

行业研究

中国刺激政策密集落地，诺贝尔奖授予 AI 领域科学家

——光大证券电子通信行业跟踪报告之三

要点

多领域一揽子政策密集落地，持续加强经济向好势头，全面重视科技股配置机会。

10月8日上午，国家 重磅发声，系统落实一揽子增量政策；10月8日下午召开经济形势专家和企业家座谈会，听取意见建议。9月26日，中共中央政治局召开会议，分析研究当前经济形势，部署下一步经济工作。中共中央

主持会议。9月24日，国务院新闻办公室举行新闻发布会，央行和证监会计划降准、降息、降低存量房贷利率、首次创设货币政策结构性政策、创设股票回购增持再贷款、降低全国二套房贷款首付比例。

AI：2024年诺贝尔物理和化学奖授予 AI 相关领域科学家。10月8日，John J. Hopfield 和 Geoffrey E. Hinton 因通过神经网络对现代机器学习作出的奠基性贡献，获得2024年诺贝尔物理学奖。10月9日，2024年诺贝尔化学奖授予三位对蛋白质结构的设计和预测作出杰出贡献的科学家，其中两位获奖者来自谷歌旗下的 AI 公司 DeepMind。

AI：英伟达 Blackwell 将于 2025 年加速成长，光铜板供应链有望深度受益。(1) 英伟达近期将 DGX B200 的首批工程版本之一送到 OpenAI 办公室，B200 比 H100 训练速度提高 3 倍，推理速度提高 15 倍。(2) 英伟达 2024 年 8 月表示 Blackwell 系列芯片正被广泛试用，生产爬坡计划在 2024Q4 开始，并持续到 2025 年；预计 Blackwell 将在 2024Q4 实现数十亿美元的收入。(3) 10月3日，OpenAI 宣布获得 66 亿美元融资，OpenAI 的投后估值突破 1570 亿美元。

数据要素：数据要素价值加速释放。坚持政府指导和市场驱动，预计 2025 年规则建立、2030 年成熟。《关于加快公共数据资源开发利用的意见》指出，要深化数据要素配置改革，扩大资源供给；鼓励应用创新、推动数据产业健康发展。

鸿蒙：HarmonyOS NEXT 于 10 月 8 日正式开启公测。HarmonyOS NEXT 是自研操作系统，于 2023 年 8 月 4 日发布，该系统是鸿蒙抛弃 Linux 内核及安卓开放源代码项目 (AOSP) 等代码的首个大版本。除了手机、平板公测“纯血鸿蒙”，未来也将会会有 PC 端鸿蒙系统。

科技行业投资建议：自主可控打造新质生产力，持续看好科技行业未来投资机会。建议聚焦半导体、AI、数据要素、鸿蒙投资主线。

1、半导体：建议关注韦尔股份、思特威、恒玄科技、澜起科技、龙迅股份、中科蓝讯、天德钰、瑞芯微、华虹公司、北方华创、中微公司等。

2、AI：建议关注寒武纪、中际旭创、新易盛、海光信息、精达股份（铜连接）、沪电股份、胜宏科技、世运电路、紫光股份、浪潮信息、中科曙光、淳中科技、金山办公等。

3、数据要素：建议关注（1）数据流通：人民网、新华网、安妮股份等；（2）数据供给：新大陆、新国都、拉卡拉、航天信息、久远银海、思特奇、太极股份等；（3）数据安全：吉大正元、信安世纪、安恒信息等。

4、鸿蒙：建议关注润和软件、芯海科技、软通动力、法本信息、九联科技、诚迈科技、神州数码、拓维信息、烽火通信、中国软件国际(H)、海量数据、中国软件、麒麟信安、太极股份。

