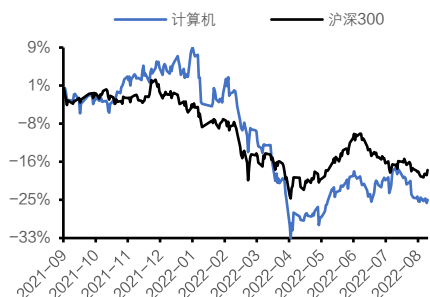


## 投资评级：看好(维持)

### 最近 12 月市场表现



#### 分析师 杨烨

SAC 证书编号: S0160522050001  
 yangye01@ctsec.com

#### 分析师 罗云扬

SAC 证书编号: S0160522050002  
 luoyy@ctsec.com

### 相关报告

1. 《否极泰来,把握产业主航道 ——计算机 2022 年中报总结》 2022-09-04

## 核心观点

- ▶ **算力是数字经济时代生产力。**全社会的数据量呈指数级增长,其中企业端占比不断提升。同时,数据的重要性正在提升,单个数据的价值下降,多维数据整合方可挖掘深度价值,进一步催生算力需求。全球的超级数据中心占比快速提升,算力呈现进一步集中。我们认为算力已经成为数字经济时代的关键生产力,目前中美两国的算力规模全球领先,根据 IDC 的研究数据,计算力和 GDP 增长显著相关。面对这一产业浪潮,我国“东数西算”工程于 2022 年 2 月启动,“数据向西,算力向东”,进一步提升国家整体算力,预计每年带动 4000 亿元投资。
- ▶ **后摩尔时代,算力从 PC 走向 AI 集群。**AI 算力需求急剧增加,传统架构难以满足,下游应用的高景气度和不断衍生的智能化需求,推动芯片革新。GPU 是当前主流数据中心端 AI 计算架构,2021 年服务器 GPU 全球市场规模达到 71.5 亿美元,三年 CAGR 高达 66%。Nvidia 2022 年上半年数据中心收入占比超过 50%,AMD 向数据中心端发力,2021Q4 已拿下服务器 CPU 的 X86 市场份额的 10.7%,并仍在持续提升。无论是 Nvidia 还是 AMD,市值均已超越传统 CPU 龙头 Intel,标志着时代算力从 PC 端走向集群 AI 端。
- ▶ **人工智能开启算力时代。**数据是 AI 模型的“汽油”,算力是 AI 模型的“发动机”,算法是 AI 模型的“大脑”,前述 AI 三要素相互耦合,生成 AI 模型。伴随场景与数据的激增,算法层面的“大模型”横空出世,深度神经网络模型为 AI 行业注入新动能,加快了传统行业的智能化转型,并正在向超大规模预训练模型演进。这一 AI 算法趋势也进一步推升了 AI 计算需求量,我们认为面对摩尔定律下单个处理器的性能提升逼近极限、计算核心数量增加可能带来的功耗、成本压力,异构计算有望打破 AI 芯片算力瓶颈。
- ▶ **投资建议:**大模型、复杂应用场景推动算力从 PC 端走向 AI 集群, AI 进入算力时代,我们建议关注海光信息(688041)、寒武纪(688256)、中科曙光(603019)、浪潮信息(000977)、宝信软件(600845)。
- ▶ **风险提示:**政策落地不及预期、疫情扰动、研发进展不及预期。

表 1: 重点公司投资评级

| 代码     | 公司    | 总市值<br>(亿元) | 收盘价<br>(09.09) | EPS (元) |       |       | PE    |       |       | 投资评级 |
|--------|-------|-------------|----------------|---------|-------|-------|-------|-------|-------|------|
|        |       |             |                | 2021A   | 2022E | 2023E | 2021A | 2022E | 2023E |      |
| 688041 | 海光信息  | 1,297.0     | 55.8           | 0.16    | 0.40  | 0.63  | 345.3 | 140.9 | 89.3  | 增持   |
| 603019 | 中科曙光  | 372.2       | 25.4           | 0.80    | 1.12  | 1.48  | 31.8  | 22.8  | 17.2  | 增持   |
| 688256 | 寒武纪-U | 288.6       | 72.0           | -2.06   | -2.35 | -1.81 | -35.0 | -30.6 | -39.7 | 未覆盖  |
| 000977 | 浪潮信息  | 332.4       | 22.7           | 1.37    | 1.67  | 2.06  | 16.6  | 13.6  | 11.0  | 未覆盖  |
| 600845 | 宝信软件  | 817.9       | 41.4           | 1.20    | 1.17  | 1.49  | 34.6  | 35.5  | 27.8  | 未覆盖  |

数据来源: Wind 数据, 财通证券研究所

注: 海光信息与中科曙光 2022-2023 年 EPS 为财通证券研究所预测, 其余为 Wind 一致预期。

## 内容目录

|   |    |
|---|----|
| 1. 算力是数字经济时代生产力 .....                               | 4  |
| 1.1. 数据成为经济发展重要生产要素 .....                           | 4  |
| 1.2. 数据重要性提升，多维数据整合挖掘深度价值，催生算力旺盛需求 .....            | 5  |
| 1.3. 算力正成为经济增长主要动力之一 .....                          | 8  |
| 1.4. 东数西算，数字经济时代新基建启航 .....                         | 10 |
| 2. 后摩尔时代，算力从 PC 走向 AI 集群 .....                      | 13 |
| 2.1. 后摩尔时代，传统计算架构亟待革新 .....                         | 13 |
| 2.2. AI 算力需求旺盛，GPU 是主流 .....                        | 14 |
| 2.3. Nvidia&AMD：市值超越 Intel，时代算力从 PC 端走向 AI 集群 ..... | 15 |
| 3. 人工智能开启算力时代 .....                                 | 18 |
| 3.1. AI 三要素：数据、算法、算力 .....                          | 18 |
| 3.2. “大模型”横空出世，人工智能进入算力时代 .....                     | 20 |
| 3.3. 异构计算突破算力瓶颈 .....                               | 21 |
| 4. 投资建议 .....                                       | 24 |
| 5. 风险提示 .....                                       | 24 |

## 图表目录

|  |    |
|--|----|
| 图 1. 计算的发展 .....                             | 4  |
| 图 2. 全球数据量规模（2016-2035） .....                | 5  |
| 图 3. 数据的创建与存储位置占比（2010-2025） .....           | 5  |
| 图 4. 数据类型占比（2010-2025） .....                 | 5  |
| 图 5. IDC 测算 2015-2025 年数据量增速（按数据重要性划分） ..... | 6  |
| 图 6. 数据按安全防护需求程度划分占比（2010-2025） .....        | 6  |
| 图 7. 数据平台技术演进 .....                          | 7  |
| 图 8. 数据中心负载任务量(2016-2021) .....              | 7  |
| 图 9. 超级数据中心数量变化(2016-2021) .....             | 7  |
| 图 10. 主要国家算力指数排名及同比去年涨幅（2021A） .....         | 8  |
| 图 11. 全球各国算力规模与占全球比例（2021） .....             | 8  |
| 图 12. 不同国家的“算力指数”与 GDP 的关系 .....             | 9  |
| 图 13. 得出“算力指数”所需的具体指标与计算口径 .....             | 9  |
| 图 14. 中国数据中心算力规模（2020-2025E） .....           | 10 |
| 图 15. 全球每年产生数据量（2016-2025E） .....            | 10 |
| 图 16. “东数西算”布局规划 .....                       | 12 |
| 图 17. “东数西算”相关政策出台历程 .....                   | 12 |
| 图 18. AI 算力需求剧增 30 万倍（2013-2019） .....       | 13 |
| 图 19. 中国云计算市场规模及增速（2016-2021） .....          | 13 |
| 图 20. 云计算市场规模相对总 IT 支出占比（2019-2021） .....    | 13 |
| 图 21. 不同技术架构的 AI 芯片特征对比 .....                | 14 |
| 图 22. 全球服务器 GPU 与 FPGA 市场规模（2019-2021） ..... | 15 |
| 图 23. 全球服务器 GPU 占服务器中市场规模比例（2019-2021） ..... | 15 |
| 图 24. Nvidia 收入分项占比（2014-2022H1） .....       | 15 |

|  |    |
|--|----|
| 图 25. Nvidia 数据收入、增速及总收入增速（2015–2022H1） .....                    | 15 |
| 图 26. X86 芯片各细分市场 AMD 份额情况（2018Q4–2021Q4） .....                  | 16 |
| 图 27. Nvidia、AMD 与 Intel 的市值之比（2006.2.28–2022.09.07，月频）<br>..... | 16 |
| 图 28. Intel/Nvidia/AMD 业绩(2021).....                             | 17 |
| 图 29. Intel/Nvidia/AMD 业绩(2022H1) .....                          | 17 |
| 图 30. Intel/Nvidia/AMD 的 PS TTM（2018.04–2022.09） .....           | 17 |
| 图 31. Intel/Nvidia/AMD 的 PE TTM (2018.04–2022.09) .....          | 17 |
| 图 32. 算力服务分类及介绍（2021 年） .....                                    | 18 |
| 图 33. AI 三要素及 AI 模型训练/推断原理.....                                  | 19 |
| 图 34. 人工智能近十年重要发展历程（2012–2021 年） .....                           | 20 |
| 图 35. AI 大模型及其参数规模（2018–2022 年） .....                            | 21 |
| 图 36. 主要人工智能模型算力需求一览（1960–2020 年） .....                          | 22 |
| 图 37. 过去五十年微处理机芯片趋势数据（1970–2020 年） .....                         | 23 |

## 1. 算力是数字经济时代生产力

### 1.1. 数据成为经济发展重要生产要素

按照 IDC 总结，全球计算数据的创建与使用总结为三个阶段

- 1980 年以前：数据集中在数据中心存储与处理，即使可通过远程访问，终端计算能力低下，也无法对数据进行深度计算，数据均为商用。
- 1980-2000：摩尔定律显威，PC 兴起，数据的产生、处理与存储更多流向终端，同时出现了音乐、电影和游戏的数字娱乐产业。
- 2000 至今：无线宽带普及推动数据进入云端，将数据与特定的物理设备脱钩。社交与流媒体的兴起使得数据中心更多承担数据交互与集中计算任务，而在不断交互过程中数据量高速增长。

