

破局摩尔桎梏， 决胜算力时代

2022 先进计算七大趋势



DEEPTECH

Copyright @DeepTech 2022.
All rights reserved.

序言

算力，又可被称为计算力，即处理数据的能力。随着数字技术与数字经济的发展，全球数据总量呈现爆发式的增长，推动了算力需求和要求的不断升级。人类社会已朝着智能化和数字化方向迈进，计算资源进入供不应求的阶段，算力的重要性也因此被提到一个前所未有的高度。

算力在推动经济发展方面发挥着巨大作用。2025年，超过40%的全球经济将由数字经济带动，接近一半的经济增长都与数据有关。如果数据是数字经济时代的核心生产要素，那么数据的底座之下，算力承担了核心生产力的角色。算力为数字经济的增长带来新的引擎。另一方面，算力构建了科技进步的正循环。算力为医药、材料、生物和能源等学科提供了研究基础设施，而新兴技术的发展将进一步推动算力提升。

当下，计算性能的提升面临来自多个维度的挑战，算力发展已经进入瓶颈期。从硬件层面来看，随着硅基芯片制程趋近极限、“内存墙”、“功耗墙”问题的凸显，CPU、GPU等计算平台性能提升乏力，摩尔定律正在逐渐失效。从网络层面来看，算力节点通过网络灵活高效调配计算资源的能力仍存在不足、数据传输的损耗与开销过大，算力资源不能得到有效的利用。从性价比层面来看，算力相关软硬件设施的投入与回报不成正比，计算领域亟需延续摩尔定律的高性价比方案。

先进计算技术指代一切从计算理论、计算架构、计算系统等层面有效提升现有算力规模、降低算力成本、提高算力利用效率的创新性技术。为了解决上述难题，打破算力危机，先进计算技术主要关注两个重点方向：对于单个计算节点性能的提升，以量子芯片、类脑芯片为代表的非冯诺依曼架构芯片的出现为计算硬件变革带来了曙光；对于算力系统的高效利用，芯片层面与数据中心层面的多元异构计算将伴随云边端一体化趋势，构建随时随地、按需随形的全新计算体系。

算力是多领域多技术融合的载体，人类对于算力的追求没有止境。未来，随着先进计算技术的不断发展，计算的效能与能效都将迎来全新的突破，算力体系将得到颠覆性地变革。在先进计算技术的支撑下，更多智慧化、智能化的应用将会涌现，绿色低碳、开放开源的算力也将成为不可逆转的趋势。

目录

新一代计算硬件

| | | |
|------|-----------|----|
| 量子芯片 | 航海梯山 无远弗届 | 04 |
| 硅光芯片 | 摩尔的尽头有“光” | 07 |
| 类脑芯片 | 来自大脑的答案 | 10 |
| DPU | 以数据为中心的计算 | 13 |

架构演变

| | | |
|------|------------|----|
| 分布式云 | 云边端协同的关键枢纽 | 16 |
| 泛在算力 | 无处不在的算力网络 | 19 |

前沿应用

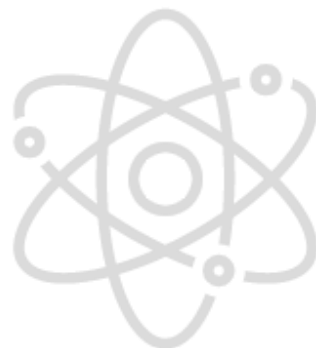
| | | |
|------|------------|----|
| 科学复兴 | 先进计算加速科学复兴 | 22 |
|------|------------|----|

量子芯片

1

航海梯山 无远弗届

当下，量子计算已经迈入 NISQ 时代。量子芯片作为量子计算机的核心，是量子计算实现规模化与商业化的关键瓶颈。在学术界与产业界的不断推动下，量子芯片在 2021 年迎来了大爆发。想要跨入通用量子计算时代，量子比特的数量与质量，比特间的连通性以及容错能力仍有待进一步突破。航海梯山，无远弗届。量子计算终将越过重重障碍，乘微观粒子以颠覆现代计算体系，为人类带来算力的腾越。



趋势 — 量子芯片 —— 航海梯山，无远弗届

量子计算是基于量子力学的全新计算模式，具有原理上远超经典计算的强大并行计算能力。量子计算以量子比特作为信息编码和存储的基本单元。与经典比特只能代表 0 或者 1 不同，具备叠加态的量子比特是 0 且 1 的存在。恰当利用量子叠加态与纠缠态，将使得量子计算机的运算效率远超传统计算机，获得指数级别的增长。

算力危机下，量子计算作为一种颠覆性的计算体系，具备帮助人类实现算力跨越式发展的无限潜力。随着“量子计算优越性”已被证明，量子计算已迈入含噪声的量子计算（NISQ）阶段。当下，科研与产业发展的重点将集中在实现解决特定问题的专用量子计算机上。

量子芯片是将量子线路集成在基片上，进而承载量子信息处理功能的处理器，是量子计算机的核心。量子计算机想要实现产业化，迈向商业阶段，量子芯片正是需要攻克的关键瓶颈。

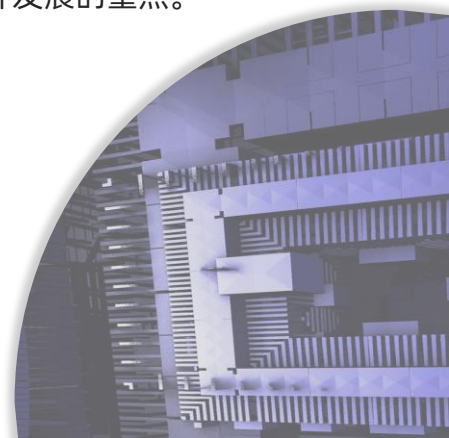
目前，包括 Google、IBM 等在内的科技巨擘和麻省理工学院、耶鲁大学、中国科学技术大学等世界知名高校已在量子芯片领域取得重大突破。

