



《泛在算力：智能社会的基石》

立场文件

2020 年 2 月



拥抱泛在算力， 拥抱智能社会的美好未来

梁华
华为公司董事长

计算是人类认知世界的一种模式。从大型机到个人计算机，从智能手机到可穿戴设备，计算能力日益成为人类能力的延伸。同时，计算模式也在不断演进，随着人工智能的发展，基于统计的计算模式将逐渐成为主流。预计 5 年后，人工智能所消耗的算力，将占到算力消耗总量的 80% 以上，计算产业正在进入一个新的智能时代。

计算和联接技术的蓬勃发展，推动智能世界万物互联已成为不可逆转的趋势。未来社会将是万物感知，万物互联以及万物智能。同时“泛在算力”将是基础。要实现“泛在算力”的这个目标，需要行业和政府达成共识和共同努力。

首先，**泛在算力是智能社会的基石**。正如人均 GDP 是衡量一个国家经济发展程度的重要指标一样，我们采用人均算力指标对主要国家进行了测算：根据目前各国算力发展情况，从低算力国家的 100 GFLOPS/人 到高算力国家的 2,500 GFLOPS/人 不等。但即便是高算力国家，当前也还处于智能社会的初级阶

段。只有当人均算力发展到 10,000 GFLOPS/人 时，才会进入智能社会的中级阶段。如同电力的普及奠定工业社会发展的基础那样，泛在的算力将成为智能社会发展的基石。

其次，**构建多元算力生态，能有效促进计算产业发展**。智能社会的应用场景多样性和数据类型的多样性对算力提出了多元架构的诉求。要构建繁荣的多元算力生态，既需要底层架构创新，也需要客户、行业伙伴、开发者携手合作，以加速生态的建设，为计算产业发展注入动力。比如，由多家国际领先服务器与云计算方案提供商以及中国高校、科研机构共同成立的“绿色计算产业联盟”，为算力架构多元生态的发展带来了显著产业资源与人才支持优势。再比如，欧洲处理器计划（EPI）作为欧洲处理器计划联盟（EPI Consortium）推动的落地计划之一，截至 2020 年 1 月，汇集了来自 10 个欧盟国家的 27 家合作伙伴，覆盖了算力芯片研发、生产、应用场景等多个产业环节，通过产业联盟的形式推动欧洲算力产业的健康发展。

最后，在算力上的投入将推动科技创新，促进经济发展，提高民生福祉，进而带动国家整体竞争力的提升。通过多个行业的案例分析发现，在算力上的投入，不仅直接带来 ICT 行业的增长，还对制造、交通、能源、零售、农业等诸多行业带来创新改变，推动经济增长。以制造业为例，在传统工厂改造为智能化工厂的场景下，算力上每 1 美元的投入，可以带动 10 美元的相关产值提升。算力投入还可以改善民生，如在教育、科研、医疗等领域的投入可有效提高居民的生活质量和幸福感，并进而提升国家的整体竞争力。因此需要把算力投入提升到战略高度进行总体规划和实施。

以中国为例，我们今天已经享受了泛在算力发展的成果，从繁忙的都市，到偏远的乡村，每天有将近 9 亿人在使用便捷的移动支付。移动支付的每一笔交易，都需要无处不在的联接和强大的算力支持：数以亿计的移动设备随时随地接入网络，发起支付交易，从扫码就餐到乘车出行，从商场购物到网上消费，都可以瞬间完成，极大的提升了支付体验，并有效保障了支付安全。

过去 30 年，基于对研发的大规模投入和持续创新，华为致力于打造世界上最好的联接；而随着智能世界的加速到来，我们从联接扩展到联接 + 计算，推动计算产业多元化发展，为世界提供充裕而经济的算力，并不断为客户创造价值，为社会发展做出贡献。

未来 30 年，随着“泛在算力”的不断实现，人工智能将全面进入千行百业，触发行业变革；人工智能还将给每个家庭、每个人带来创新的体验，进而深刻地改变我们的生活方式，人类社会也终将进入高度发达的智能社会。

千里之行，始于足下。要最终实现“泛在算力”的美好未来，需要行业的每一个人都全力参与，躬身入局。华为也愿意与计算产业的合作伙伴一起，为实现“泛在算力”而努力！

目 录

执行摘要	02
第一章：算力的定义及发展趋势	05
1.1 算力的定义及历史发展	05
1.2 由云－边－端组成的泛在算力架构是技术发展至今的必然选择结果	07
1.3 随着社会的智能化进程，对泛在算力提出了 4A 的具体要求	14
第二章：算力衡量指标体系	26
2.1 算力衡量指标体系介绍	27
2.2 算力指标衡量结论	35
2.3 随着社会智能化进程的演进，对算力水平的需求呈指数级增长	38
第三章：算力对各国家 / 区域的效益分析	47
3.1 对算力的投入可以带来可观的直接性经济回报	47
3.2 对算力的投入可以对众多行业带来极大的杠杆效应，创造巨大的延伸性经济增长	48
3.3 此外，对算力的投入亦能从创新、公共服务、民生等三方面提升国家整体竞争力	51
第四章：算力发展面临的挑战与政府的建议举措	54
4.1 算力的发展面临来自于功耗、网络和生态的现实挑战，并需重视基础设施建设及安全问题	54
4.2 各国政府可采取规划倡议、设施 / 服务提供等抓手，针对性的推动算力发展	62

执行摘要

从中国古代依靠人力计算的算盘，到当下无处不在的智能设备，算力经过了不同阶段的演变，正向着愈发性能强大和生态多样化的未来迈进。在芯片时代，算力的发展经历了单核、多核、网络化三大阶段：首先，由于技术和商业化可行性低，单核硅基芯片计算能力将在 3 纳米到达极限。与此同时，由于算力散失效应带来的多核堆叠性能提升受限，以及核心数增加带来的单位算力功耗提升，共同使得依靠核心数增加换取算力提升的步伐将在 128 核由于经济性原因迅速放缓，推动算力从单设备多核阶段向网络化阶段演进。而在智能社会对算力需求始终不断增长的大背景下，受限于网络技术发展及网络带宽成本，边缘端的算力部署将成为必然的演进方向，来填补算力需求的错配，最终形成云-边-端泛在算力部署的理想架构。

而从需求侧来看，未来人类社会的智能化进程不可避免地需要人工智能、物联网、区块链、AR/VR 等核心技术加持，以多样化的智能应用场景改善生产水平、民生福祉、政务效率，而这些技术的应用亦将成为未来算力发展的核心驱动力，带动对计算设备（如云计算中心、边缘计算设备、高算力终端）的巨大需求。其中人工智能将全面赋能各行各业，预计到 2030 年将需要超过 16,000 EFLOPS¹ 的算力支撑，相当于近 1,600 亿颗当下旗舰手机芯片（骁龙 855）NPU² 所能提供的算力；物联网技术在工

厂和家居环境的应用拉动云端算力和终端算力提升，2030 年需求将超 8,500 EFLOPS，相当于 79 亿台当下高端物联网边缘设备所用芯片（EPYC 7401）提供的算力；区块链技术在数字货币等领域的应用拉动全网算力提升，在 2030 年发展到超过 5,500 EFLOPS 的算力，相当于 13 亿台当下入门级矿机（蚂蚁 V9）的算力总和；AR（增强现实）/VR（虚拟现实）技术拥有极大的增长空间，通过需求拉动和普及的方式拉动算力增长，2030 年对于算力的需求将超过 3,900 EFLOPS，相当于 21 亿台当下顶级游戏机所（SONY PS4）提供的算力总和。而这些需求的实现将共同对算力形成随时、随地、随需、随形 (Anytime、Anywhere、Any Capacity、Any Object) 的能力要求。

而为了对算力拥有更为科学、具象的认识，我们设计了算力衡量指标体系，以有效衡量各国的算力水平。在算力的具体衡量方式上，由于算力的物理承载集中在云端、边缘端、终端三个环节，算力水平应基于云、边、端各层面典型设备的产品形态进行测算。在云端侧主要关注超级计算机与数据中心的算力水平，在测算其服务器数量及平均芯片计算能力的基础上综合考虑多核心损失效应以及网络传输损失效应；在边缘侧主要考量边缘服务器的算力水平，以 CDN³、智能网关等部署方式作为牵引，衡量各场景下的边缘服

¹ EFLOPS（exaFLOPS）为每秒浮点运算，是对计算机性能的一种衡量方式。1 EFLOPS = 10^9 GFLOPS = 10^{18} FLOPS

² Network Processing Unit (NPU) 是为人工智能应用，特别是人工神经网络、机器视觉和机器学习而设计的一类微处理器或计算机系统。

³ CDN 的全称是 Content Delivery Network，即内容分发网络。CDN 是构建在现有网络基础之上的智能虚拟网络，在基于空间分布的基础上致力于提供终端客户更好的网络服

务器数量及平均算力，综合考量网络传输损失效应；在终端侧主要考量各类智能设备的算力水平，以智能手机、笔记本电脑、智能可穿戴设备等形态作为牵引，衡量其数量及平均算力水平。而通过对全球 27 个国家基于以上方法的测算，算力水平的高低与一国的经济发展水平形成了紧耦合的绑定关系，是综合国力发展的重要表现。更重要的是，算力是未来智能社会阶段性演进中的重要标志。在智能社会的起步阶段，各产业的智能场景覆盖率较低，同时智能设备在产业内的应用刚刚起步，仅能实现自动化生产等初级应用，目前主要国家 / 地区的人均算力⁴预估在约 100-2500GFLOPS⁵的量级；而当智能社会来到发展阶段，在制造、驾驶等产业将实现垂直化的高度智能行业覆盖及渗透，从而引领产业生态及商业模式的全面升级与重塑，具有主动分析能力的 AI 将大行其道，届时主要国家 / 地区的人均算力水平将达到约 10,000 GFLOPS 的量级；当智能社会进入最终的成熟阶段时，智能场景将实现所有行业的全方位覆盖及大部分行业内高渗透率，届时 AI 将具备自我进化能力，主要国家 / 地区的人均算力将超过约 29,000 GFLOPS 的量级。

通过对全球各个国家算力水平的进一步分析，我们发现算力对国家可以带来可观的直接性经济回报，同时带动延伸性经济增长，并且推动政务高效和民生福

祉。首先，作为具有高附加值的高科技产业，算力已为 ICT 产业带来了杠杆效应极强的直接性经济增长。通过在芯片、服务器、数据中心、智能终端、高速网络等领域进行研发、制造、使用、人才培育，每投入 1 美元即可以带动约 4.7 美元的直接产业产值增长。其次，由于算力投入带来的数字化智能技术加持，亦为包括制造、交通、零售在内多个行业带来延伸性效益，具体表现在产业产值增长、生产效率提升、商业模式创新、用户体验优化等方面。以制造业为例，在传统工厂改造为智能化工厂的场景下，每 1 美元的投入，可以带动 10 美元的相关产值提升。算力投入也可以在推动科技创新、政府高效和民生福祉方面发挥积极作用，带动国家整体竞争力提升。针对创新领域进行算力投入，可进一步提升国家科学技术水平，提高国家整体科研实力。针对公共服务业进行算力投入，可进一步提高政府工作效率，提高各级政府公共服务能力，进而创建平安和谐的社会环境。针对民生进行算力投入，在教育、科研、医疗、就业、社区服务等领域不断加大建设数字化基础设施和生态系统，可有效提高居民的生活质量，提升居民的幸福感和获得感。即使智能计算本身存在的“边际效应”将导致单位算力投入带来的效益提升会随着社会智能化程度的成熟而下降，但所带来的收益仍然可观。

算力的好处是不言而喻的，但是算力在发展的过程中

⁴ 人均算力 = 国家总算力 / 国家人口总数

⁵ gigaFLOPS

亦将遇到 3 大挑战和 1 个问题的影响，政府在其中需要积极承担重要角色，推动社会向智能化发展。

- 首先，为奠定算力发展基础，政府应积极布局算力基础设施建设、同时推动云计算、AI 等关键应用技术的普及与发展。
- 未来，针对未来算力功耗激增的挑战，政府的核心目标应在于引导算力基础设施的部署，并同时推动节能型“绿色算力”发展，以实现智能化社会的可持续发展。可采取的政策手段包括 1) 倡议行业联盟设节能目标；2) 设立新能源投资计划，以直接政府投资带动新能源规模提升；3) 建立能源碳排放交易机制，通过以高额成本换取有限排放配额机制，引导企业主动减排。
- 针对网络质量提升的挑战，政府部门的核心目标应在于利用各种政策手段及资源，提高国家整体网络基础水平，为泛在算力部署的实现打下坚实基础。核心政策抓手包括 1) 形成地理覆盖与网速布局的矩阵式监控体系，持续推动网络建设进程；2) 利用公私合营（PPP）的模式，以少量政府财政投入牵引社会资本投入网络基建。

- 针对生态多样化发展的挑战，政府的核心目的应在于拓展多样化的企业类型及吸引高端人才，以确保多样化的智能应用场景可实现，且算力资源可充分利用。可采取的抓手包括 1) 政府规划算力生态企业多样性发展目标、自身推动移动化场景发展，拉动多样化架构的市场占有率，切实牵引多种算力生态的发展；2) 可通过建立多样化生态联盟，助力各类架构缔结产业上下游资源、同时为产业发展提供人才支持；3) 放宽算力相关市场准入门槛，优化相关政务治理流程与效率，切身支持行业发展。

除以上挑战外，算力的安全问题亦尤为重要，各国政府还需关注算力安全漏洞的风险防范，重点关注智能社会创建与运转的安全红线，提升对于算力风险漏洞的监测及预防能力，可用政策包括 1) 各国政府联合共同制定国际化标准安全规范；2) 通过公共财政提供专项资金，建设安全监测专业机构，且支持第三方安全机构持续监测运营；3) 针对未遵循算力相关法规者建立并完善究责机制。

第一章

算力的定义及发展趋势

第一章主要内容

主要内容

1.1 算力的定义及历史演进

- 算力是指设备通过处理数据，实现特定结果输出的能力
- 算力发展经历人力化、机械化、电气化、集成电路化及移动化，出现在生活各个角落

1.2 算力发展三阶段及未来部署形式

- 单核受技术及成本限制，未来预计在3纳米达到极限：硅基制程提升带来的量子隧穿效应和较低商业化可行性
- 同时，操作系统内核等原因引发的散失效应使多核核数在128核逼近极限：存储、操作系统、软件及单位算力功耗限制影响多核算力持续提升
- 算力网络化成为必然，但需与边缘设备共筑“云、边、端”泛在算力：受网络条件和带宽成本因素限制，需要更加灵活及低成本的边缘设备补充