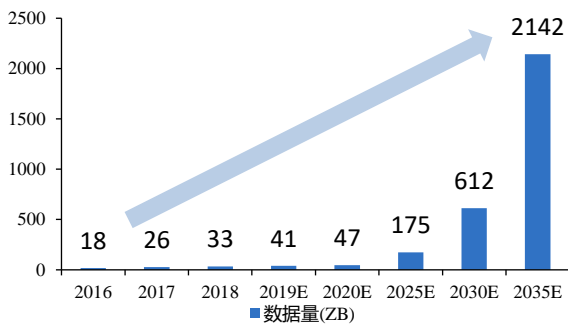
图 1. 计算的发展



数据来源：IDC，财通证券研究所

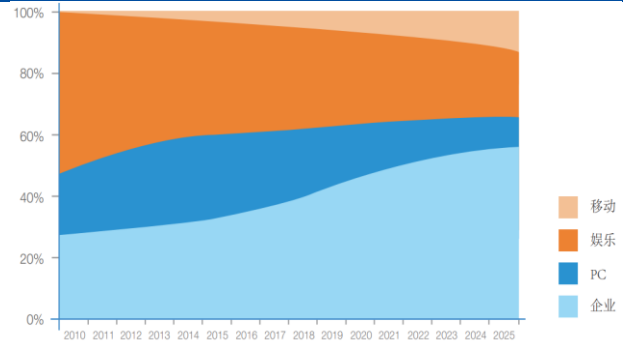
数据量呈现指数级增长，企业端占比不断提升。根据中国信通院，到 2035 年，全球数据量将达 2142ZB（ZB：Zettabyte，1ZB 约十万亿亿字节），是 2020 年所创建数据量的 45-46 倍。而由于进入云时代，数据在本地存储的需求逐步减少，企业在云端可为客户提供实时的数据和服务。

图 2. 全球数据量规模 (2016-2035)



数据来源: 中国信通院, 财通证券研究所

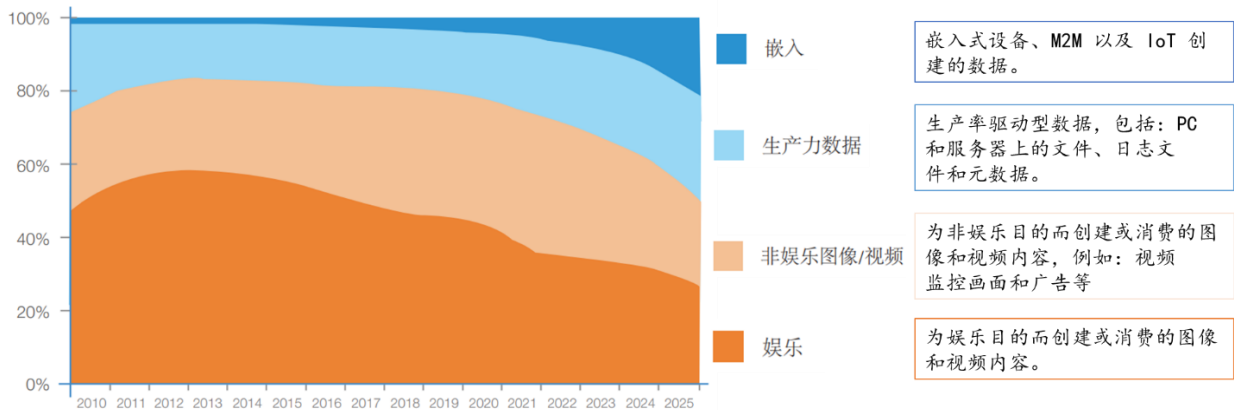
图 3. 数据的创建与存储位置占比 (2010-2025)



数据来源: IDC 《data age 2025》, 财通证券研究所  
注: IDC 该报告发布于 2017 年 3 月

数据逐渐成为经济发展的重要生产要素。基于云能够快速访问的数据运用方式, 数据日益影响企业和日常生活, 例如商业航空旅行、自动驾驶、医疗应用、控制系统和遥测技术等社会生产力紧密相关的数据占比持续提升。IDC 预计到 2025 年, 娱乐相关的消费型数据占比将从 2012 年接近 60% 下降至 30% 左右, 余下约 70% 将是非娱乐化的图像/视频、生产力数据、工业生产等嵌入式数据。

图 4. 数据类型占比 (2010-2025)



数据来源: IDC 《data age 2025》, 财通证券研究所  
注: IDC 该报告发布于 2017 年 3 月

### 1.2. 数据重要性提升, 多维数据整合挖掘深度价值, 催生算力旺盛需求

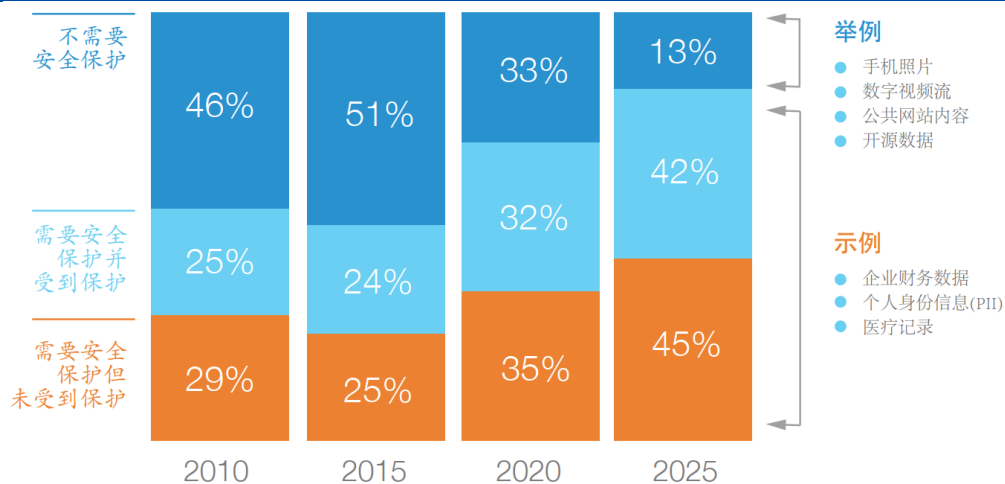
重要数据量快速增长。不同数据间的重要性相差较大, 例如: 医疗类数据重于流媒体数据, 不同数据运行出现问题所造成的影响亦不同, 例如私用 PC 宕机造成文件丢失和自动驾驶数据错误造成人员伤亡。IDC 预计未来关键数据量增速将高于数据量总体增长, 2025 年需要安全防护的企业经营/医疗记录等数据占比将达 87%。

图 5. IDC 测算 2015–2025 年数据量增速（按数据重要性划分）

| 数据类型  | 数据描述  | 2015–2025 CAGR |
|-------|---|----------------|
| 所有数据  | 包括全球数据圈中的所有数据。  | 30%            |
| 比较关键的 | 对于用户日常生活的顺利而言必不可少的数据。   | 37%            |
| 关键的   | 对于用户日常生活的持续而言必不可少的数据。   | 39%            |
| 超关键的  | 能够对用户的健康和幸福造成直接和即时影响的数据。（例如商业航空旅行、医疗应用、控制系统和遥测技术数据。这种类型的数据在元数据和嵌入式系统数据中的占比较大。 | 54%            |

数据来源：IDC，财通证券研究所

图 6. 数据按安全防护需求程度划分占比（2010–2025）



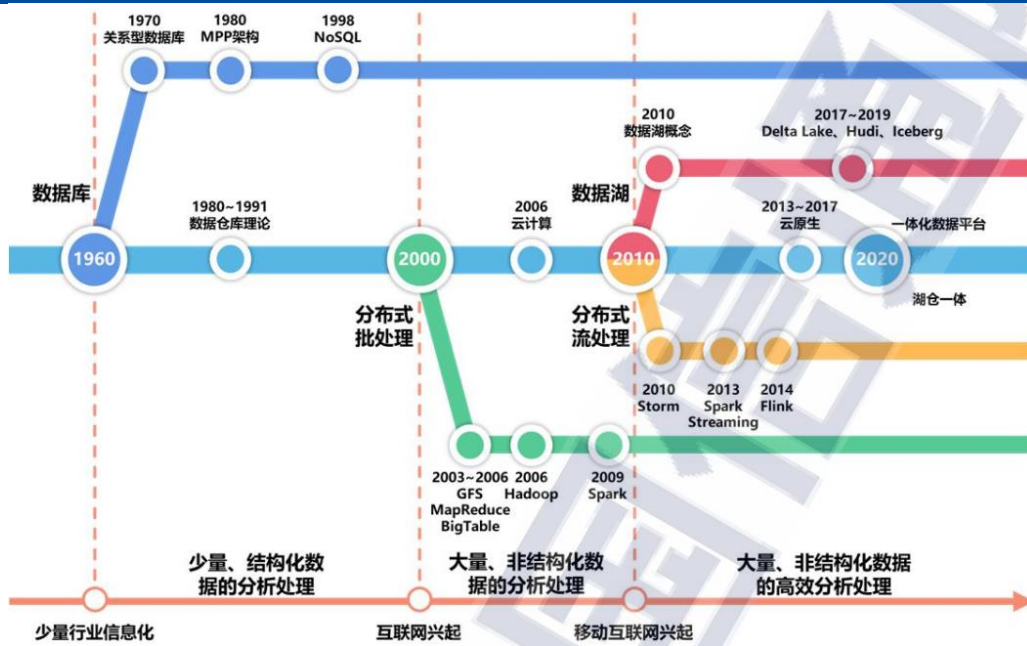
数据来源：IDC《data age 2025》，财通证券研究所

注：IDC 该报告发布于 2017 年 3 月

单个数据的价值下降，多维数据整合方可挖掘深度价值，进一步催生算力需求。当前数据指数级增长的同时，数据类型也更加多元化，在交通、工业、商业运营等领域，少量、单一化的数据的价值较低，只有将大量、多维度的数据进行综合分析才能产生应有的价值。例如交管领域对于黑车的识别，需要将车辆行驶轨迹、车辆图像识别、人像识别与对比、车牌登记数据比对等多维度数据进行分析才能准确筛选。对于庞杂的不同类型数据（尤其非结构化的）统一分析与存储的需求催生了数据湖概念，同时随着云计算技术的深入应用，带来资源集约化和应用灵活性优势的云原生概念也产生，大规模集群计算需求旺盛。