超导量子芯片，由于其电子线路设计和制备的工艺与现代集成电路技术相兼容，被认为是具有清晰的商业前景、能较快实现产业化的一种量子芯片技术路径。Google、IBM、麻省理工学院、中国科学技术大学、浙江大学和初创公司本源量子等在此技术上均有布局。



图 | 全球部分量子芯片初创公司

2021 年底，IBM 公司推出全球首个 127 超导量子比特芯片“Eagle”，突破超导量子芯片百位量子比特的大关。但量子芯片的规模与性能，不仅仅由名义上的量子比特数目决定。逻辑门保真度、退相干时间、量子比特间的连通性等也是评价量子芯片性能的重要指标。接下来，超千位比特的超导量子芯片有望在三年内问世。在进一步提高比特数量的同时保证比特的质量，将是超导量子芯片发展的重点。



趋势 — 量子芯片 —— 航海梯山，无远弗届

作为量子计算的另一种代表性技术路径，光量子芯片因其工作环境要求低、集成度高等特点也受到广泛关注。包括美国 PsiQ、加拿大 Xanadu 和中国图灵量子等在内的光量子集成芯片领跑者，在 2021 年贡献了丰富的研究与产业成果，为光量子计算的商业化奠定了基础。

2021 年 3 月，Xanadu 推出 X8 光量子处理器。这是一款可编程、可扩展、可执行多种算法的光量子芯片，能够集成到现有的基于光纤的电信基础架构中，有效降低运营成本。2021 年 7 月，图灵量子金贤敏领导的团队提出了首个基于光子集成芯片的物理系统可扩展的专用光量子计算方案。2022 年初，图灵量子发布首个商用科研级专用光量子计算机，利用定制化的三维光量子芯片研究和解决各类实际问题。下一步，可靠的高品质光源技术是光量子计算迎来应用爆发前必须克服的难题。

实验室引领了量子计算技术革新，而市场助力了量子计算生态的建立。根据 ICV 统计，2021 年度量子计算硬件及软件公司共完成 42 笔融资，融资总额约 26.91 亿美元，超过以往的总和。

想要实现通用量子计算机需要具备三个前提 —— 百万量子比特的相干操纵能力、低环境要求和高集成度。同时，对于大规模的量子比特系统，通过量子纠错来维持系统的正确运行是必然要求，这也是想要迈入容错量子计算时代所面临的主要挑战之一。目前，量子纠错主要是通过增加逻辑量子比特内自由度的冗余来达到降低净误差的目的。

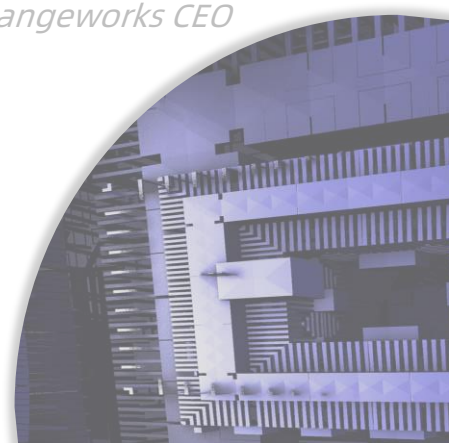
虽然国际学术界一般认为真正实现可编程通用量子计算机还需 15 年甚至更久，但我们有理由相信，在政策的推动、相关产业的升级和资本的不断加持下，规模化与商业化的量子计算将加速到来。

未来，量子芯片将赋能先进计算，为人工智能、密码分析、气象预报、量子化学和药物设计等领域的大规模计算难题提供颠覆性的解决方案。

*“在量子领域，我们正在看到越来越多过去被认为是需要 10 — 20 年才能出现的突破在不到 3、4 年内发生。”**

——William Hurley
Strangeworks CEO

*Source: Author-related. (2022.04). Facing tomorrow's quantum hackers today. MITTR & TII



硅光芯片

2

摩尔的尽头有“光”

硅光芯片兼具微电子与纳米光子的技术优势，既具有微电子尺寸小、成本低、集成度高的特点，又将光子高带宽、高速率、多通道的长处发挥地淋漓尽致，实现了两种物理体系的高效整合。随着云计算与人工智能等应用的爆发，硅光芯片进入了技术与产业的高速发展期。目前，硅光芯片已在大型数据中心内部数据传输等领域迎来商业化落地。未来，硅光芯片与电子芯片将互补共存，深化光电融合的发展模式。



趋势二

硅光芯片——摩尔的尽头有“光”

随着集成电路微电子器件的进一步微缩化，电子芯片在提高算力的同时所面临的能耗和数据传输带宽等问题成为了难以逾越的障碍。光子芯片作为一种在数据运算与传输方面更具优势的集成技术，在后摩尔时代发出闪耀“光”芒。

与电子芯片利用电子作为信息载体不同，光子芯片是利用光子来生成、处理并传输信息。光子是一种玻色子，其静止质量为零，所以在信号传输的过程中由质量的惯性所带来的能量损耗较小，光子器件传输 1 比特信息所需要的能量仅为电子器件的千分之一。同时，光子的信息容量比电子高出 3~4 个数量级，极高的信息密度与抗干扰能力是光子作为信息载体的先天优势。光子芯片能够在信息传输的同时完成计算，在与成熟的硅基半导体加工工艺有机融合后，能够释放光子计算的无限潜力。

半导体集成电路需要将大量微电子原件（如晶体管、电阻、电容等）放在一块基板上，硅光芯片则利用硅的强大光路由能力，通过施加电压产生持续的激光束驱动硅光子原件，来实现光信息的传输、计算等功能。硅基光电子芯片

在微电子技术和光电子技术之间取长补短，既具有微电子尺寸小、成本低、集成度高等特点，又将光电子高带宽、高速率、多通道的优点发挥地淋漓尽致，实现了两种物理体系的高效整合。另一方面，硅光芯片对于工艺制程的要求较低，45 nm 或 65 nm 的 CMOS 工艺线即可满足现阶段光计算的绝大部分要求，这使得硅光芯片在制备流程上具备极高的性价比。