1.3 技术推动算力需求增长，对算力提出4A要求

- 人类社会必然向智能化社会演进，背后核心驱动是算力
- 未来智能社会算力的核心特征是随时、随地、随需、随形（4A，即Anytime、Anywhere、Any Capacity、Any Object），由人工智能、物联网、区块链、AR/VR等技术牵引

资料来源：案头研究；罗兰贝格

图 1.1 第一章主要观点介绍

1.1 算力的定义及历史发展

算力作为近年来的一个热门话题引发了全社会的大量讨论，然而其定义仍然众说纷纭。其中，2018 年诺贝尔经济学奖获得者 William D. Nordhaus 在《计算过程》一文中对算力进行定义：“算力是设备根据内部状态的变化，每秒可处理的信息数据量”。结合众多权威的定义中算力的特征，我们定义算力是设备通过处理数据，实现特定结果输出的能力。

算力载体发展演变图



资料来源：案头研究；罗兰贝格

图 1.2 随技术的演进，算力的承载形态逐渐多样化

从古至今，算力的载体发展经历了以算盘和机械计算器为代表的时代（如图 1.2）。其中算盘作为最早出现的基于人脑的算力载体，是中国古代劳动人民发明创造的一种简便的计算工具，被广泛应用于人们的日常生活。算盘初步实现了计算功能，但由于完全依靠人力开展而导致效率低下。1642 年，法国科学家 Blaise Pascal 引用算盘的原理，发明了第一部机械式计算器，解放了人力。之后，算力发展经历了以电子和集成电路计算机为代表的时期。1937 年 John V. Atanasoff 和 Clifford Berry 设计了世界上第一台基于电子的计算机——“阿塔纳索夫-贝瑞计算机”。依托于物理介质的改变，算力的计算效率实现了大幅提升。贝尔实验室发明晶体管（1947）和 Jack Kilby 和 Robert Noyce 发明集成电路（1958）后，算力载体的物理体积得以大幅度压缩，同时性能进入指数级增长轨道，出现集成电路计算机。而 1993 年，IBM 公司推出全球第一款智能手机 Simon，正式标志着算力载体可携带，且不再被物理有源电缆所束缚，算力正式移动化。在过去 20 年，随着算力载体的丰富程度得到了极大提升，从用于架设互联网基础平台的服务器，到用于日常办公及娱乐的手机和电脑，再到融入提升居住环境与健康水平的智能可穿戴设备，算力已经融入到了人类生活的每个角落，呈现多样化发展趋势。

算力架构演变

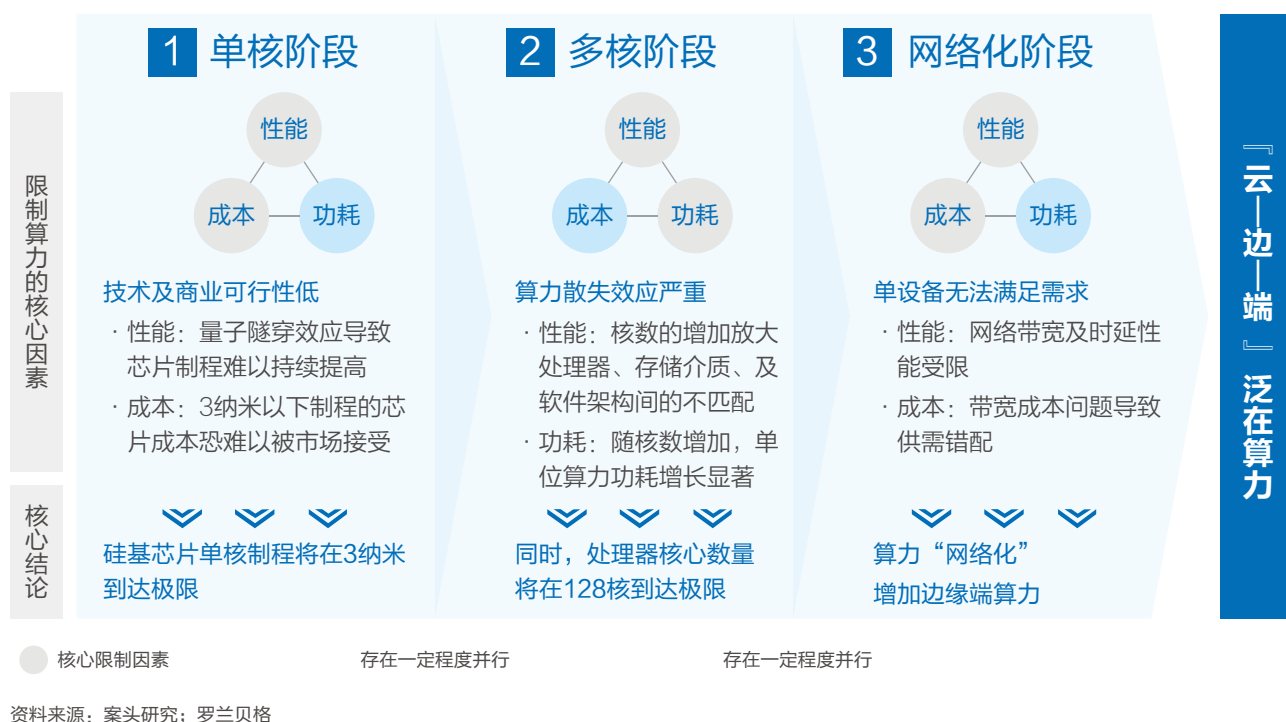


图 1.3 技术的限制驱动了算力的架构形态 – 从单核至多核及网络化演进

1.2 由云—边—端组成的泛在算力架构是技术发展至今的必然选择结果

算力自 20 世纪 60 年代进入集成电路时代后，影响硅基芯片算力发展主要有性能、成本、功耗三大因素。性能因素，即影响算力处理及输出数据能力的因素，如物理效应、制造工艺、封装等；成本因素，即影响获取单位算力所需经济投入的因素，如芯片设计成本、芯片制造成本等；功耗因素，即影响使用单位算力所需能耗的因素，如芯片功耗、单位算力功耗比等。

受这三大因素影响，硅基芯片算力的发展大致经历了从单核走向多核，再到目前的网络化的三个阶段。首先，由于技术和商业可行性低，单核芯片计算能力将在芯片制程为 3 纳米到达极限，推动芯片由单核逐步向多核发展。同时，随着核数增加放大处理器、存储介质、操作系统与软件间的不匹配而导致的算力散失效应（算力无法随着多核核心数增加而成比例上升）和难以继续提升的单位算力功耗，导致多核芯片核心数量在 128 核时接近上限。随后，由于芯片的单核算力上限和多核数量走向极限，在算力需求持续增长的背景下，需要“网络化”算力对需求缺口进行补充。然而，由于网络技术限制导致的算力潮汐效应需要，更为灵活、低成本的边缘端设备可解决需求错配问题。所以，未来算力必然构建“云—边—端”泛在部署架构，满足社会智能化发展带来的算力需求（如图 1.3）。

1.2.1 受硅基纳米制程提升引发的量子隧穿效应（性能）和先进制程芯片商业化可行性低（成本）影响，硅基单核芯片制程将在 3 纳米达到极限

1) 从性能方面来看，由于量子隧穿效应等硅基材料物理极限的存在，通过更低纳米制程工艺获得单核芯片性能增长的路径面临艰巨挑战。

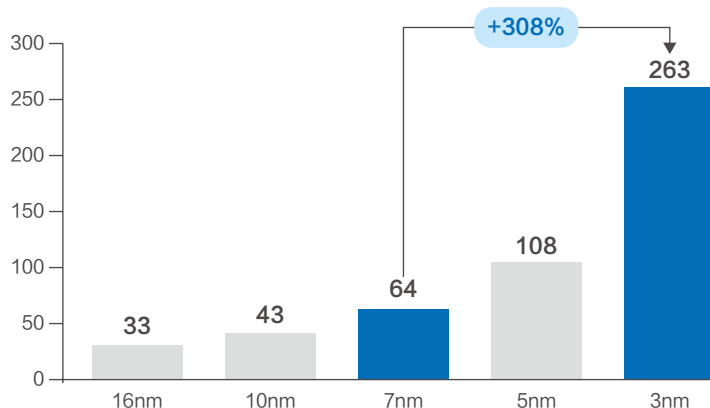
通常所说的 12 纳米、7 纳米制程工艺指的是 CPU 或 GPU 芯片上栅极的宽度，被称为栅长。栅长越短，在相同尺寸的硅片上集成的晶体管越多。为了实现更强的单核芯片性能，芯片制造商必须不断降低纳米制程，缩小晶体管的尺寸，通过在芯片上塞入更多晶体管来提升芯片算力。但随着晶体管尺寸缩小，其栅长也随之缩短，硅基芯片晶体管的栅长在低于 3 纳米时极易发生量子隧穿效应（其原理为，当栅长缩小到一定程度的时候，即使没有加电压，源极和漏极都接近互通，晶体管便失去了开关的作用，因而无法实现逻辑电路）。因此，通过更加先进的纳米制程工艺提升单核芯片性能将面临技术上的严峻挑战。

2) 从成本方面来看，推动 3 纳米制程芯片所需的巨额投入将急剧拉高单芯片成本，导致其难以被大众市场接受，导致商业可行性低

手机SoC芯片成本增长



旗舰级手机SoC单芯片成本[美元]，
以高通骁龙855为例



资料来源：IBS，罗兰贝格



评论

成本显著上升

· 在千万级出货量假设下，3纳米制程旗舰级手机SoC单芯片（以高通骁龙855为例）成本较7纳米增加了约200美元

市场难以接受

· 芯片成本激增将导致采用3nm芯片终端产品价格大幅上扬，大众消费市场无法承受



“摩尔定律最多还能引领半导体界两代，也就是7纳米和5纳米。到了3纳米之后，将遇到极大的困难与瓶颈。”

——联发科董事长 蔡明介

图 1.4 3 纳米制程旗舰级手机端 SoC 单芯片成本显著提升

我们分析，在千万级芯片出货量的假设下，3 纳米制程手机端旗舰级 SoC 单芯片（以高通骁龙 855 为例）成本较 7 纳米显著增加约 200 美元，恐难以被消费者接受。增加的成本主要来源于先进纳米制程工艺带来的巨额的芯片研发、设计和制造成本。以半导体制程中关键的光掩模开模成本为例，在 28 纳米制程下光掩膜版开模不超过 100 万美元，但 7 纳米至少需要 1000 万美元，增加 10 倍。大幅上涨的芯片成本导致的终端产品价格暴增难以被主流市场用户接受，仅有极高端的用户能够接受。其较低的商业可行性带来的狭小市场规模无法

支撑芯片制造商进行技术投入，导致技术制程止步不前。2018 年，全球前五大晶圆代厂商中联电和格罗方德已先后对外宣布退出 7 纳米之后的高端制程竞赛。

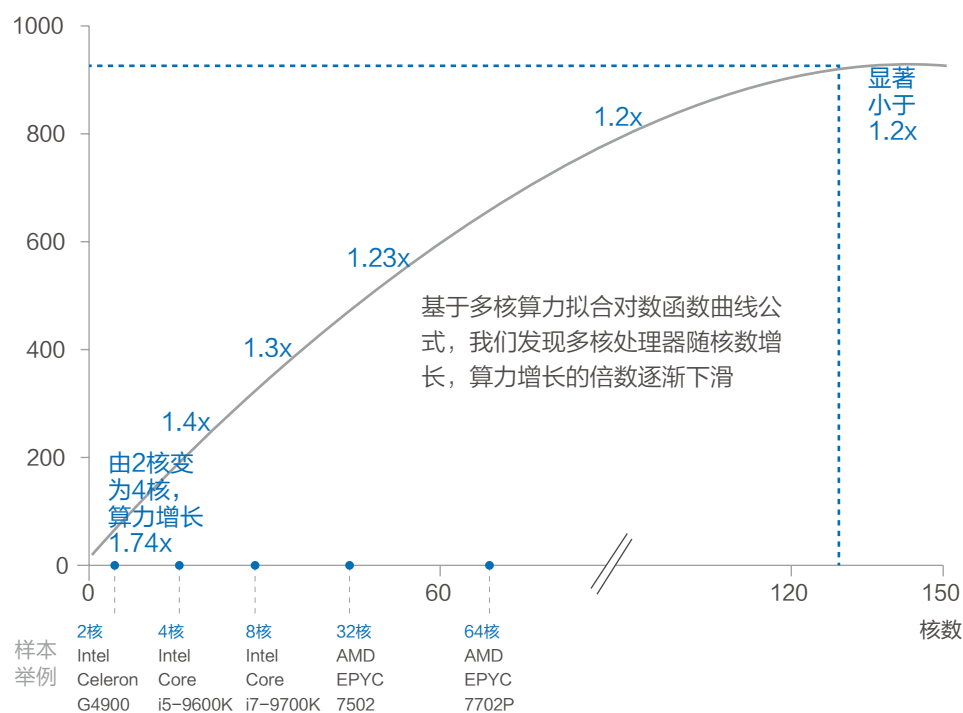
1.2.2 同时，受存储、系统、软件限制（性能）和单位算力功耗显著上升（功耗）两大因素影响，芯片核心数量将在 128 核达到上限，推动算力从多核阶段向网络化阶段过渡

多核处理器是指在一颗集成多个内核，能够进行并行计算的芯片，其天然适合以图像识别、自然语言理解为代表的 AI 训练以及 3D 建模、游戏渲染、视频编辑等高算力消耗类计算任务。然而，多核处理器的算力受到多种限制因素影响，单纯靠堆积核心数量增加算力也并非没有极限。

1) 从性能方面来看，多核处理器受存储介质调用限制、操作系统限制与应用软件限制，随核数增加算力散失效应越发严重，导致多核处理器核心数量在超越 128 核后无法满足经济性要求，难以继续增长。

多核处理器算力增长预测

算力/GFLOPS



资料来源：罗兰贝格

图 1.5 通过增加核数提升算力的模式经济性无法满足要求

计算方法

- 根据拟合后的公式，当核心数翻倍由128核走向256核时，多核算力增长显著低于1.2倍，我们认为其经济性无法满足要求

核心原因

- 多核CPU瓶颈效应
- 芯片处理能力与存储传输能力不匹配
- 操作系统内核设计限制
- 串行任务必然存在，多核能力存在必然约束

在现有的冯·诺依曼架构下，通过拟合不同核心数量的芯片计算能力样本数据，我们发现多核处理器随核数增长，算力增长的倍数快速下滑：从 4 核增至 8 核时，总算力可提升 1.74 倍，而当核数由 128 核增至 256 核时，总算力水平仅能提升已跌破 1.2 倍（1.16X），已显著丧失经济性。（如图 1.4）。其核心原因在于存储介质调用限制、操作系统内核原理、串行任务必然存在等三大因素。在存储介质调用限制方面，多核 CPU 的算