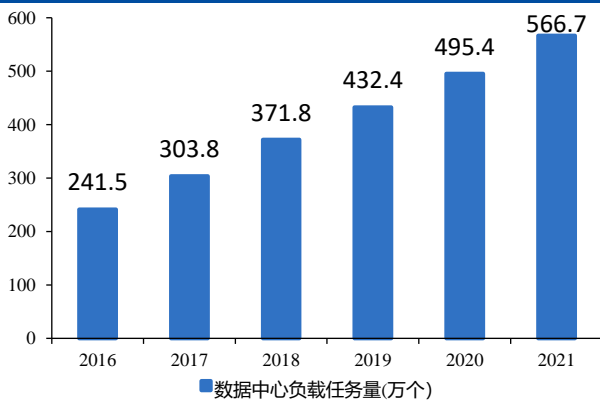
图 7. 数据平台技术演进



数据来源: IDC, 财通证券研究所

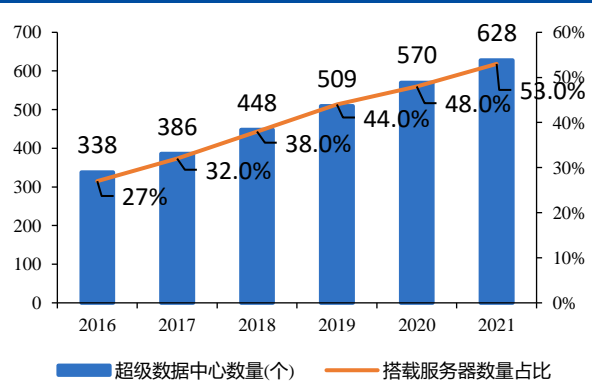
全球超级数据中心占比快速提升, 算力进一步集中。近五年全球数据中心任务量增长 135%, 大数据挖掘与应用、人工智能深度学习需求旺盛, 随着摩尔定律逐渐走向极限, 超级数据中心成为数据中心主要增量。根据 Cisco 数据, 全球超级数据中心打造服务器数量占所有数据中心搭载服务器比例由 2016 年的 27% 提升至 2021 年的 53%。

图 8. 数据中心负载任务量(2016-2021)



数据来源: Cisco Global Cloud Index、财通证券研究所

图 9. 超级数据中心数量变化(2016-2021)

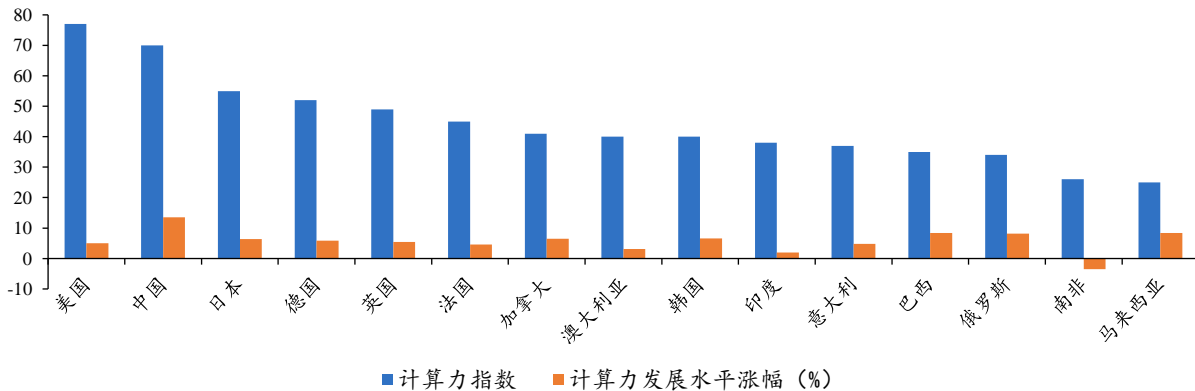


数据来源: Cisco Global Cloud Index、财通证券研究所

### 1.3. 算力正成为经济增长主要动力之一

算力已成为数字经济时代的关键生产力要素。在传统的西方经济学理论中，驱动社会经济发展的生产要素包括了劳动、资本、土地、企业家的才能等。而在一系列新兴信息数字技术的兴起与广泛应用的数字经济时代中，与计算能力的形成直接相关的算力资本投入（数据、算力、算法），如同农业时代的水利、工业时代的电力，已成为数字经济发展的核心生产力，是国民经济发展的重要基础设施。国务院印发的《“十四五”数字经济发展规划》中也明确提及到 2025 年我国数字经济核心产业增加值占 GDP 比重将从 2020 年的 7.8% 上升到 10%，数字经济将为经济社会持续健康发展提供持续动力。

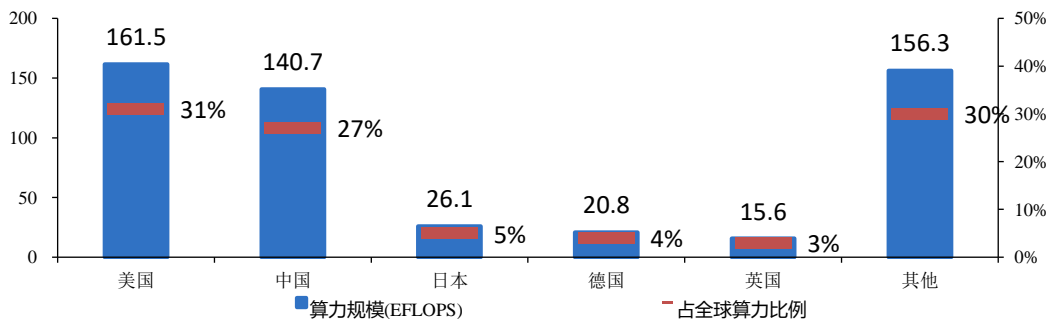
图 10. 主要国家算力指数排名及同比去年涨幅（2021A）



数据来源：IDC，财通证券研究所

中美算力规模全球领先。截至 2021 年底，全球数据中心算力总规模达 521EFLOPS (EFLOPS: 每秒进行  $10^{18}$  次浮点运算)。中美两国算力规模分别约为 161EFLOPS 和 140EFLOPS，占全球总算力份额约为 31% 和 27%。

图 11. 全球各国算力规模与占全球比例（2021）

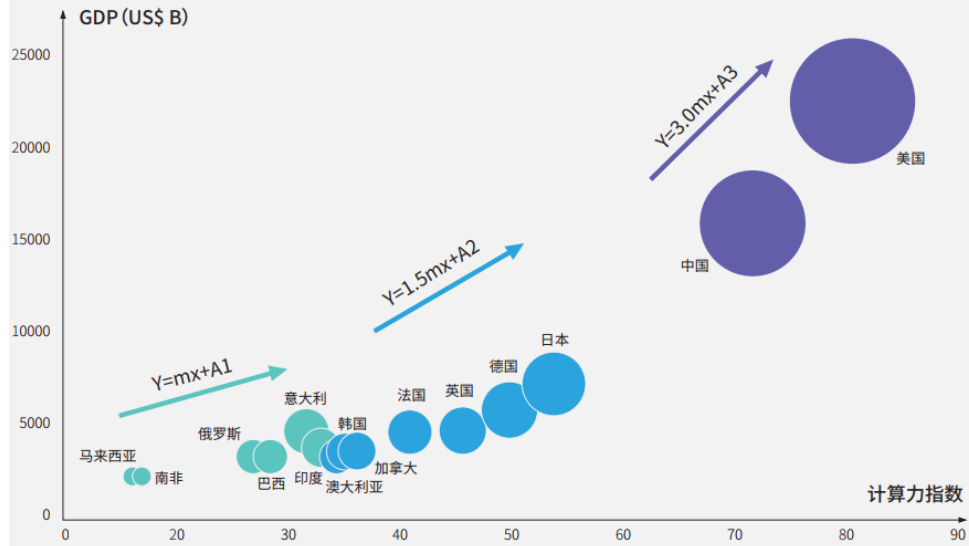


数据来源：中国信通院，财通证券研究所



计算力和 GDP 增长显著相关。根据 IDC 等机构的研究，全球平均来看，他们构建的“计算力指数”每提升 1 个点，国家数字经济和 GDP 将分别增长 3.5‰和 1.8‰。同时，“计算力指数”越高，对 GDP 的拉动作用越显著。