风险分析：半导体需求不及预期风险；宏观经济不如预期风险；行业竞争加剧风险。

电子行业 买入（维持）

行业与沪深 300 指数对比图



资料：Wind

目 录

1、多领域一揽子政策密集落地，持续加强经济向好势头	3
1.1 国新办举行新闻发布会，介绍金融支持经济高质量发展有关情况	3
1.2 中共中央政治局召开会议，分析研究当前经济形势和经济工作	3
1.3 国家 重磅发声，系统落实一揽子增量政策	4
1.4 召开经济形势专家和企业家座谈会，听取意见建议	4
2、半导体：2024 年全面复苏，看好边缘算力 SoC 等领域	5
2.1 半导体公司 24Q2 整体业绩向好	5
2.2 半导体多个细分领域 24 年营收有望实现双位数增长	6
2.3 边缘算力 SoC：AIoT 智能终端的大脑，端侧算法部署的核心	7
3、AI：诺贝尔授予 AI 领域科学家，Blackwell 将于 2025 年加速成长	9
3.1 2024 年诺贝尔物理和化学奖授予 AI 相关领域科学家	9
3.2 字节跳动豆包发布首款 AI 智能体耳机 Ola Friend	10
3.3 Blackwell 新品逐步应用，大规模部署有望加速	11
3.4 OpenAI 官宣新融资，估值突破 11000 亿	12
3.5 英伟达：FY25Q2 持续高增长，B 系列机柜将成 2025 年出货主力	13
3.6 光模块：AI 供应链最核心受益赛道	15
3.7 AI 产业链季度收入和净利润梳理	16
4、数据要素：重磅政策相继落地，数据要素价值加速释放	18
5、鸿蒙：中国自主可控的智能终端操作系统	20
5.1 鸿蒙系统：未来将沿着智能化、空间化、一体化三大主线不断迭代	20
5.2 HarmonyOS NEXT 于 10 月 8 日正式开启公测	21
5.3 鸿蒙誓师大会召开：曙光在望，待到山花烂漫时	21
6、投资建议	22
7、风险提示	22

1、多领域一揽子政策密集落地，持续加强经济向好势头

1.1 国新办举行新闻发布会，介绍金融支持经济高质量发展有关情况

国务院新闻办公室 2024 年 9 月 24 日（星期二）上午 9 时举行新闻发布会，中国人民银行行长潘功胜、国家金融监督管理总局局长李云泽、中国证券监督管理委员会主席吴清介绍金融支持经济高质量发展有关情况。

中国人民银行行长潘功胜指出，根据中央的决策部署，为进一步支持经济的稳定增长，人民银行将坚定支持性的货币政策立场，加大货币政策调控的强度，提高货币政策调控的精准性，为经济稳定增长和高质量发展创造良好的货币金融环境。

人民银行行长宣布，将于近期在人民银行网站上逐项发布政策文件或公告，内容包括：

(1) 降低存款准备金率和政策利率。近期将下调存款准备金率 0.5 个百分点，向金融市场提供长期流动性约 1 万亿元；在今年年内还将视市场流动性的状况，可能择机进一步下调存款准备金率 0.25-0.5 个百分点。降低中央银行的政策利率，即 7 天期逆回购操作利率下调 0.2 个百分点，从目前的 1.7% 调降至 1.5%，同时引导贷款市场报价利率和存款利率同步下行，保持商业银行净息差的稳定。

(2) 降低存量房贷利率和统一房贷最低首付比例。引导商业银行将存量房贷利率降至新发放贷款利率的附近，预计平均降幅大约在 0.5 个百分点左右。统一首套房和二套房的房贷最低首付比例，将全国层面的二套房贷款最低首付比例由当前的 25% 下调到 15%。将 5 月份人民银行创设的 3000 亿元保障性住房再贷款，中央银行资金的支持比例由原来的 60% 提高到 100%，增强对银行和收购主体的市场化激励。将年底前到期的经营性物业贷款和“金融 16 条”这两项政策文件延期到 2026 年底。原来这两份文件是今年年底到期，人民银行和金融监管总局一起，把这两份文件延长到 2026 年底。

(3) 创设新的货币政策工具支持股票市场稳定发展。第一项是创设证券、基金、保险公司互换便利，支持符合条件的证券、基金、保险公司通过资产质押，从中央银行获取流动性，这项政策将大幅提升机构的资金获取能力和股票增持能力。第二项是创设股票回购增持专项再贷款，引导银行向上市公司和主要股东提供贷款，支持回购和增持股票。

1.2 中共中央政治局召开会议，分析研究当前经济形势和经济工作

中共中央政治局 9 月 26 日召开会议，分析研究当前经济形势，部署下一步经济工作。中共中央 主持会议。

会议强调，（1）要加大财政货币政策逆周期调节力度，保证必要的财政支出，切实做好基层“三保”工作。（2）要发行使用好超长期特别国债和地方政府专项债，更好发挥政府投资带动作用。（3）要降低存款准备金率，实施有力度的降息。（4）要促进房地产市场止跌回稳，对商品房建设要严控增量、优化存量、提高质量，加大“白名单”项目贷款投放力度，支持盘活存量闲置土地。（5）

要回应群众关切，调整住房限购政策，降低存量房贷利率，抓紧完善土地、财税、金融等政策，推动构建房地产发展新模式。（6）要努力提振资本市场，大力引导中长期资金入市，打通社保、保险、理财等资金入市堵点。（7）要支持上市公司并购重组，稳步推进公募基金改革，研究出台保护中小投资者的政策措施。

