

引用格式: 洪学海. 人工智能驱动信息技术体系重构与再造——兼评《中国至2050年信息科技发展路线图》. 中国科学院院刊, 2025, 40(1): 149-161, doi: 10.16418/j.issn.1000-3045.20240919002.

Hong X H. Artificial intelligence driven information technology system restructuring and reconstruction—Review of *Information Science & Technology in China: A Roadmap to 2050*. Bulletin of Chinese Academy of Sciences, 2025, 40(1): 149-161, doi: 10.16418/j.issn.1000-3045.20240919002. (in Chinese)

人工智能驱动信息技术体系 重构与再造

——兼评《中国至2050年信息科技发展路线图》

洪学海

中国科学院计算技术研究所 北京 100190

摘要 《中国至2050年信息科技发展路线图》和《信息科技：加速人—机—物三元融合》，对我国信息科技领域的长期发展进行了预判。文章回顾了近15年来信息科技领域的发展变化，并与当初的预判进行了比较，指出了当初的一些重要判断符合预期。近年来，全球信息技术重大创新的整体节奏在放缓，而新一代人工智能快速演进，将为未来10年信息技术体系重构和再造带来新机遇，并将加速信息技术体系创新进程。文章对计算技术、数据空间、网络空间和智能空间的技术体系重构与再造进行了系统分析，指出了挑战性问题，并提出了发展对策。

关键词 人工智能，算力空间，数据空间，网络空间，智能空间，重构与再造

DOI 10.16418/j.issn.1000-3045.20240919002

CSTR 32128.14.CASbulletin.20240919002

2009年，以李国杰院士为组长的中国科学院信息科技路线图专家组，出版了《中国至2050年信息科技发展路线图》（以下简称《路线图》）^[1]，并于2013年出版的《科技发展新态势与面向2020的战略选择》中

发表《信息科技：加速人—机—物三元融合》（以下简称《三元融合》）^[2]。《路线图》和《三元融合》前瞻性、战略性分析了我国信息科技未来几十年发展的大趋势和战略重点，为当时我国信息科技发展方向和

修改稿收到日期：2025年1月10日

战略重点圈定了相关重点领域，对我国信息科技领域的重大科研活动起到了一定的指引作用。近15年来，全球信息科技发展已经发生了一系列重大的变化，尤其是近几年来，以大模型为代表的新一代人工智能技术极速发展，为信息技术体系重构与再造创新带来了重大的机遇，将加速驱动信息技术体系创新进程。因此，评估十几年前出版的《路线图》和《三元融合》中对信息科技相关战略预判的一些结论与影响，并在新形势下，进一步前瞻未来10年信息科技变化趋势，为抢占信息领域科技制高点，到2035年实现科技强国等战略目标，具有重要的现实意义。

1 《路线图》和《三元融合》的重要预判与近15年发展的比较

回顾近15年来，信息科技发展变化与《路线图》和《三元融合》给出的预判，现在看来，很多当初的预判至今依然适用。

1.1 关于信息科学理论发展的长期预判符合预期

《路线图》认为：信息技术不会变成以增量改进为主的传统产业技术，而是面临一次新的信息科学革命。信息技术的基础理论大部分是在20世纪60年代以前完成的，近40年信息科学没有取得重大突破。上一次基本创新（即基于科学突破的重大发明）的高峰期是在20世纪40年代，现在已有大量的知识积累，按照经济与技术发展长波规律的推测，21世纪20—30年代可能出现基本创新的高峰。2020年以后什么技术将成为新的主流技术就会逐步明朗；2020—2035年将是信息技术改天换地的大变革期。预计21世纪上半叶将兴起一场以高性能计算和仿真、网络科学、智能科学、计算思维为特征的信息科学革命。在网络科学和智能科学取得重大突破以后，21世纪下半叶，基于信息科学的新的信息技术将取得比20世纪下半叶更大的发展。

近几年，人工智能（AI）突飞猛进，验证了《路

线图》的基本预判。AI现有重大技术发明，如Transformer深度学习框架，是基于Geoffrey Hinton等科学家长期对神经网络模型的基础理论研究成果。深度学习的黑盒模型为AI科学研究提出了急需解决的科学问题，需求的牵引必将引发科学的突破。当前AI技术的大发展，预示人类已经处在进入智能时代的前夜，目前的技术离实现真正的通用人工智能（AGI）还有一定的距离，再经过10—20年的努力，大概率有基于重大科学突破的基础发明，如同20世纪40—60年代一样，发明计算机和集成电路以后，人类就进入了信息时代。

1.2 关于信息技术发展的长期预判基本符合预期

1.2.1 人—机—物“三元融合”预判正在成为现实

《路线图》和《三元融合》指出，人—机—物三元融合强调的是物理空间、信息空间和社会空间的有机融合，物理空间分别与信息空间、社会空间源源不断地进行信息交互，而信息空间与社会空间则进行着认知属性和计算属性的智能融合。

近15年来，人—机—物三元融合正在加速。快速发展的移动互联网、物联网、4G/5G高速接入网及边缘智能等，为实现人—机—物三元融合准备好了物质条件，数据智能化为智能融合提供了纽带，计算机系统的基本模式正在从人机共生向人—机—物三元融合世界发展。移动互联网实现了人与人的互联、融合，物联网（传感网）实现了人与环境的互联与融合，工业互联网实现通过网络连接各种工业设备和系统，实现工业数据的实时传输、共享与智能化处理，并通过新一代智能模型，改善、提升工业生产效率和质量，以及成本控制等。人—机—物三元融合最显著的是脑机融合，脑机神经连接是重要的科技突破。埃隆·马斯克的“神经连接”公司，继首例人脑设备植入手术顺利完成后，2024年7月迎来第2名人类植入者。通过在人脑皮层植入NeuroLink相关芯片并采用大约10 bits/s的通信速率，使得瘫痪病人能够通过思考来控制