数字化的时代背景下，5G、人工智能、物联网等领域下游应用的兴起导致了数据量的井喷。根据 IDC 的预测数据，2025 年全球数据总量将达到 175 ZB。

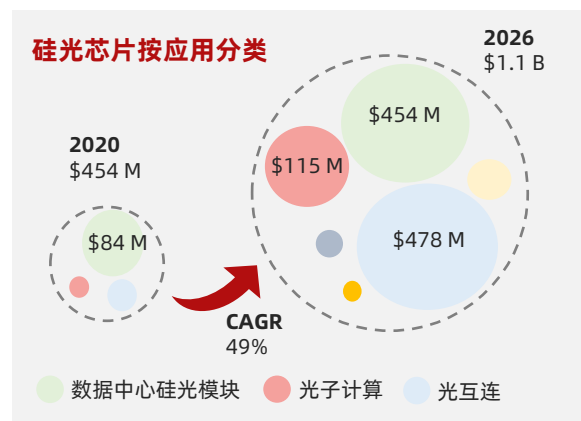


图 | 2020-2026 硅光芯片市场规模 (来源: Yole)

趋势 二

硅光芯片——摩尔的尽头有“光”

立足于光子的传输速率和计算密度等优势，硅光芯片在信息传输与信息处理方面的作用逐渐凸显。

在信息传输方面，数据中心内大规模的分布式计算使得服务器之间需要进行频繁的大量数据交换，数据的访问与搬运性能成为与计算性能一样重要的指标。传统的高速光模块在通信网络设备成本中占比高达 50% - 60%，其高昂的成本阻碍了数据传输速率的发展。而硅光芯片能够在较低的成本下，提供数据中心集群设备间、服务器与服务器间、甚至是芯片内部的超高速光互连。未来，硅光芯片将成为数据中心内超高带宽数据互联的标配。

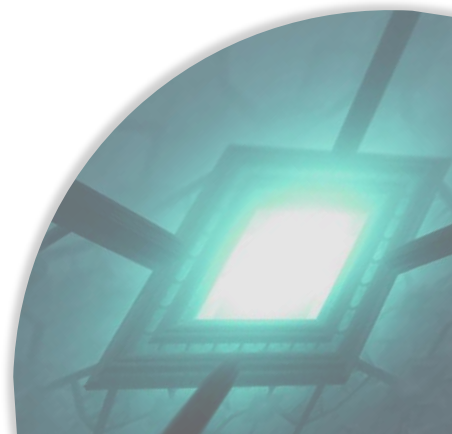
在信息处理方面，硅光芯片具备相比电子芯片更快的矩阵运算速率。光子计算公司曦智科技于 2021 年底发布的第二代高性能光子计算处理器“PACE”，集成了超过 1 万个光子器件，其在特定循环神经网络下的运行速度可达目前最先进的电子芯片的数百倍。人工智能正在从“Model Centric”向“Data Centric”转变，基于海量数据的大模型背后，将会是硅光芯片大放异彩的赛场。

当下，硅光芯片仍然面临着产业链与工艺水平的制约，还未形成系统化的设计与标准化的制造工艺，光学器件的密度有待进一步的提升。并且光计算的精度低于电子芯片，限制了硅光芯片的应用场景。只有打通了设计、工艺、封装、产品等完整的产业链条，硅光芯片才能真正释放性能与成本的巨大潜力。

硅光芯片的商业化是一个持续且漫长的过程。未来 3 年，硅光芯片的商业化产品将首先出现在对算力、延时和功耗有特定需求的应用场景，如量化交易、自动驾驶和大数据中心。未来 10 年，随着技术的不断演进，光计算将会在硅光芯片和电子芯片长期互补共存的基础下走向大规模普及。

“不久的将来，大家可能会看到，服务器中、板卡上、芯片里和计算系统内各个尺度的光电进一步融合。”

—— 沈亦晨 曦智科技创始人



类脑芯片

3

来自大脑的答案

人脑是自然界中最完美的信息处理系统，类脑计算借鉴人脑处理信息的方式，以更少的器件、更低的能耗与更高的能效比颠覆传统计算范式。类脑芯片是类脑计算落地应用的硬件载体，具备“事件驱动”与“存算一体”等特性。未来，人工智能的发展必将从狭隘的弱人工智能走向更具鲁棒性与通用性的通用人工智能。脑科学与计算机科学的深度融合，将为通用人工智能的实现提供必要的条件。



趋势 三

类脑芯片 —— 来自大脑的答案

人工智能时代，存算分离的冯诺依曼架构逐渐显现瓶颈。人脑是自然界中最完美的通用智能体，类脑计算借鉴人脑处理信息的方式，以更少的器件、更低的能耗与更高的能效比解决冯诺依曼架构所面临的“内存墙”与“功耗墙”等诸多挑战。

人工神经网络，是基于统计学递归原理所构建的计算架构。而类脑计算是基于借鉴脑神经元工作原理的新型架构，是一种时域加空域的计算，在具备稀疏性的同时能够有效体现人脑处理信息时候的高效性、近似性与时空连续性。

类脑芯片，是类脑计算落地应用的硬件载体。类脑芯片具备“事件驱动”的特性，即只有输入信号达到一定的阈值，神经元才会被激活并发出脉冲，通过人工突触将信号传递到下一个神经元。在输入信号不足以刺激神经元发出脉冲时，其他非活跃神经元自身能够保持极低的运行功耗。与此同时，类脑芯片还具备“存算一体”的优势，能够很好地满足边缘计算领域低延时场景的需求。目前，类脑芯片存在基于数模混合电路、异步数字电路和同步数字电路等多种技术方案，类脑计算的研究也正处于飞速发展期。

在实验室领域，包括斯坦福大学、曼彻斯特大学、清华大学以及浙江大学在内的世界知名高校均在类脑计算方向有所建树。而产业界中，IBM、Intel、阿里巴巴等大型科技公司和众多初创公司也正在积极布局类脑芯片。