图 12. 不同国家的“计算力指数”与 GDP 的关系



数据来源：IDC，浪潮集团，清华全球产业研究院，财通证券研究所

图 13. 得出“计算力指数”所需的具体指标与计算口径

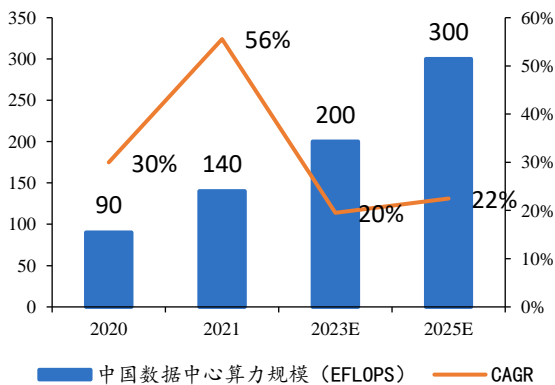
| 一级指标   | 二级指标                             | 三级指标   | 计算口径  |
|--------|----------------------------------|--|---|
| 计算能力   | 通用/科学/AI/<br>终端/边缘计算能力           | 通用/AI/边缘计算硬件支出规模<br>手机与 PC 支出规模与保有量<br>全球 Top500 超算数量及排名 | 通用/AI/边缘算力支出规模/该国当年 GDP<br>该国超级计算机在全球 TOP500 排名加权得分<br>智能手机及 PC 支出规模/该国当年 GDP 总量<br>智能手机及 PC 的保有量 |
| 计算效率   | 新技术使用率<br>云计算渗透率<br>CPU/内存/储存利用率 | 新技术使用率<br>云计算渗透率<br>服务器 CPU/内存/存储平均利用率                   | 新技术(SSD/SCM/异构)平均使用率<br>云计算支出规模/该国当年 GDP 总量<br>服务器 CPU/内存/存储设备平均利用率                               |
| 计算应用水平 | 大数据、人工智能、物<br>联网、区块链、机器人         | 相关软件、硬件、服务整体支出规模   | 支出规模/GDP 总量   |
| 基础设施支持 | 数据中心软件和服务                        | 数据中心软件和服务支出规模  | 数据中心软件和服务支出规模/该国当年 GDP  |
|        | 数据中心规模                           | 超大规模数据中心数量   | 超大规模数据中心(机架数 10000+)数量  |
|        | 数据中心效率                           | 数据中心平均 PUE   | 数据中心平均 PUE 值  |
|        | 网络基础设施                           | 网络支出规模<br>电信支出规模   | 网络支出规模/该国当年 GDP<br>电信支出规模/该国当年 GDP  |
|        | 储存基础设施                           | 存储设备支出规模<br>存储出货量(TB)                                    | 存储设备支出规模/GDP<br>存储设备出货量 TB  |

数据来源：IDC，财通证券研究所

#### 1.4. 东数西算，数字经济时代新基建启航

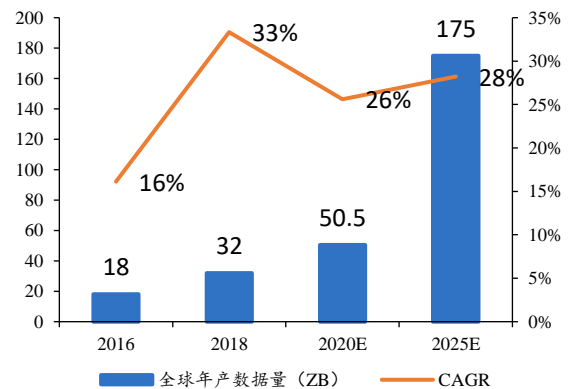
大数据时代，算力与数据增长齐头并进。根据《中国算力白皮书（2022年）》的数据，2021年我国算力总规模达到140Efllops（每秒一万四千亿亿次浮点运算，包含通用算力、智能算力、超算算力，边缘算力暂未纳入统计范围），全球占比约为27%，近五年年均增速超30%。展望未来，工信部印发的《“十四五”信息通信行业发展规划》指出2025年我国数据中心算力总规模将增长到300EFLOPS，CAGR达22%；另一方面，伴随5G、人工智能、物联网等技术的应用普及，数据流量增长速率也在不断加快。根据IDC的预测，全球数据总量在2020年将达50ZB，而这一数据到2025年有望达到175ZB，CAGR达28%。与此同时，根据中央网信办的数据统计，2019年度中国移动互联网数据接入量为1,655.50亿GB，预计2024年将达到5,680.90GB，CAGR也高达28%。因此，我们认为在当前数字经济大时代下，适度超前建设以数据中心为首的新型基建具有明确的战略意义。

图 14. 中国数据中心算力规模（2020-2025E）



数据来源：工信部，财通证券研究所

图 15. 全球每年产生数据量（2016-2025E）



数据来源：IDC，财通证券研究所

“东数西算”正式启动，预计每年带动 4000 亿投资。今年二月，国家发展改革委、中央网信办、工业和信息化部、国家能源局联合印发通知，同意在京津冀、长三角、粤港澳大湾区、成渝、内蒙古、贵州、甘肃、宁夏等 8 地启动建设国家算力枢纽节点，并规划了 10 个国家数据中心集群，标志着“东数西算”工程正式全面启动。与“西气东输”“西电东送”“南水北调”等工程相似，“东数西算”是一个国家级算力资源跨域调配战略工程，针对我国东西部算力资源分布总体呈现出“东部不足、西部过剩”的不平衡局面，引导中西部利用能源优势建设算力基础设施，“数据向西，算力向东”，服务东部沿海等算力紧缺区域，解决我国东西部算力资源供需不均衡的现状。“东数西算”工程对我国的长远发展有重要的意义：

- **有利于提升国家整体算力水平：**通过全国一体化的数据中心布局建设，扩大算力设施规模，提高算力使用效率，实现全国算力规模化集约化发展。
- **有利于促进绿色节能：**数据中心属于高耗能行业，又被称为“不冒烟的钢厂”。将数据中心从能耗指标紧张的东部地区迁往清洁能源富集、电力更加廉价的西部，将大幅提升绿色能源使用比例，降低数据中心运营的成本。
- **有利于扩大有效投资：**数据中心产业链条长、投资规模大，带动效应强。发改委预计“十四五”期间大数据中心投资还将以每年超过 20% 的速度增长，每年预计各方面投资额达 4,000 亿元。
- **有利于推动区域协调发展：**通过算力设施由东向西布局，将带动相关产业有效转移，促进东西部数据流通、价值传递，延展东部发展空间，推进西部大开发形成新格局。

图 16. “东数西算” 布局规划



数据来源：国家发改委，财通证券研究所

图 17. “东数西算” 相关政策出台历程

| 时间          | 政策                            | 相关目标  |
|-------------|-------------------------------|---|
| 2020 年 12 月 | 《关于加快构建全国一体化大数据中心协同创新体系的指导意见》 | 到 2025 年，全国范围内数据中心形成布局合理、绿色集约的基础设施一体化格局。东西部数据中心实现结构性平衡，大型、超大型数据中心运行电能利用效率降到 1.3 以下。                           |
| 2021 年 5 月  | 《全国一体化大数据中心协同创新体系算力枢纽实施方案》    | 引导超大型、大型数据中心集聚发展，构建数据中心集群，推进大规模数据的“云端”分析处理，重点支持对海量规模数据的集中处理；在城市城区内部，加快对现有数据中心的改造升级，提升效能。                      |
| 2021 年 7 月  | 《新型数据中心发展三年行动计划（2021-2023 年）》 | 到 2023 年底，全国数据中心机架规模年均增速保持在 20% 左右，平均利用率力争提升到 60% 以上，总算力超过 200EFLOPS，高性能算力占比达到 10%。                           |
| 2022 年 1 月  | 《“十四五”数字经济发展规划》               | 加快构建算力、算法、数据、应用资源协同的全国一体化大数据中心体系；加快实施“东数西算”工程，推进云网协同发展；推动智能计算中心有序发展，打造智能算力、通用算法和开发平台一体化的新型智能基础设施              |
| 2022 年 2 月  | “东数西算”工程正式启动                  | 同意在京津冀、长三角、粤港澳大湾区、成渝、内蒙古、贵州、甘肃、宁夏启动建设国家算力枢纽节点，并规划了张家口集群等 10 个国家数据中心集群。至此，全国一体化大数据中心体系完成总体布局设计，“东数西算”工程正式全面启动。 |

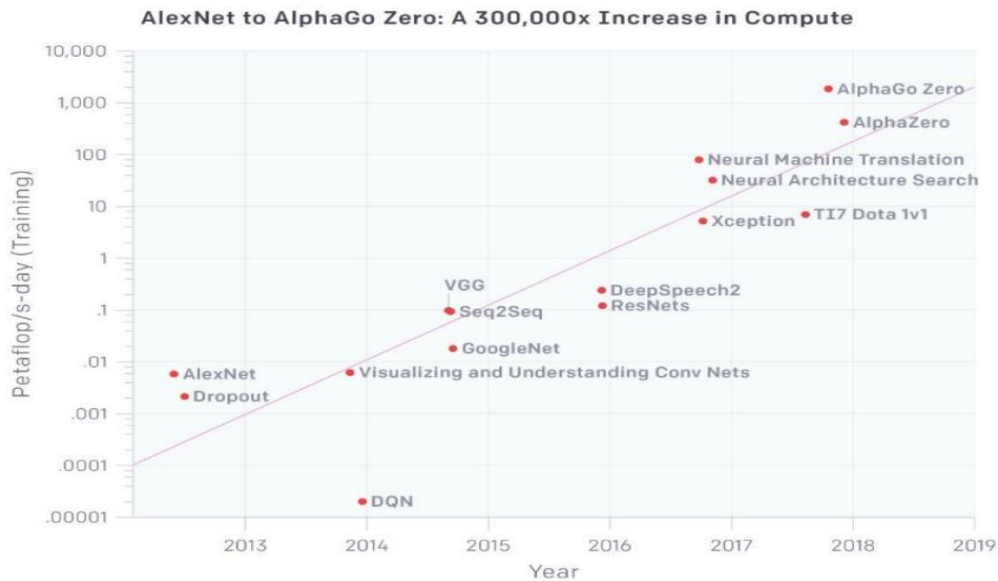
数据来源：国务院，国家发改委，财通证券研究所

## 2. 后摩尔时代，算力从 PC 走向 AI 集群

### 2.1. 后摩尔时代，传统计算架构亟待革新

AI 算力需求急剧增加，传统架构难以满足。根据摩尔定律，集成电路上可以容纳的晶体管数目大约每 18 个月增加一倍，而 AI 训练算力自 2012 年开始以平均每 3.43 个月翻倍的速度实现指数增长。在芯片制程达到个位数纳米级的当下，传统芯片算力提升速度难以赶上计算量增速。