力输出能力会遇到瓶颈效应，导致随核心数量增长的算力无法被完全使用。瓶颈效应指的是当芯片与存储介质间的总线带宽无法与 CPU 计算能力相匹配时，便会产生传输的数据量小于 CPU 可处理的数据量，导致 CPU 算力输出能力受限。在 CPU 与存储介质间存在着总线，CPU 通过总线从存储介质中调度数据进行处理。而随着芯片核心数的增加，CPU 计算能力也随之增加，但存储介质与芯片之间的总线带宽与芯片核心数量的增速存在差异（核心原因在于理论上核心数越多，其对内存等存储介质的并发请求数也越多，相应存储介质的总线带宽需要被多核心共享，导致处理器数据处理时间与存储器访问时间的差距更加突出，瓶颈效应越发严重），导致了多核芯片算力无法得到全量发挥，算力散失效应明显。在操作系统限制方面，主要来自与操作系统设计时并未对多核处理器进行针对性设计，导致大量算力被浪费。由于多核处理器内部有多个核心，因此存在任务分配、调度、仲裁以及平衡负载等问题。所以，多核之间的任务调度是充分利用多核处理器性能的关键，目前的以 Windows、Linux 为代表等主流操作系统所采用的全局队列调度任务调度算法仍然针对于 8-16 核主流芯片优化。操作系统对大于 32 核的处理器的高效任务运行缺少优化，难以满足超多核心处理器的数据实时处理的要求，导致多核处理器的算力大量闲置，无法有效利用多核处理器多核心算力。在串行计算任务限制方面，并不是所有的程序及任务都是可以并行化切割的，导致该类任务天然难以充分利用多核处理器算力。在实际情况中，大量任务的运算过程存在前后依赖关系，即下一次运算依赖上一次的结果。这种计算任务只能采取串行计算，多核心芯片无法在此场景下发挥其并行计算优势。以典型的 RTS 游戏场景（实时战术游戏）为例，玩家所操纵单位的每一次动作、效果都是基于上一次的操作信号输入、受打击情况等计算得出，没有这些先决的计算结果，完全无法进行后续的演算。在这种场景下对任务进行并行化没有意义，无法充分利用为并行计算而设计的多核芯片的性能。

2) 从功耗方面来看，随着核心数量增加，单位算力所需功耗显著提升，在算力需求显著增加且单设备承载功耗有限的条件下，需要通过“网络化”算力对缺口进行补充

随着多核处理器的核心数量增加，其单位算力所需要消耗的功耗亦在显著增加，而在单设备功耗无法持续增长的前提下，通过堆核实现设备算力增长的方式必然会达到瓶颈。导致部分算力需求缺口需要通过“网络化”算力补充。以英特尔两款相同 14 纳米制程工艺、3.6GHz 相同主频的芯片为例，8 核的 i7-9700k 相较 6 核的 i5-8600k 单位算力功耗从 1.11 W/GFLOPS 上升 11% 至 1.23 W/GFLOPS。然而，在电池技术与散热技术未取得突破性进展的前提下，单设备所能承载的功耗有限，无法满足算力需求的不断提升，需要通过其他方式对算力需求进行补充。以分布式计算为代表的“网络化”算力存在形式应运而生。“网络化”算力是指通过网络传输手段将分布在不同位置的算力进行连接以实现分布式计算。在单设备功耗有限但算力需求持续提高的背景下，可以通过此种形式对算力需求缺口进行有效补充。

在实际应用场景中，分布式计算可以通过网络将世界各地成千上万台计算设备所具备的算力串联，形成算力池，然后通过分布式计算模式分配计算任务，完成需要庞大计算量的项目。例如世界最大的粒子物理实验室－欧洲核子研究组织(CERN)通过大型强子对撞机夸克实验项目⁶，利用网络将全球超级计算机进行连接组件“网络化”算力池，帮助科学家们从5年积累的约100PB海量实验数据（相当于6百万张蓝光碟能存储的数据）中证明希格斯玻色子⁷的存在。其发现希格斯粒子的过程依赖于世界各地的研究所通过高效的网络互连与分布式算力分配实现计算资源共享。该项目充分利用DCN（数据通信网络）⁸构建计算节点内部的网络并利用DCI（数据中心网络）⁹推动计算节点之间的互联，都对大量高性能计算设备释放算力起到至关重要的作用。同时，该项目的算力分配模式采取多层级的金字塔式，从位于CERN的零级站点(Tier-0)通过最高可达100Gbps的高速网络链接全世界12个一级站点(Tier-1)进行数据分发，近200个二级站点(Tier-2)再从一级站点接收数据。

再比如从1999年5月17日开始正式运行的国际项目SETI@home¹⁰，充分利用个人计算机闲置算力组成分布式计算网络寻找外星智慧。该项目通过将射电望远镜采集到的海量数据拆分成零碎的数据包上传到互联网，分配给每一台安装了SETI@home软件的计算机，利用计算机闲置算力对数据进行分析。短短5年间该项目借助闲置的处理器资源积累了近200万年的CPU运行时间，处理了超过13亿个数据单元，进行了近 5×10^{21} 次浮点运算，是算力“网络化”的成功实践。

1.2.3 在算力未来需求持续增加的情况下，网络带宽和时延限制（性能）和网络带宽成本限制（成本）导致的算力需求错配需要通过在边缘端部署算力进行支撑，构成“云－边－端”泛在算力部署方案

在“网络化”算力有效补充了单设备无法满足的大部分算力需求后，仍然有部分算力需求受不同类型网络带宽及时延限制，仍然无法满足低时延、大带宽、低传输成本的算力需求场景，如智慧安防网络、CDN加速等。近年来，许多场景中出现了边缘计算（即通过将算力部署在更加靠近终端设备的位置来满足算力需求）的形式，以实现更快、更低时延、更低成本的算力输出。英特尔公司高级专家公开表示，“数据在不断增加，如果全部传输到云端进行处理，显然当前的带宽和传输速度都难以满足要求，因此将数据处理从云端迁移到边缘端十分

⁶ 大型强子对撞机(LHC)是世界上最大、最强大的粒子加速器。它于2008年9月10日首次启动，至今仍是欧洲核子研究中心(CERN)加速器综合体的最新成员。大型强子对撞机由一个27公里长的超导磁体环组成，环上有许多加速结构，可以在运行过程中提高粒子的能量

⁷ 希格斯玻色子是希格斯场的量子激发。希格斯粒子的衰变能会生成耦合实粒子

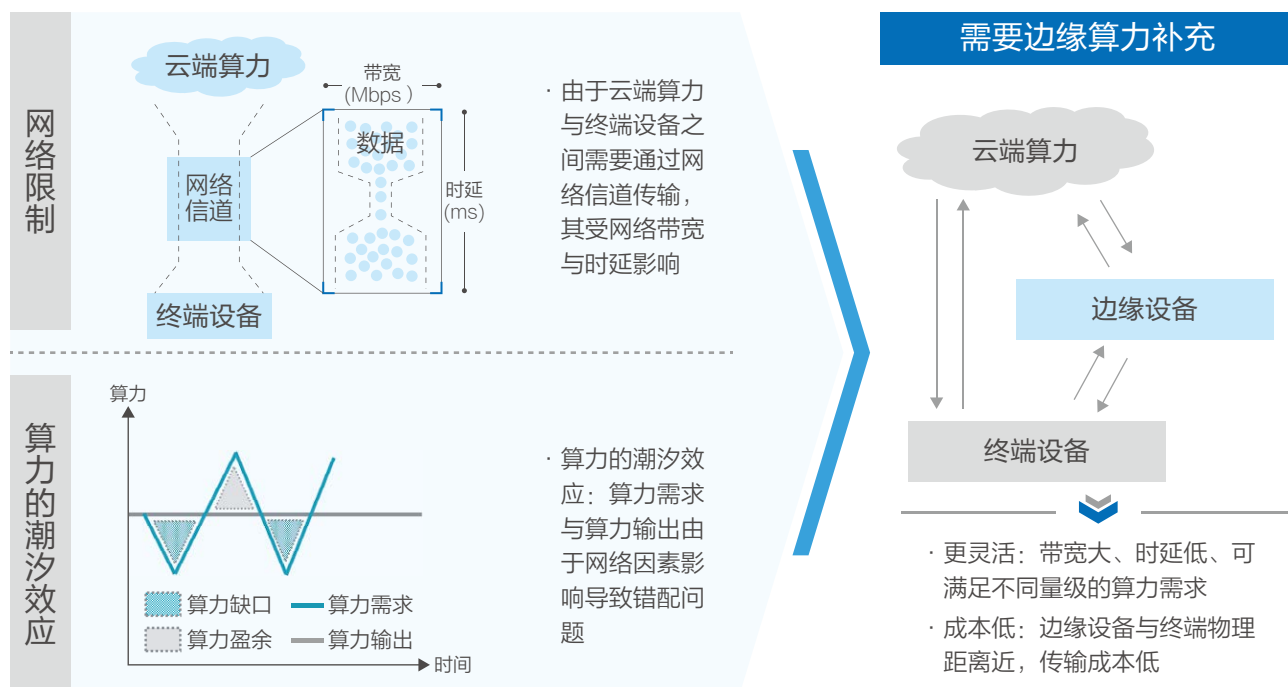
⁸ 数据通信网络(DCN)利用各种有线和无线通信信道将数字数据从一台计算机传输到另一台计算机。互联网是该网络的代表

⁹ 数据中心网络(DCI)技术是通过高速分组光连接将两个或多个数据中心在短、中、长距离上连接在一起

¹⁰ SETI@home(“SETI at home”)是一个基于互联网的公共志愿者计算项目，采用了伯克利SETI研究中心创建的BOINC软件平台，由加州大学伯克利分校空间科学实验室主办。它的目的是分析无线电信号，寻找地外文明的迹象，因此是作为世界范围SETI进行的活动之一

有必要”¹¹。同时，将算力部署在边缘端也能显著节省网络带宽成本，是更加经济的算力部署模式，引入边缘计算后，计算、存储、网络成本可以节省约 30% 以上¹²。

1) 从性能方面来看，由网络带宽对云端算力的全量发挥的限制和网络时延的影响所带来的算力错配要通过增设边缘端来弥补，因此未来形成“云¹³、边¹⁴、端¹⁵”多层级算力部署方案是必然趋势



资料来源：案头研究，罗兰贝格

图 1.6 网络传输带宽和时延导致算力错配，需要边缘算力支撑

网络带宽及网络时延共同决定了网络信道的传输质量，影响到网络算力的全量发挥（如图 1.6）。具体而言，“网络化”算力的使用需要经历终端与云端的数据双向传输过程，这段过程是通过网关、基站、数据中心等不同网络节点之间的信道所实现的，其中信道的容量决定了传输的速率（带宽），信道的长度与材质决定了数据传输的时延，两者共同影响了数据传输的效率，而数据传输的效率则决定了云端算力的输出功率，即数据传输的效

¹¹ 与非网《边缘计算市场不断扩张，算力远没走到尽头》2018

¹² 中国电子技术标准化研究院《边缘云计算技术及标准化白皮书》2018

¹³ 云端的算力是由 GPU 或 NPU 等芯片产生的，通过在虚拟平台调度服务器上处理复杂的数据来实现合理使用

¹⁴ 边端的算力由 CPU、FPGA 等芯片产生，通过边端服务平台的实时数据过滤和响应，可以保证网络的稳定性且降低延迟

¹⁵ 终端的算力是由 CPU、GPU 或 DSP 等芯片产生的。更好的用户体验可以通过操作系统上的软硬件资源管理来实现

率越高，云端算力就能够发挥更大的效用。但由于目前网络技术的限制，带宽受传输网类型、传输方式等因素影响，时延受网络技术、传输距离、干扰等因素影响，导致云端算力始终无法做到全量输出（如图 1.7）。

受两大网络因素影响，算力供给与算力需求间会产生错配，部分算力需求需要通过增设边缘算力满足。智能化社会发展过程中由新兴技术、应用、场景带来的数据量持续增长，对整体算力提出了持续的要求。然而，受两大网络限制因素与成本限制，网络化的算力与算力需求间将会产生算力错配，在不同的时间会出现算力盈余和算力缺口。所以，通过部署更灵活、成本更低的边缘计算设备，完善在云、边缘、现场终端不同层级的算力，可以减弱云端和终端长距离数据传输的网络限制问题。如在 5G 时代的 MEC¹⁶ 部署可以降低传输距离，满足如智慧工厂、智慧港口、智慧交通等场景下的超低时延和大带宽的网络要求。同时减小低价值流量占用网络（重复性数据）或者超大规模流量（视频图像）对于网络资源的消耗。以智慧港口为例，对起重机的控制级通信操作，需要其网络满足超低时延（10-20 毫秒以内），及超过 99.999% 的可靠性，当前的 4G 网络远远无法满足。在云端，算力由 GPU、NPU 等芯片产生，通过虚拟平台调度服务器设备进行复杂的数据处理，实现高性能算力的合理运用。在边缘端，算力由 CPU、FPGA 等芯片产生，依托于网关设备，通过边缘服务平台在边缘节点实现数据筛选和实时响应，保障数据传输的稳定性和低延时性。在终端，算力由 CPU、DSP 等芯片产生，操作系统可进行软硬件资源管理，使终端设备拥有更流畅的用户体验。