1.3 国家 重磅发声，系统落实一揽子增量政策

国务院新闻办公室 10 月 8 日（星期二）上午 10 时举行新闻发布会，国家

主任郑栅洁和副主任刘苏社、赵辰昕、李春临、郑备介绍“系统落实一揽子增量政策 扎实推动经济向上结构向优、发展态势持续向好”有关情况，并答记者问。

针对当前经济运行中的新情况新问题，党中央、国务院科学决策、果断出手，在有效落实存量政策的同时，围绕加大宏观政策逆周期调节、扩大国内有效需求、加大助企帮扶力度、推动房地产市场止跌回稳、提振资本市场等五个方面，加力推出一揽子增量政策，推动经济持续回升向好。

一揽子增量政策体现了三个“更加注重”，就是更加注重提高经济发展质量，更加注重支持实体经济和经营主体健康发展，更加注重统筹高质量发展和高水平安全。这些政策的谋划制定，也体现了四个“坚持”。（1）坚持目标导向，锚定全年目标任务，用好用足政策空间，努力激发更大发展潜能。（2）坚持问题导向，聚焦经营主体和社会关切，紧扣经济社会发展和企业生产经营中的困难问题，采取针对性更强的措施。（3）坚持系统施策，统筹财政、金融、消费、投资、房地产、股市、就业、民生等各领域政策，增强宏观政策取向一致性，强化政策合力。（4）坚持长短结合，统筹做好今明两年政策的衔接，推动经济持续平稳健康发展，确保“十四五”顺利收官。

1.4 召开经济形势专家和企业家座谈会，听取意见建议

中共中央政治局 、国务院总理 10 月 8 日下午主持召开经济形势专家和企业家座谈会，听取对当前经济形势和下一步经济工作的意见建议。专家和企业家认为，今年以来，面对复杂严峻的外部环境，我国经济运行总体平稳、稳中有进。近期出台的一揽子增量政策力度大、针对性强，市场预期明显改善。大家还就推动经济持续回升向好提出意见建议。

强调，做好下一阶段经济工作十分重要，各方面要增强责任感紧迫感，深入贯彻党中央决策部署，努力完成全年经济社会发展目标任务。抓紧落实好稳经济一揽子增量政策，对已出台的政策加快实施、确保早落地早见效，对正在研究的政策尽快拿出具体方案。跟踪评估政策传导效果，该优化的优化、该加力的加力。同时，结合谋划明年经济工作，研究储备一批稳经济、促发展的政策举措，并根据经济形势变化适时推出。稳经济的关键是稳企业。要切实做好助企纾困工作，指导各地用好各项惠企政策，坚决治理乱罚款、乱检查、乱查封等问题，不断优化营商环境，激发企业活力。要按照党的二十届三中全会部署，深入落实各项改革任务，努力以改革的办法破解发展难题。

2、半导体：2024 年全面复苏，看好边缘算力 SoC 等领域

2.1 半导体公司 24Q2 整体业绩向好

2024 年二季度半导体行业上市公司整体业绩向好。24Q2 营收增速前五名的子行业分别是存储、CIS、第三代、后道设备、MEMS。24Q2 归母净利润增速前五名的子行业分别是 CIS、第三代、数字、封测、后道设备。

表 1：半导体细分领域 2024Q2 营业收入增速排名（单位：亿元，%）

子行业	营业收入（亿元）					同比增速（%）				
	23Q2	23Q3	23Q4	24Q1	24Q2	23Q2	23Q3	23Q4	24Q1	24Q2
存储	104.46	122.31	141.63	145.25	190.85	-26%	22%	41%	93%	83%
CIS	62.40	82.16	84.76	77.71	95.68	-19%	32%	23%	38%	53%
第三代	17.84	29.55	23.60	21.62	27.05	-3%	56%	-18%	51%	52%
后道设备	12.20	9.24	12.48	10.11	18.17	-22%	-23%	-23%	17%	49%
MEMS	4.37	7.95	7.32	4.77	6.46	7%	109%	60%	54%	48%
前道设备和零部件	115.80	141.76	172.81	127.10	158.02	40%	34%	28%	33%	36%
数字-GPU 和 CPU	19.59	15.73	29.88	18.46	25.51	-3%	-8%	22%	30%	30%
模拟	53.10	60.54	66.93	57.55	68.44	-11%	9%	21%	29%	29%
封测	202.10	226.04	244.30	205.07	245.79	-6%	-6%	3%	13%	22%
数字	167.25	169.96	193.40	174.33	202.56	-10%	0%	7%	17%	21%
功率半导体	101.89	102.22	104.61	95.07	116.33	6%	10%	15%	7%	14%
射频	16.54	22.78	28.29	17.56	18.54	13%	75%	123%	53%	12%
材料	111.78	114.88	114.12	113.75	125.00	-12%	-13%	-8%	6%	12%
EDA	4.39	4.33	6.99	3.39	4.72	44%	15%	16%	38%	8%
代工	40.21	43.12	47.13	41.60	42.85	-17%	-2%	28%	41%	7%
红外	17.52	15.28	18.53	15.81	18.19	40%	26%	-1%	21%	4%