他们的手机或电脑^①。这真正实现了人一机一物三元融合。Neurolink成为现实，也预示着《路线图》指出的，“今后几十年内模拟计算可能又会成为受人重视的研究方向。我们在重视数字技术的同时，还要探索模拟量处理的新途径以及数模混合处理的新方法”，这一预判得到一定程度的验证。

1.2.2 计算与服务平台的总体结构依然呈现出“集中与分散交替为主”的发展规律

《路线图》指出，计算与服务平台的总体结构几乎是每隔20年左右有一个重大的变革，呈现出计算资源集中与分散交替为主的“三国定律”。目前，在“AI大模型+Agent^②”“AI大模型+RAG^③”等模式的加持下，未来这种集中与分散交替为主的发展规律将继续延续。装载有轻量化AI模型+Agent或+RAG或+Copilot^④的AI PC，将进一步改变人们使用计算机和互联网的模式。一方面，计算与服务模式分散化的AI

PC将进一步解耦人们对集中化的计算资源的需求，以及紧耦合的服务模式；另一方面，集成式的多云系统、联邦云系统、超算互联网等，将进一步提升计算资源的最大化利用（图1）。因此，《路线图》所指出的计算与服务平台的集中—分散交替的“三国定律”，在未来还会继续演进。这种周期性的变化，使得信息技术领域始终保持着创新的活力。

1.2.3 集成电路领域延续创新变革发展态势

《路线图》指出，集成电路领域将延续摩尔定律（more Moore）、扩展摩尔定律（more than Moore）和超越CMOS（beyond CMOS）3个路径发展。

在延续摩尔定律方面，晶圆级的大芯片已经成为芯片结构横向扩展发展的重要成果。2012年后登拉德缩放比定律受限，摩尔定律发展速度虽然放缓，但围绕缩小COMS工艺特征尺寸、提高集成度，以及通过新材料的应用和器件结构的创新来改善电路性能的努力

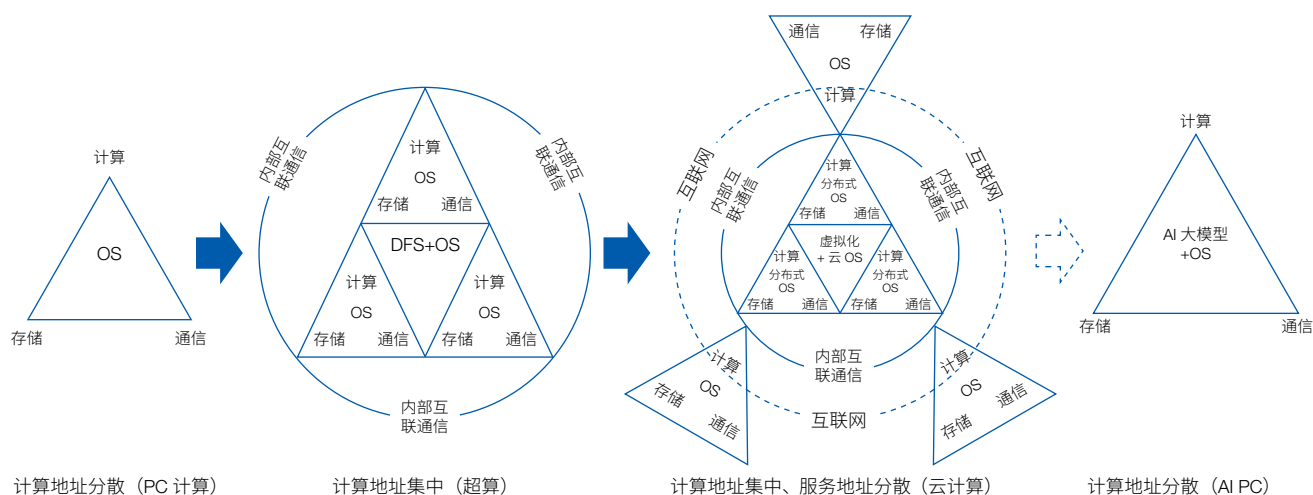


图1 计算与服务平台总体结构“集中与分散”简略图

Figure 1 Schematic diagram of overall structure of computing and service platform, including “centralization and decentralization”

① 解局|全球首例脑机接口设备人体移植！马斯克的“心灵感应”灵不灵？. (2024-01-30)[2025-01-12]. <http://www.xinhuanet.com/20240130/d15db20a738b4a3881158fcff3d5a5e/c.html>.

② 人工智能的Agent(智能体)是指能够感知环境、进行决策和执行动作的智能实体。

③ RAG(Retrieval Augmented Generation), 是一种结合检索和生成技术的模型。

④ AI Copilot是指依托于大模型(LLM)的智能助手。

力仍然在继续, FinFET^⑤结构和工艺技术在半导体制程发展到22—5 nm过程中, 发挥了重要的作用。2020年, 在同行都在将晶圆分割成数百颗独立芯片之时, 美国Cerebras Systems公司则是选择将整个晶圆做成一颗芯片, 其发布的WSE-2二代晶圆级芯片, 采用CMOS 7 nm工艺制程, 集成了1.2万亿个晶体管、40万个核心^⑥。这也是FinFET结构和工艺发展的重要成果之一。近年来, 我国由于在集成电路领域受到“卡脖子”, 与FinFET同时代发明的FD-SOI^⑦技术也被重新捡起来, 成为破解之道之一。

在扩展摩尔定律方面, 近些年来, 基于先进封装技术的Chiplet发展起来。Chiplet技术允许将整个芯片拆分成多个较小的、可以用不同工艺制造的不同模块, 然后通过高速互连方式集成到一个封装中, 实现全功能的芯片系统, 从而优化性能、功耗和成本。因此, Chiplet技术被视为一种可以拓展摩尔定律的方式, 延续了集成电路行业提高性能和降低成本的趋势。咨询机构Markets.us研究报告称, 在2023年, CPU Chiplet占据了主导市场地位, CPU Chiplet市场份额超过41%^⑧。虽然GPU Chiplet市场份额低于CPU Chiplet, 但在专业应用领域发挥了关键作用。此外, 通过传统微电子工艺, 实现光电子器件和微电子器件的单片集成的硅基光电子集成技术, 有效解决了集成电路芯片目前金属互联的带宽、功耗和延时等问题, 也实现了扩展摩尔定律。