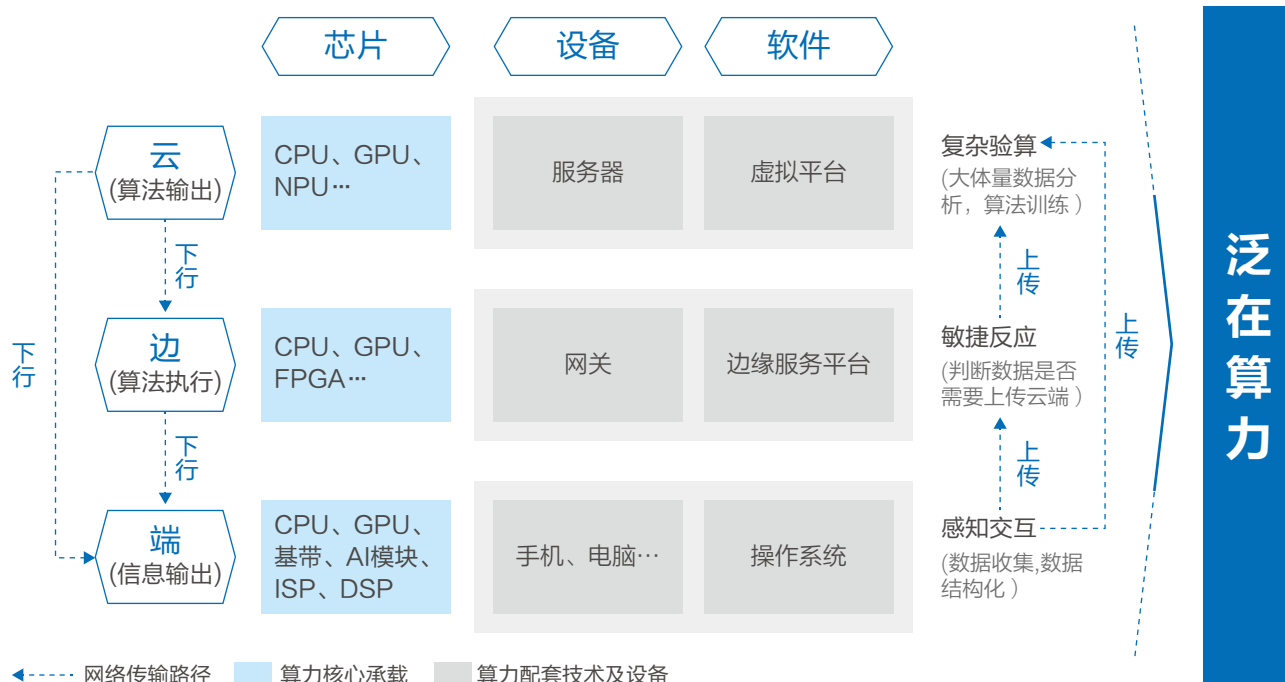
2) 从成本方面来看，在边缘端部署算力能有效节省网络带宽成本，是更加经济性的算力部署方案。

尽管网络技术的发展将提升网络质量，但仍然存在使用网络服务的带宽成本。部署边缘计算可将一部分算力放在靠近接入网的机房，在降低传输时延的同时更能有效降低网络的带宽成本。2018 年，阿里云和虎牙直播¹⁷共同建设了边缘节点服务，不仅实现了边缘节点网络连接小于 5 毫秒的低时延，更节省了 30% 以上的网络带宽成本。

¹⁶ 多接入边缘计算 (MEC)，也可称为移动边缘计算，是 ETSI 赋予的网络架构概念，支持云计算能力和蜂窝网络或任何网络边缘的 IT 服务环境

¹⁷ 虎牙是一个集游戏、娱乐、体育为一体的视频直播平台

算力架构概览



资料来源：专家访谈，罗兰贝格

图 1.7 泛在算力架构由芯片、设备、软件组成

综上所述，算力在进入网络化阶段后，受到网络技术发展和网络带宽成本限制，增设边缘侧设备成为必然趋势，未来将形成云端侧负责大体量复杂计算、边缘侧负责简单计算执行、终端侧负责感知交互的泛在算力部署形式（如图 1.7）。

整体来看，算力从单核到多核再到网络化，算力最终走向云 - 边 - 端多层次泛在部署的理想方案。在此期间，无论是集成电路时代大规模生产制造对于计算设备的需求，还是信息化时代全球互联互通对于移动端计算能力的依赖，都足以证明算力需求的持续高速提升来自于社会智能化程度的提高。

1.3 随着社会的智能化进程，对泛在算力提出了 4A 的具体要求¹⁸

¹⁸ “4A”是指“Anytime, Anywhere, Any Capacity, Any Object”，即随时、随地、随需、随形

算力发展的核心驱动因素



资料来源：罗兰贝格

图 1.8 分布领域、底层技术和应用场景共同支持智能社会发展

未来社会将逐步走向智能化，其背后的关键技术（人工智能 / 物联网等），是算力需求发展的核心驱动因素（如图 1.8）。具体来看，每一次的生产力变革，都带来了全社会的飞跃式发展。随着以人工智能、物联网、区块链、AR/VR 等关键信息技术的逐渐成熟，我们认为未来这些技术可推动大量的智能场景的实现。人工智能技术将推动无人驾驶、智能办公、智慧医疗等场景的有效落地，物联网技术将推动智能消防、智慧工厂、智慧农场、智能家居等场景落地，区块链技术将推动应用于数字证书、信息加密等场景落地，AR/VR 技术则可推动智慧商场、游戏、智慧课堂等场景落地。这些场景未来将在产业领域实现跨越式发展、助力各产业创新、增强产业数字化程度并增强市场活力，在政务领域帮助政府提升运行效率、提高城市管理水平、加强居民生活幸福度，在民生领域推动社会民生保障、创造宜居空间、实现可持续化发展，共同推动社会向智能社会发展。

关键技术对算力的具体要求

		人工智能： 以无人驾驶为例	物联网： 以智慧工厂为例	区块链： 以数字货币为例	AR/VR： 以VR游戏为例
随时	速响应： 提供低时延且不间断的算力支持	· 无人驾驶级别越高对网络要求越高，L4 需带宽 >100Mbps, 时延 5~10ms	· 物联网设备精准生产要求连续性算力支撑，带宽 >1Gbps, 时延 5~10ms	· 区块链对于带宽和时延要求较低	· 为达到舒适用户体验，带宽需 >180Mbps，时延 <20ms
随地	广覆盖： 提供不受地理范围限制的算力支持	· 2040 年，所有新车配备自动驾驶功能，算力需求遍布全球	· 覆盖家庭、工厂、电网等多种环境	· 赋能包括金融、政务、零售、支付、物流等多种场景	· 应用于游戏、工作室、智慧商场等多种场所
随需	大量级： 提供能满足波峰与波谷的算力需求	· 2030 年较 2018 年算力需求增长约 390 倍 · 需要满足出行高峰算力需求	· 2030 年较 2018 年算力需求增长约 110 倍 · 2025 年，全球 IoT 设备量达 215 亿个	· 2030 年较 2018 年算力需求增长约 2000 倍	· 2030 年较 2018 年算力需求增长约 300 倍
随形	载体丰富： 能够装载在随形的载体上	· 无人驾驶汽车 · 边缘计算设备 · 云计算中心	· IoT 设备 · 边缘计算设备 · 云计算中心	· 云计算中心 · 终端设备，如采用 ASICS 或 FPGA 架构矿机	· AR/VR 设备 · 手机电脑 / 设备 · 云计算中心

资料来源：IDC，宽带发展联盟；罗兰贝格

图 1.9 算力在智能社会可以做到随时、随地、随需、以随形存在

未来人工智能、物联网、区块链、AR/VR 技术发展将带来对算力随时、随地、随需、随形 (Anytime、Anywhere、Any Capacity、Any Object) 的 4A 能力要求（如图 1.9）：

在随时层面，未来算力发展将关注低时延与连续性，低时延要求算力对数据处理的及时反馈，而连续性要求算力不间断的支撑。以无人驾驶汽车场景为例，面对复杂多变的路况信息，无人驾驶汽车需要确保针对行驶过程中的任何突发性状况进行实时反映；同时在行驶的旅途中始终如一地保证算力对于路况、车况监控的“专注度”。未来 L4 & L5 级别对网络带宽的需求将大于 100Mbps，时延要求达到 5-10 毫秒的水平。在随地层面，高速增长的技术应用场景与终端设备使得算力需要大量应用在不同地点。因此在未来社会中，算力将因为智能应用场景的不断涌现而无处不在。无论是应用在政务领域的人脸识别、智能安防，民生领域的智能出行、移动支付，亦或是产业领域的智慧工厂、智慧农业，所有智慧场景都与算力密不可分。在随需层面，未来的算力应能够满足任意量级场景与终端设备的算力需求，并按照实际需求进行自适应化的调配。如，无人驾驶领域的算力需能够同时满足车辆出行低峰期与高峰期的算力需求。在随形层面，未来算力形态将不仅限于人脑，同时算力将在以大型数据中心、数据中心为代表的云端，以智慧路桩、边缘计算单元为代表的边缘端，以人脑、无人驾驶汽车、IoT 设备、个人电脑、AR/VR 一体机为代表的终端，形成算力“云、边、端”泛在算力部署架构。

1.3.1 人工智能技术的发展对于算力的驱动及要求

人工智能技术将全面赋能各行各业。预计到 2025 年，人工智能涉及的场景规模将达到 2,081 亿美金¹⁹，并在无人驾驶、智慧金融、智慧医疗、智慧零售、文娱等领域大显身手。人工智能技术对于算力的核心拉动点在于未来各应用场景内单设备芯片算力的增长和人工智能技术的行业渗透率的进一步提升，带动对云计算中心、边缘设备和终端 NPU 的巨大需求。整体预计在 2030 年，人工智能相关领域对于算力的需求将达到 ~16,000 EFLOPS，相当于 1,600 亿颗高通骁龙 855 内置的人工智能芯片所能提供的算力。

人工智能算力



资料来源：IDC，罗兰贝格

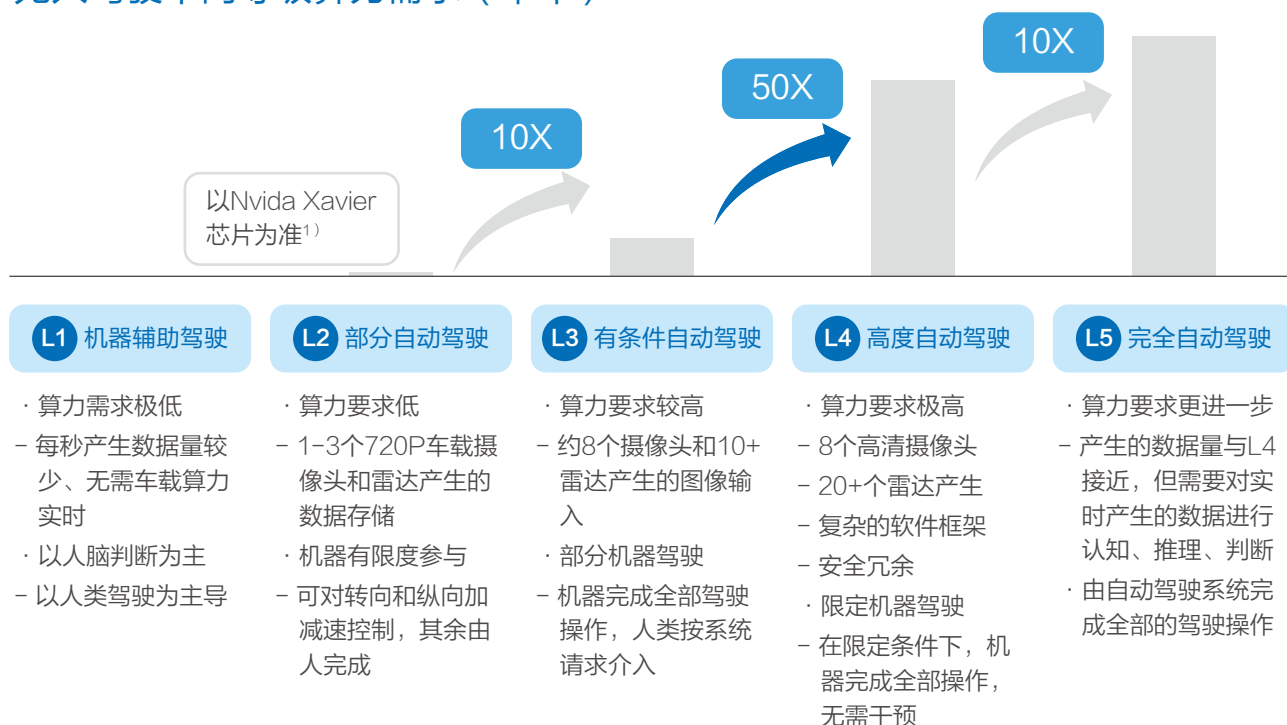
图 1.10 人工智能技术对算力需求逐步提升

以无人驾驶场景为例，人工智能技术对算力的核心拉动方式主要有 3 点：

- 1) 单车算力需求几何级数增加：随着无人驾驶的等级提升，L4/5 级无人驾驶汽车的数据量产生速率需要具有强大的车载芯片算力增长 5,000 倍予以匹配。

¹⁹ IDC 《Worldwide Semiannual Artificial Intelligence System Spending Guide》2019

无人驾驶不同等级算力需求（单车）



1) Xavier 系统级芯片是Nvidia为满足L2级别自动驾驶推出的芯片，是DRIVE AutoPilot系统的核心

资料来源：罗兰贝格

图 1.11 L1 到 L5 级别无人驾驶汽车对算力需求成倍增长

在无人驾驶初级阶段，如 L2 阶段（如图 1.11）对于算力的要求较低，仅需要满足 1-3 个 720P 车载摄像头、1-2 个雷达所产生的数据量即可，且无需对所有进行实时分析。在无人驾驶的中级阶段，以目前完全达到 L3 级别的最新特斯拉汽车（搭载 Full Self-driving Computer 系统）为例，其使用的 AI 自动驾驶芯片上的神经网络处理器（NPU）通过提供强大算力支持，在匹配其专属的 Autopilot 3.0x 系统后，能够同时处理安装在车头和车身上的总计 8 个高清摄像头和 10+ 雷达产生的数据量，使得特斯拉汽车可以对行驶路线、车辆状态、路况安全等情况进行快速分析并持续作出反应，实现有条件自动驾。在无人驾驶的高级阶段，以能够满足 L4 级别无人驾驶算力需求的华为 MDC 系统（Mobile Data Center）为例，其通过整合最新的 Ascend（昇腾）CPU 芯片、AI 芯片、ISP 芯片与 SSD 控制芯片在提供低时延（小于 200 毫秒的端到端延时；小于 1 毫秒的 ROS 内部延时；小于 10 微秒的内核调度延时）的背景下，能够满足 16 个摄像头、16 个超声波雷达、8 个激光雷达、6 个毫米波雷达等车载传感器同时工作而产生的庞大数据流量。除此之外，英伟达预测，L4 级别的无人驾驶汽车需要面临更为复杂的驾驶条件、更冗杂的安全纠错机制、更多层级的软件框架，在无法完全靠优化算法的情况下，对车载算力约有 50 倍的提升要求。最后，在无人驾驶的终极阶段，即完全自动驾驶阶段（L5），车载芯片将会面临来自非标准公路、山路、复杂路况下的路况苛刻识别需求，在与 L4 级别汽车的数据传输量相似的情况下，对车载芯片的复杂数据处理能力和快速分析和决策能力提出了更高要求，预计有约 10 倍的计