资料来源：Wind、光大证券研究所整理；注：各细分板块内组分公司参照光大证券电子行业重点公司分类。

表 2：半导体细分领域 2024Q2 归母净利润增速排名（单位：亿元，%）

子行业	归母净利润（亿元）					同比增速（%）				
	23Q2	23Q3	23Q4	24Q1	24Q2	23Q2	23Q3	23Q4	24Q1	24Q2
CIS	0.20	2.89	2.65	6.02	9.92	-99%	扭亏	扭亏	1269%	4870%
第三代	0.14	1.82	-2.05	2.22	2.06	-75%	935%	-488%	扭亏	1360%
数字	9.15	2.05	-56.47	7.28	20.06	-63%	-80%	-22279%	扭亏	119%
封测	7.47	9.43	13.09	5.08	13.22	-58%	-41%	29%	251%	77%
后道设备	2.28	0.03	1.46	-0.01	3.84	-51%	-99%	-58%	-104%	68%
功率半导体	8.88	8.20	14.70	5.86	9.18	-57%	-56%	-12%	-53%	3%
前道设备和零部件	28.87	22.50	30.52	18.77	28.23	88%	3%	25%	18%	-2%
材料	13.30	9.78	4.74	7.42	11.06	-20%	-28%	-57%	-31%	-17%

数字-GPU 和 CPU	1.80	-1.19	2.42	-0.24	1.44	85%	N/A	扭亏	N/A	-20%
EDA	0.83	0.86	0.79	-0.52	0.51	46%	-27%	-55%	-309%	-38%
射频	1.53	3.75	3.30	1.11	0.71	-17%	129%	529%	扭亏	-54%
红外	2.33	1.29	-2.66	0.67	-0.11	77%	-19%	N/A	-15%	-105%
MEMS	-0.39	0.61	1.42	-0.33	0.05	-224%	184%	扭亏	N/A	扭亏
代工	-1.86	-0.57	-2.22	-1.50	-1.91	-115%	-110%	N/A	N/A	N/A
存储	-1.48	-2.63	0.34	12.39	13.68	-108%	-129%	扭亏	扭亏	扭亏
模拟	-1.92	-2.59	3.36	-1.51	1.59	-116%	-141%	107%	-230%	扭亏

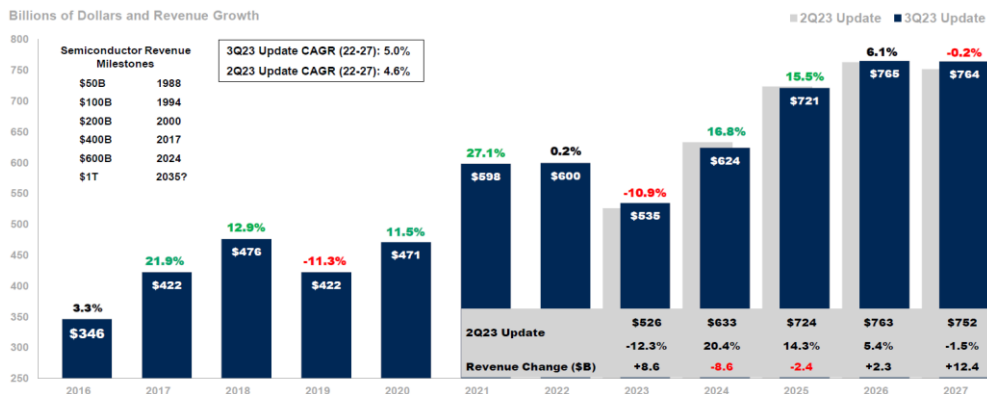
资料：Wind、光大证券研究所整理；注：（1）各细分板块内组分公司参照光大证券电子行业重点公司分类。（2）正负数据之间的净利润同比增速用“扭亏”表示，并不参与排名