在超越CMOS方面, 碳纳米管晶体管已经展现出超越商用硅基晶体管的性能和功耗潜力, 碳基集成电路技术成为重要发展对象。2019年8月, 美国麻省理工学院的Gage Hills等^[3]在*Nature*发表论文, 报告了碳纳米管芯片制造领域的一项重大进展: 一个利用14 702个碳纳米管晶体管构成的16位RISC-V指令集微处理器RV16X-NANO, 该处理器采用兼容CMOS工艺制造, 证明可以完全由CNFET^⑨打造超越硅的微处理器, 为先进微电子装置中的硅带来一种高效能的替代品。2024年7月, 北京大学基于碳纳米管晶体管新型器件技术, 结合高效的脉动阵列架构设计, 成功制备了世界首个碳纳米管基的张量处理器芯片, 可实现高能效的卷积神经网络运算^[4]。

此外, 近15年来, 集成电路芯片发展的“牧本周期”^⑩还在延续, 能够实现更好性价比的领域专用架构(DSA)处理器, 如NPU、TPU、DPU^⑪等各种“XPU”, 应运而生。国内研发的深度学习系列处理器寒武纪、类脑天机芯、天眸芯^[5]和市场上自动驾驶芯片等属于DSA范畴。同时, 核心数增多的通用架构的多核和众核处理器仍然在发展, 如X86系列CPU处理器和NVIDIA的系列GPU处理器。

1.2.4 软件工程的摩尔定律日益显现

《路线图》指出, 软件工程的发展走势将类似于摩尔定律, 今后几十年内如果能够使得软件业和服务业也产生摩尔定律现象, 无疑将会引发一场革命。目

⑤ FinFET, 即Fin Field-Effect Transistor, 鳍式场效应晶体管。

⑥ 你见过晶圆大小的芯片嘛? 这家AI独角兽推新品对标英伟达H100. (2024-03-14)[2025-01-12]. <https://baijiahao.baidu.com/s?id=1793494526161632350&wfr=spider&for=pc>.

⑦ FD-SOI即Fully Depleted Silicon-On-Insulator, 全耗尽型绝缘体上硅。

⑧ 半导体行业观察. Chiplet市场规模, 高速增长. (2024-01-22)[2025-01-12]. <https://t.cj.sina.com.cn/articles/view/3856710564/e5e0bba401901j84i>.

⑨ CNFET, 即Carbon Nanotube Field-Effect Transistor, 碳纳米管场效应晶体管。

⑩ “牧本周期”是指芯片类型有规律地在定制化和通用化之间不断交替演化。牧本周期大约每十年波动一次, 反映了设计工具和方法学的成本效益变化。也称之为“半导体行业之摆”。

⑪ NPU(neural processing unit)神经网络处理器, TPU(tensor processing unit)张量处理器, DPU(data processing unit)数据处理器。

前，基于LLM大模型的AI编码生成助手，为软件工程的摩尔定律提供了依据。如谷歌首席执行官曾透露，该公司通过在代码自增长工具中集成大模型，生成了这家科技公司超过1/4的新代码，包括自动导入包、自动生成构造函数等。目前AI辅助编码方面，出现了两大发展方向：① AI编码助手或者AI代码生成器大量涌现，如美国GitHub与Open AI合作推出的GitHub Copilot、亚马逊的CodeWhisperer等；② 传统的低代码/无代码工具，大量引入AI辅助功能，如低代码开发旗舰公司OutSystems通过Mentor新型生成式AI驱动的“数字工作者”改变了整个软件开发生命周期。统计显示，目前软件工程师基于大模型开发应用软件，时间上可以节约20%—30%。随着面向软件开发的专用大模型能力日益增强，在软件工程领域的摩尔定律也将有望成为现实。

2 未来10年信息技术体系重构与再造的创新机遇与挑战

《路线图》预测，2020—2035年将是信息技术改天换地的大变革期，将可能出现基本创新的高峰。自2019年以来，AI大模型的大发展表明，信息科技已经进入到基本创新突破期的前夜。在全球信息技术创新进入缓慢期的背景下，AI将加速信息技术体系的创新进程。因此，未来10年将是信息技术体系重构与再造的创新机遇期。一方面，对于原理还不太清楚的AI科学，一定还会有大的突破；另一方面，信息领域的科学与技术融合发展将成为大趋势，并且信息技术将成为信息科学发展的主要推动力。更重要的是，新一代AI将加速驱动计算技术体系、数据空间技术体系、网络空间技术体系和智能空间技术体系的重构与再造。

2.1 信息技术体系重构与再造创新的历史演进

人类对信息技术体系重构与再造创新一直没有停滞。

在处理器方面，从4位微处理器到64位处理器，

从复杂指令集（CISC）到精简指令集（RISC），从一级缓存到二级、三级，执行从顺序指令执行到乱序指令执行，从单一核心架构到多核心架构、众核心架构，从通用功能架构到功能专用架构并存，从单个芯片到异构集成封装多核心、多功能的芯片等。目前，在AI等领域对算力提出更大需求的牵引下，晶圆级大芯片也被开发出来。

在计算机体系结构方面，20世纪50—60年代，为实现资源的最大化利用，通过分时操作系统提高指令执行效率的批处理计算机得到快速发展。此后，为提高指令并行性进一步提高计算机运行速度，指令流水线技术取得重大突破；为了进一步突破计算性能瓶颈，超标量与超长指令字两种计算机体系结构应运而生。再后来，多任务、多个处理器并行执行的并行计算体系在一系列超级计算机中得以广泛应用。近年来，面向大数据、云计算和AI等发展需求，按照数据流动规律进行组织和管理的数据流体系结构重新得到重视，成为计算机体系结构发展的“老树新芽”技术。

