将为公司未来进一步拓展通用市场提供强有力的产品支撑。

景嘉微GPU系列产品



景嘉微7系列GPU示意图



3.3.1 东芯股份：多类别存储芯片助力企业具备先发优势

- ◆ **东芯股份拥有自主知识产权，努力打造中国领先的存储设计企业**：公司是一家Fabless（无晶圆厂）芯片企业，主要从事芯片设计和销售业务，将晶圆制造、封装测试等生产环节委托第三方完成，拥有自主知识产权，聚焦于中小容量NAND/NOR/DRAM芯片的研发、设计和销售，是目前国内少数可以同时提供NAND/NOR/DRAM设计工艺和产品方案的存储芯片研发设计公司。公司愿景是成为中国领先的存储设计企业，使命为提供可靠高效的存储产品及设计方案。
- ◆ **担当本土存储“芯”使命，研发前瞻性产品—存算一体化芯片&DTR NAND**：东芯半导体持续研发和优化产品性能，致力于研发1×nm NAND Flash芯片，聚焦于高附加值产品，研发前瞻性产品：存算一体化芯片&DTR NAND。东芯半导体在已有的多类别存储技术的基础上，叠加新的研发方案，在存算一体布局中具有很高的先发优势。

存储产品之间差异性

比较项目	非易失性		易失性
	NAND Flash	NOR Flash	DRAM
存储原理	浮栅型	浮栅型/电子俘获型	电容充放电型
读取速度	较慢	较快	极快
擦除/写入速度	快	较慢	极快
存储容量	高 (Gb/Tb)	中 (Mb/Gb)	中 (Mb/Gb)
擦写次数	十万级别	十万级别	-


公司产品工艺流程



3.3.2 兆易创新：身处“集成电路设计”行业的IC设计企业

- ◆ **在整个产业链中处于重要地位并拥有核心竞争力：**兆易创新成立于2005年4月，主要业务为存储器、微控制器和传感器的研发、技术支持和销售。公司产品广泛应用于工业、消费类电子、汽车、物联网、计算、移动应用以及网络和电信行业等各个领域，助力社会智能化升级。公司作为 IC 设计企业，自成立以来一直采取 Fabless 模式，专注于集成电路设计、销售和客户服务环节，将晶圆制造、封装和测试等环节外包给专门的晶圆代工、封装及测试厂商。
- ◆ **作为全球化芯片设计公司，兆易创新致力于存储器、控制器及周边产品的设计研发。**公司存储器产品包括：闪存芯片（NOR Flash、NAND Flash）和动态随机存取存储器（DRAM），公司以存储为主，控制器及周边产品为辅，多赛道多产品线的组合布局多元化布局助力穿越周期影响，技术和产品优势不断增强。

存储器主要参数

Model	Length	Width	Height(Max)	Pitch
 USON6 1.2*1.2 mm	1.20mm	1.20mm	0.40mm	0.40mm
 USON8 1.5*1.5 mm	1.50 mm	1.50 mm	0.50 mm	0.40 mm
 USON8 3*2 mm (0.45mm)	3.00 mm	2.00 mm	0.50 mm	0.50 mm
 USON8 3*4 mm	3.00 mm	4.00 mm	0.60 mm	0.80 mm
 USON8 4*4 mm	4.00 mm	4.00 mm	0.50 mm	0.80 mm

存储器产品介绍

英文名称	中文名称	介绍	应用
NOR Flash	代码型闪存芯片	主要用来存储代码及少量数据	公司 NOR Flash 产品广泛 应用于物联网、工业及汽车电子、穿戴式设备、人工智能、网络通信、安防监控产品、PC 主板、移动设备、数字机顶盒、路由器、家庭网关等领域
NAND Flash	数据型闪存芯片	分为两大类：大容量 NAND Flash 主要为 MLC、TLC 2D NAND或 3D NAND，擦写次数从几百次至数千次，多应用于大容量数据存储；小容量 NAND Flash 主要是 SLC 2D NAND，可靠性更高，擦写次数达到数万倍以上。	公司 NAND Flash 产品属于SLC NAND，为移动设备、机顶盒、数据卡、电视、汽车电子等设备的多媒体数据存储应用提供所必需的大容量存储。
DRAM	动态随机存取存储器	是当前市场中最为重要的系统内存，在计算系统中占据核心位置，广泛应用于服务器、移动设备、PC、消费电子等领域。因极高的技术和资金壁垒，DRAM 领域市场处于高度集中甚至垄断态势。	公司首款自有品牌 DRAM 产品已于 2021 年 6 月推出，实现了从设计、流片，到封测、验证的全国产化，在满足消费类市场强劲需求的同时，助力 国产自主供应生态圈的发展构建。该产品主要面向消费类、工业控制类及汽车类等市场领域，应用于机顶盒、电视、监控、网络通信、智慧家庭、平板电脑、车载影音系统等诸多领域