2.2 半导体多个细分领域 24 年营收有望实现双位数增长

根据 Gartner 数据，2023 年全球半导体销售收入为 5350 亿美元，同比下降 10.9%；2024 年全球半导体销售收入有望迎来复苏，Gartner 预计 2024 年全球半导体收入为 6240 亿美元，同比增长 16.8%。

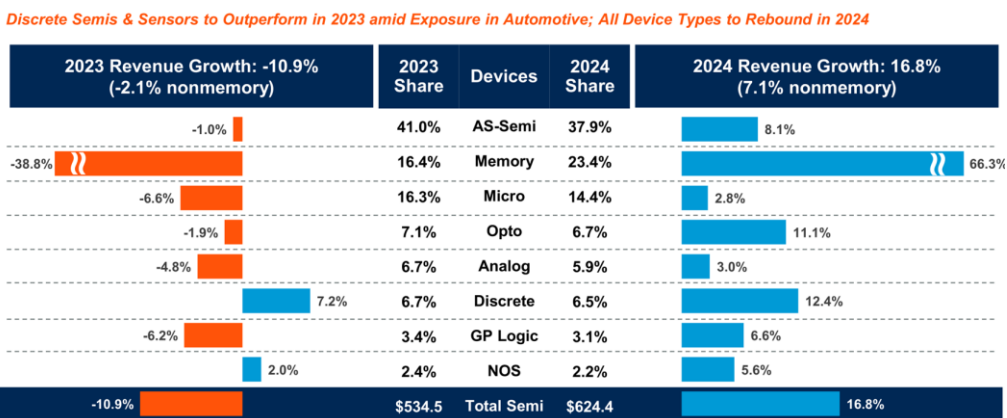
多个细分领域 2024 年营收有望实现双位数增长。根据 Gartner 数据，在细分领域中，2024 年存储行业收入增速第一，同比增长为 66.3%，非存储行业收入同比增速为 7.1%。

图 1：全球半导体收入预测



资料：Gartner 2023Q3 预测 注：2023—2027 年为 Gartner 2023Q3 预测

图 2：2023 和 2024 年半导体行业细分领域市场份额和同比增速（单位：%）

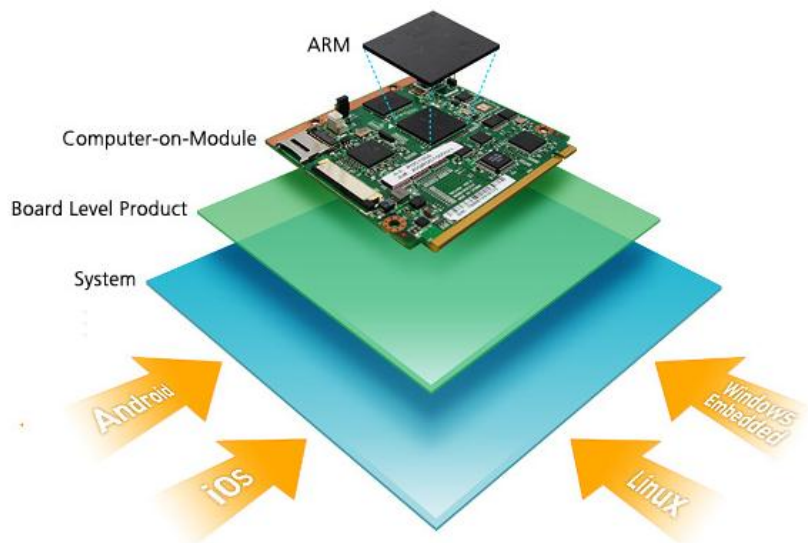


资料：Gartner 《Gartner 2023 年 10 月 12 日：Semiconductor & Electronics Forecast Update 3Q23》，光大证券研究所。备注：2023-2024 年为 Gartner 预测数据

2.3 边缘算力 SoC：AIoT 智能终端的大脑，端侧算法部署的核心

芯片系统（System-on-Chip, SoC）是一种集成电路，将一个系统所需的所有组件压缩到一块硅片上。芯片内部设置高速总线负责各个处理器和外部接口的数据传输。配备闪存接口、动态存储器接口、显示接口、网络接口以及各种高速、低速外部设备接口。SoC 的关键技术主要包括总线架构技术、IP 核可复用技术、软硬件协同设计技术、SoC 验证技术、可测性设计技术、低功耗设计技术、超深亚微米电路实现技术，以及嵌入式软件移植、开发研究，是一门跨学科的研究领域。SoC 意味着在单个芯片上实现以前需要多个芯片才能实现的系统功能，克服了多芯片板级集成出现的设计复杂、可靠性差、性能低等问题，并且在减小尺寸、降低成本、降低功耗、易于开发等方面也有突出优势。SoC 对研发设计、制造工艺以及软硬件协同开发技术的要求较高，主要体现在芯片验证和测试难度的提高，以及 IP 复用、混合电路设计的困难加大。任何 SoC 的设计都是性能、功耗、稳定性、工艺难度几方面的平衡。完整的 SoC 系统解决方案，除了提供硬件参考设计外，还需要提供系统级的软件参考设计，包括驱动软件、大型 OS（Linux、Android、ChromeOS、国产 OS 等）的移植、针对性的算法、中间件和上层应用软件的适配等，并通过严格的可靠性、兼容性测试。