在信息技术产品生产方面，从早期的由一家厂商包揽了一个计算机的软件、硬件设计与制造的垂直体系，向多厂商分散、配合生产各自优势产品的扁平化体系方向发展。例如，早期的美国IBM公司生产的大型计算机，其操作系统、CPU、存储器等都是由IBM自己设计、生产和制造。随着微软Windows操作系统、Intel CPU等的成熟商用，计算机生产商不再做垂直化产品研发的工作，而是将其他厂商的产品进行扁平化分工、整合，由此形成了“Windows+Intel”（Wintel）计算产品体系，并形成了牢固的产品迭代节奏，至今形成了“Wintel”、“ARM+Android”和“ARM+iOS”体系。目前，由于GPU和Transformer架构在AI大模型领域的成功，“NVIDIA+Transformer”体系已经基本形成。

2.2 计算技术体系重构与再造

近年来,全球信息技术重大创新的整体节奏在放缓。未来10年,在人工智能发展驱动下,计算技术体系创新将快速推进。一方面,进入后摩尔时代,为追求算力增长和功能丰富,计算处理器芯片将继续向低成本、高良率的Chiplet堆叠集成芯片和3D芯片方向发展,异构集成创新成为产业技术主要发展方向,晶圆级大芯片将成为重点发展方向之一;同时,功能丰富的低世代工艺智能芯片、新型硅光电子芯片、碳基芯片,以及非冯·诺依曼体系结构量子计算芯片,也将快速向前发展(图2);另一方面,算力空间将向算力聚合与服务能力提升方向前进,将由超级计算、云计算,向云联邦、超算智算融合、算网融合等方向发展(图3)。

2.2.1 处理器技术体系重构与再造

在处理器设计环节,芯片设计的复杂性日益呈指数级增长,导致设计周期长、成本高,传统的设计方法难以应对这种复杂性。未来10年,AI技术将成为优

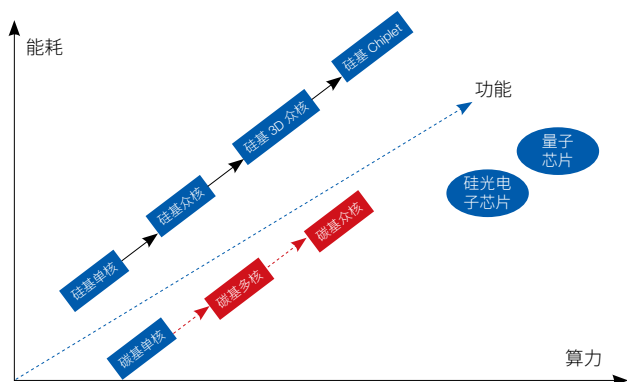


图2 算力芯片重构与再造演进路线图简图

Figure 2 Schematic diagram of evolution roadmap for computing chip reconstruction and reconstruction

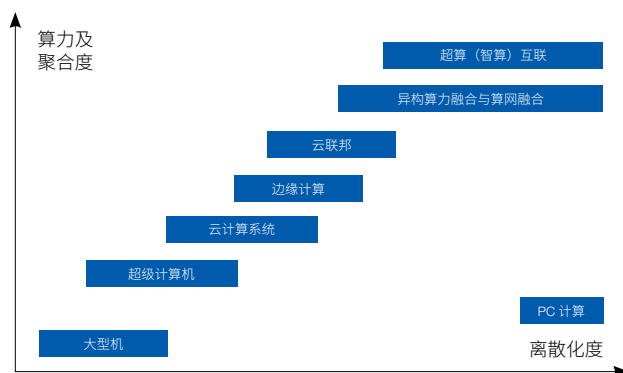


图3 算力空间重构与再造演进路线图简图

Figure 3 Schematic diagram of evolution roadmap for computing space reconstruction and reconstruction

化处理器芯片设计过程的关键工具,提高设计效率和质量,并生产出性能更高的芯片。2024年3月,美国白宫科技政策办公室发布《国家微电子研究战略》,明确提出加强将AI和机器学习及基于物理的方法集成到EDA工具中。在2024年全球芯片领域顶会Hotchips年会上,多个报告展示了AI辅助芯片设计的重要成果。其中,美国Synopsys公司报告了强化学习(reinforcement learning)在芯片设计中的应用,其Synopsys.ai套件,在大语言模型支持下,在整个EDA堆栈中充分利用生成式人工智能(AIGC),用于RTL(register-transfer level)设计、验证及其他辅助资料创建的生成功能等^⑫。在处理器制造环节,比利时微电子研究中心(IMEC)发布的工艺路线图指出^⑬,FinFET晶体管结构将在3 nm走到尽头,然后过渡到新的GAA(gate all around)结构,集成电路工艺尺度将进入埃米^⑭阶段。此路线图提出的CMOS 2.0范式愿景指出,CMOS 2.0是通往真正3D芯片的道路。但这种方法面临的挑战在于将严重依赖后端供电网络

⑫ 6 Synopsys. AI-Assisted Chip Design Tutorial. (2024-08-25)[2025-01-12]. <https://www.hc2024.hotchips.org/assets/program/tutorials/3-HC24.synopsys.SteliosDiamantidis.v03.pdf>.

⑬ 半导体行业观察. 未来十年的芯片路线图. (2023-05-29) [2025-01-07]. <https://baijiahao.baidu.com/s?id=1767226748342488547&wfr=spider&for=pc>.