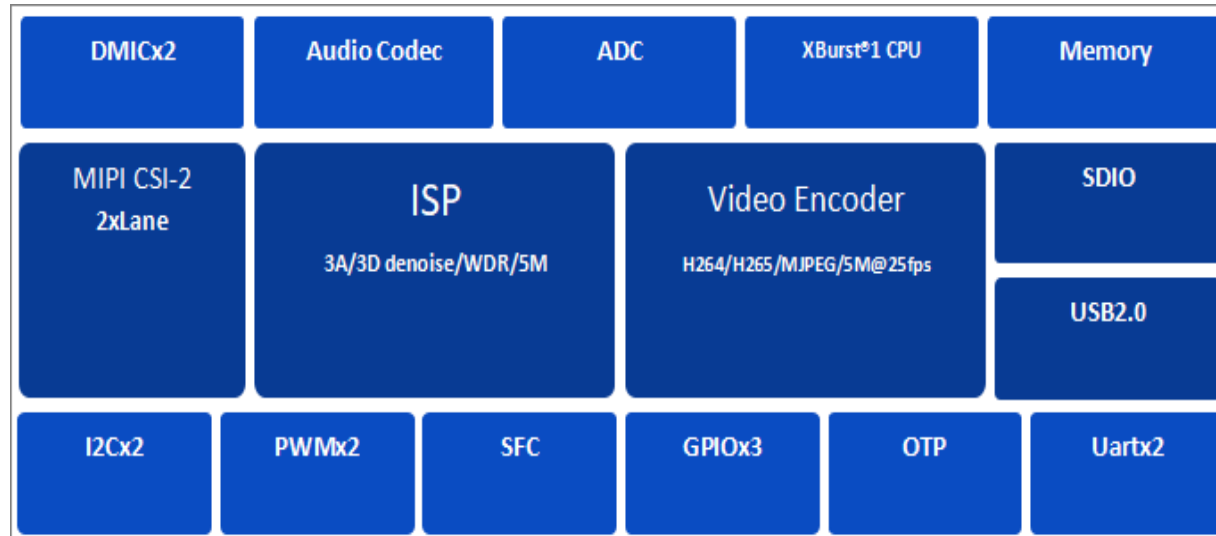
3.3.3 北京君正：多年深耕自主创新

- ◆ **公司拥有较强的自主创新能力**：多年来在自主创新CPU技术、视频编解码技术、图像和声音信号处理技术、SoC芯片技术、软件平台技术等多个领域形成多项核心技术。公司已形成可持续发展的梯队化产品布局，基于自主创新的XBurst CPU和视频编解码等核心技术，公司推出了一系列具有高性价比的微处理器芯片产品和智能视频芯片产品，各类别的芯片产品分别面向不同的市场领域。
- ◆ **2020年，君正完成对北京矽成（ISSI）及其下属子品牌Lumissil的收购，并拥有其100%股份。**其中，ISSI存储部门有高速低功耗SRAM，低中密度DRAM，NOR/NAND Flash，嵌入式Flash pFusion，及eMMC等芯片产品。模拟和互联部门Lumissil有LED驱动、触控传感、音频驱动、微处理器、电源管理和互联等芯片产品。

Xburst系列CPU Core参数

	XBurst1	XBurst2
Base ISA	MIPS32 R5	MIPS32 R5
SIMD Extension	MXU2.0 - 128bits SIMD	MSA128 - 128bits SIMD MXA128 - 128bits SIMD MXU3.0- 512bits SIMD
Micro-Architecture	9 stage pipeline Single issue	Dual-Issue In-Order 2 Hardware Threads SMT
Coremark	2.3	3.6 (single thread)
Power Consumption	0.07mW/MHz, 40nm	0.13mW/MHz, 28nm

芯片框图



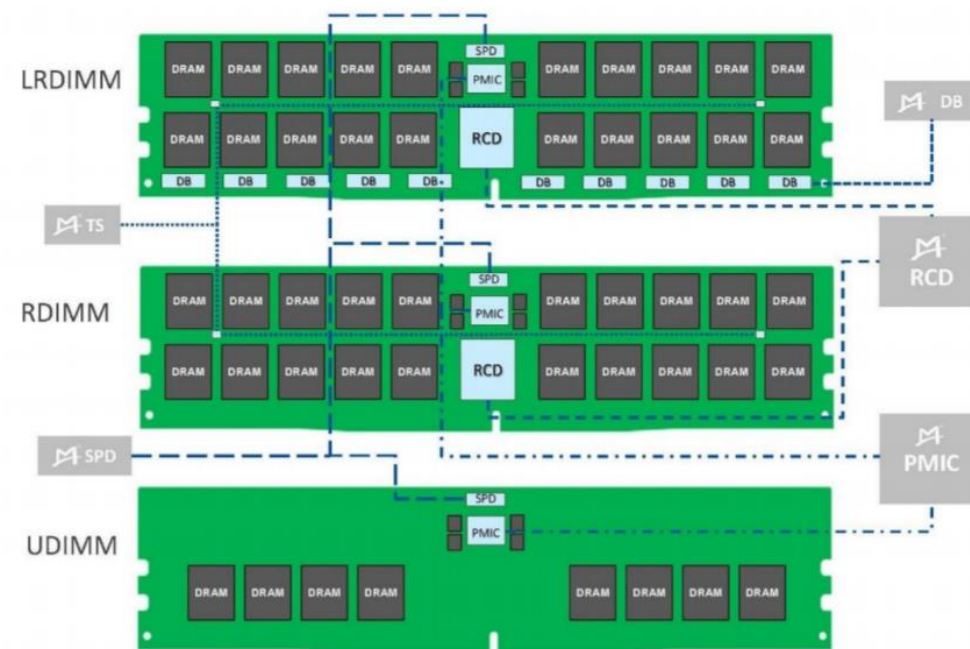
3.3.4 澜起股份: 业界领先的集成电路设计公司

- ◆ **业界领先的集成电路设计公司，澜起股份为全球仅有的3家内存接口芯片供应商之一。** 主要经营模式为Fabless模式，公司的两大产品线为互连类芯片产品线（主要包括内存接口芯片、内存模组配套芯片、PCIe Retimer 芯片、MXC 芯片等）和津逮®服务器平台产品线（包括津逮®CPU 和混合安全内存模组（HSDIMM®））。同时，公司正在研发基于“近内存计算架构”的 AI 芯片。
- ◆ **公司发明的DDR4全缓冲“1+9”架构被采纳为国际标准。** 现已成为全球可提供从DDR2到DDR4内存全缓冲/半缓冲完整解决方案的主要供应商之一，在内存接口芯片市场位列全球前二及内存模组配套芯片。