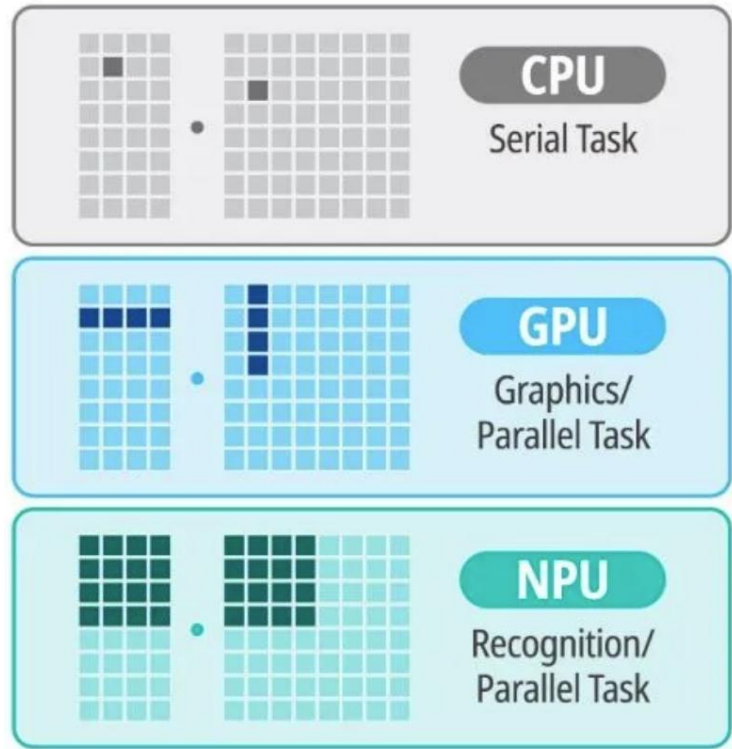
图 3：SoC 示意图



资料：华强电子网

AI 将提升社会劳动生产率，特别是在有效降低劳动成本、优化产品和服务、创造新市场和就业等方面为人类的生产和生活会带来革命性的转变。AI 技术必须具备三个要素：算法、数据、算力。CPU 是通用处理器，设计用于执行广泛的计算任务。它具有强大的灵活性和可编程性，但可能在特定任务（如 AI 计算）上效率不高。GPU 最初设计用于处理图形和视频渲染，它擅长处理并行计算任务，因此在 AI 领域也得到了广泛应用。然而，GPU 并非专门为 AI 计算设计，它在处理某些类型的 AI 任务时可能不如 NPU 高效。CPU 是线性、串行任务（指令）执行，效率较低，通用性较高；GPU 是并行处理和专用图形并行处理，效率更高；而 NPU 则是“并行认知处理”，在 AI 机器学习方面，效率更高。