⑭ 1 nm 等于10 埃米,即10 Å。

(BPDN), 会将所有供电改到晶体管的背面, 并需通过系统技术协同优化 (STCO), 重新思考设计过程, 改变设计方法。相对于硅基集成电路计算芯片, 碳基芯片具有良好的低功耗、与硅基工艺兼容良好等特点。如前文所述, 目前碳基芯片已经取得积极进展^[4], 将成为未来与硅基芯片并行发展的主要方向之一。

从更长期来看, 处理器芯片重构与再造创新发展趋势体现在如下 5 个方面: ① **电路理论方面**, 从电子电荷向分子、极化、强电子相关态、自旋方向等方向发展; ② **材料方面**, 从硅基向碳基、宏观分子材料、纳米结构、复合金属氧化物等方向发展; ③ **器件结构方面**, 从 CMOS 器件向分子器件、自旋器件、铁磁性器件、量子器件等方向发展; ④ **数据载体方面**, 从模拟量、数字量向量子位等方向发展; ⑤ **系统结构方面**, 由冯·诺依曼架构、多核众核架构, 向可重构、量子、神经形态学计算架构等方向发展。目前, 量子计算还处在探索阶段, 主流方案包含超导、离子阱、光量子、超冷原子、硅基量子点和拓扑等多条技术路线, 基本都沿着量子计算优越性由专用量子计算向通用量子计算的路线图发展。

对我国而言, 在 FinFET 时代, 集成电路制造正在面临美国联合其盟友的打压、封锁, 向更先进制程工艺芯片制造发展受阻, 同时 X86、ARM 指令架构也存在需要授权的制约。随着集成电路制造工艺向 3D 方向发展, 前道工艺光刻机的作用将会减弱, 刻蚀机、薄膜机和电子束检测设备等作用将会增加, 我国发展集成电路技术与产业有了新机遇。因此, 我国一方面要积极发展集成电路先进制程制造工艺技术, 研发先

进制程的前道工艺极紫外光刻 (EUV) 设备; 另一方面, 要积极推动发展刻蚀机、薄膜机、电子束检测设备等, 向满足 3D 芯片制造工艺方向发展。此外, 基于先进封装的 Chiplet 技术, 以及与 FinFET 同时代产生的 FD-SOI^⑮ 技术, 也可能为我国计算芯片发展提供新的选择。FD-SOI 是不同于 FinFET 的技术与工艺路线, 是一种平面工艺技术, 具有减少硅几何尺寸同时简化制造工艺的优点, 在 22/12/10 纳米 FD-SOI 技术性能与 14/7/5 纳米 FinFET 技术相当; 并且, FD-SOI 相比 FinFET, 具有 PPA^⑯ 平衡性好等特点。但目前主要集成电路制造设备与工艺是围绕 FinFET 生态的; FD-SOI 面临的挑战是还没有建立完整的产业链生态, 扩大市场难。

2.2.2 算力空间技术体系重构与再造

为适应人工智能等领域对更大算力和多种资源协同服务的需求, 未来 10 年, 面向应用的计算将与智能计算深度融合, 表现为临近互联、封装集成、大规模向量化等, 并期待 RISC-V+AI Core^⑰ 指令接口的标准化。在 AI 发展驱动下, 算力空间重构与再造的一个重要方向是优化冯·诺依曼架构, 即在一个计算机系统内部区域做工作, 想办法通过减少内存和处理单元之间的路径, 实现高内存带宽及较低的访问开销, 打破冯·诺依曼架构“存储墙”和“功耗墙”等问题, 实现更高效能的计算。

在 AI 计算方面, 由于不同模型对算力和带宽的需求不同, 如基于 Transformer 的 LLM 计算中, 性能瓶颈常常是在带宽而非计算, 即带宽跑满、算力闲置。为此, 存算一体、存内计算 (PIM/PNM^⑱) 技术路线

⑮ FD-SOI, fully depleted silicon-on-insulator, 全耗尽型绝缘体上硅, 是一种在 SOI 衬底上制作全耗尽型 CMOS 器件的技术。采用 SOI 结构, 在一块绝缘层上生长硅层, 从而减少晶体管中漏电流。

⑯ PPA 指: 功耗 (power)、性能 (performance)、面积 (area)。

⑰ AI Core 是指一种专门为深度学习算法优化设计的计算单元, 通常包括矩阵计算单元 (Cube Unit)、向量计算单元 (Vector Unit) 和标量计算单元 (Scalar Unit) 三种基础计算资源, 分别对应张量、向量和标量三种常见的计算模式。

⑱ PIM/PNM, Processing In/Near Memory, 存内/近存计算。

被提出,利用片内的高带宽,处理常见的矩阵运算和部分向量运算。具体是通过TensorCore(张量计算核心)的乘累加单元阵列提供算力,通过HBM^①的高带宽使数据能源源不断到达TensorCore。在AI计算中,虽然PIM/PNM可以减少数据从内存读取的时间,提高计算效率,从而加速模型的训练和推理过程^②,但其面临众多技术挑战,包括将计算单元集成到内存芯片中的复杂性、功耗和散热问题,以及数据一致性和可靠性问题等。

在片上和集群的算力互联方面,使用片上光网络(ONoC^③)连接多个同构的处理单元,如互联CPU、AI Core等,这是将光集成电路(PIC)作为NoC与AI Core的硅片集成,封装成一个AI芯片,直接在芯片上做光电转换输出到光缆,实现了芯片之间的互联。基于光网络的互联方案与动态调度方面,如Google TPU4 AI训练集群的光互联方案,是将64颗TPUv4以4×4×4的方式构成一个三维立体结构Cube,并且这个AI训练集群的拓扑互联方案的重构,可根据需求实时对AI计算资源缩容和扩容^④。

面向人一机一物三元融合的广泛应用场景,AI等算力与服务需求以及高性能计算机(超算)从E级向Z级发展面临诸多难题,未来将跨网域、多异构的算力与服务聚合是一条重要的发展途径。将超算、智算和量子计算等多种异构算力资源融合,并结合算力网实现算网融合是一个大胆的创意,但多种异构算力资源的融合是一个难题。例如,AI算力主要由基于GPU、NPU等芯片,但AI的应用通常会同时用到CPU、GPU、TPU等,怎么解决CPU运算与GPU、TPU等运算衔接是个难题。因此,需要解决多种算力资源虚拟化、接口标准统一、高效协同计算、应用任务分布与调度、编程模型等难题。同时,异构算力如