2019年8月，清华大学自主研发的异构融合类脑芯片“天机芯”登上《Nature》杂志封面。“天机芯”总共可实现4万个神经元，兼容人工神经网络与脉冲神经网络，能达到普通AI芯片10倍以上的能效比。目前，“天机芯”芯片已完成从科研到产业应用的转变。2021年9月，Intel发布最新一代类脑芯片Loihi 2。该芯片在非激活状态下能够实现毫瓦级别的功耗。Loihi 2作为一种研究型芯片已被应用在机械臂、嗅觉传感、人造皮肤等领域的研究上。



图 | 全球类脑计算研究部分顶尖高校

趋势 三

类脑芯片 —— 来自大脑的答案

类脑芯片在实验室内已经取得了重要技术突破，而在市场上也能够看到前沿的应用落地。智能感知终端设备由于需要实时处理多模态、非结构化的数据并做出正确的决策，是类脑计算施展拳脚的绝佳场景。目前，已有类脑芯片企业针对智能机器人、智能安防等具体应用对生物神经网络进行优化，并将算法变成芯片架构，提供定制化的类脑智能解决方案，推动了类脑芯片的产业化与商业化落地。

当下，类脑芯片仍面临着来自材料体系的难题。目前能够实现量产的类脑芯片，基本还是采用电路模拟人工神经突触的方式来进行类脑计算，这样做不仅对芯片制造工艺要求较高，生产效率也十分低下。理想的类脑芯片材料，应该既具备在器件层面模拟神经元和突触的功能，又能够于系统层面实现大规模的集成。努力探寻满足上述条件的材料，是类脑芯片实现突破的关键卡口。

尽管类脑计算行业还处在突围阶段，其市场前景仍受到广泛的认可。根据 Yole Développement 预测，2035 年类脑计算市场将占人工智能市场总收入的 15%-20%，市场规模约 200 亿美元。

长路漫漫，始于足下。从短期来看，类脑芯片需要以端侧智能为利基市场，凭借自身超低功耗、高能效比的优势，挖掘和创造更多的应用场景。荆棘载途，云程万里。从长期来看，类脑芯片想要迎来商业价值的释放，不仅需要产业端关注制造成本的降低，更需要科研端从系统结构入手，发展类脑计算完备性，拓展类脑芯片应用于通用计算的可能性。

未来，人工智能的发展必将从狭隘的弱人工智能走向更具鲁棒性与通用性的通用人工智能。类脑计算是通往通用人工智能的必经之路，但绝不是独木桥。只有将计算机的刚性与脑的弹性进行有机结合，在现有的硅基计算机架构基础上载以具备时空复杂性的类脑芯片，才能引领人工智能的下一轮变革。

“人工通用智能是一个必然的趋势，而异构融合类脑芯片正是这把钥匙。”

—— 施路平 清华大学教授



DPU

4

以数据为中心的计算

随着云计算的普及，数据中心在快速扩容的同时也面临着来自性能、规模、能耗与安全的问题。为更好地解决数据中心的瓶颈并实现云服务的降本增效，DPU（数据处理芯片）替代传统的智能网卡，成为了完美的解决方案。数据是贯穿计算、网络和存储三要素的一根主线。未来，DPU、CPU、GPU将成为大型数据中心内的三驾马车，DPU 作为整个数据中心的大脑，通过数据流链接并管理着整个计算系统。



趋势 四

DPU —— 以数据为中心的计算

过去 20 年，智能化和数字化的浪潮加速了云计算与数据中心的发展。作为信息时代的 IT 基础设施，数据中心在快速扩容的同时也面临着来自性能、规模、能耗与安全的问题。

一方面，数据中心的计算资源被复杂的基础设施操作所占据，留给业务处理的计算空间正在逐渐缩减。数据显示，处理 25G 带宽的网络服务，将耗费 CPU 约 20% 的算力，而当带宽变为 100G，这个数字将提升到 70%。另一方面，服务器集群的规模越来越大，如何高效地统筹各服务器间的计算和存储资源是数据中心想要实现降本增效必须克服的难题。

针对上述问题，从智能网卡演变而来的 DPU 成为了共识的解决方案。DPU 集合了高性能的网络接口与灵活可编程的多核 CPU，能够实现算力卸载、降低网络拥塞、网络虚拟化和安全加速等多种功能，为数据中心业务与基础设施操作的分离提供了强有力的支持。

DPU 最根本的功能是对 CPU 任务的卸载。DPU 将数据中心基础设施操作中占据 CPU 资源较多的网络、存储、虚拟

化和安全等任务都卸载并集成到一个平台。在此基础之上，CPU 算力能够被最大规模地释放给上层应用，从而提升可供售卖的计算资源，为数据中心带来经济效益。

DPU 另一个重要的使命是对数据中心进行集群管理。DPU 作为数据中心的大脑，利用快速的数据迁移能力实现基础设施硬件的池化，包括算力的池化、存储的池化以及将来内存的池化。池化后的资源能够得到更高效的利用和管理，从而有效降低数据中心的 TCO（整体拥有成本）。与此同时，DPU 作为数据处理和运维的中枢，负责将不同的数据任务调度给最合适的 CPU、GPU 或 DPU 自身进行处理，加速了异构计算到来的步伐。

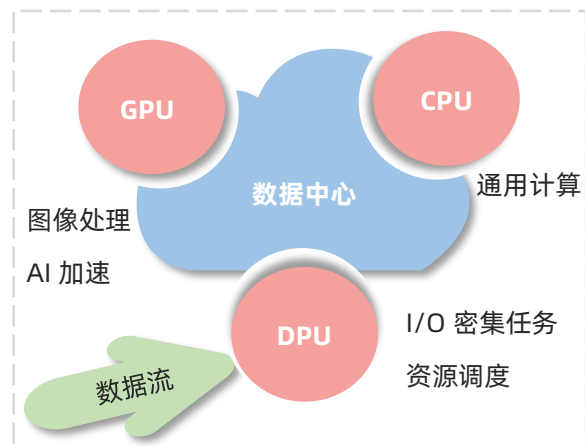


图 | CPU、GPU、DPU 成为数据中心“三驾马车”