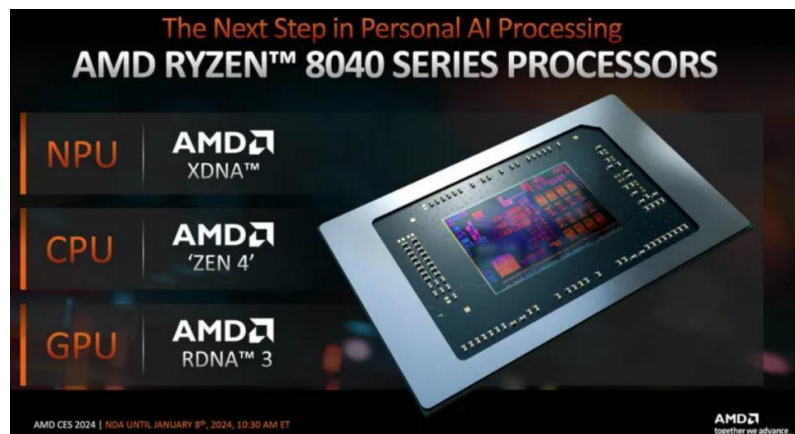
图 4: CPU、GPU 和 NPU 的对比



资料：公众号【电脑报】《你知道 CPU、GPU，那 NPU 是什么？看完这篇秒懂》

英特尔（Intel）、AMD 和高通（Qualcomm）等主要厂商均将 NPU 集成到其最新处理器中。

图 5: AMD RYZEN 8040



资料：AMD CES 2024

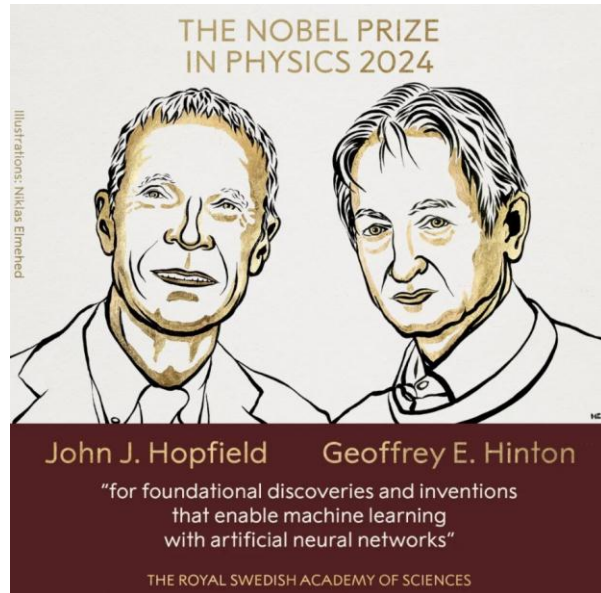
AI 新浪潮未来将提升海量 IoT 设备的边缘算力需求。AIoT 时代拥有海量 IoT 终端。“智能”将是物联网时代最核心的生产力，AI 技术将渗透到云、边、端和应用的各个层面，与 IoT 设备进行深度融合。在训练方面，AI 模型的训练需要海量数据支持，SoC 中的 NPU 提取视觉和语音的特征数据，为云端 AI 提供大数据支撑。在推理层面，各类带有 NPU 的边缘侧芯片 SoC 将提供丰富的 AI 算力，经过压缩的轻量级 AI 模型可以在音箱、摄像头等边缘侧部署。

3、AI：诺贝尔授予 AI 领域科学家，Blackwell 将于 2025 年加速成长

3.1 2024 年诺贝尔物理和化学奖授予 AI 相关领域科学家

2024 年诺贝尔物理学奖授予约翰·J·霍普菲尔德(John J. Hopfield)和杰弗里·E·辛顿(Geoffrey E. Hinton)，表彰他们在使用神经网络进行机器学习的基础性发现和发明。辛顿以霍普菲尔德的“霍普菲尔德网络”为基础，发明了玻尔兹曼机 (the Boltzmann machine) —— 通过输入机器运行时很可能出现的示例来训练机器。玻尔兹曼机可用于对图像进行分类，或创建训练模式类型的新示例。辛顿在此基础上继续发展，帮助开启了机器学习的爆炸式发展。他在神经网络方面的开创性研究为 ChatGPT 等 AI 系统铺平了道路。辛顿是众所周知的“AI 教父”，OpenAI 联合创始人、前首席科学家伊利亚·苏茨克维 (Ilya Sutskever) 是辛顿的学生。