算力增长需求。

2) 随时、随形：无人驾驶场景落地实现需要在不同位置实现对多层次、多来源的环境信息进行随时响应，需要边缘服务器部署以应对高级别无人驾驶阶段毫秒级时延响应要求。在未来无人驾驶的场景中，如果行车安全的核心信息，如路况环境、行人密度等需传输到数百公里的云端服务器进行处理，但是基于目前的网络技术现状和发展目标，在可预见的未来解决传输距离带来的延迟问题，且可靠性难以保证。然而，无人驾驶汽车在高速行进的状态中，重要信息的传输延迟将造成车辆损毁、车道堵塞、安全事故等严重后果。从无人驾驶汽车的等级提升对延迟的要求角度来看。L3 级别的自动驾驶对于传输时延的要求在 10-20 毫秒，在进入 L4 & L5 级别后，对于传输时延的要求进一步提高到 10 毫秒以下。因此基于单车智能、依赖车网通信的自动驾驶路线图在 5G 网络发展背景下仍然无法保障传输时延与稳定性，所以未来以车路通信（边缘侧部署 RSS、边缘计算平台等计算单元，RSS 提供路测融合感知，边缘计算平台协同数据收集、路由和分发）为代表的算力边缘化部署解决方案将为无人驾驶场景提供计算能力和传输时延的双重保障。

3) 随地、随需：2040 年，搭载强大 NPU 芯片的高算力 L4/L5 级无人驾驶汽车的普及将拉动无人驾驶领域整体算力快速增长。借助人工智能技术无人驾驶汽车可以避免因为驾驶员人为失误而造成的交通事故，有效提升道路的安全性；通过与周围车辆、道路、信号灯等环境信息的实时互联，实现更加及时的路线预判和路况识别，降低交通拥堵。全球自动驾驶渗透率将快速提升，2020 年 L1/L2 级功能渗透率将达 40%；L3 级功能也逐渐开始量产，2020 预计渗透率将达 5%；L4 级别无人驾驶汽车将在 2025 年左右实现量产。至 2040 年，所有新车都将配备不同级别的自动驾驶功能，其中 L4/L5 级别自动驾驶在新车销售中的渗透率将超越 30%。未来将有约 2 亿辆搭载具有不低于 300-500 TFLOPS²⁰ NPU 芯片的无人驾驶汽车上路，驱动整体算力需求增长。同时，由于交通存在高峰时间（如早晚高峰时期与节假日）与低峰时间（如深夜和正常工作时间），这要求部署在车载芯片外的算力在不同的时间段提供支持。以上海市的车辆行驶情况为例，高峰上路车辆（8 万辆²¹）是低峰上路车辆（5 万辆²²）数量的 1.6 倍，在人们出行方式不发生巨大改变的前提下，每天不同时间的无人驾驶车载算力巨大需求波动需要大量的云端和边缘端计算设备提供弹性的算力支持。

²⁰ TeraFLOPS= 每秒数兆次浮点运算

²¹ 罗兰贝格分析

²² 罗兰贝格分析

1.3.2 物联网技术的发展对于算力的驱动及要求

物联网技术将覆盖工厂、家居等多个领域。预计到 2025 年，物联网技术涉及的场景规模将达到 1580 亿美元²³，主要通过低算力物联网设备的普及、配套云端计算中心和边缘端计算单元的增加共同拉动算力增长。整体预计在 2030 年，物联网技术相关领域对于算力的需求将达到 ~8,500 EFLOPS，相当于 79 亿台当下高端边缘设备使用的 EPYC 7401 芯片所提供的算力。

物联网算力



资料来源：IDC，罗兰贝格

图 1.12 物联网技术对算力需求亦在不断强化

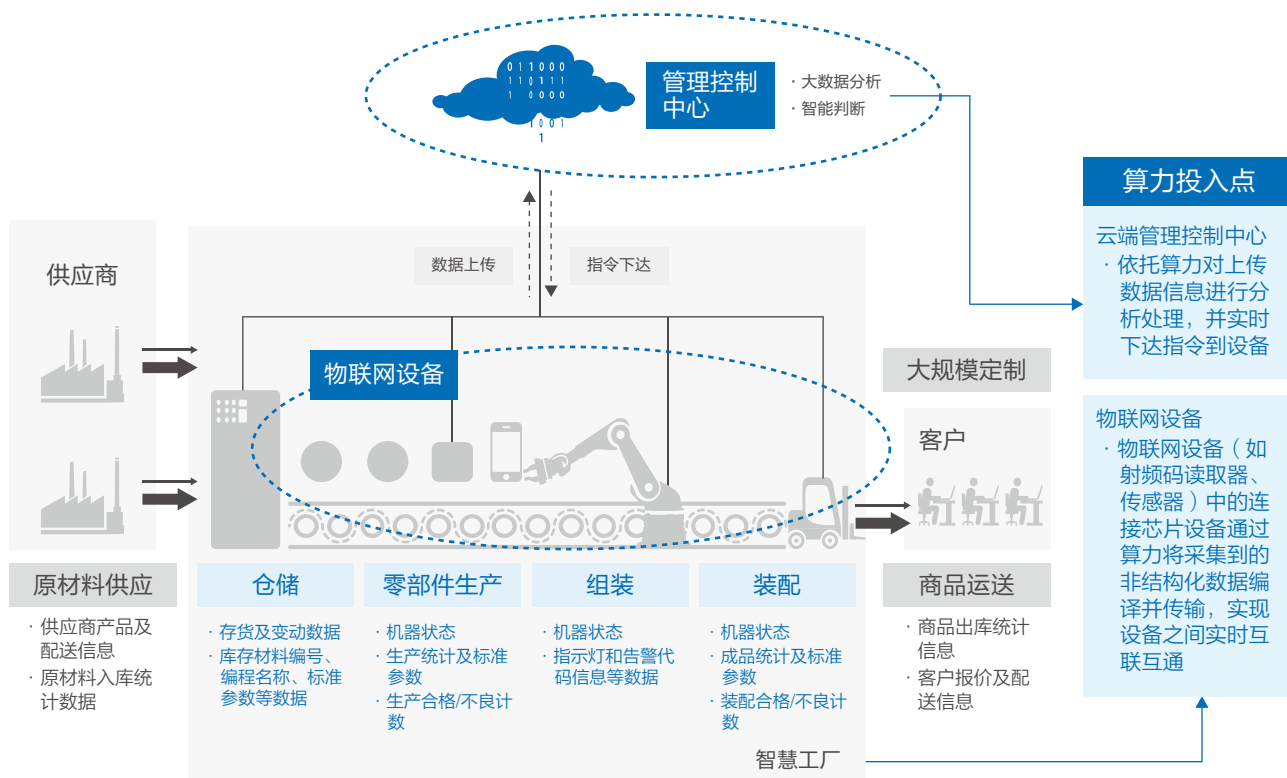
以智慧工厂场景为例，物联网技术对算力的核心拉动方式主要有 3 点：

1) 随地、随形：215 亿台物联网设备大范围的普及将推动整体算力快速增长。以智慧工厂场景为例，其包括的物联网设备有红外传感器，光电传感器，超声波传感器等，需要以华为 Boudica 150²⁴ 为代表的物联网芯片提供算力支撑。到 2025 年，应用在工业物联网相关领域上的智能设备将达到 137 亿台，其余物联网设备分布在家庭、电网等多种不同类型的场景下。

²³ ZION 《IoT Devices Market by Device Type for Building and Home Automation, Energy and Utilities, Manufacturing, Connected Logistics, BFSI, Transportation, Connected Health and Others: Global Industry Analysis, Size, Share, Growth, Trends, and forecast 2016 – 2024》2019

²⁴ 华为新推出的高度集成的 V150 NB-IoT SoC，可以降低开发成本，缩短物联网设备的开发周期

智慧工厂算力投入点



资料来源：罗兰贝格

图 1.13 智能工厂算力投入点包括云端管理控制中心、物联网设备

2) 随需：智慧工厂内部大量物联网设备产生的数百 GB 的数据需要云服务器提供算力支撑。仍以智慧工厂为例（如图 1.13），只有具有强大算力的云端服务器和边缘侧服务器才能对工厂内制造生产的各个环节中数十万物联网设备每天产生的数百 GB 数据（包括但不限于机器名称、材料编号、编程名称、标准参数、生产统计、生产合格/不良计数、机器状态、指示灯和告警代码信息等）进行集中分析处理。强大的算力可以为智慧工厂运营带来两大效益：第一是通过网络实时反馈到相关屏幕上，帮助工厂管理人员掌控整体工厂运行情况，及时处理问题；第二是作为工厂自动化的基础，为整个生产线从原材料供应、仓储、零部件生产、组装、装配到最终的商品运送环节提供基本数据及机器状态等关键数据，结合云端数据分析与智能判断能力，实现智慧生产。

3) 随时：大量物联网数据信息的实时上传对车间内部网络的时延和带宽要求较高。在生产制造的过程中，任何设备节点上网络传输的延迟都可能造成整条生产线的停机，影响生产延续性和生产效率，甚至会威胁到操作人员的安全。同时智慧工厂内除了数据本身量级的加大（数千上万台物联网设备），数据交互的频率也较以往大幅提升，对带宽要求进一步增强，因此，建设智慧工厂对车间内部的局域网络带宽需要达到 Gbps 级别，最高时延控制在 5ms-10ms 以内。

1.3.3 区块链技术的发展对于算力的驱动及要求

区块链技术将赋能各行各业。预计到 2025 年²⁵，区块链技术涉及的场景规模将达到 385 亿美金，在数字货币、供应链金融、信息共享、版权保护、跨境支付、物流链等领域大显身手，因安全问题要求的算力持续增长和应用场景的快速普及将带动以云服务器为主的算力增长。整体预计在 2030 年，区块链技术相关领域对于算力的需求将达到 ~5,500 EFLOPS，相当于 13 亿台入门级矿机蚂蚁 V9 所能提供的算力。

区块链算力



资料来源：罗兰贝格

图 1.14 区块链技术的普及提升对算力需求

以比特币网络为例，区块链技术对算力的核心拉动方式主要有 2 点：

1) 按需、随形：在未来，算力池（即维持比特币体系运作的计算力）需要持续不断的增长以保证区块链系统的机制安全。

区块链技术本质上解决的是双花问题，其以分布式账本的形式存在，每一笔被记录的交易都是公开透明、不可篡改、不可伪造的。账本上记录了每个人相关的数字资产的交易，就是所谓“加密数字货币”。以比特币网络为例，通过参与到区块的生产中，提供工作量证明（PoW），即可获得新的区块奖励（即 1 个比特币）。

²⁵ IDC《Worldwide Semiannual Blockchain Spending Guide》2019

在比特币的安全机制中，工作量证明解决了一个“集体决策”的“大多数”问题。网络通过竞争记账机制给予记账过程中工作量最多的个人记账权力。在极端情况下，当个人拥有全网 51% 的算力时便可掌握绝对的记账权利，进而可以在区块链内故意排除或篡改交易顺序，使同样的交易发生两次，从而对整个区块链系统安全产生威胁。因此，这就要求未来以云端计算中心等形式支撑的算力持续增长，以避免人为对区块链机制内算力池实现控制，影响到整个区块链的安全机制。

2) 随时、随地：区块链技术的核心加密机制将全面应用到金融、政务、零售等行业内多个场景，相关算力亦需同步增长以满足区块链的发展要求。作为构建未来智能社会的信任基础，区块链未来将不仅限于传统的对数字货币加密具有极高要求的金融领域，作为一种底层信任机制，可以同时为包括政务、零售、支付、物流在内的广泛应用场景赋能。：1) 信息共享，通过统一的中心进行信息分发，基于区块链保持各个节点的数量一致性原则确保信息共享的准确性，以腾讯“公益寻人链”为例；2) 版权保护，以区块链技术为基础提高版权保护手段及提高法务存证公信力，如电子身份证、时间戳保护、广州“仲裁链”；3) 物流链，借助区块链的不可篡改性对海外商品进行溯源，如阿里、京东、腾讯海外商品进口溯源；4) 供应链金融，通过区块链技术取代传统仓单、合同、票据等纸质文档提高数据可靠性；5) 跨境支付，借助搭建跨境区块链账本，降低支付流程、减少结算周期、降低结算成本，如 Ripple、Circle 等；6) 资产数字化，通过区块链技术将在实体资产证券化后进行数字化，实现便于分割、流通方便等优势，如腾讯的微黄金应用。以上应用场景所涉及的区块链技术均离不开以云端计算中心提供的强大算力在创立、验伪、交易环节上对场景进行支持。

1.3.4 AR（增强现实）/VR（虚拟现实）技术的发展对于算力的驱动与要求

AR/VR 技术将全面赋能游戏及商务等应用领域。预计到 2025 年，AR/VR 涉及的场景规模将达到 596 亿美元²⁶，设备端计算能力的提升、VR/AR 设备的普及和普及需要的云计算中心和边缘计算设备算力配套将共同推动整体算力的增长。整体预计在 2030 年，物联网技术相关领域对于算力的需求将达到 ~3,900 EFLOPS，相当于 21 亿台索尼顶级游戏机 PS4 所提供的算力。

²⁶ Marketsandmarkets 《Virtual Reality Market by Offering, Technology, Device Type, Application and Geography – Global Forecast to 2024》2019

AR/VR算力



资料来源：Markets&Markets，罗兰贝格

图 1.15 VR/AR 技术发展离不开算力支撑

以 VR 应用为例，AR/VR 技术对算力的核心拉动方式主要有 3 点：

1) 单设备芯片算力需求：走向 4K/8K 的 VR 视频设备需要更为强大的算力支持。随着以 4K 和 8K 为代表的显示领域创新落地，未来 VR 设备需要渲染的图形量将增加 4-16 倍（如从 1920*1080 增加至 3840*2160 或 7680*4320）后，需要单设备算力或相配套的云、边算力同步提升，才能满足未来 VR 设备的显示要求。同时，以双路渲染、透镜矫正、复杂交互、物理特效为代表的下一代特效算法的大规模利用也对 VR 设备的算力提出更高要求。最后，VR 设备紧贴人眼的特性要求视频图像的显示参数极高（如低于 20ms 延时、75Hz 以上刷新率、1K 以上陀螺仪刷新率）。综上所述，大渲染量、高级特效、低延时要求对负责输出 VR 视频的 GPU 算力和相配套的云端、边缘设备的算力提出了近乎苛刻的要求。以目前显卡制造商英伟达推出了 GTX GeForce VR Ready 计划为例，在满足目前基本 1080P 水准的 VR 渲染要求下，其推荐 GPU 英伟达 GeForce GTX 970 显卡算力已达到 2.44 TFLOPS。而据英伟达管理层称，如果未来 VR 设备要满足以上述的种种要求，图形处理器 GPU 性能要达到 >100TFLOPS，是现有旗舰级 GPU 算力的 40 倍。