趋势 四

DPU —— 以数据为中心的计算

根据英伟达的预测，未来用于数据中心的 DPU 数量将达到与数据中心服务器数量同等的级别，DPU 的全球市场规模有望在三年内超过 100 亿美元。DPU 广阔的市场空间，也吸引了世界范围内的众多玩家。DPU 领域的玩家主要分为三大类。第一类是包括亚马逊、谷歌、阿里巴巴和华为在内的大型云服务厂商；第二类是英特尔、英伟达和博通等传统半导体巨头；第三类则是以美国 Fungible、Pensando（被 AMD 以 19 亿美金收购）和中国头部企业云豹智能为代表的 DPU 初创公司。

当下，DPU 芯片的核心竞争要素是性能与可编程能力的极致平衡。想要打造算力大、通用可编程的 DPU，必须深入了解云服务厂商的需求与数据中心的应用场景，根据需求与场景去抽象和定义产品，在带来较低的迁移成本的同时也敏捷地支持业务变化。作为中国 DPU 赛道的领头羊之一，云豹智能将包括 DPU 芯片在内的整板功耗控制在 75W 以下，为数据中心裸金属、虚拟机和容器应用场景提供一体化的 DPU 解决方案，在成本降低的同时获得数倍以上的性能提升。未来，“应用定义软件、软件定义芯片”的软硬件融合思路将贯穿芯片设计的全流程。

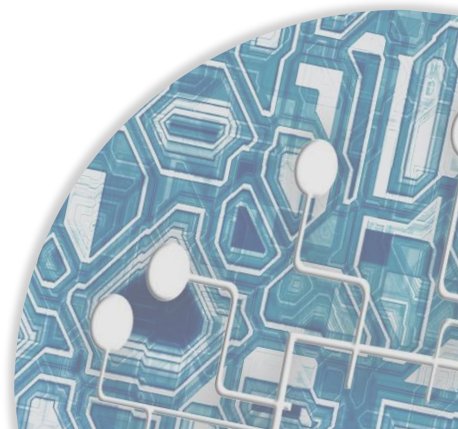
除了在数据中心和云计算领域发光发热，DPU 在未来会承载更加丰富的应用场景。包括网络安全、通信、边缘计算和数据存储等都将是 DPU 的目标市场。

如果说 CPU 是“以计算为核心，指令控制流驱动计算”的平台，那么 DPU 则是“以数据为中心，数据流驱动计算”的平台。数据中心正在从“CPU Centric”的时代走向“Data Centric”的时代。数据是贯穿计算、网络和存储三要素的一根主线。数据中心内部，DPU 好比交通信号灯，通过数据流链接并管理着整个系统，从而实现更高速率和更低延时的业务响应。

算力不仅需要提升，更需要变得可以驾驭。DPU 的兴起正推动着计算基础架构的变革，为更高效的数据处理和计算解决方案铺平了道路。

“未来，云服务厂商的创新与优化都将集中在 DPU 上。”

—— 萧启阳 云豹智能 CEO



分布式云

5

云边端协同的关键枢纽

分布式云靠近数据的源头，搭载在数据中心与终端设备之间的边缘基础设施之上，能够提供低延时、广连接、大带宽的云服务。随着硬件设备的发展，分布式云节点将具备部分中心云的能力，但并不能完全取代中心云，而是以云边端深度协同的形式构建更加完备的云计算架构。未来，分布式云将服务于自动驾驶、工业制造、智慧城市等对时延和连接有苛刻要求的应用，释放更多边缘智能的价值。



趋势五 分布式云——云边端协同的关键枢纽

从电子计算机的诞生到客户端电脑的普及再到云计算的兴起，计算架构经历了从集中式到分布式再到集中式的演变。5G 与 IoT 浪潮之下，智能终端数量及其产生的数据总量呈现指数级别的增长，集中式的云计算难以满足带宽、时延与成本的需求，算力架构开始再次向分布式迁移。

分布式云搭建在数据中心与终端设备之间的边缘基础设施上，是云计算能力向边缘节点下沉的表现。分布式云靠近数据源头，广泛分布在不同地理位置，覆盖各种数据热点区域和客户场景，就近提供计算、网络、存储、安全等云能力。基于以上物理层面的优势，分布式云是一种满足广连接、大带宽、低延时和碎片化需求的精细化云服务。

分布式云作为中心云的延伸，承担着“神经末梢”的计算任务。与物流配送“最后一公里”类似，分布式云最显著的特点就是能够针对计算任务提供敏捷的响应。由于终端设备与大型数据中心地理距离的限制，传统云服务难以进一步降低如自动驾驶和虚拟现实等应用的响应时延。由于分布式云更靠近边缘侧，能够满足实时数据分析与处理的数据产生和使用的需

求，从而为以传感器网络为基础的各类应用提供了可靠的计算基础。随着分布式云覆盖密度的提升，更多的边缘智能与端侧智能将成为现实。

分布式云能够提供更高效和安全的服 务。中心云的模式下，大量终端设备的数据需要回传到云端，极易造成网络通道的堵塞和带宽容量的浪费。而分布式云的架构，每一个计算节点都有能力承担数据预处理的工作，从而实现数据的分层传输，减缓大规模并发对于骨干网络的冲击。另一方面，由于缩减了数据的传输链路，分布式云能够减少传输过程中数据隐私泄露的风险，并满足特殊客户关键数据不离场的需求。未来，数据安全相关法律法规的实行将使得分布式云成为企业的必要选择。