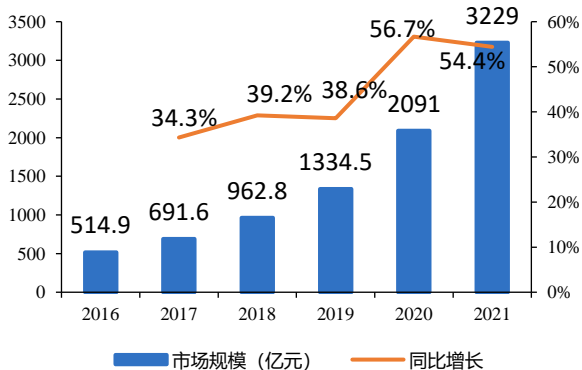
图 18. AI 算力需求剧增 30 万倍（2013-2019）



数据来源：Open AI、财通证券研究所

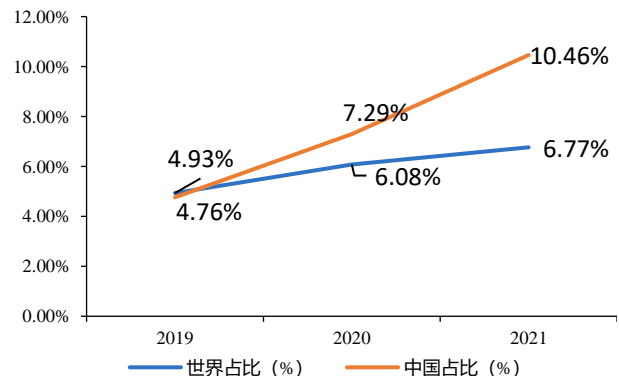
下游应用的高景气度和不断衍生的智能化需求，推动芯片革新。中国云计算市场规模增长迅速，2021 年达到 3229 亿元，近 5 年 CAGR 达 44.6%。云计算作为人工智能云端训练芯片的主要应用场景，推动了芯片架构迭代。

图 19. 中国云计算市场规模及增速（2016-2021）



数据来源：中国信通院、财通证券研究所

图 20. 云计算市场规模相对总 IT 支出占比（2019-2021）



数据来源：中国信通院、Gartner、财通证券研究所



## 2.2. AI 算力需求旺盛，GPU 是主流

GPU 是当前主流数据中心端 AI 计算架构。按技术架构分类，AI 芯片可分为图形处理器（GPU）、现场可编程门阵列（FPGA）、专用集成电路（ASIC）、类脑芯片。GPU 起初强调图形处理，随着强大的并行计算能力被发掘，逐步进入通用计算领域；FPGA 以半定制化为特征，注重于服务垂直领域；ASIC 则是针对客户应用场景需求进行定制；类脑芯片尚处于起步阶段。

图 21. 不同技术架构的 AI 芯片特征对比

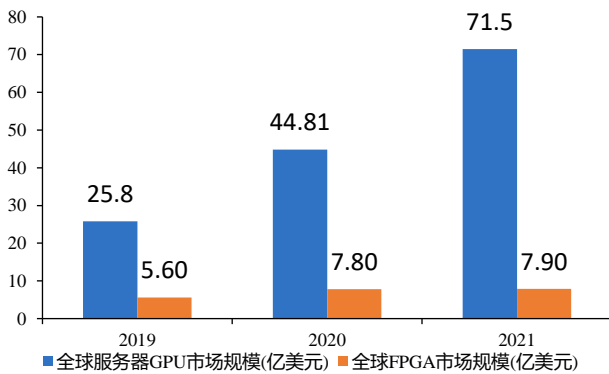
|      | GPU                                  | FPGA                                     | ASIC                           |
|------|--------------------------------------|--|--------------------------------|
| 基本架构 | 多核并行计算的基础结构；核心数多                     | 无指令、无需共享内存的体系结构；寄存器和片上内存（BRAM）属于各自的控制逻辑  | 多以 CPLD（复杂可编程逻辑器件）和 FPGA 为基础架构 |
| 主要功能 | 图像处理、密集型并行运算                         | 可根据定制编程反复改写（半定制化）                        | 可根据算法定制（全定制化）                  |
| 成熟度  | 较为成熟                                 | 正在发展                                     | 正在发展                           |
| 优点   | 比 CPU 更强的并行计算能力；比 FPGA 和 ASIC 更强的通用性 | 灵活编译芯片硬件层；计算速度快                          | 可实现全优化定制设计，适用于量产降低成本，充分利用产能    |
| 缺点   | 并行运算能力在推理端无法完全发挥                     | 复杂算法开发难度大；单个单元计算能力弱，电子管仍有冗余              | 初始成本过高；编程架构设计难度大；通用性弱          |
| 成本   | 高                                    | 较高                                       | 低                              |
| 功耗   | 大                                    | 较大                                       | 小                              |
| 试用场景 | 云端训练、云端推理                            | 云端推断、终端推理                                | 云端推理、云端推理、终端推理                 |
| 代表厂商 | AMD、NVIDIA、景嘉微                       | Altera（被 Intel 收购）、Xilinx（被 AMD 收购）、安路科技 | 谷歌、寒武纪                         |

数据来源：中商产业研究院、财通证券研究所

智能算力需求旺盛，GPU 充分受益。从市场规模来看，2021 年服务器 GPU 全球市场规模达到 71.5 亿美元，2019-2021 年复合增速 66%，占服务器整体市场规模比例快速提升；同期 FPGA 全球市场规模为 7.9 亿美元，相比 2020 年基本持平；而 ASIC 主要用于终端推理。我们认为，中期看 GPU 仍将是数据中心端 AI 训练等加速计算的主流芯片，充分受益于智能算力高景气。

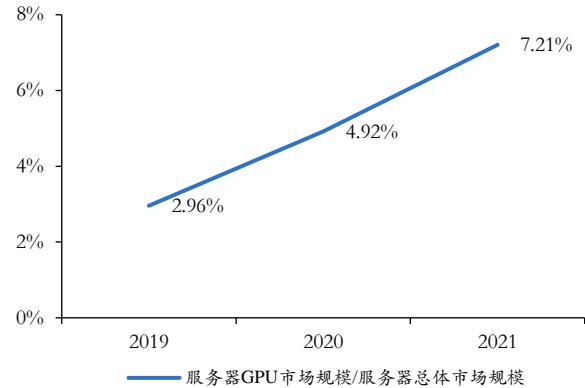


图 22. 全球服务器 GPU 与 FPGA 市场规模 (2019-2021)



数据来源: 中国信通院, 财通证券研究所

图 23. 全球服务器 GPU 占服务器中市场规模比例 (2019-2021)

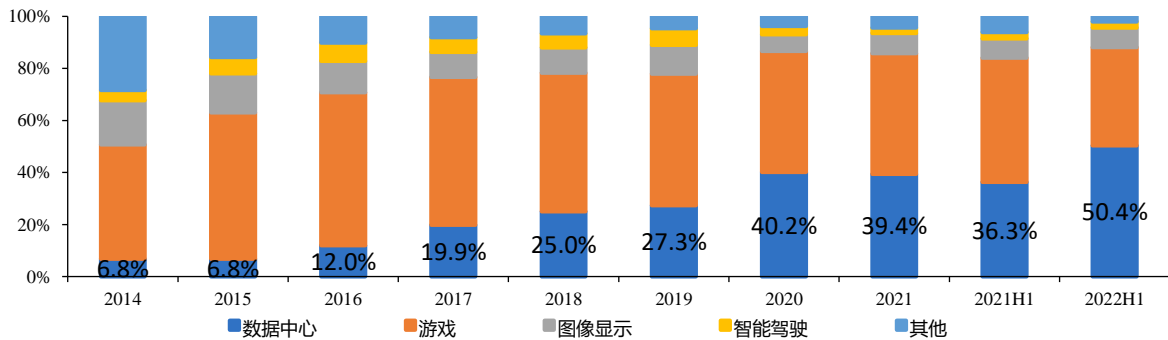


数据来源: IDC, 财通证券研究所

### 2.3. Nvidia&AMD: 市值超越 Intel, 时代算力从 PC 端走向 AI 集群

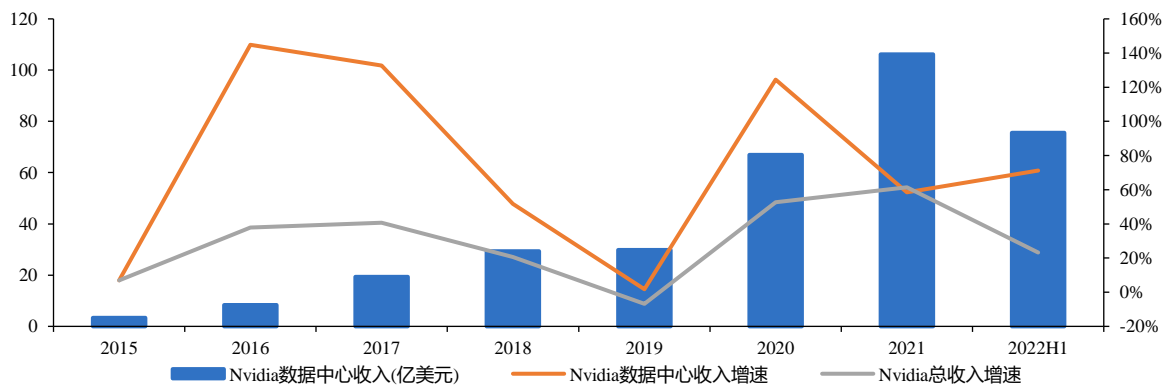
Nvidia: 2022 年上半年数据中心收入占比超过 50%。受益于智能算力高景气, Nvidia 作为全球 GPU 龙头, 数据中心业务占比历年快速提升, 由 2014 年的 6.8%, 提升至 2022 年上半年的 50.4%, 是主要收入增长动力。

图 24. Nvidia 收入分项占比 (2014-2022H1)