- 2) 随地、随需、随形：VR/AR 设备普及率提高拉动算力整体增长。伴随以 Oculus、Hololens 为代表的高清设备和以优质应用的涌现，未来 VR/AR 设备在商业、医疗、教育、影视娱乐等行业的渗透率会急剧增长。据 IDC 预测，2019 年 VR 全球终端出货量超过 700 万台，预计到 2023 年终端出货量将达到 3670 万台，年均复合增长率超过 51.3%。同时，就像过去观看视频是先下载后观看，随着 4G 网络普及，在线点播已经成为主流。未来随着 5G 网络的普及，大量高清 VR 内容将会实现像现在网络视频一样的在线点播，需要云端提供编码和渲染支持、边缘端算力提供编码、拼接、渲染支持。因此未来 VR/AR 设备的普及将带动云计算中心和边缘计算设备等配套算力的快速提升，拉动 VR/AR 领域的整体算力快速增长。
- 3) 随时：高质量的虚拟现实体验对于要求带宽和时延能够满足至少 180 Mbps 带宽和 20 毫秒时延。如需在三维全景下达到理想的沉浸式体验效果，VR/AR 设备需要实现更高的视网膜分辨率、更广的观察视角、更精准的位置感应（如 3D 音频动态跟踪和手势位置动态追踪等）等，这对带宽的需求至少为 180Mbps。此外，在 VR/AR 游戏中，端到端的时延至少需要小于 20 毫秒，才能保证在使用过程中避免感知到明显的图像滞后而导致的眩晕。

第二章 算力衡量指标体系

第二章的主要内容

主要内容

2.1 算力衡量指标体系介绍

- 为科学有效衡量算力的水平，我们开发了算力衡量指标体系，基于云、边、端各层面典型设备的产品形态进行测算，同时考量算力散失效应、网络技术限制效应等调整系数对结果进行校正

2.2 算力指标衡量结论

- 我们基于算力水平衡量模型选取全球一部分国家进行算力水平评估，发现人均算力水平的高低与一国的经济发展水平形成了紧耦合的绑定关系，更是一国智能化、数字化发展水平的集中体现，是一切数字化应用建设及发展的底层基础

2.3 随着智能社会阶段性演进，对算力水平的需求呈指数级增长

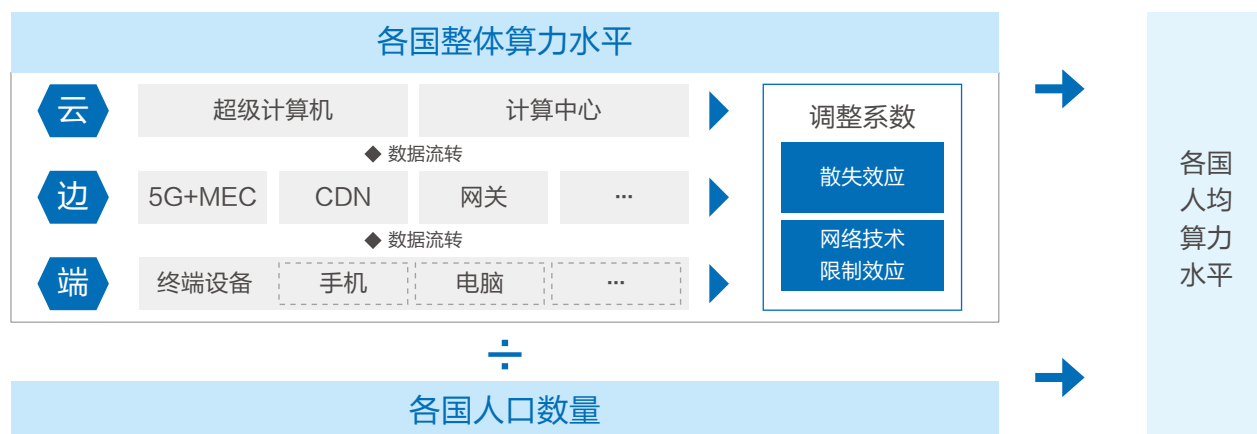
- 算力会随着智能社会的发展而显著增高。主要国家/地区目前人均算力均低于3,000 GFLOPS，未来达到发展阶段后的人均算力约为10,000 GFLOPS以上，达到成熟阶段后的人均算力约为29,000 GFLOPS以上
- 随着智能社会阶段的演进，对泛在算力部署架构的比重也产生了需求变化：云端和边缘端将会在未来实现更为显著的增长，从而在整体算力中占据更高的比例

资料来源：罗兰贝格

图 2.1 第二章的主要观点介绍

2.1 算力衡量指标体系介绍

估算模型



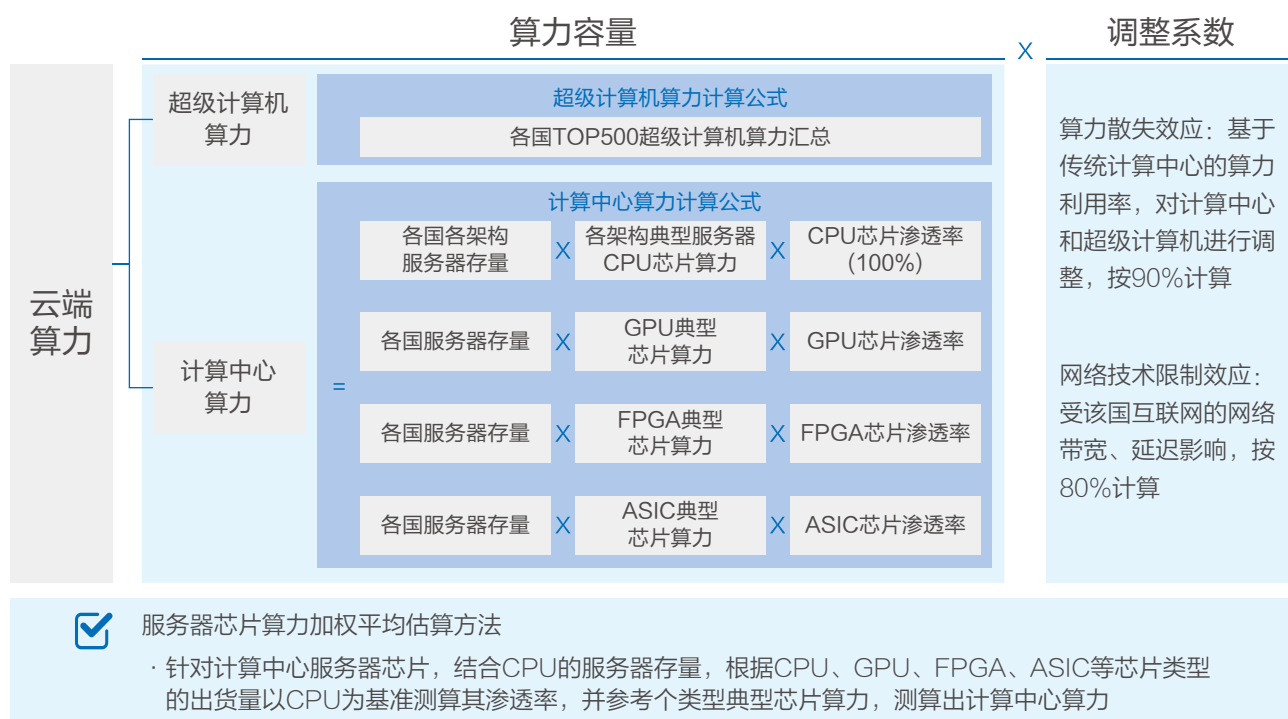
- 基于云、边、端三个层级的相应产品设备形态，对各国整体算力进行估算，同时考量算力散失效应、网络技术限制效应等调整系数对于算力的影响
- 将各国整体算力水平除以各国人口数量，得到各国人均算力水平

资料来源：罗兰贝格

图 2.2 算力衡量指标体系定义

指标体系定义：由于算力的物理承载集中在云端、边缘端、终端三个环节，算力水平应基于云、边、端各层面典型设备的产品形态进行测算，同时考量算力散失效应、网络技术限制效应对于算力的影响，应利用相应调整系数对测算结果进行合理校正（如图 2.2）。首先，算力物理承载集中在云、边、端三个层级，应针对各层级产品形态分别进行算力估算。其次，在测算结果的基础上考量算力散失效应、网络技术限制效应对于算力的影响，利用相应调整系数对算力水平的测算结果进行合理校正（算力散失效应指的是，对于多核处理器，多核心共同计算时的算力输出将小于单核性能相加之和，其主要原因是核数越多，其对共享资源的争抢概率也越高，这些共享资源包括 L3 缓存、内存、QPI 总线等，共享资源的争抢和调度导致了部分算力的衰减；网络技术限制效应是指，云中心本身具有强大的处理性能，能够处理海量的数据，但由于带宽、时延等网络技术限制的影响，海量数据上行与下载遇到瓶颈，算力出现衰竭）。最后，将各国整体算力水平除以该国人口数量，即可得到各国人均算力水平。

2.1.1 针对云端算力，超级计算机和计算中心服务器设备是核心的关注对象，应基于这两种产品形态分别展开算力水平的估算。



资料来源：Gartner, Microway, Nvidia, Ittco, 案头研究；罗兰贝格

图 2.3 云端算力详细估算方法介绍

概括的来看应为：（如图 2.3）

- 1) 超级计算机算力：各国 Top500 超级计算机算力汇总 * 算力散失效应系数 * 网络效率
- 2) 计算中心算力：（CPU 芯片渗透率 × 各国 X86 服务器存量 × 典型 X86 服务器 CPU 算力 + CPU 芯片渗透率 × 各国 ARM 服务器存量 × 典型 ARM 服务器 CPU 算力 + CPU 芯片渗透率 × 各国其他 RISC 服务器存量 × 典型其他 RISC 服务器 CPU 算力 + CPU 芯片渗透率 × 各国 z system 服务器存量 × 典型 z system 服务器 CPU 算力 + CPU 芯片渗透率 × 各国其他架构服务器存量 × 典型其他架构服务器 CPU 算力 + GPU 芯片渗透率 × 各国服务器存量 × GPU 典型芯片算力 + FPGA 芯片渗透率 × 各国服务器存量 × FPGA 典型芯片算力 + ASIC 芯片渗透率 × 各国服务器存量 × ASIC 典型芯片算力）× 算力散失效应系数 × 网络效率

针对超级计算机的算力估算，通过统计世界各国的 Top500 超级计算机的算力总值，计算各国超级计算机的整体算力水平。在实际的计算过程中，亦考量了芯片计算的散失效应（即多核心共同计算时的算力输出将导致 $1+1<2$ 的效果）²⁷，因此假设衰竭比例为 10%；而在网络带宽层面，假设由带宽及延时带来的算力输出损失比例为 20%。

²⁷ 英特尔、高通等公司专家研讨

针对计算中心的算力估算，目前全球共有计算中心 44 万座（包括中小企业设备机房），所安装服务器存量总数约为 5,500 万台²⁸。其中，根据 IDC 公布的各国 x86 服务器历史出货量占比²⁹，估算出各国的计算中心 X86 服务器当前的存量数量。此外，根据各架构服务器的市占率³⁰推算出其他架构服务器的存量。同时当前计算中心服务器主要使用 CPU、GPU、FPGA、ASIC 四类芯片，因此基于四类芯片的市场份额和平均价格，可推算出四类芯片的不同市场渗透率。通过选取四类芯片的典型芯片算力，进而最终估算计算中心服务器的整体算力水平。而在调整系数方面，芯片计算散失效应和网络技术限制带来的损失比例同样假设为 10% 与 20%。其中，应针对不同价格区间的芯片算力和对应的芯片渗透率，估算不同种类芯片的平均算力，如：

- 1) 针对计算中心 CPU 芯片，通过选取不同架构典型芯片计算 CPU 平均算力。四大芯片类型中 CPU 市场渗透率为 100%³¹，安装在所有的计算中心服务器中。同时因搭载 X86 架构芯片的服务器历史较长，早已演化出了不同阶层的性能差异化分群，因此，需要针对 CPU 进行价格区间的细分，通过加权方法算出 CPU 的平均算力（如图 2.3.）。基于 2018 年服务器世界各国分架构存量（X86 架构进一步划分价格区间³²），分别选取典型 CPU 芯片类型（如 Intel Xeon E5、IBM Z14、华为鲲鹏 920 等）估算相应的 CPU 芯片算力，最终通过加权平均的方式估算出计算中心 CPU 芯片的整体算力。（如图 2.4）

²⁸ 中泰证券，2018

²⁹ IDC 《Worldwide Server Market Revenue Declined 11.6% Year Over Year in the Second Quarter of 2019, According to IDC》2019

³⁰ ITCantor, 2018

³¹ AMR 《Artificial Intelligence Chip Market by Chip Type, Application, Technology, and Industry vertical – Global Opportunity Analysis and Industry Forecast, 2018–2025》2019

³² IDC 《Worldwide Server Market Revenue Increased 12.6% Year Over Year in the Fourth Quarter of 2018, According to IDC》2019

服务器CPU算力

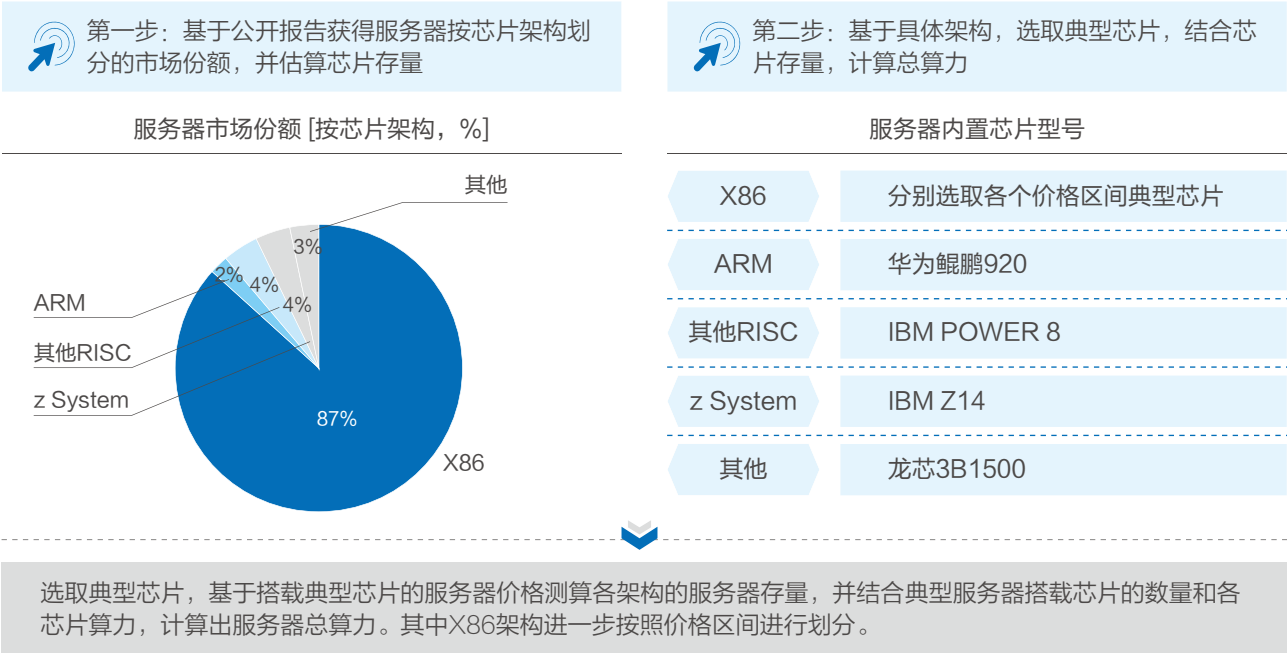


图 2.4 服务器 CPU 算力估算方法介绍

2) 针对计算中心的 GPU、FPGA、ASIC 芯片，由于更为适用于 AI 大规模神经网络并行运算场景于近年逐步兴起，因此 CPU 市场渗透率较低，仅安装于部分计算中心服务器中。GPU，FPGA，ASIC 芯片在当下服务器市场的渗透率分别为 41.7%、20.4% 以及 1.6%³³，同时各品类内不同芯片类型性能相近，因此无需切分价格区间分群进行估算，故在 GPU 领域采用 Nvidia Telsa P4，在 FPGA 领域采用 Intel Stratix 10 GX/SX，在 ASIC 领域采用海思 Ascend 910 型号芯片作为演算基础。