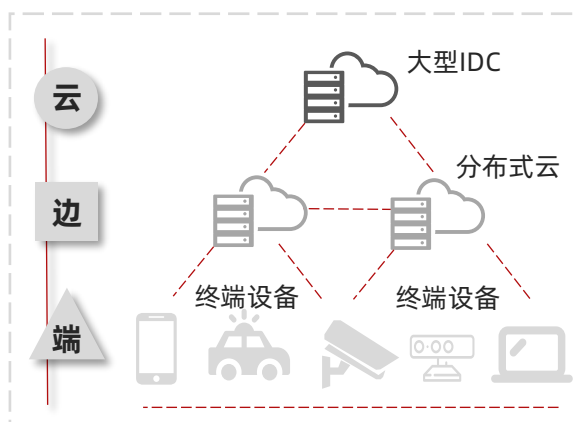


图 | 云-边-端协同架构

趋势五 分布式云——云边端协同的关键枢纽

面对差异化的计算需求，分布式云能够基于物理终端的位置、计算节点的位置和算力系统的状态，实现算力的智能调度与分配，在保证业务响应低时延的同时有效提升计算资源的利用率。云资源的智能调度依托于强大的基础设施软件，专注于软件定义边缘的分布式云厂商秒如科技，正是凭借全栈自研的异构调度引擎，得以提供强大的跨云、弹性可伸缩的计算资源调度和编排。从需求侧来看，海量应用促使全球算力需求不断攀升。从供给侧来看，算力网络的结构会随着分布式云的大规模普及而呈现颗粒化、下沉的趋势。基于供需两端的发展情况可以判断，未来的计算资源将是无处不在且位置无感的。

不同于中心云是资源定义的云服务，分布式云是业务场景定义的云服务。近几年，除了数据中心的成本和效能等因素，云服务厂商的核心竞争将集中在用户体验之上。分布式云基于中心云所不具备的低延时、低带宽成本等优势，能够提供给客户更优质的云体验。进入快速成长阶段，分布式云的进一步发展依托于更多上游应用场景的落地，而分布式云厂商竞争的关键在于对客户和业务场景理解的深度。

当下，分布式云仍然还面临着部署成本、软硬件异构管理、安全体系等多方面的难题。下一个五年，在突破上述瓶颈后，分布式云能够推动全面云化的进程。根据 Gartner 预测，到 2025 年将有超过 50% 的组织在其选择的地点使用分布式云，从而实现业务的转型与升级。

随着硬件设备的发展，分布式云节点将具备部分中心云的能力，但并不能完全取代中心云，而是以云边端深度协同的形式构建更加完备的云计算架构。作为云边端协同的关键枢纽，分布式云需要在业务、应用管理、数据和资源等方面与中心云实现协同，并在应用层面为自动驾驶、工业制造和智慧城市等场景下的终端设备带来实时智能互连。未来，分布式云将推动边缘计算、分布式云原生、分布式数据库、分布式存储等行业分布式应用的革命。

“企业 IT 基础设施的发展在经历了从硬件定义到软件定义再到服务定义的演进过程后，无不处在的分布式云将成为未来最为关键的 IT 基础设施。”

——孙思清 浪潮云首席技术官

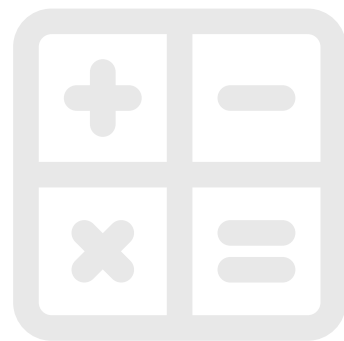


泛在算力

6

无处不在的算力网络

在数据爆发式增长与算力单点性能极限之间的矛盾日益突出的背景下，以网络连接泛在算力，构建云边端一体的立体算力网络迫在眉睫。未来，算力将呈现泛在分布、多元异构、位置无感的趋势。随着网络技术与计算技术的并行发展、算网的深度融合，智能可调度的泛在算力网络将加速到来。下一个 10 年，算力将如同水电煤等基础资源，渗透进人类生产生活的点滴，成为智能社会运行的核心要素。



趋势六

泛在算力 —— 无处不在的算力网络

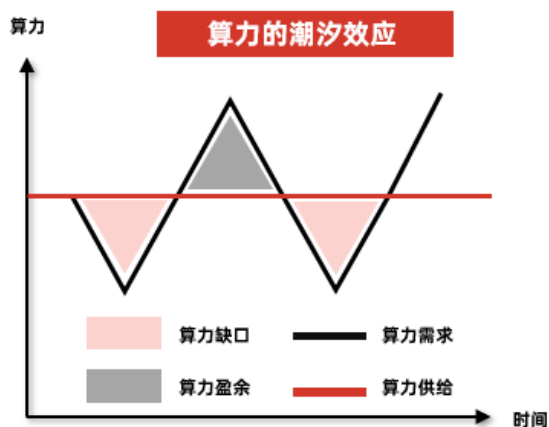
先进计算与网络技术的发展，推动了智能世界万物互联的进程。未来社会，智能化设备无处不在，算力也将无处不在。迈入算力时代，算力将突破云的边界，向万物蔓延，呈现出云边端立体化泛在分布的趋势。泛在算力，将如同工业社会里石油的角色一样，成为智能社会的基石。

进入后摩尔时代，受到性能、成本和功耗等因素的影响，单个芯片算力的提升逐渐放缓。与此同时，多核芯片数量及性价比也趋于极限，推动了算力从单点向网络化阶段的过渡。在算力需求不断增长的背景下，网络化的算力能够弥补局部算力的缺口和算力供给不均衡的现象。通过网络连接泛在算力，突破单点算力的性能极限，成为打破算力危机的关键。

网络化的算力并不仅仅是简单地将云端数据中心与数据中心之间、数据中心与算力需求端之间互相连接起来，而是构建云边端一体的多层级算力架构。在面对一些如自动驾驶、智慧安防等有着低时延、大带宽算力需求的场景时，网络信道的传输质量影响了数据传输的效率，从而限制了云端算力的全量输出，产生算力供给与算力需求间的错配。为了解决上述问题，

将部分数据处理工作从云端迁移至边缘端，引入广泛分布的边缘算力，是建设网络化算力体系的支柱。边缘算力的引入，不仅能对计算需求起到“削峰填谷”的作用，显著提升算力网络的效能，更能有效节省网络带宽成本，提升算力网络的经济性。未来，云端算力是主引擎，边缘算力是涡轮增压器，终端算力是倍增器，共同打造无处不在的算力网络。