公司DDR4内存接口芯片自带产品及其应用情况

DDR4 内存接口芯片产品	应用
Gen1.0 DDR4 RCD 芯片	DDR4 RDIMM 和 LRDIMM, 支持速率达 DDR4-2133
Gen1.0 DDR4 DB 芯片	DDR4 LRDIMM, 支持速率达 DDR4-2133
Gen1.5 DDR4 RCD 芯片	DDR4 RDIMM 和 LRDIMM, 支持速率达 DDR4-2400
Gen1.5 DDR4 DB 芯片	DDR4 LRDIMM, 支持速率达 DDR4-2400
Gen2 DDR4 RCD 芯片	DDR4 RDIMM 和 LRDIMM, 支持速率达 DDR4-2666
Gen2 DDR4 DB 芯片	DDR4 LRDIMM, 支持速率达 DDR4-2666
Gen2 Plus DDR4 RCD 芯片	DDR4 RDIMM、LRDIMM 和 NVDIMM, 支持速率达 DDR4-3200
Gen2 Plus DDR4 DB 芯片	DDR4 LRDIMM, 支持速率达 DDR4-3200



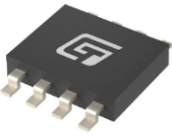
公司DDR5内存接口芯片及内存模组配套芯片



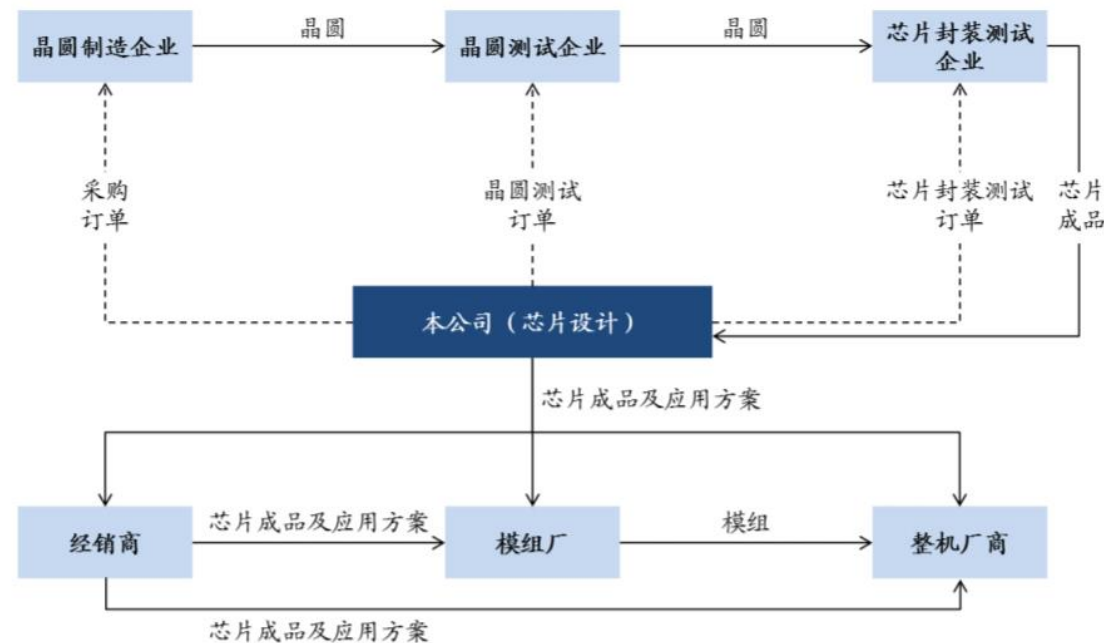
3.3.5 聚辰股份：十年“创芯”发展，精准洞悉市场新需求

- ◆ 作为一家全球化的芯片设计高新技术企业，聚辰半导体长期致力于为客户提供存储、模拟和混合信号集成电路产品并提供应用解决方案和技术支持服务。公司目前拥有EEPROM、音圈马达驱动芯片和智能卡芯片三条主要产品线，产品广泛应用于智能手机、液晶面板、蓝牙模块、通讯、计算机及周边、医疗仪器、白色家电、汽车电子、工业控制等众多领域。
- ◆ 公司已在智能手机摄像头、液晶面板、计算机及周边等细分领域奠定了领先优势，未来公司将持续以市场需求为导向，以自主创新为驱动，对EEPROM、音圈马达驱动芯片、智能卡芯片等现有产品线进行完善和升级，并积极开拓NOR Flash、电机驱动芯片等新产品领域。