何与算力网进行融合也是个难题,需要突破原有的并行计算、云计算等思想,进行颠覆性的创新。为此,需要进行原理性创新和技术实验的重大科技基础设施来支撑。自2023年以来,美国IBM、微软、英伟达、谷歌等公司分别合作,在构建量子—超算异构融合算力平台方面取得进展,如IBM利用127比特量子云平台与“富岳”超算的结合,实现了包含28个原子的FeS团簇分子计算。这为多种算力融合技术发展路径提供了很好的借鉴。

2.3 数据空间技术体系重构与再造

数据已经成为AI大模型发展的三大核心要素之一。中国工程院发布的《数据空间发展战略蓝皮书(2024)》(以下简称《蓝皮书》)定义的未来数据空间是:人一机一物互联,产生大量数据,通过社会再生产,数据又作用于人一机一物,这一实践活动最终形成了人类活动的新空间。即:在AI发展驱动下,数据空间将由单一计算机系统的数据空间、基于互联网的数据中心空间,发展到人一机一物融合的人类智能活动空间。因此,未来10年,构建面向人类智能活动的的数据空间技术体系,是一项重要任务。

在AI发展需求驱动下,构建面向人一机一物三元融合的人类智能活动的的数据空间,还面临众多挑战,主要包括:① 理论范式问题。基于什么理论来支撑数据空间构建,是基于自组织理论还是他组织理论?基于自组织理论要研究数据空间构建是怎样从混沌无序的状态向稳定有序的终态的演化,以及系统内部各要素之间的协同机制。基于他组织理论,要研究数据空间构建需要设计怎样的制度、政策及机制等外部力量,推动数据空间的建立。此外,对于数据空间建立的复杂度,是基于机械论还是耗散结构理论(即是强调将数据空间系统是由数据孤岛构成,通过系统组织

① HBM, High Bandwidth Memory, 高带宽存储器。

② ONoC, Optical Network-on-Chip, 是一种基于光学技术的片上网络架构。它利用光学信号在芯片内部进行数据传输,旨在解决传统电子片上网络(ENoC)在带宽、延迟和功耗方面的问题。

形成“整装数据”结构),还是强调引入负熵来形成有序数据空间结构?等等。② **技术问题**。数据空间的架构是什么?各个数据孤岛是通过什么方式进行连接?联接协议是什么?接口是什么?有什么标准?如何保护各个数据源的数据隐私与安全?数据空间中的数据如何进行统一标识?如何封装?数据空间如何与算力空间进行耦合?数据空间的各类数据如何统一标识和表示,以便于进行综合挖掘和智能应用?等。目前,《蓝皮书》为此提供了一些参考,包括把数据空间看作一个数据要素场、构建数联网根服务体系,以数联网、数据标识、数据向量化、深度神经网络学习及AI大模型等为技术应用体系等。

2.4 网络空间技术体系重构与再造

几十年来互联网体系结构保持相对稳定,但依然存在两大问题:鲁棒性差,脆弱,不安全;适应性弱。为此,国内外对互联网体系结构的创新努力一直在持续,基本上采取渐进式和变革性两条技术路线,即:采用“自下而上”打补丁的方法,升级更新具体技术适应新业务和新终端;采用“自上而下”完善体系结构,从根本性解决问题和克服挑战^[8]。变革性的技术路线是采用“推倒重来”的思路设计全新网络,国内外先后开展了New Arch、Clean Slate、GENI、FIND、FIA、FIRE、新一代互联网体系结构理论、面向服务的未来互联网体系结构与机制等研究计划或项目,先后提出了Open Flow、SDN、NDN、CCN等互联网体系结构的新设想和新技术。

未来10年,AI将驱动网络空间技术体系从硬件链路到网络协议,乃至网络应用的全体系的创新。除了在新型网络架构、寻址路由、内生安全等领域继续突破创新外,互联网体系结构还将向智能化、敏捷化、网存算一体化不断发展,从而支撑超大规模、人一机一物融合、跨时空壁垒的智能连接与服务。随着AI技

术的发展,互联网将围绕4个方面进行创新^②:① **网络设备方面**,将AI能力下沉到网络设备。通过感知业务质量,感知信道性能等,让网络设备具有更强的感知能力,以便更好地优化网络,提升网络为特定业务的服务能力。② **路由协议方面**,通过增强数据面网络层的能力,简化控制面和管理面,满足应用需求的多样性,并提升边缘能力来降低对中间节点的要求。通过协议的简化和优化,使得网络可感知、可溯源、可定位,实现更好的性能及内生安全要求。③ **网络管理方面**,首先是网络本身的智能化,随着AI大模型在网络管理中的应用,提升网络智能化水平,让网络能够具有自动驾驶、自动发现、自动配置、自动维护的能力;其次,网络需要能够适应更多业务发展,具有更好的开放性和安全性,为云网融合、算网融合实现资源一体化调度。④ **安全方面**,改变过去网络安全被动式响应和防御的方式,未来要从网络设备的内生安全,到端到端的云、网、边、端协同防护体系的构建,持续提升网络安全防护能力。

2.5 智能空间技术体系重构与再造

人类大脑是一个既能处理专用任务又能处理通用任务,并具有“自觉意识”的低能耗智能空间。信息领域的智能空间是人工智能融合的空间,是能处理各种模态信息的智能体的集合,其愿景是构建像人类大脑的智能体,使得处理各种专用任务的智能与处理通用任务的智能进行深度融合,解决智能碎片化的问题。

2024年诺贝尔物理学奖和化学奖都颁发给机器学习领域的科学家,表明以机器深度学习为代表的AI,成为未来10年最有前景的技术。近年来,以大模型为代表的新一代人工智能的研究和应用取得突破性进展,催生AI新的理论 and 应用范式,推动AI理论和应用进入新的发展阶段。美国Open AI公司的ChatGPT