泛在算力是多元异构且位置无感的。除通用计算外，随着高性能计算和人工智能加速器的兴起，算力的内核不断朝着GPU、FPGA、DPU等多元异构化的方向演进。算力网络将通过统一的基础设施层



资料来源：华为《泛在算力：智能社会的基石》

趋势 六

泛在算力 —— 无处不在的算力网络

对多元化的硬件进行管理和编排，并对外提供多样性算力的统一供给。同时，算力网络的智能调度能力将淡化算力供给来源的意识，让算力需求方更专注于上层应用的执行效果，实现算力的处处可用、计算资源的位置无感。算力应用与算力基础设施的解耦，将会催生更多智能应用和智慧场景的诞生。

根据华为预测，到 2030 年，全球数据量年新增 1 YB，通用算力增长 10 倍达到 3.3 ZFLOPS，人工智能算力增长 500 倍超过 100 ZFLOPS。经历过“泡沫之巅”与“绝望之谷”，AI 已经迈入了稳步发展的阶段。AI 是下一个十年的时代主旋律，泛在的算力能够为人类带来无处不在的感知和无所不及的 AI。未来，AI 将在算力引擎的带动下，走向通用化、普惠化，持续在智慧城市、智慧医疗、智能制造和科学研究等领域释放价值。

在通往泛在算力的道路上，仍然存在着诸多维度的挑战。其中，网络技术是需要面对的第一座大山。当前各算力节点通过网络灵活高效调配算力资源的能力仍存在不足，算网协同和全局调度能力较弱。网络技术的发展需要追赶计算技术的步伐，

进一步提升网络链接的实时性、可靠性与承载能力，为实现“算网融合”打下坚实的基础。再者，不同算力节点的标准机制挑战和系统的安全可信挑战也不容忽视。终端、边缘端及云端间海量数据交互的安全与隐私问题如何保证，来自于广泛分布的边缘算力节点的网络攻击如何预防，不同节点的算力体系与管理标准如何拉通，将是构建开放、高效、可信的泛在算力网络必须要解决的症结。

下一个 10 年，算力将如同水电煤等基础资源，渗透进人类生产生活的点滴，成为智能社会运行的核心要素。泛在的算力为各行各业带来的效率提升能够释放出新的生产力，从而推动社会经济的发展。机器与设备因泛在算力网络的存在获得更高的输出效能，人类的工作负担将得到进一步的减轻，从而有机会从事更多富有创造力的工作。泛在算力是大势所趋，虽道阻且长，但行则将至。

“算力已然朝着网络化的方向演进。未来，算网融合下的泛在算力将成为推动数字经济与智能社会发展的主引擎。”

—— 王晓东 天翼云解决方案部总经理



科学复兴



先进计算加速科学突破

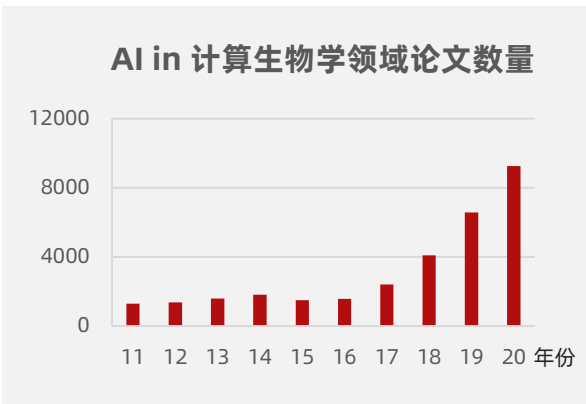
科技发展呈螺旋上升之势，永不停滞。底层学科的突破不断扩张着人类认知的边界。在知识的版图之上，科学家与工程师们建造了科技应用结晶的高楼。高楼之下，正是计算技术灌溉的基石。站在高楼之上，人类得以再次瞭望更广阔的远方。计算能力的每一次革新都助推着科技的发展。迈入算力时代，以量子计算、光计算为代表的先进计算将为人工智能等工具搭载全新的引擎，加速人类科学复兴。



趋势七 科学复兴——先进计算加速科学突破

科学的突破不仅仅需要敏锐的观察力、丰富的想象力、严谨的推理能力，更需要强大的计算能力。从手稿纸上基于人类智力的演算，到图灵机代替人脑进行机器运算，再到基于大数据的机器学习和逐渐工具化的量子计算，人类每一次计算能力的革新都助推着科技的发展。迈入算力时代，以量子计算、光计算为代表的先进计算技术将为人工智能等工具搭载全新的引擎，加速人类科学复兴。

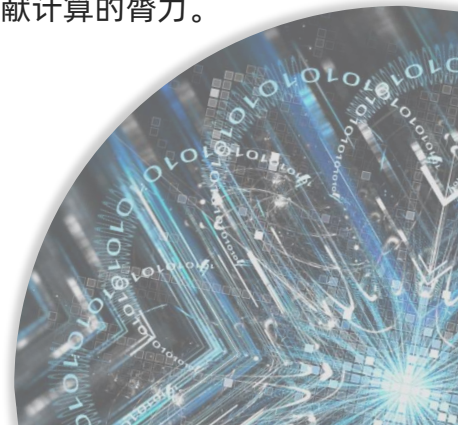
过去二十年，随着算力的不断提升，人工智能系统正变得足够强大。AI 除了可以应用于解决现实世界的问题，更能为科学发现本身带来突破。当下，人工智能已经成为科学研究人员的有效工具，开启了众多全新的科学探究领域，并不断促成跨学科的合作。



数据来源：Web of Science

如果说数学是物理学的正确描述语言，那么 AI 在生物学中也扮演着类似的角色。由于长链氨基酸的三维扭曲、折叠和交织，蛋白质结构预测成为困扰生物学界 50 余年的难题。2022 年 12 月，人工智能公司 DeepMind 发布 AlphaFold2，这是一种能将蛋白质的形状预测到接近原子尺度的 AI 工具。AlphaFold2 的预测结果，帮助了人类进一步阐明蛋白质的作用，也为了解癌症和新冠肺炎等重大疾病提供了宝贵的参考。