数据来源: Nvidia 财报, 财通证券研究所

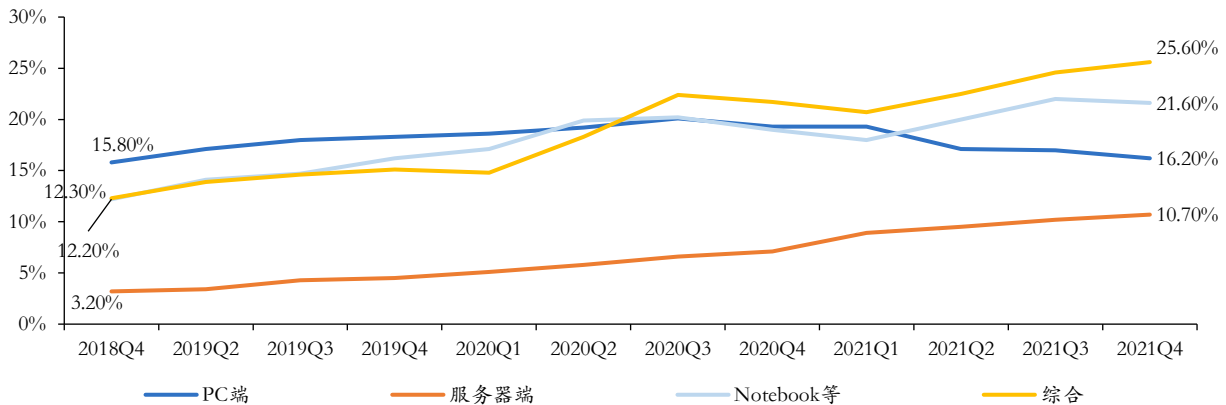
图 25. Nvidia 数据收入、增速及总收入增速 (2015-2022H1)



数据来源: Nvidia 财报, 财通证券研究所

**AMD：数据中心端发力，抢占 X86 市场。**AMD 于 2016 年发布 Zen 架构，相比 Intel 的 IDM 模式，AMD 的 fabless 模式在当前芯片制程接近极限背景下更具优势，借助全球领先 foundary 厂领先于 Intel 自身的制程优势，性能持续追赶。尤其在服务器 CPU 方面，是 AMD 抢占 intel 市场的主要发力点，在 X86 芯片的市场份额持续提升至 2021Q4 的 10.7%。

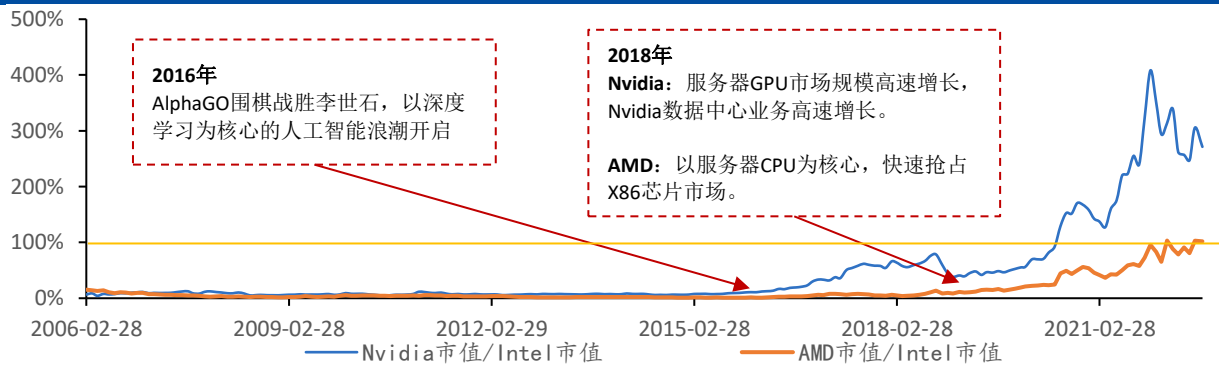
图 26. X86 芯片各细分市场 AMD 份额情况（2018Q4–2021Q4）



数据来源：tom's HARDWARE，财通证券研究所

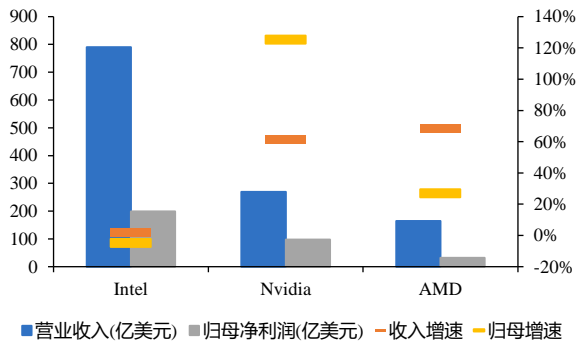
**Nvidia 和 AMD 市值双双超越 Intel，时代算力从 PC 端走向集群 AI 端。**得益于 Nvidia 在数据中心市场的强劲表现和 AMD 在服务器 CPU 市场的发力，目前它们的市值均已超越 Intel，虽然 Intel 目前的收入和利润体量明显高于 Nvidia 和 AMD，但是资本市场给予后两者估值溢价明显，标志着时代算力正从传统的 PC 端走向服务器的 AI 集群，产业浪潮不可逆转。

图 27. Nvidia、AMD 与 Intel 的市值之比（2006.2.28–2022.09.07，月频）



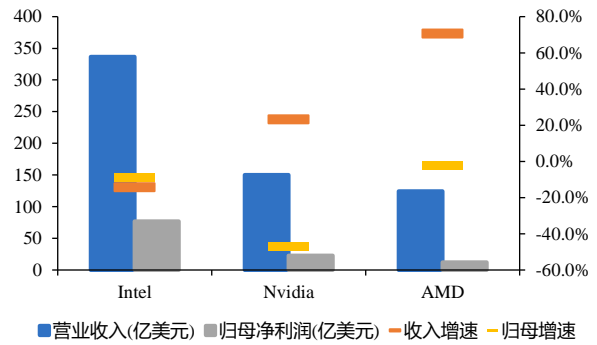
数据来源：Wind，财通证券研究所

图 28. Intel/Nvidia/AMD 业绩(2021)



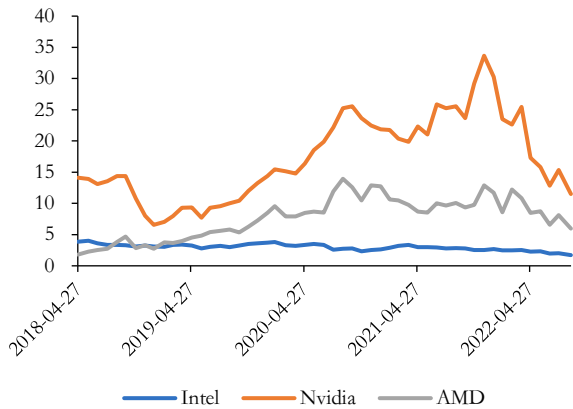
数据来源: Wind, 财通证券研究所

图 29. Intel/Nvidia/AMD 业绩(2022H1)



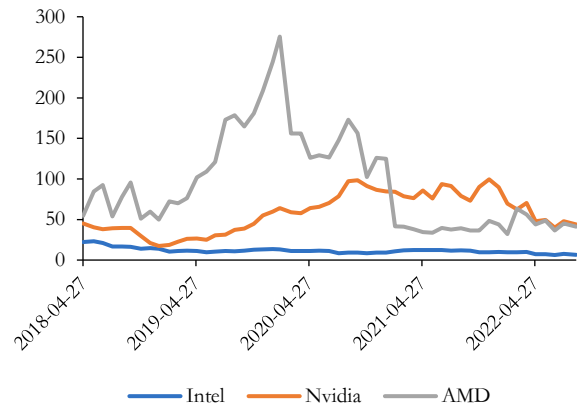
数据来源: Wind, 财通证券研究所

图 30. Intel/Nvidia/AMD 的 PS TTM (2018.04-2022.09)



数据来源: Wind, 财通证券研究所

图 31. Intel/Nvidia/AMD 的 PE TTM (2018.04-2022.09)



数据来源: Wind, 财通证券研究所

### 3. 人工智能开启算力时代

#### 3.1. AI 三要素：数据、算法、算力

AI 行业快速发展，智能算力需求提升。根据使用设备和提供算力强度的不同，算力可分为基础算力、智能算力与高端算力三大类。随着深度学习技术的快速发展，以及互联网和云计算时代海量数据和高效计算能力的支撑，计算机视觉技术、语音技术、自然语言理解技术等人工智能技术取得了突破性进展，并解锁多个行业的人工智能场景，驱动了人工智能行业相关的计算量快速增长。根据《中国算力白皮书(2022年)》的数据统计，2021 年全球智能算力总规模达 113EFLOPS，占全球总算力规模的 22%。我们认为，伴随人工智能技术的复杂性不断增加，人工智能计算能力的需求将呈指数级增长。

图 32. 算力服务分类及介绍（2021 年）

| 算力分类 | 分类标准   | 应用领域               | 全球总规模 (EFLOPS) |
|------|--|--------------------|----------------|
| 通用算力 | 主要基于 CPU 芯片的服务器提供的计算能力                         | 电子邮箱、数据存储等         | 395            |
| 智能算力 | 主要基于 GPU、FPGA、ASIC 等芯片的加速计算平台提供的人工智能训练和推理的计算能力 | 人工智能相关计算           | 113            |
| 超算算力 | 主要基于大规模集群提供高密度计算                               | 工业仿真、气象、生物信息、石油物探等 | 10             |