聚辰股份主要产品一览

产品种类	细分种类	型号	特性	产品示意图
EEPROM	传统应用领域	I2C系列	EEPROM（电可擦除可编程只读存储器）是一类通用型的非易失性存储芯片，在断电情况下仍能保留所存储的数据信息，可以在计算机或专用设备上擦除已有信息重新编程，耐擦写性能至少100万次。聚辰的EEPROM产品具有高可靠性、宽电压、高兼容性、低功耗等特点。公司EEPROM产品线包括I2C、SPI和Microwire等标准接口的系列EEPROM产品，以及主要应用于计算机和服务器内存条的SPD/SPD+TS（温度传感器）系列EEPROM产品。	
		SPI系列		
		Microwire系列		
		SPD/SPD+TS系列		
EEPROM	车规级	A2	具备高可靠性和低失效率等优势，擦写次数最高可达400万次以上，其温度适应能力强，数据可存储100年。新产品SPI NOR Flash也按照车规等级125°C设计。产品主要应用于车载摄像头、液晶显示、娱乐系统等外围部件，并逐渐延伸到新能源车三电系统、车身控制、底盘传动及微电机、智能座舱等核心部件。	
		A1		
NOR Flash		1.65~3.6V	NOR Flash以其合适的容量、灵活的存取操作、及其非易失性产品特性，非常适合作为智能设备的指令程序存储器。随着5G、IOT、AMOLED、TDDI、TWS及汽车电子等应用市场快速发展，NOR Flash的需求保持持续增长动力。聚辰半导体以其领先的存储器设计技术，推出SPI NOR Flash产品，可以覆盖从消费级，到工业级，直至汽车级的所有应用，产品在可靠性，功耗，温度和速度等关键性能指标方面的技术领先性。	
		2.7~3.6V		
		1.65~1.95V		
		1.1~2.0V		

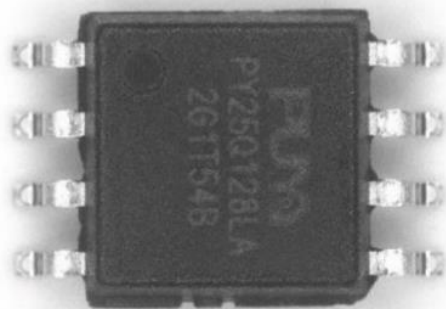
聚辰股份整体业务流程



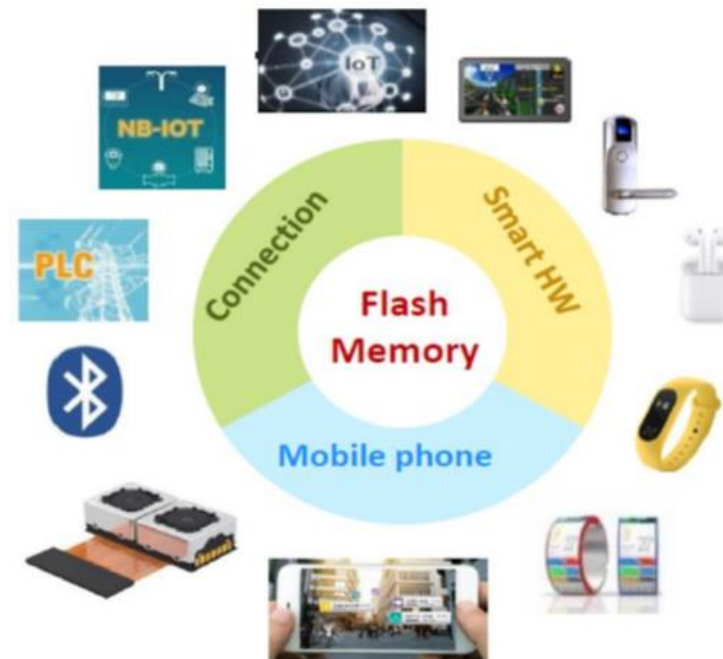
3.3.6 普冉股份：深耕非易失性存储，建立多方稳定合作

- ◆ **凭借低功耗、高可靠性的产品优势，在下游客户处积累了良好的品牌认可度：**普冉股份是国内低功耗SPI NOR Flash存储芯片和高可靠性IIC EEPROM存储器芯片NOR Flash和EEPROM的主要供应商之一。公司已经和汇顶科技等主控原厂、深天马等手机屏幕厂商、舜宇等行业领先的手机摄像头模组厂商建立了稳定的业务合作关系，产品应用于三星、OPPO、华为、小米、联想、惠普等品牌厂商。
- ◆ **普冉的产品覆盖Flash、EEPROM、Analog、MCU、良品裸晶圆、晶圆级芯片封装。**普冉产品在下游应用广泛，包括车用电子、工业用电子、通讯用电子、手机用电子、消费类电子、物联网。现阶段公司正积极开拓海外市场，寻求和其他品牌厂商的潜在合作机会。

普冉股份NOR Flash产品



普冉股份产品体系发展示意图



3.4.1 江波龙：存储新维度提升企业核心竞争力

- ◆ **江波龙聚焦NAND Flash闪存应用和存储芯片定制、存储软硬件开发，处于国内存储器领先地位**：江波龙电子成立于1999年，主要从事Flash及DRAM存储器的研发、设计和销售。拥有嵌入式存储、移动存储、固态硬盘及内存条4条产品线，提供消费级、工规级、车规级存储器以及行业存储软硬件应用解决方案，拥有行业类存储品牌FORESEE和国际高端消费类存储品牌Lexar（雷克沙）。公司深耕存储行业20余年，专注于半导体存储产品与应用，形成固件开发、存储芯片测试、集成封装设计、存储产品定制等核心竞争力。
- ◆ **构建存储新维度，迈向存储综合服务商**：1、公司2022年成功发布第一款自主研发SLC NAND Flash，在Flash技术应用取得了重大突破。2、通过新型材料在存储产品中的创新应用，解决存储散热、数据安全、工艺可靠性等一系列封装材料问题。3、已经拥有覆盖eMMC、UFS、SD&SSD等存储产品核心标准专利储备，消除了产品出海核心专利壁垒问题。4、积极参与共建ITM专利池，联合协会伙伴推动NM2.0的技术标准研究及产品落地。