虽然 AlphaFold2 预测人类蛋白质的覆盖率很高，达到了 98.5%，但在某些生物制药或疾病治疗领域，对氨基酸的结构精度需要真正达到原子级别。从这个角度看，AlphaFold2 还有很多的提升空间。与此同时，AlphaFold2 模型在运行时，仍需要耗费几个小时甚至一整天。未来，随着 AI 加速器的不断进步和先进计算技术的引入，人工智能底层的算力将会迎来更大的腾跃，彼时的 AlphaFold2 将会以更高的功率、更精的准确度，帮助人类扩张已知蛋白质的数据库，并为靶向药物的快速研发和迭代贡献计算的算力。



趋势七 科学复兴——先进计算加速科学突破

除了应用于生物学领域，AI 在物理和数学的突破中也扮演着重要的角色。2022 年 2 月，DeepMind 公布 AI 应用于核聚变的成果，提供了一种全新的控制等离子体的智能方法。虽然距离核聚变的商业化还至少需要 20 年，但 DeepMind 的尝试无疑加快了通往可控聚变反应堆的进度。机器学习工具也被多位科学家证实能够帮助计算数学常数和发现数学规律。

二十一世纪以来，量子计算处在高速发展期，而量子计算机的能力也已经达到了可以用来进行广泛技术研究的阶段。来自美国多所知名大学的 Q4Q 团队就一直致力于用量子计算加速科学研究。

Q4Q 团队的目标是在量子计算发展的基础之上，利用量子算法和软件包来重建经典计算机难以实现的物理系统的量子模拟。量子模拟能够为预测分子和固体的电子和振动光谱提供前所未有的准确性，以此在能源科学、制药科学、材料制造等领域产生巨大影响。目前，团队研究的方向包括固体的能带结构、分子电子/振动的特性等。接下来，Q4Q 团队将进一步探究周期性系统的量子模拟，推动量子计算在选定材料和分子性质方面的应用。

先进计算不仅能显著提升科研的效率，更是科学领域创造力的倍增器。科学突破离不开人类的创造力，而创造力的本质是对知识的重新组合。大脑的时空复杂性赋予了人类创造力，使得我们能够在无数的组合中抓住灵光一现。面向更复杂、更多维信息的先进计算技术，能够用秩序驯服混沌，让思维的源泉从一川从入海走向百川赴海，进而加速人类知识海洋的聚积。

当“Computing for Everything”从口号变为现实，当“无穷大谜题”被解开，当创造力也可以被创造...人类终将坐上算力的相位加速器，在科学复兴的时代背景下，努力追赶宇宙膨胀的光锥。

*“我们将看到一种全新的科学复兴，这些 AI 技术将继续变得更加复杂，并被应用到广泛的科学领域。” **

—— Demis · Hassabis
DeepMind CEO

*Source: Will Douglas Heaven. (2022.02). DeepMind's AI can control superheated plasma inside a fusion reactor. MITTR



致谢

冯黎

秒如科技 CEO

孙思清

浪潮云首席技术官

高天垚

联想之星合伙人

王光熙

联想创投合伙人

华宝洪

灵汐科技 CMO

王晓东

天翼云解决方案部总经理

孔令国

和利资本创始管理合伙人

萧启阳

云豹智能 CEO

乔宁

时识科技 CEO

杨光

耀途资本创始合伙人

沈亦晨

曦智科技 CEO

*感谢上述学者及行业人士接受访谈调研；按照姓名首字母排序。

About Us

DeepTech 成立于 2016 年，是一家专注新兴科技的资源赋能与服务机构，以科学、技术、人才为核心，聚焦全球新兴科技要素的自由链接，为产业、政府、高校、科研院所、资本等科技生态的关键角色提供服务，通过科技数据与咨询、出版与影响力、科创资本实验室三大业务板块，推动科学与技术的创新进程。

About The Report

作为智能社会的底层、数字经济的引擎，算力的重要性已经被提到一个前所未有的高度。当下，算力的提升仍然面临着来自多个维度的挑战。不论是硬件层面还是架构层面，算力发展都亟需变革。本报告围绕计算能力与效能的提升，通过科研数据分析、专家访谈等方式洞悉计算领域发展趋势，探寻具备技术颠覆性、有商业化前景的先进计算技术，并提炼出 2022 年先进计算技术及应用七大趋势，以飨读者。

先进计算技术涉及众多领域，许多技术仍处在产业化和商业化的早期阶段。此篇报告希望以技术观察者的角色提供给科研工作者、资本及产业链相关企业一个前沿的视角。

Please use the following to reference the report:

《破局摩尔桎梏，决胜算力时代 —— 2022 先进计算七大趋势》，2022. DeepTech 2022 Insights. China.

Disclaimer

本报告由 DeepTech 发布，其版权归属北京演绎科技有限公司（DeepTech），DeepTech 对此报告拥有唯一著作权和解释权。没有经过 DeepTech 的书面许可，任何组织和个人不得以任何形式复制、传播等。任何未经授权使用本报告的相关商业行为，DeepTech 将依据中华人民共和国相关法律、法规追究其法律责任。

本报告所载数据和观点仅反映 DeepTech 于发出此报告日期当日的判断。DeepTech 对报告所载信息的准确性、完整性或可靠性做尽最大努力的追求，但不作任何保证。在任何情况下，本报告中的信息或表述均不构成任何投资等建议，本公司对该报告的数据和观点不承担法律责任。不同时期，DeepTech 可能会发布其它与本报告所载资料、结论不一致的报告。同时 DeepTech 对本报告所载信息，可在不发出通知的情形下做出修改，读者应自行关注。

Find Out More

<https://www.deeptechchina.com/>

Contact Us

research@deeptechchina.com

Head Office

中国北京市朝阳区建国路甲92号

Shimao Tower, No. 92 Jia Jianguo Road, Chaoyang District,

Beijing 100022, China

DEEPTECH

SCIENCE

×

TECHNOLOGY

×

TALENTS



相信科学的力量

IN SCIENCE WE TRUST