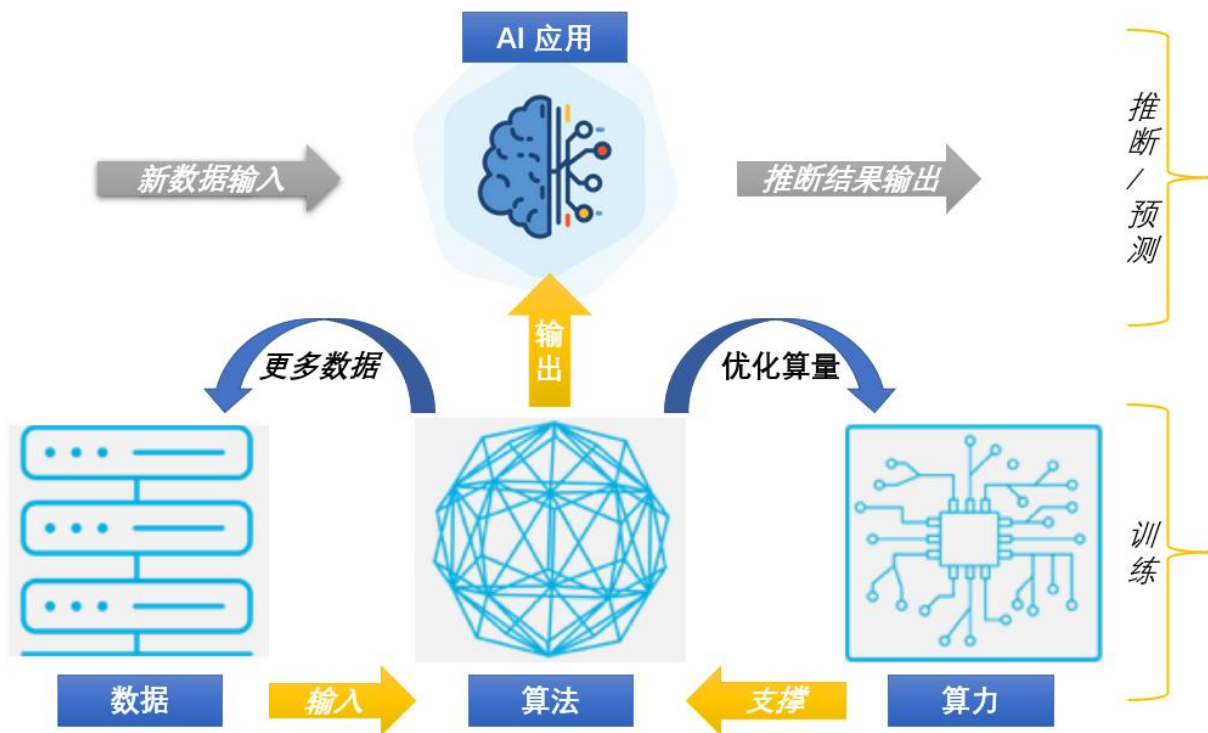
数据来源：《中国算力白皮书（2022 年）》，财通证券研究所

**AI 三要素相互耦合，共同生成 AI 模型。**一个传统的 AI 模型包括训练和推断（预测）两大部分。训练环节指将训练数据（通常为现有的历史数据）输入进算法中，通过 AI 芯片（GPU、FPGA 等）提供算力支撑，以及数据工程师的分析调参，最后生产满足特定功能的 AI 应用模型。推断环节指通过向训练完成的 AI 应用模型中输入实际应用场景中的新数据，并生产对应的推断结果。在这一过程中，数据、算法和算力扮演着同等重要的角色，三要素的耦合关系是探索 AI 未来发展道路的重要基础：

- **数据是 AI 模型的“汽油”：**数据是一切人工智能的基础。数据因其可具象性强，也是最容易被理解的竞争壁垒（特斯拉在自动驾驶数据的积累、科大讯飞在智慧教育的题库数据积累等）。未来数据的突破口在于 1）数据积累的行业下沉（智能化渗透率的提升，传感技术的升级等）；2）现有数据的打通（实现将不同行业，政府与企业间的数据互联互通）；因此，在特定行业具备数据积累先发优势和跨行业数据整合能力的公司有望形成保持领先。

- **算力是 AI 模型的“发动机”**：算力是最容易被直观量化的指标（英伟达每年推出的新 GPU 参数），但也是目前最大的瓶颈。算力的瓶颈并不体现在算力的绝对大小，而在于实现该算力的成本。特别是在算法场景众多、迭代速度较快的 AI 领域，如何设计出同时满足通用性和高算力的 AI 芯片仍是当下炙手可热的话题。因此，具备由单一芯片模式往融合异构多芯片模式发展能力的公司有望率先受益。
- **算法是 AI 模型的“大脑”**：算法是 AI 实现技术跃迁的根本，也是最难以被直观理解的部分。从 AlexNet 重新复兴神经网络到 Transformer 开启大模型时代，人工智能的每一次里程碑事件都伴随着算法层面的突破与创新。往后展望，一个好的算法除了能更好的完成设定的任务外，还需具备 1) 更强的通用性(激活更多的可用数据);2)更优化的计算原理(减少算力的负担)。因此，我们认为在 AI 领域具备科研资源和资金实力的公司将有望拔得头筹。

图 33. AI 三要素及 AI 模型训练/推断原理



数据来源：BING，财通证券研究所

### 3.2. “大模型”横空出世，人工智能进入算力时代

神经网络模型为 AI 行业注入新动能。英国数学家阿兰·图灵在 1950 年提出了“机器能思考吗”这一跨世纪的命题，人类从未停止过对智能化的思考与探索。自 1956 年达特茅斯会议上首次提出人工智能概念以来，人工智能技术与应用已经发展 60 多年，经历了多次高潮和低谷。过去的十年间，我们有幸见证了深度学习的兴起为行业的发展注入的惊人的活力：

- 2012 年 AlexNet（一种卷积神经网络模型）引入了利用 GPU 并行运算，以压倒性的准确率夺得了当年 ImageNet 图像识别大赛的冠军，带来了深度神经网络的又一次复兴。
- 2016 年 DeepMind 开发的人工智能 AlphaGo 打败人类顶尖棋手李世石，开启了人工智能发展的新纪元。
- 2017 年 Google 团队提出的 Transformer 模型横空出世，成为了日后自然语言学习（NLP）、计算机视觉（CV）的架构标准；基于 Transformer 的预训练模型 BERT 更是将 NLP 模型的精准度和泛化能力带上了新的台阶。
- 2020 年有 1750 亿参数的 GPT-3（Generative Pre-trained Transformer 3）的诞生，标志着“大模型”正在成为迈向强人工智能的重要一步。

图 34. 人工智能近十年重要发展历程（2012-2021 年）

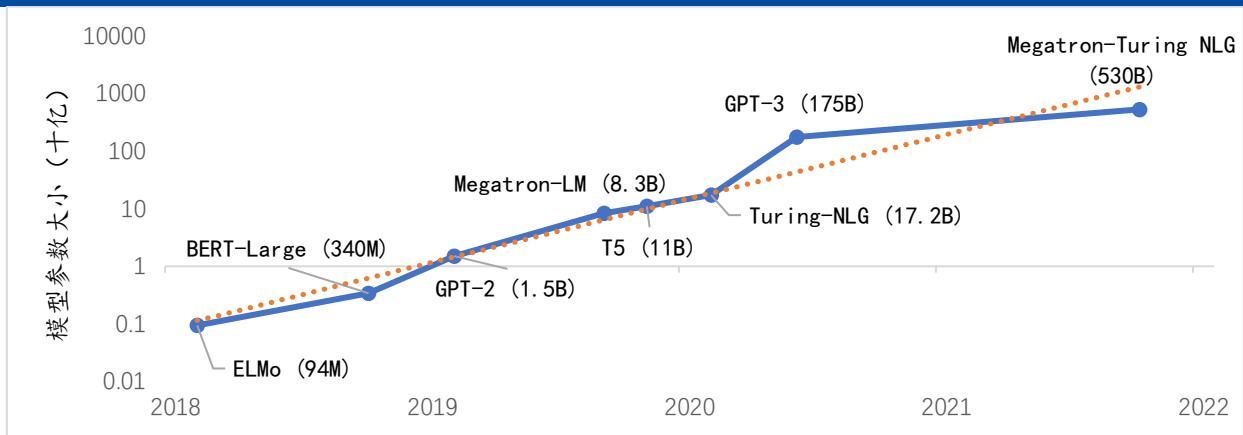


数据来源：CSDN，财通证券研究所



伴随场景与数据的激增，大模型或许成为规模化创新的基础。大模型指通过超大规模的参数设置和数据输入，通过大算力和大数据，以及强大的算力支撑与训练，而得出的基础的、具有通用性的、综合的大模型。大模型浪潮最早起源于预训练模型，Google 发布的 BERT 模型即是自然语言处理领域最为典型的预训练模型。OpenAI 则提出了 GPT 模型，尤其是 2020 年发布的 GPT-3 模型参数量即达到 1750 亿，在全球掀起大模型的浪潮。大模型可以使人工智能具备处理语言、视觉、机器人、推理、人际互动等各类相关任务的能力。因此这类模型将赋能各行各业，加快传统行业的智能化转型，在法律、医疗、教育等领域都会带来积极影响。例如 Meta 推出的拥有 150 亿个参数的 ESMfold 模型可以在原子大小的精度上预测蛋白质的三维结构，推测速度比 AlphaFold2 快出一个数量级。可以说，在 AI 模型开发领域，经历了预训练模型-大规模预训练模型-超大规模预训练模型的演进。