自主研发SLC NAND Flash



自主研发SLC NAND Flash

从无到有，自主研发设计，更懂Flash技术应用

多样性产品覆盖
容量：512Mb~8Gb，中国大陆首个512Mb SPI NAND
封装：TSOP/BGA/WSON

高性能高可靠
数据可靠性：撰写10万次/数据保存10年
JEDEC & AEC-Q100可靠性考核

领先高效测试技术
自研存储专用测试机台，支持同测数1200+
低成本高覆盖率量产测试

全过程品质控制
全过程SPC、SYL、SBL管控
颗粒级过程追溯能力

新型材料在存储产品中的创新应用



新型材料在存储产品中的创新应用

FORESEE eMMC UFS2.2 UFS3.1

封装散热材料
产品过X-Ray或其他辐射条件时，存储数据不致损坏

导热系数	1000
热膨胀系数	2000
热冲击耐受度	137-233
弹性模量	237
重量	401

数据保护材料

封装工艺材料
保证SIP产品工艺性能和可靠性要求

结构外壳材料
3D打印材料；三颗和抗信号干扰材料等

读取速度快
500MB/s, 1000MB/s, 1900MB/s

单颗容量大
128GB(2X), 256GB(4X), 1TB(8X)

封装尺寸同
11.5*13(mm), 11.5*13(mm), 11.5*13(mm)

创新NM Card



创新NM Card

Nano SIM + Lexar NM Card 256GB 90MB/s

存储与SIM二合一

便捷一键换机

信息安全存储

社交情景全记录

NM2.0 双卡合一

3.4.2 佰维存储：先进芯片设计和封装技术有望打开市场

- ◆ **佰维存储专注于存储芯片研发与封测制造，是国家高新技术企业，国家级专精特新小巨人企业：**公司整合了存储器研发设计、固件算法开发、先进封装、测试设备研发与算法开发、品牌运营等，构筑了研发封测一体化的经营模式。公司存储芯片产品广泛应用于移动智能终端、PC、行业终端、数据中心、智能汽车、移动存储等信息技术领域，主要包括嵌入式存储、消费级存储、工业级存储、先进封测服务四大板块，是国内率先进入全球科技巨头供应链体系的存储器企业。
- ◆ **持续加大存储器核心领域研发投入，打造全国一流半导体厂商：**佰维存储当前处于高速成长阶段，半导体存储器领域，公司是国内半导体存储器率先进入顶级品牌供应链的企业。在封装测试领域，佰维多层叠Die、超薄Die、多芯片异构集成等先进工艺等核心封装测试指标均达到业内一流、国际领先的水平。

佰维存储产品集合

嵌入式芯片

- > **嵌入式存储**
 - eMMC
 - eMCP
 - UFS
 - ePOP
 - BGA SSD
 - DMMC
- > **内存**
 - LPDDR5
 - LPDDR4
- > **小容量存储**
 - SPI NAND

存储模组

- > **固态硬盘**
 - 2.5寸 SATA SSD
 - Half Slim SSD
 - U.2 SSD
 - M.2 SSD
 - mSATA SSD
- > **内存产品**
 - DDR5 SODIMM
 - DDR5 UDIMM
 - DDR4 SODIMM
 - DDR4 UDIMM
- > **存储卡**
 - CF 卡
 - CFast卡
 - CFexpress Card
 - NM卡
 - SD Card

信创产品

- > **信创SSD**
 - 2.5寸SATA SSD
 - M.2 SATA SSD
 - M.2 NVMe SSD
- > **信创内存模组**
 - DDR4 SO-DIMM
 - DDR4 U-DIMM

封测服务

- > **可封测产品**
 - 嵌入式存储芯片

佰维存储封装形式



BGA



LGA



QFN



SIP



TSOP



制程能力













快封服务

3.4.3 德明利：芯存无限可能

- ◆ **集中于闪存主控芯片设计、研发，存储模组产品应用方案的开发、优化，以及存储模组产品的销售**，德明利以闪存主控芯片的自主设计、研发为基础，结合主控芯片固件方案及量产工具开发、存储模组测试等形成完善的存储模组应用方案，高效实现对NAND Flash存储颗粒应用性能提升和数据管理。
- ◆ **积极布局人机交互触控领域完成初步业务并完成自研触摸控制芯片投片**，公司在此基础上形成针对不同应用场景的触控模组一体化解决方案，已实现小批量试产出货，为客户提供 6.5 寸至 21.5 寸的小、中、大多尺寸显示屏的触控芯片产品，并逐步导入智能家电领域、后装车机领域和中大屏商显领域等市场。

德明利主要存储产品一览

产品种类	细分种类	型号	特性	示意图
移动存储器	USB	USB MODULE	高可用度的Flash资源，满足消费级标准、满足高可靠性与耐用性要求。支持USB2.0/3.0，2GB-256GB	
		USB Drive	持有USB 3.2 Gen 1性能，容量：16/32/64/128GB，接口：USB-C，Type-C	
	存储卡	SD	SD2.0/3.0，容量：2GB-1TB	
		MICRO SD	兼容SD5.1，支持UHS-I总线模式，容量：8GB-128GB	
固态驱动器	SSD	2.5英寸SATA	符合SATA Revision 3.1标准。规格参数：2.5英寸。容量：960GB，1TB	
		Half Slim SATA	符合SATA Revision 3.1标准。规格参数：Half Slim。容量：120GB，128GB	
		M.2 SATA 2242	符合SATA Revision 3.1标准。规格参数：M.2。容量：120GB，128GB	
		M.2 SATA 2280	符合SATA Revision 3.1标准。规格参数：M.2。容量：120GB，128GB	
		M.2 Pcle 2280	符合Pcle Gen3，NVME1.3标准。规格参数：M.2 2280。容量：120GB，128GB	
		mSATA	符合SATA Revision 3.1标准。规格参数：Msata。容量：120GB，128GB	