② 引自郭贺铨院士在第三届中国IPv6创新发展大会上的报告。

成为新一代人工智能发展的分水岭。AI大语言模型技术不断成熟，率先在文本AIGC产生大规模应用，并催生图像生成模型蓬勃发展，如DALL-E2、Stable Diffusion等。大模型从“一专多能”向“多专多能”前进，带动新一轮应用范式创新。AI大模型技术也使信息领域的科技创新和产业生态发生巨变，促进了从底层芯片到应用的穿透式的技术与产业生态重构，并将逐渐重构数据空间、算力空间、开发框架、算子库，乃至基础芯片、系统及开源和应用服务模式等。AI for science正在重构科学研究范式（AI4R），加速驱动科学研究进程，产生颠覆式突破。当前，AI for science已经在生物领域取得显著成效。例如，2024年5月，美国Google旗下DeepMind公司发布的新一代AI模型AlphaFold3，能够预测蛋白质、DNA、RNA、小分子等的几乎所有生物分子结构和相互作用；AI for Science已经在内嵌物理模型的神经网络（PINNs）、约束条件下组合优化问题求解、偏微分方程求解等领域取得初步成效，证明其具备解决科学研究与计算问题的强大潜力。未来，AI for science将继续驱动生命科学、医药研发、物理装置控制、数学发现、材料发现与合成等领域创新发展，并将在跨界交叉领域进一步驱动产业创新和更广泛的落地应用。同时，多模态大模型开启了通用人工智能（AGI）发展道路，使AGI进展到“临界点”。

未来10年，智能空间技术体系重构与再造面临诸多挑战。①目前的智能感知可适应性差、认知机理不明、泛化能力弱等问题已经开始制约AI的更广泛应用，亟待深入探索类人多模态感知、人机混合智能理论与方法，以及突破多源信息复杂场景和自主无人系统等决策、行为智能等。②当前AI研究已经从深度

学习时代走向“自监督+深度学习+强化学习”的大模型时代，以深度学习为基础的人工智能理论已经遇到天花板，急需探索AI新的理论。Minyoung Huh等^[9]最近发表的“柏拉图表征假说”，证明了神经网络训练，正趋向于在它的表征空间中收敛成一个共享的现实“世界统计模型”。Max Tegmark团队发表的一项颇具颠覆性的研究^[10]，揭示了LLM中竟存在类似于人类大脑的脑叶分区结构，显示出模型内部的几何结构与人类大脑的功能分区有着惊人的相似性。这一发现表明，AI在某种程度上模仿了人类大脑的信息处理方式。这类研究不仅提供了理解LLM内部运作的新视角，也对AI的潜力有了更为深刻的认识，为未来AI的发展提供了新的理论支持。随着研究的深入，未来或许能在大模型的基础上，开发出更智能、更类人的AI。③当前的多模态大模型面临数据资源耗尽、能耗极大、成本极高和商业模式等问题。更重要的是，GPT-5研发受阻，文本域里规模定律（Scaling Law）可能已遇到天花板，而探索新型Scaling Law的预训练多模态大模型、空间智能是不是就是通用人工智能（AGI）的发展技术路径，还有待进一步的验证。未来智能空间重构一个可能的途径是在大模型、空间智能的基础上，构建更多各种功能的智能体Agent，与大模型等进行深度融合与功能拓展，整合各种人工智能为一个集成的智能环境，如发展具身智能。但正如李国杰院士指出的，具身智能还有许多认知问题需要解决^②，这预示着未来通过具身智能发展AGI还面临众多挑战。另外一个可能的途径是Open AI发现的扩展测试时计算技术，即：在大模型推理阶段，扩展思考和计算，进行反向多步推理，让大模型能够自主学习策略、拆解任务、识别并纠正错误。得益于思维链

② 李国杰院士在2024松山湖论坛指出具身智能两个“元问题”：①具身智能是集成各种人工智能技术的综合性智能，还是所谓接近动物智能的“行为主义智能”？②具身智能是不是通用智能？要不要重点研究适应各种环境的通用具身智能？人工智能究竟是一种不解决具体问题的原始能力，还是针对明确任务和有限环境边界的问题求解技术？

(CoT) 推理的引入, Open AI o1 模型是测试时计算的突破, 证明了模型可以进行更深入的推理和解决更复杂的问题, 为解决规模定律天花板问题和发展 AGI 提供了全新的思路。当然, 还有其他可能的技术路径, 这些努力将加速 AGI 的到来。

3 发展对策

3.1 聚焦 AI 驱动的信息领域重大科技问题、突破关键理论与技术

我国信息科技领域既面临短期破解局部“卡脖子”和建立自主技术体系与生态的长期需求, 又要面临后摩尔时代颠覆性技术创新突破、抢占科技制高点的历史使命。因此, 要持续加强信息科技领域的中长期战略研究, 聚焦新一代 AI 的理论、技术与新模型产品, 聚焦发展 AI 之根基的各类高性能 XPU 算力芯片的设计与制造、高质量数据集和新型网络的构建, 以及高效能计算基础设施建设等, 将技术创新与工程实现和技术产品与市场生态建设等统一起来, 组织科研院所与市场龙头企业进行协同创新。一方面, 聚焦“卡脖子”真问题, 除芯片制造前道工艺 EUV 光刻机外, 未来集成电路向 3D 芯片发展, 前道工艺光刻机作用在减弱, 高精度、高可靠与高性能的刻蚀机、薄膜机及电子束检测设备等成为芯片制造关键设备, 急需布局研制; 同时自主架构高性能及开源 RISC-V 指令集的各种 XPU 芯片也需加强研发, 并用相当长的时间打造我国信息技术与产业自主生态。另一方面, 要抓住 AI 驱动全球信息技术领域正在进入重构与再造的创新机遇, 系统研究计算技术、数据空间、算力空间、数据空间、网络空间和智能空间的技术体系重构与再造面临的挑战性问题, 协同市场龙头企业, 组织技术创新研发与工程实现队伍, 推进我国信息科技领域的整体创新。