图 35. AI 大模型及其参数规模（2018-2022 年）

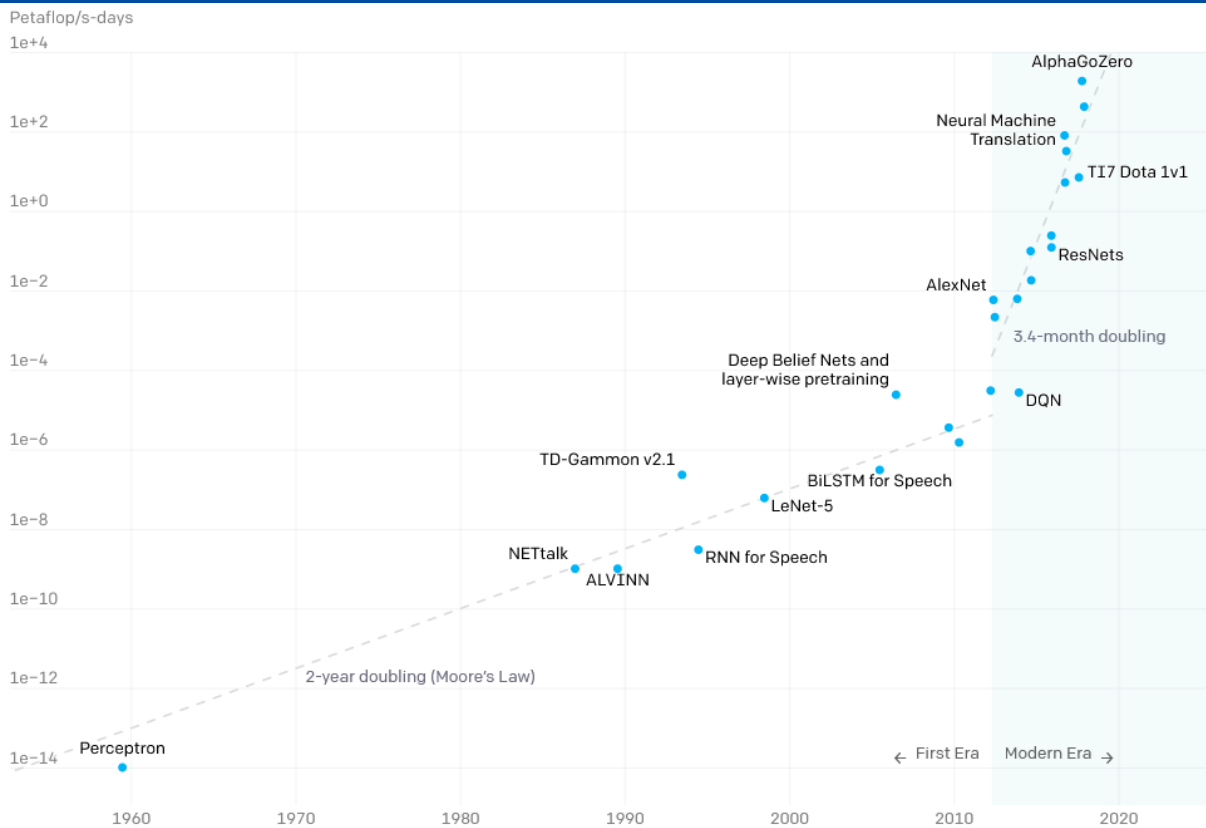


数据来源：机器之心，财通证券研究所

### 3.3. 异构计算突破算力瓶颈

大模型问世，AI 计算量需求指数增长。如第二章所述，AI 训练算力自 2012 年开始以平均每 3.43 个月翻倍的速度实现指数增长，但摩尔定律逼近极限，同期算力增长仅 7 倍，远低于 AI 模型计算量的 30 万倍。由于实现人工智能所需的深度学习算法需要很高的内在并行度、大量浮点计算能力以及强大的矩阵运算能力，其与图形渲染有着相似之处，这为使用 GPU 训练深度神经网络打下了理论基础。自 2012 年 AlexNet 面世之后，GPU 及 AI 加速器（DSA）的广泛应用使得硬件计算能力的不足得以被快速弥补，人工智能行业迎来了又一次加速发展。

图 36. 主要人工智能模型算力需求一览（1960-2020 年）



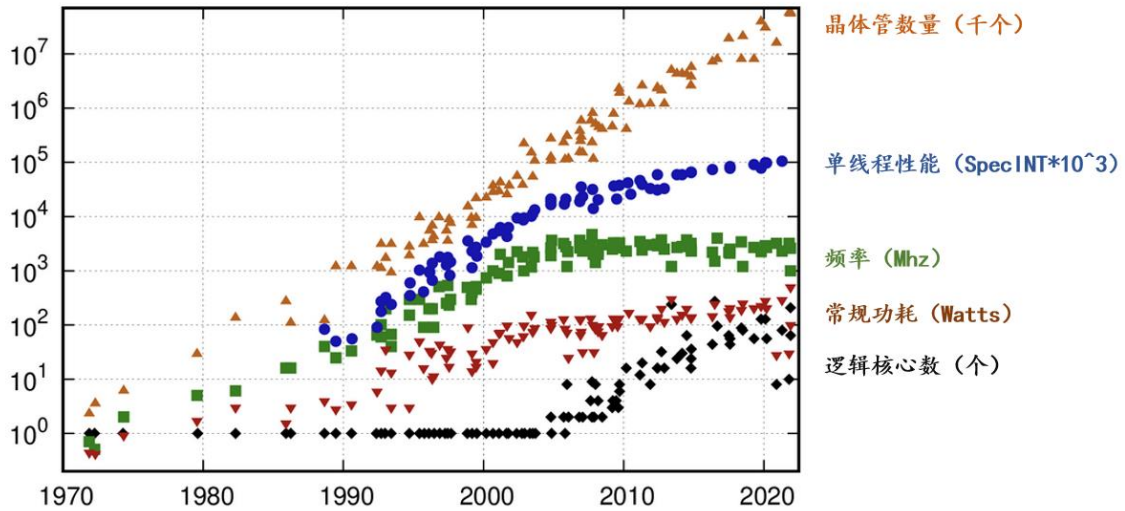
数据来源：OpenAI, 财通证券研究所

**异构计算打破 AI 芯片算力瓶颈。**大模型训练、自动驾驶等复杂应用场景对算力的追求没有止境，尽管过去 10 年晶体管的数量依然随摩尔定律增长，但单核处理器性能已经出现明显瓶颈，晶体管的数量增长主要来自于核心数量的增加，算力提升需要寻求新的突破点：

- **单个处理器的性能提升：**目前集成电路制程工艺的边际提升极其困难，考虑到硅原子的直径约为 0.1nm，在 3nm 制程之后量子隧穿效应（Quantum tunnelling effect）会越来越明显。因此，单纯通过改进工艺以提高晶体管密度，从而提升芯片性能的方法已难以满足大数据和人工智能时代下的业务规模爆炸性增长。
- **处理器的核心数量增加：**核心数的增加会提高芯片功耗、成本，并且需要软件端适配多核。尽管“东数西算”等项目能从宏观层面增加全社会的总算力，但 AI 芯片大规模落地的前提是满足 AI 场景快速迭代的通用性，否则新增的算力依然会出现利用率低、成本高。
- **处理器的灵活度提升：**随着算力需求不断攀升，单一计算类型/架构的处理器无法应对多样的数据处理任务。当前，GPGPU 作为一种异构计算方案，灵

活可编程能力与 AI 算法的需求相对匹配，是目前 AI 算力的主流技术路线。我们认为，随着 Chiplet 技术的逐步成熟，异构并行计算将向架构更加复杂的超异构计算演进，创造出更加兼顾算力和灵活性的新可能。

图 37. 过去五十年微处理机芯片趋势数据（1970-2020 年）



数据来源: GitHub, 财通证券研究所

#### 4. 投资建议

大模型、复杂应用场景推动算力从 PC 端走向 AI 集群，AI 进入算力时代，我们建议关注：

##### 国产芯片：

- **海光信息 (688041)**：脱胎于 AMD，兼具性能优越与自主可控，在 CPU 领域是国内 X86 技术架构领航员。深算一号 DCU 达到国际上同类型 GPU 高端产品水平，主要面向数据中心市场，是国产化、数字化浪潮主要受益者。
- **寒武纪 (688256)**：专注于人工智能芯片 (ASIC) 产品的研发与技术创新，云端智能芯片及加速卡也已应用到国内主流服务器厂商的产品中，2021 年 11 月发布了第三代云端 AI 芯片思元 370，并于 2022 年 4 月新款训练加速卡 MLU370-X8。

##### 国产整机：

- **中科曙光 (603019)**：国内超算领域龙头，受益行业信创潜在放量，叠加海光信息大股东身份带来的芯片分配优势，公司业绩高增长确定性强。
- **浪潮信息 (000977)**：全球第三、国内第一的服务器龙头，开创与互联网客户联合开发 JDM 模式，渠道优势显著，是国产芯片走向商业化的重要平台。

##### IDC 运营商：

- **宝信软件 (600845)**：背靠宝武集团，拿地与成本优势显著，乘“东数西算”之风，宝之云全国布局顺利开展，成长确定性强。

#### 5. 风险提示

**政策落地不及预期**：“东数西算”等国家政策的落地受到宏观经济、国际政治局势等外界因素的影响，未来存在进展不及预期的风险。

**疫情扰动风险**：新冠疫情的反复扰动可能会导致人工智能、数据中心等产业发展不及预期。

**研发进展不及预期**：人工智能是高研发投入产业，存在研发进展不及预期的风险。

## 信息披露

### 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，并注册为证券分析师，具备专业胜任能力，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。本报告清晰地反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响，作者也不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

### 资质声明

财通证券股份有限公司具备中国证券监督管理委员会许可的证券投资咨询业务资格。

### 公司评级

买入：相对同期相关证券市场代表性指数涨幅大于 10%；

增持：相对同期相关证券市场代表性指数涨幅在 5%~10%之间；

中性：相对同期相关证券市场代表性指数涨幅在-5%~5%之间；

减持：相对同期相关证券市场代表性指数涨幅小于-5%；

无评级：由于我们无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使我们无法给出明确的投资评级。

### 行业评级

看好：相对表现优于同期相关证券市场代表性指数；

中性：相对表现与同期相关证券市场代表性指数持平；

看淡：相对表现弱于同期相关证券市场代表性指数。

### 免责声明

本报告仅供财通证券股份有限公司的客户使用。本公司不会因接收人收到本报告而视其为本公司的当然客户。

本报告的信息来源于已公开的资料，本公司不保证该等信息的准确性、完整性。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他投资标的邀请或向他人作出邀请。

本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的价格、价值及投资收入可能会波动。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。

本公司通过信息隔离墙对可能存在利益冲突的业务部门或关联机构之间的信息流动进行控制。因此，客户应注意，在法律许可的情况下，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。在法律许可的情况下，本公司的员工可能担任本报告所提到的公司的董事。

本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议。在任何情况下，本公司不对任何人使用本报告中的任何内容所引致的任何损失负任何责任。

本报告仅作为客户作出投资决策和公司投资顾问为客户提供投资建议的参考。客户应当独立作出投资决策，而基于本报告作出任何投资决定或就本报告要求任何解释前应咨询所在证券机构投资顾问和服务人员的意见；

本报告的版权归本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制、发表或引用，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。