德明利UDStore亮相CFMS峰会，共话存储解决方案



3.5.1 中科曙光：液冷存储系统为存储降耗打开新思路

- ◆ **中科曙光是国内高性能计算领域领军企业，亚洲第一大高性能计算机厂商：**中科曙光于1996年在中国科学院大力推动下组建成立，经过20多年的发展，公司在高端计算、存储、安全、数据中心等领域拥有了深厚的技术积淀和领先的市场份额，在全国各省、自治区和直辖市均设立分支机构，拥有国际领先的5大智能制造生产基地和7大研发中心。公司未来会继续致力于核心技术的研发，推动产业进步，助力“数字中国”发展。
- ◆ **公司具备完善的存储产品体系，持续聚焦核心技术自研：**中科曙光存储产品包括分布式统一存储、多控统一存储、高密度存储服务器，完善的产品体系能够轻松应对数据量快速增长和业务系统转型升级带来的挑战。公司坚持全自研路线与核心技术突破，不断将创新技术应用用于产品开发。且针对不同的客户开发定制化解决方案，成功覆盖气象、能源、金融、生物信息等丰富的应用场景。
- ◆ **曙光ParaStor液冷存储系统采取“冷板式液冷”技术，实现节能降耗，提升存储性能：**存储芯片研发人员在实现对存储设备节能降耗、提升芯片性能等方面从未停止努力，中科曙光发布的业界首款液冷存储——曙光ParaStor液冷存储系统，实现将液冷技术与存储技术深度结合，显著降低内存、HDD存储关键部件的工作温度，大幅降低散热功率。相比于风冷，液冷存储系统性能提升超过20%。

中科曙光存储产品分类



曙光ParaStor液冷存储系统



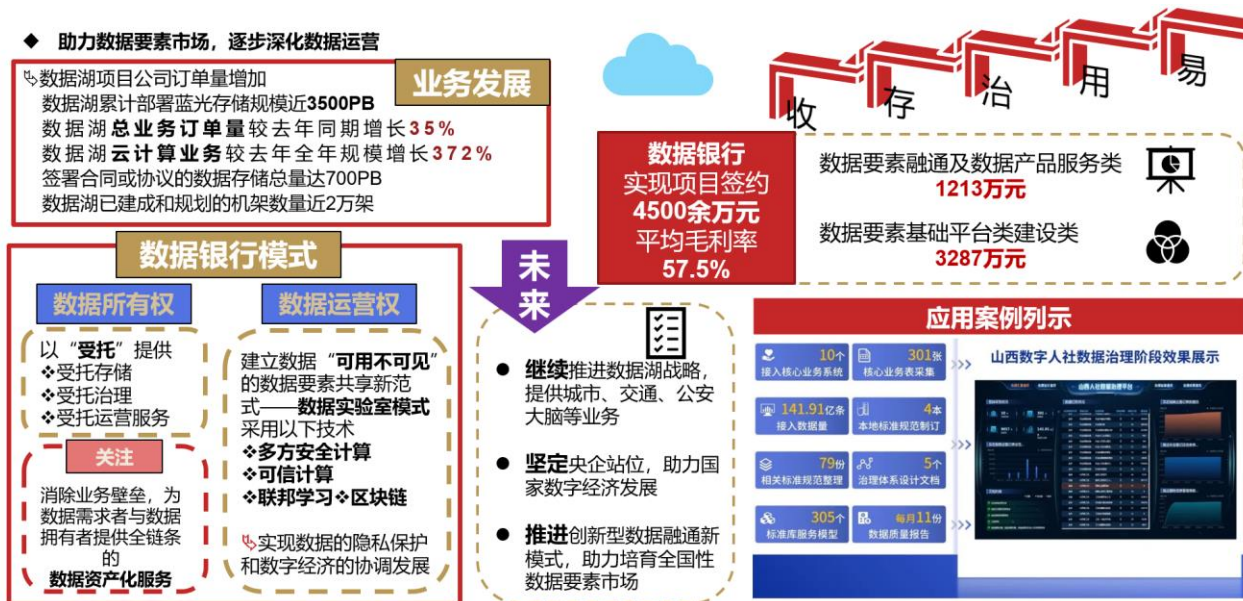
3.5.2 易华录：数据要素先行者

- ◆ **易华录数据要素先行者**：易华录转型成为数字经济基础设施的建设和运营商，业务主要围绕政企数字化与数据运营服务，坚定以数据湖战略为中心，围绕超级储存和数据变现两大业务主线提升数据资产化服务能力。
- ◆ **旗下蓝光存储产品与“数据湖”优势显著**：蓝光存储架构体系光盘技术，于2016年实现300G光盘量产 2020年500G小规模量产，1TB产品增加多值化处理技术，正在全新 IC 开发；数据湖优势为：**1、存算分离**，融合 Hadoop、MPP、图数据库等主流计算引擎，与自主研发超级存储引擎对接，可实现存算分离；**2、高兼容性**，遵循 DAMA3.0\DCMM标准，与华为等主流大数据平台适配兼容；**3、高独立性**，不依赖特定平台，完全适配华为 ARM 架构服务器且兼容鲲鹏处理器指令集，已取得兼容性证书。