2.1.2 针对边缘端算力，5G+MEC、CDN、路由器服务器、智能监控节点、工业物联网边缘服务器设备是核心的关注对象，应基于这几种产品形态分别展开算力水平的估算

对边缘端算力的估算,首先需要明确边缘计算设备在网络中的位置及部署形式(如图2.5)。边缘计算由智能家居、智能医疗、智能工厂、安防监控、智能能源、游戏娱乐等不同类型的智能应用场景所牵引，比如：1) 在工业互联网场景中，边缘计算与云计算协同工作，在边缘计算环境中安装和连接的智能设备能够处理关键任务数据并实时响应，延时几乎为零；2) 在智慧家居场景中，智能网关与智能终端可作为边缘计算节点对大量异构数据进行处理，再将处理后的数据统一上传到云平台，实现用户对全域智能家居终端的控制；3) 在智慧交通场

³³ AMR 《Artificial Intelligence Chip Market by Chip Type, Application, Technology, and Industry vertical – Global Opportunity Analysis and Industry Forecast, 2018–2025》2019

景中，借助云边协同，可实现全方位车车，车路实时信息交互，保证交通安全，提高出行效率。当下，技术实现方式包括 MEC、边缘云等，而背后本质上需要通过网络侧及现场侧的设备联动方可实现。因此，需要基于边缘计算主要的解决方案形态来测算边缘计算的算力（如图 2.4），网络侧主要通过 CDN 服务器节点等加载边缘服务器，现场侧主要通过智能网关和智能终端进行算力输出（其中智能终端因与人直接交互，移至终端侧进行算力估算）。



资料来源：案头研究，罗兰贝格

图 2.5 边缘算力估算涵盖的目标介绍

所以，基于边缘计算的实现形式，应针对包括 5G+MEC、CDN、智能网关、智能监控、工业物联网在内的多种边缘侧解决方案形态分别进行算力预估。具体的计算逻辑为（如图 2.6）：

- 1) 基于 5G+MEC 形式的边缘计算算力：5G 技术的发展将大幅驱动边缘计算的应用场景的增加，然而当下 5G 网络仍在部署初期，应用尚不成熟，算力暂时按照 0 来进行估算；
- 2) 基于 CDN 形式的边缘计算算力：各国 CDN 节点数量 × 单位节点平均服务器数量 × 典型服务器芯片计算能力 × 网络效率；
- 3) 基于智能路由器形式的边缘计算算力：各国路由器数量 × 智能路由器占比 × 典型服务器芯片计算能力 × 网络效率；
- 4) 基于智能监控形式的边缘计算算力：各国监控节点 × 单位节点的平均服务器数量 × 典型服务器芯片计算能力 × 网络效率；

5) 基于工业互联网形式的边缘计算算力：各国工业互联网覆盖产线 × 单位产线的平均服务器数量 × 典型服务器芯片计算能力 × 网络效率。



资料来源：案头研究，罗兰贝格

图 2.6 边缘算力详细估算方法介绍

针对 CDN 形式的算力估算，目前全球 CDN 节点所部署服务器数量 1,200,000 个³⁴。而在实际研究中，边缘计算解决方案的渗透率约为 100%，单位节点所部署服务器数量为 110 个³⁵。基于世界前 25 家 CDN 厂商的世界各地区 CDN 节点数³⁶，以及各国数字化产业规模占比³⁷，估算出各国的 CDN 节点所部署服务器数量。在实际的研究过程中，选取知名芯片厂商 AMD 厂商的 EPYC7401 作为 CDN 服务器芯片算力的典型代表，网络效率层面的损失比例假设为 5%。

³⁴ GlobalDots 《Content Delivery Network Explained》2019

³⁵ CDN 专业专家研讨

³⁶ GlobalDots 《Content Delivery Network Companies》2019

³⁷ Euromonitor

针对智能路由形式的算力估算，目前全球共有路由器 930,000,000 个³⁸。而随着人工智能、大数据、云计算和物联网等行业的蓬勃发展，当前采用边缘计算方案的智能路由器比重显著上升，当前智能路由器占整体路由器份额的比例约为 15%³⁹，并选取 Dlink OTT 网关 / 路由器作为边缘计算服务器芯片算力的典型代表，网络效率层面的损失比例假设为 5%。

针对智能监控节点形式的算力估算，目前全球监控摄像头数量约 300,480,000 个⁴⁰，根据专家估计，平均每 400 个监控路由数由一个监控节点控制，同时每个监控节点约配置 4 台边缘服务器节点，所以部署服务器数量约 3,000,000 个⁴¹。基于各国数字化产业规模占比⁴²，估算出各国的监控节点所部署服务器数量。在实际的研究过程中，选取海思的 Hikey 970 作为服务器芯片算力的典型代表，网络效率层面的损失比例假设为 5%。

针对工业物联网形式的算力估算，信通院估计，目前全球工业互联网市场规模初步估算达到 32.7 亿美元⁴³，根据专家建议，如平均部署一条工业物联网产线需 150 万美元，且平均每条产线需 1 台边缘服务器。基于各国数字化产业规模占比⁴⁴，估算出各国的工业物联网部署边缘服务器数量。在实际的研究过程中，选取 Intel Xeon Phi 5120D 作为服务器芯片算力的典型代表，网络效率层面的损失比例假设为 5%。

2.1.3 针对终端算力，我们的产品估算对象包括手机、平板、笔记本电脑、台式电脑、智能音箱、可穿戴设备、智能电视、智能机顶盒、智能车载、无人机、游戏机、矿机、智能摄像头等各类智能设备，整体计算逻辑为（如图 2.7）：

智能终端算力：各国设备拥有量 × 各芯片厂商市场份额 × （高端芯片算力 × 高端芯片市占率 + 中端芯片算力 × 中端芯片市占率 + 低端芯片算力 × 低端芯片市占率）

³⁸ IDC《IDC Trackers Show Moderate Growth in Worldwide Ethernet Switch and Router Markets in Q2 2019》2019

³⁹ TPLINK 等行业专家研讨

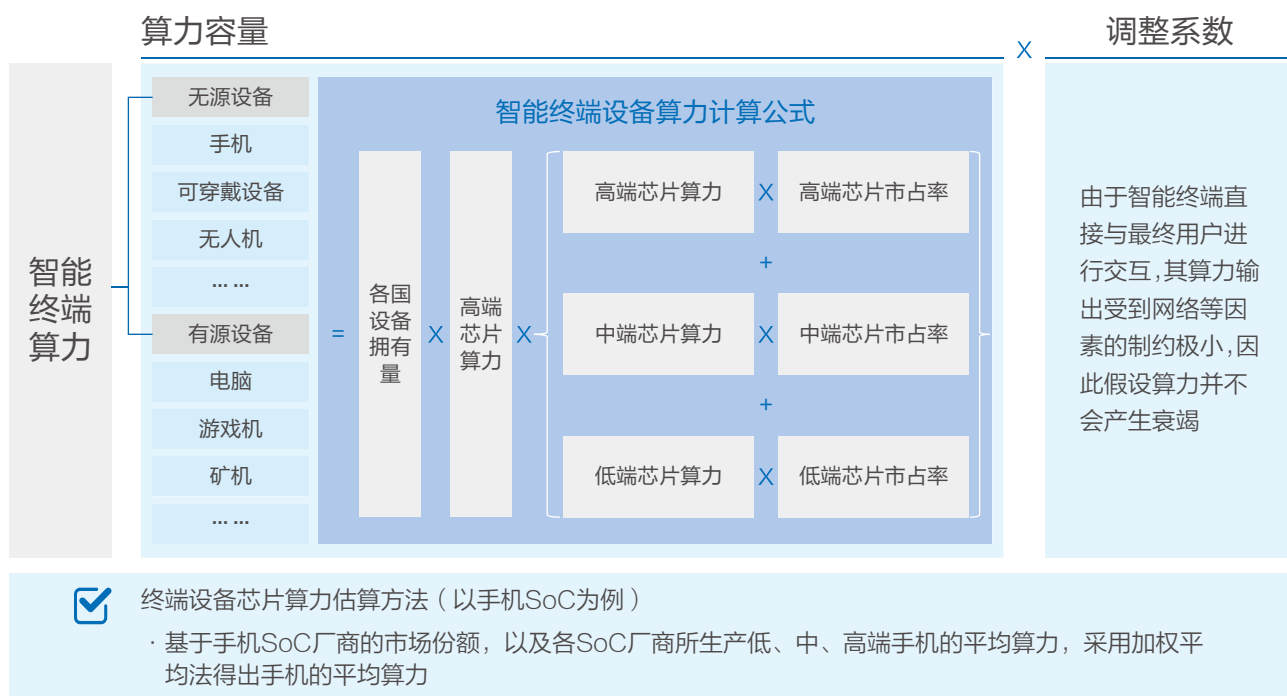
⁴⁰ Euromonitor

⁴¹ 行业专家研讨

⁴² Euromonitor

⁴³ 《工业互联网平台白皮书（2019）》，信通院

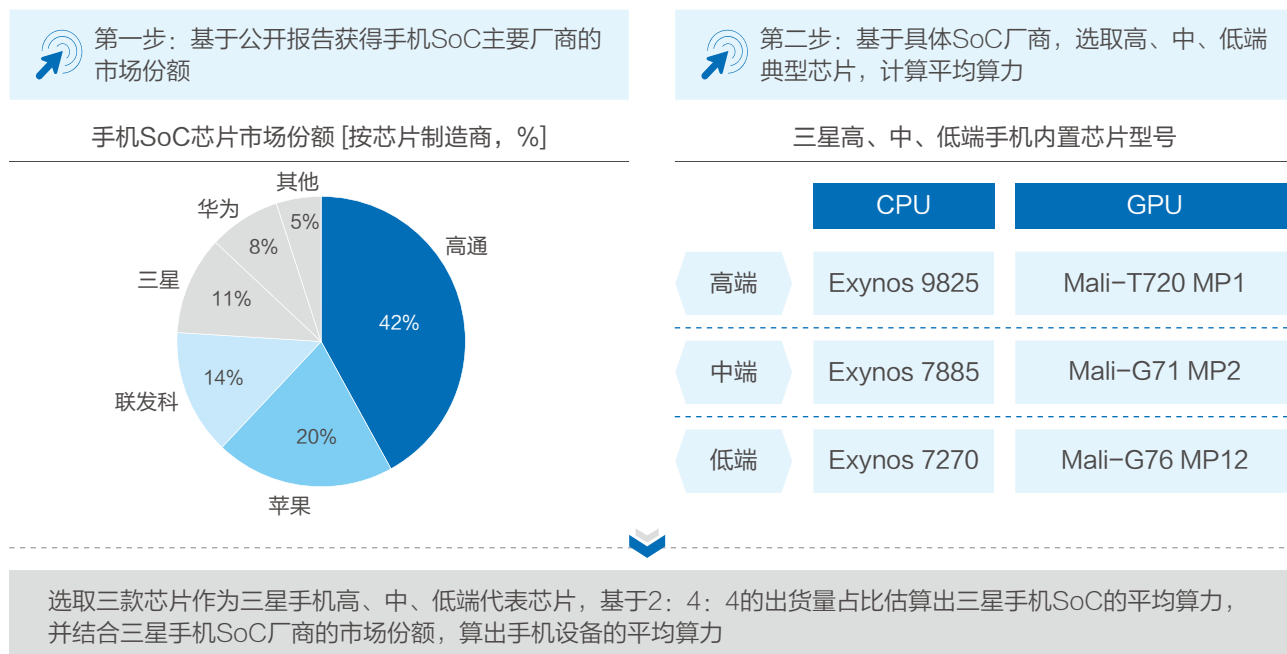
⁴⁴ Euromonitor



资料来源：案头研究，罗兰贝格

图 2.7 终端算力详细估算方法介绍

手机平均算力



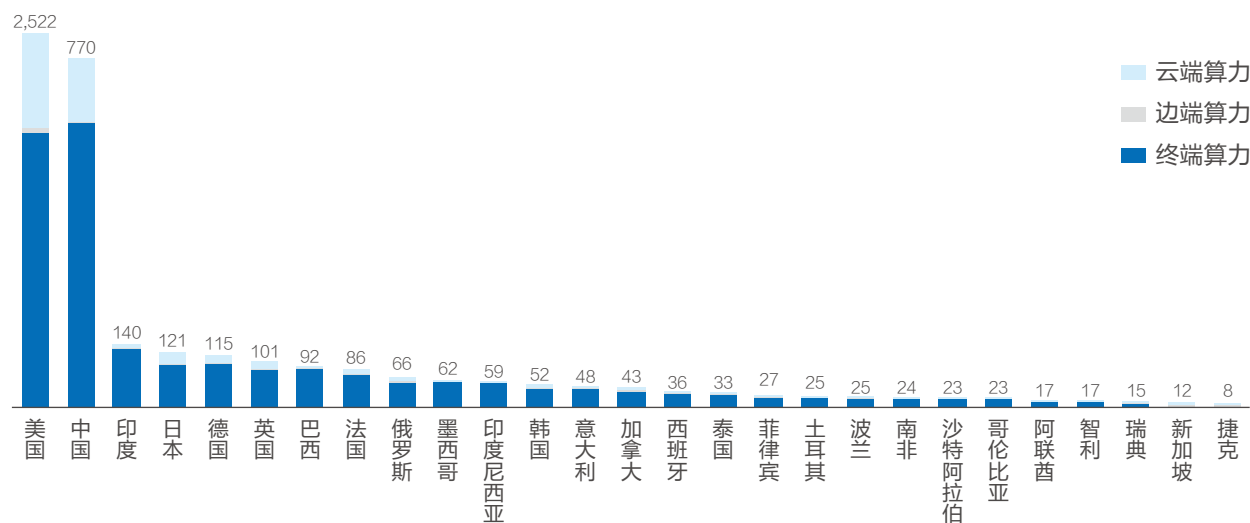
资料来源：案头研究，罗兰贝格

图 2.8 以智能手机为例，终端设备算力考量了芯片市场份额与档次

同时，在针对不同品类的智能终端进行算力估算时，应针对高、中、低端终端设备芯片算力和对应的厂商份额进行考量，得出终端设备加权平均算力水平。在实际演算时，首先得出各国的智能终端设备保有量⁴⁵。其次需要关注各智能设备品类内部，不同核心芯片供应商的保有量占比，并针对不同芯片供应商内部切分高、中、低端的芯片等级，综合考量各自的典型芯片算力，加权平均得出整体的算力水平。以手机为例，全球市场主流的手机 SOC 厂商包括高通、苹果、联发科、三星、华为等。以三星为例，根据芯片计算性能高低不同分别选取 Exynos 9825、Exynos 7885、Exynos 7270 作为高、中、低端手机 SOC 算力的典型代表，同时结合高通、三星等手机 SOC 芯片专家输入为基准，对其高、中、低端的芯片市占率进行了 2:4:4 的结构假设，估算出三星手机 SOC 的整体算力。最后，结合各 SOC 厂商的市场份额，采用加权平均的方式，得出智能手机的整体算力。除手机外，可穿戴设备、无人、矿机等其他智能终端，皆可采取同样方法进行估算。而在调整系数方面，由于智能终端直接与最终用户进行交互，其算力输出受到网络等因素的制约极小，因此假设算力并不会产生损失（如图 2.8）。