3.2 布局信息领域重大科技基础设施

没有大的工程牵引、重大科技基础设施支撑, 信

息科学不会有大的发展。信息科技发展历史上, 重大科技基础设施为信息技术创新奠定了物质基础, 产生了一系列重大技术发明。例如, 在互联网领域, 美国早期支持 ARPANET 的基础设施, 为人类创造了互联网。围绕未来网络, 美国组织了“从零开始”(“Clean Slate”)的革命性未来网络研究, 先后支持下一代互联网研究基础试验设施(GENI)和从零开始设计新的互联网架构(FIND)等项目。在云计算领域, 美国也支持了全球信息网格(GIG)和云计算测试床(CloudLab)等项目。重大科技基础设施的建设为美国等在信息科技领域领先, 创造了极大的优势。为抓住全球信息技术发展进入到缓慢期以及计算技术、算力空间、数据空间、网络空间和智能空间等技术体系重构与再造的机遇期, 我国应加快布局信息领域的重大基础设施, 抢占信息科技制高点。

3.3 一体化布局信息科技领域“装置群”

在自然科学领域, 往往围绕一个或两三个重大科学问题, 构建一个大型的基础设施, 进行重大科学探索与发现的实验研究。信息领域与之不同, 信息领域的重大科技问题特征是: 基础原理、基础共性技术和领域应用 3 类问题并存, 许多重大科技问题分散在信息技术体系框架的各个部分; 并且, 信息科技的基本原理、共性技术等局部问题一旦突破, 将能够影响整个领域和其他领域的发展。我们不能指望通过一个集中化、实验型的基础设施来解决信息领域各个环节面临的问题。因此, 信息领域的重大科技基础设施应该是一个“装置群”, 以应对各个子领域的科技问题。同时, 信息领域的重大科技基础设施应该是一个柔性可组合的离散、分布式“大装置”, 要进行一体化的布局。

致谢 感谢李国杰院士、孙凝晖院士提出的宝贵意见和建议。

参考文献

- 1 中国科学院信息领域战略研究组. 中国至2050年信息科技发展路线图. 北京: 科学出版社, 2009.
Strategic Research Group in Information Field, Chinese Academy of Sciences China's Roadmap for Information Technology Development to 2050. Beijing: Science Press, 2009.
- 2 李国杰. 信息科技: 加速人—机—物三元融合// 中国科学院. 科技发展新态势与面向2020年的战略选择. 北京: 科学出版社, 2013.
Li G J. Information Technology: Accelerating the Ternary Integration of Human-Machine-Object// New Trends in Scientific and Technological Development and Strategic Choices for 2020. Beijing: Science Press, 2013.
- 3 Hills G, Lau C, Wright A, et al. Modern microprocessor built from complementary carbon nanotube transistors. *Nature*, 2019, 572: 595-602.
- 4 Si J, Zhang P P, Zhao C Y, et al. A carbon-nanotube-based tensor processing unit. *Nature Electronics*, 2024, 7(8): 684-693.
- 5 Yang Z Y, Wang T Y, Lin Y H, et al. A vision chip with complementary pathways for open-world sensing. *Nature*, 2024, 629: 1027-1033.
- 6 Lee D, Hyun B, Kim T, et al. PIM-MMU: A memory management unit for accelerating data transfers in commercial PIM systems. *arXiv*, 2024, doi: 10.48550/arXiv.2409.06204.
- 7 Jouppi N, Kurian G, Li S, et al. TPU v4: An optically reconfigurable supercomputer for machine learning with hardware support for embeddings// *Proceedings of the 50th Annual International Symposium on Computer Architecture*, Tokyo: Sigarch, 2023, doi: 10.1145/3579371.358935.
- 8 罗洪斌, 张珊, 王志远, 等. 对互联网体系结构创新的认识与思考. *电子学报*, 2024, 52(4): 1411-1420.
Luo H B, Zhang S, Wang Z Y, et al. Understanding and thinking about the innovation on Internet architecture. *Acta Electronica Sinica*, 2024, 52(4): 1411-1420. (in Chinese)
- 9 Huh M, Cheung B, Wang T Z, et al. The Platonic representation hypothesis. *arXiv*, 2024, doi: abs/2405.07987v5.
- 10 Li Y X, Michaud E J, Baek D D, et al. The geometry of concepts: Sparse autoencoder feature structure. *arXiv*, 2024, doi: abs/2410.19750v1.

Artificial intelligence driven information technology system restructuring and reconstruction

—Review of *Information Science & Technology in China: A Roadmap to 2050*

HONG Xuehai

(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

Abstract The *Information Science & Technology in China: A Roadmap to 2050* and *Information Technology: Accelerating the Integration of Human Machine Things* have predicted the long-term development of China's information technology field. This study reviews the development and changes in the field of information technology in the past 15 years, and compares them with the initial predictions, pointing out that some important judgments made at the beginning were in line with expectations. In recent years, the global development of information technology has entered a slow period, while the rapid evolution of the new generation of artificial intelligence will bring new opportunities for the reconfiguration and reconstruction of the information technology system in the next 10 years, and will accelerate the innovation process of the information technology system. This study systematically analyzes the technological system reconfiguration and reconstruction of computing technology, data space, network space, and intelligent space, points out challenging issues, and proposes development strategies.

Keywords artificial intelligence, computing power space, data space, cyberspace, intelligent space, reconfiguration and reconstruction

洪学海 中国科学院计算技术研究所研究员, 中国科学院计算机网络信息中心客座研究员, 博士生导师。《中国至2050年信息科技发展路线图》专家组成员。主要研究领域: 高性能计算+人工智能、分布式系统、大数据、信息技术与信息化发展战略等。
E-mail: hxh@ict.ac.cn

HONG Xuehai Professor and Ph.D. Supervisor at the Institute of Computing Technology, Chinese Academy of Sciences (CAS). Adjunct Professor and Ph.D. Supervisor at the Computer Network Information Center, CAS. Member of the expert group for *China's Information Science & Technology: A Roadmap to 2050*. Main research areas include HPC+AI, distributed system, big data, information technology and information technology development strategy, etc. E-mail: hxh@ict.ac.cn

■责任编辑: 文彦杰