公司蓝光存储产品架构



公司“数据银行”收益模式



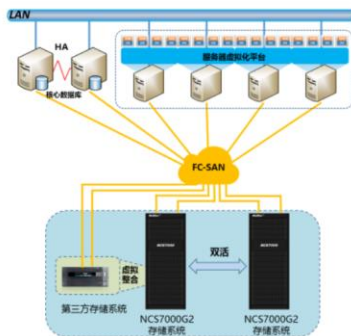
3.5.3 同有科技：深耕存储几十年

- ◆ **作为中国存储第一股**，同有科技是业界少数拥有超过百项自主知识产权的专业存储厂商，也是国内完全拥有存储架构、存储管理软件及存储核心算法的自主研发企业。公司通过“内生+外延”战略，持续整合存储产业链资源，以最大的协同效应带动中国存储行业整体技术进步。
- ◆ **数字经济政策红利持续释放，带来存储行业新增长点；闪存、分布式存储技术成熟发展，推动市场空间快速扩大**：基于闪存技术的固态硬盘（SSD）具备高性能、高可靠、低延迟、低功耗快速响应企业用户需求，全闪存存储系统可以提供百万级的 IOPS 以及毫秒级的延迟，具备较高的安全性和稳定性，已在金融联机交易系统、医疗 HIS 系统、运营商 BOSS 系统等关键重载业务领域得到较好的应用。

双活存储解决方案

方案概述

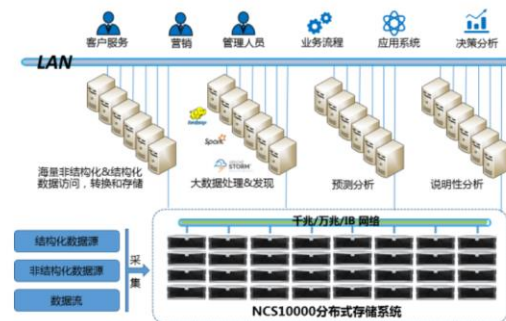
随着IT应用的快速发展，政府、金融、电信等行业越来越多用户要求核心业务7*24小时持续运行。针对以上用户的需求，同有科技推出双活存储解决方案。无论是主机、网络，还是磁盘阵列等主要部件出现故障，该方案都能保证业务系统的正常运行，从而实现对用户关键业务的系统连续性保护。



大数据分析存储解决方案

方案概述

为多种大数据计算平台提供有效支撑，为深度数据挖掘提供数据保障



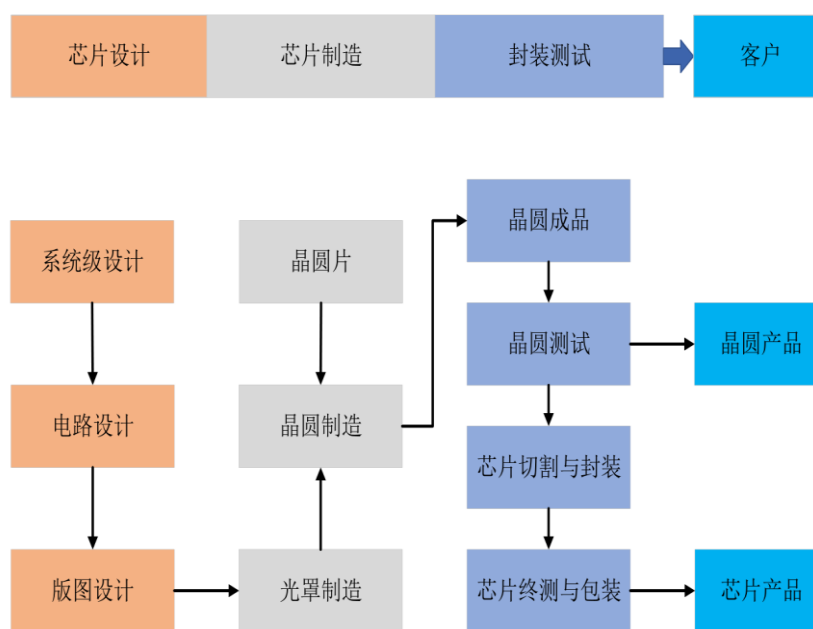
3.5.4 恒烁股份：CiNOR存算一体推理芯片处于研发阶段

- ◆ **恒烁股份拥有高质量核心团队，专注集成电路核心技术研发**：恒烁股份是主营业务为存储芯片和MCU芯片研发、设计及销售的集成电路企业，其核心团队来自美国硅谷和国内存储器顶尖企业，现有主营产品包括NOR Flash存储芯片和基于Arm Cortex -M0+内核架构的通用32位MCU芯片，同时，公司还在致力于开发基于NOR闪存技术的存算一体终端推理AI芯片，并提供边缘计算的完整解决方案。
- ◆ **公司致力于开发CiNOR存算一体终端推理AI芯片**：CiNOR存算一体AI推理芯片拟用于小算力、低功耗应用场景，主要专注在边缘和终端设备上，专注以下四种终端市场应用场景：人脸识别、语音关键词识别、心电图检测及电力设备故障声纹检测，与ChatGPT等云端应用差异较大。目前尚处于研发阶段，还未实现量产。