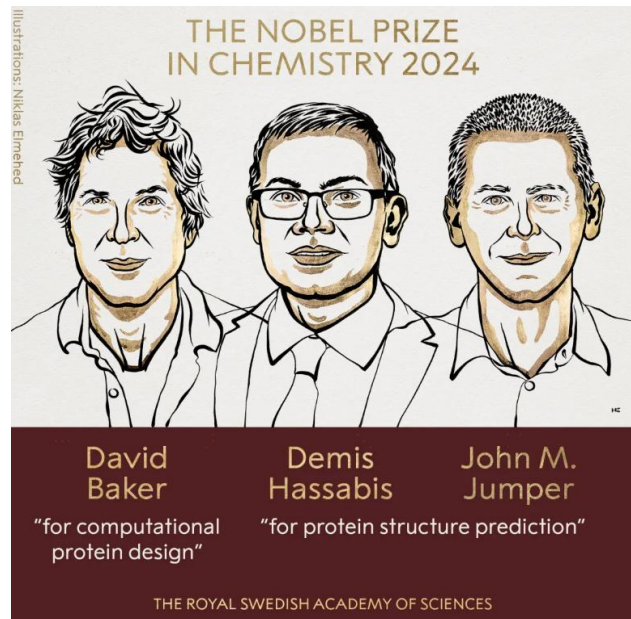
图 6：2024 年诺贝尔物理学奖获得者



资料：瑞典皇家科学院诺贝尔奖委员会

2024 年诺贝尔化学奖授予大卫·贝克(David Baker)、戴米斯·哈萨比斯(Demis Hassabis)和约翰·江珀(John M. Jumper)，以表彰他们在蛋白质设计和蛋白质结构预测领域作出的贡献。诺贝尔奖委员会表示，哈萨比斯和江珀创建了一个 AI 模型“AlphaFold”，以预测研究人员已经确定的几乎所有 2 亿种蛋白质的结构。用 AI 预测蛋白质结构已成为生物领域的重要方向。该领域的开源算法包括 DeepMind 的 AlphaFold2 以及任职华盛顿大学大卫·贝克团队研发的 RoseTTAFold。2021 年 7 月，AlphaFold 的论文首次发表在《nature》杂志，缔造者之一约翰·江珀被评为当年度《nature》十大科学人物之一。2024 年 5 月份，DeepMind 推出 AlphaFold 3 模型，其能够以更高的准确性预测不同大分子之间复合物的结构，及大分子、小分子和离子之间的相互作用，以帮助科学家更精确地针对疾病机制，从而开发出更有效的治疗药物。

图 7：2024 年诺贝尔化学奖获得者



资料：瑞典皇家科学院诺贝尔奖委员会

3.2 字节跳动豆包发布首款 AI 智能体耳机 Ola Friend

2024 年 10 月 10 日，字节跳动豆包发布了首款 AI 智能体耳机 Ola Friend。该款耳机接入了豆包大模型，并与豆包 App 深度结合。用户戴上耳机后，无须打开手机，便能通过语音唤起豆包进行对话。在形态和重量方面，该款耳机为开放式耳机，单只仅重 6.6g，是目前同类中最轻薄的耳机，其主体部分最宽 10mm，最薄 7mm。

图 8：字节跳动豆包首款 AI 智能体耳机 Ola Friend



资料：雷科技

据字节方面介绍，为了做好语音识别，该款耳机接入了字节豆包大模型的 Seed-ASR（语音识别）技术模型。该模型可以高精度识别中英文、口音，甚至能通过上下文识别各类消息。

在该款耳机的宣传片中，豆包特别强调了 Ola Friend 能够在信息查询、旅游出行、英语学习及情感交流等场景为用户提供帮助。

天眼查信息显示，字节跳动于 2024 年 4 月收购了开放式耳机品牌 Oladance（主体公司：深圳大十未来科技有限公司），目前已实现对其 100% 控股。Oladance 是一个创新驱动的消费电子品牌，由前 Bose 高管李浩乾联合多位 Bose 工程师于 2019 年创立，产品以开放式耳机为主。

对于 Ola Friend 的 AI 功能，豆包相关负责人表示，豆包的各种能力会在后续持续迭代，为用户在生活中各个场景提供帮助。

在全球当下最受关注的 AI 领域里，不乏有科技巨头和字节作出了同样的尝试，让大模型技术在硬件设备上落地，选择的产品形态多以眼镜、耳机和头显为主。例如，Meta 在 2023 年末发布内置 Llama 大模型的智能眼镜 Ray-Ban Meta；苹果、Midjourney 等公司正在研究头显和 AI 的结合。

3.3 Blackwell 新品逐步应用，大规模部署有望加速

Microsoft Azure 率先运行配备英伟达 Blackwell 系统 GB200 的 AI 服务器，并展示了其下一代 NVIDIA 训练和推理平台。该机架需要巨型冷却解决方案，冷却剂分配单元 (CDU) 所占用的通道空间是 GB200 计算机架的两倍。Microsoft Azure 所展示的服务器中，左侧为 GB200 机架，右侧宽度是左侧的两倍，包含液体-空气热交换器，以及泵、电源和监控装置。

在 FP8/INT8 性能方面，Nvidia 的 B200 GPU 的性能比 H100 处理器高出 2.5 倍 (4,500 TFLOPS/TOPS vs 1,980 TOPS)。借助 FP4 数据格式，Nvidia 的 B200 提供了高达 9 PFLOPS 的性能，可以训练极其复杂的 LLM，从而为 AI 提供新的模型。

预计微软将在年度 Ignite 会议上分享有关其基于 Blackwell 机器人和人工智能项目的更多细节。Blackwell 服务器预计将在 2024 年年底或 2025 年初大规模部署安装。

图 9：Microsoft Azure NVIDIA GB200 及 Rack CDU



资料来源：Microsoft X 推文

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/397021115105006161>