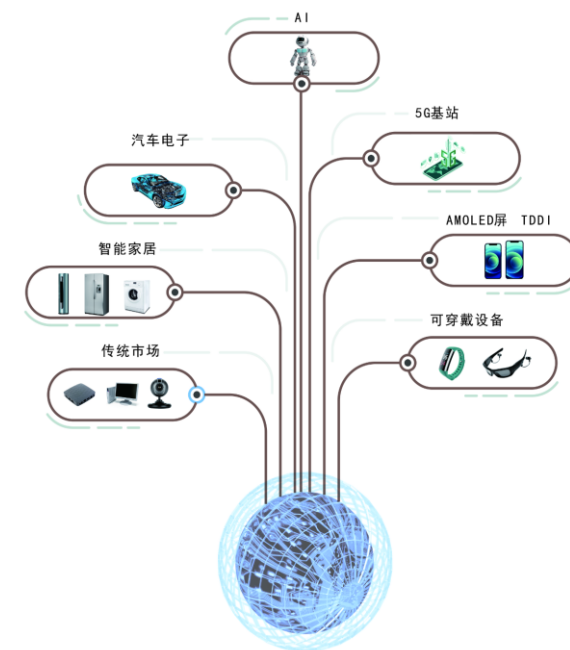
公司NOR Flash主要产品系列型号

电压	产品图示	制程	主要型号	应用领域
低电压 1.65-2V		65nm 50nm	ZB25LD20、ZB25LD40、ZB25LD80 ZB25LQ80、ZB25LQ16、ZB25LQ32、ZB25LQ64、ZB25LQ128	手机 TDDI/AMOLE D屏, 4G CAT1 模组、智能穿戴 设备
高电压 2.3-3.6V		65nm 50nm	ZB25D20、ZB25D40、ZB25D80、ZB25D16 ZB25VQ40、ZB25VQ80、ZB25VQ16、ZB25VQ32、ZB25VQ64、ZB25VQ128	智能电表、电视 机、机顶盒、PC BIOS、安防产 品、网通产品、 POS机、银行安 全芯片等
宽电压 1.65-3.6V		65nm	ZB25WD20、ZB25WD40、ZB25WD80 ZB25WQ40、ZB25WQ80、ZB25WQ16	物联网 IoT 产 品、TWS 耳机、 BLE 蓝牙及其 他电池驱动通 讯模组

公司主要产品业务流程图



公司产品应用场景





04 风险提示

风险提示

- ◆ **核心技术水平升级不及预期的风险:** AIGC相关产业技术壁垒较高，公司核心技术难以突破，进程低于预期，影响整体进度。
- ◆ **AI伦理风险:** AI可能会生产违反道德、常规、法律等内容。
- ◆ **政策推进不及预期的风险:** 受到宏观经济、财政、疫情影响，政策推进节奏不及预期。
- ◆ **中美贸易摩擦升级的风险:** 供应链存在部分海外提供商，容易受到美国“卡脖子”制裁，导致产品研发不及预期。

分析师与研究助理简介

刘泽晶（首席分析师）2014-2015年新财富计算机行业团队第三、第五名，水晶球第三名，10年证券从业经验。

分析师承诺

作者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，保证报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求客观、公正，结论不受任何第三方的授意、影响，特此声明。

评级说明

公司评级标准	投资评级	说明
以报告发布日后的6个月内公司股价相对上证指数的涨跌幅为基准。	买入	分析师预测在此期间股价相对强于上证指数达到或超过15%
	增持	分析师预测在此期间股价相对强于上证指数在5%—15%之间
	中性	分析师预测在此期间股价相对上证指数在-5%—5%之间
	减持	分析师预测在此期间股价相对弱于上证指数5%—15%之间
	卖出	分析师预测在此期间股价相对弱于上证指数达到或超过15%
行业评级标准		
以报告发布日后的6个月内行业指数的涨跌幅为基准。	推荐	分析师预测在此期间行业指数相对强于上证指数达到或超过10%
	中性	分析师预测在此期间行业指数相对上证指数在-10%—10%之间
	回避	分析师预测在此期间行业指数相对弱于上证指数达到或超过10%

华西证券研究所：

地址：北京市西城区太平桥大街丰汇园11号丰汇时代大厦南座5层

网址：<http://www.hx168.com.cn/hxqz/hxindex.html>

华西证券股份有限公司（以下简称“本公司”）具备证券投资咨询业务资格。 。
本公司不会因接收人收到或者经由其他渠道转发收到本报告而直接视其为本公司客户。

本报告基于本公司研究所及其研究人员认为的已经公开的资料或者研究人员的实地调研资料，但本公司对该等信息的准确性、完整性或可靠性不作任何保证。本报告所载资料、意见以及推测仅于本报告发布当日的判断，且这种判断受到研究方法、研究依据等多方面的制约。在不同时期，本公司可发出与本报告所载资料、意见及预测不一致的报告。本公司不保证本报告所含信息始终保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者需自行关注相应更新或修改。

在任何情况下，本报告仅提供给签约客户参考使用，任何信息或所表述的意见绝不构成对任何人的投资建议。市场有风险，投资需谨慎。投资者不应将本报告视为做出投资决策的惟一参考因素，亦不应认为本报告可以取代自己的判断。在任何情况下，本报告均未考虑到个别客户的特殊投资目标、财务状况或需求，不能作为客户进行客户买卖、认购证券或者其他金融工具的保证或邀请。在任何情况下，本公司、本公司员工或者其他关联方均不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告而导致的任何可能损失负有任何责任。投资者因使用本公司研究报告做出的任何投资决策均是独立行为，与本公司、本公司员工及其他关联方无关。

本公司建立起信息隔离墙制度、跨墙制度来规范管理跨部门、跨关联机构之间的信息流动。务请投资者注意，在法律许可的前提下，本公司及其所属关联机构可能会持有报告中提到的公司所发行的证券或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。在法律许可的前提下，本公司的董事、高级职员或员工可能担任本报告所提到的公司的董事。

所有报告版权均归本公司所有。未经本公司事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容，如需引用、刊发或转载本报告，需注明出处为华西证券研究所，且不得对本报告进行任何有悖原意的引用、删节和修改。

THANKS





前沿报告库是中国新经济产业咨询报告共享平台。行业范围涵盖新一代信息技术、5G、物联网、新能源、新材料、新消费、大健康、大数据、智能制造等新兴领域。为企事业单位、科研院所、投融资机构等提供研究和决策参考。



扫一扫免费
获取海量报告