2.2 算力指标衡量结论

世界各国总算力水平估算结果（部分示例）[单位：10⁹ GFLOPS]



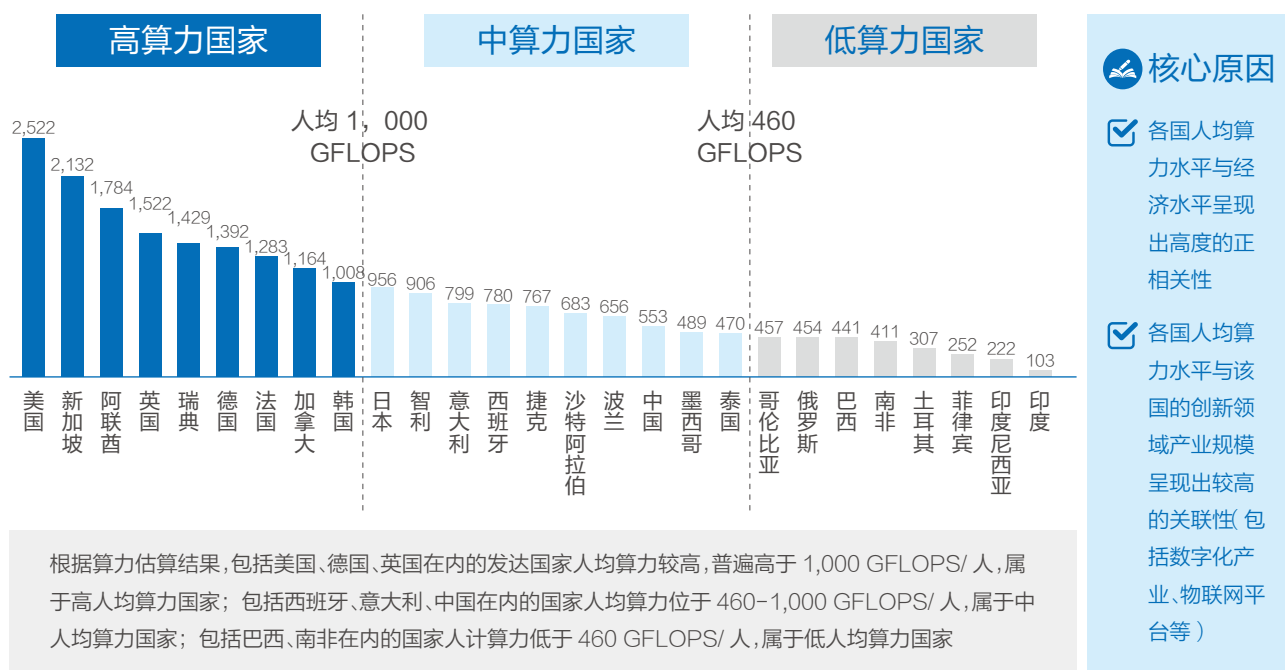
资料来源：案头研究；罗兰贝格

本次报告我们选取了全球一部分国家，对其算力水平进行估算。从总算力来看，美国与中国居首，遥遥领先与其他国家。（如图 2.9）

图 2.9 世界主要国家总算力估算结果

⁴⁵ Euromonitor

世界各国人均算力水平估算结果（部分示例）[单位：人均GFLOPS]



资料来源：案头研究；罗兰贝格

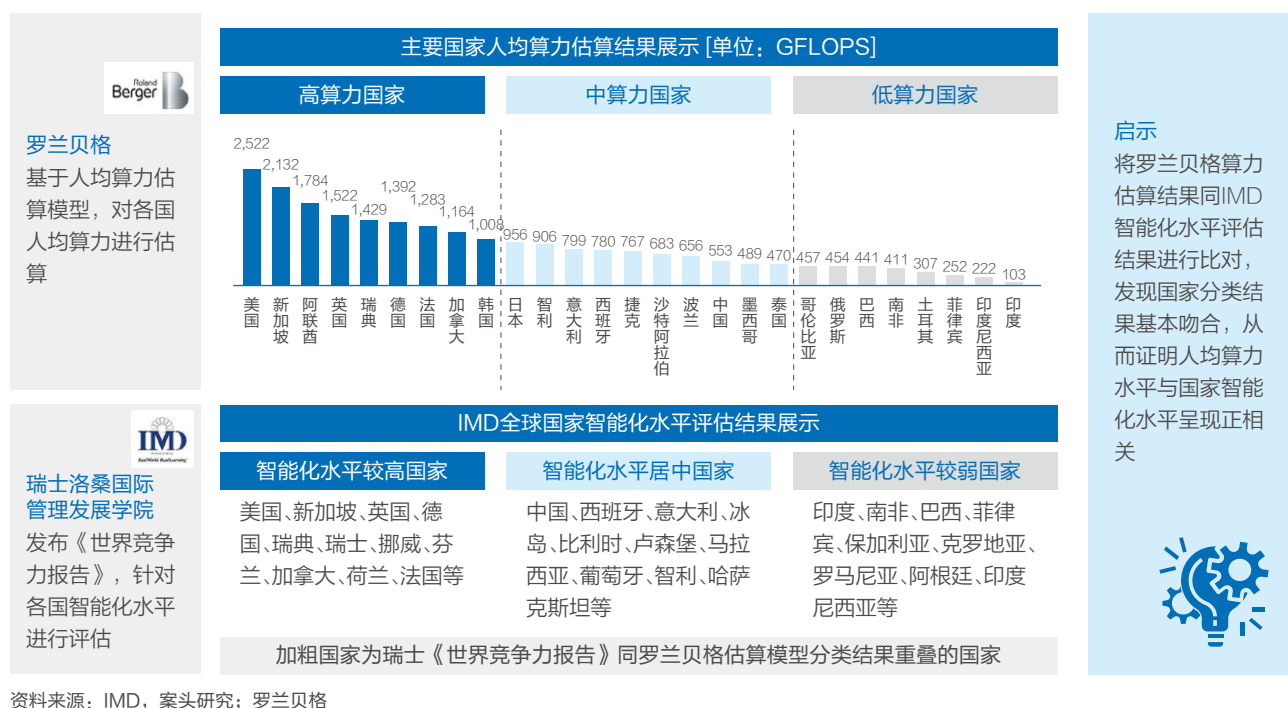
图 2.10 世界主要国家人均算力估算结果呈现出高、中、低的分布

从人均算力水平结果上来看（如图 2.10），新加坡、美国、日本在内的发达国家人均算力较高，普遍高于 1,000 GFLOPS/人，属于高算力国家。该结果与全球超算 500 强榜单的成果接近。在 2019 年 6 月发布的最新一期榜单中，美国的超级计算机算力占比 37.1%，位列第一。其“顶点”和“山脊”以 148.6 PFLOPS 和 94.6 PFLOPS 占据前两名⁴⁶。西班牙、智利、意大利、中国在内的国家人均算力位于 460~1,000 GFLOPS/人，属于中算力国家。其中，中国的超算水平较为突出。2019 年有 228 台超算进入 500 强榜单（排名第一），占 500 强总算力的 32.3%（排名第二）。俄罗斯、巴西、哥伦比亚、南非在内的国家人均算力低于 460 GFLOPS/人，属于低算力国家。这些国家都仅有个位数的超级计算机进入榜单。

从实操性的角度看，该算力水平衡量模型可评估出全球大部分国家的人均算力水平，并显著形成高、中、低的分类结果。通过估算结果可以看出，人均算力水平较高的国家，其经济发展水平位居世界领先水平，同时在包括数字化产业、物联网平台在内的前沿领域产业规模方面同样位居全球前列。以人均算力水平排名前列的美国为例，其经济发展水平世界领先，美国 2018 年人均 GDP 达到 62,641 美元，人均收入达到 63,390 美元⁴⁷；在产业规模方面，2018 年美国每千人数字化产业规模约为 5,685 美元⁴⁸，美国 2018 年在物联网平台的

⁴⁶ Top 500 《China Extends Lead in Number of TOP500 Supercomputers, US Holds on to Performance Advantage》2019⁴⁷ 世界银行《GDP per capita (current LCU) - United States》2018⁴⁸ Euromonitor

市场规模为 270 亿美元⁴⁹，其经济发展水平和前沿领域产业规模均位居国际领先水平。反观人均算力水平排名较后的菲律宾，其经济发展水平相对较为落后，2018 年人均 GDP 仅达 3,103 美元，人均收入达到 10,720 美元；同时在产业规模较为滞后，2018 年每千人数字化产业规模约为 16 美元⁵⁰，菲律宾 2018 年物联网产业规模仅为 1.9 亿美元⁵¹，其经济发展水平和前沿领域产业规模均处于世界较为落后水平。



资料来源：IMD，案头研究；罗兰贝格

图 2.11 世界各国智能化水平与人均算力息息相关

此外，我们发现各国人均算力水平衡量结果与瑞士洛桑学院发布的全球竞争力报告呈现出了高度正相关的关系：瑞士洛桑国际管理发展学院 (IMD) 是国际权威科研机构，其于 1980 年创建国际竞争力评价体系，每年通过发布《全球竞争力报告》对全世界 63 个经济体的智能化水平进行了评估和排名。智能化水平是指一个经济体开发和应用数字技术，对政府实践、商业模式和整个社会进行转型的能力，测度模型包括 3 个一级指标、9 个二级指标和 50 个三级指标，其中一级指标分别为知识、技术和未来准备度三个模块。将 IMD 对于世界各国智能化水平的评估结果，同我们估算的全球各国人均算力水平相比对，可以发现人均算力水平与国家智能化水平呈现高度正向相关：包括美国、新加坡、英国在内的当前智能化水平较高的国家，人均算力水平同样呈现较高的水平；而泰国、印度尼西亚等较低智能化水平的国家，人均算力水平同样位于中等以下（如图 2.11）。

⁴⁹ Statista 《Internet of Things in the U.S. – Statistics & Facts》2019

⁵⁰ Euromonitor

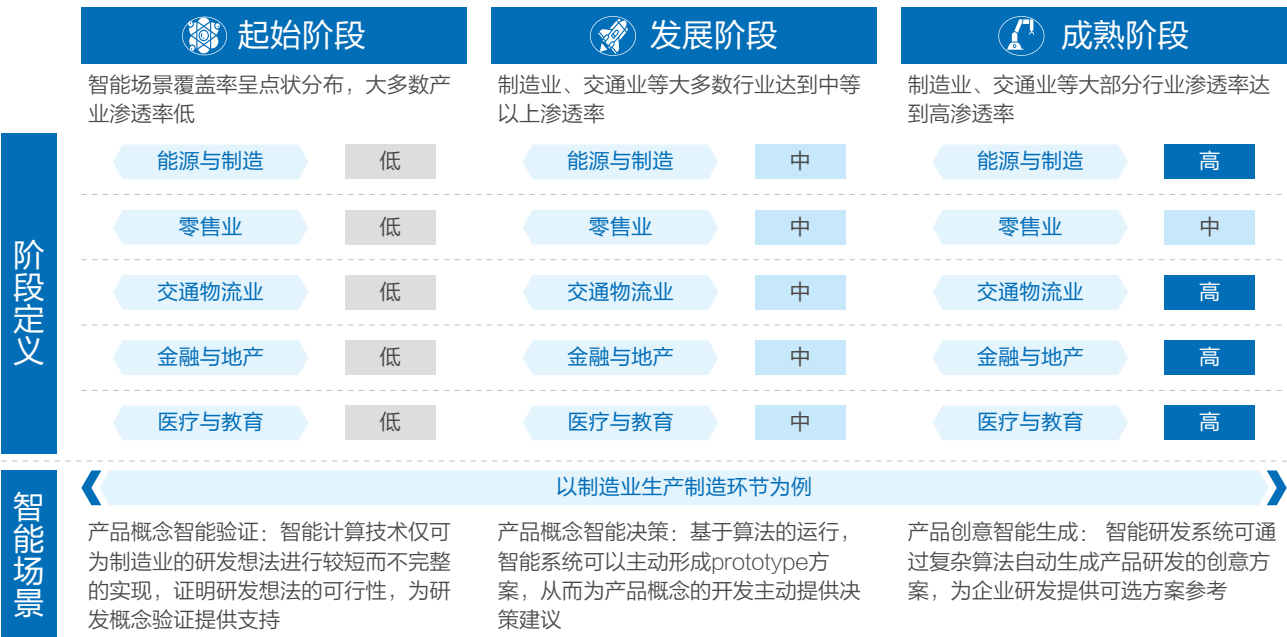
⁵¹ Frost & Sullivan

从中，已可初步看出人均算力水平的高低与一国的经济发展水平形成了紧耦合的绑定关系，是综合国力发展的重要表现；同时，算力水平更是一国智能化、数字化发展水平的集中体现，是一切数字化应用建设及发展的底层基础。

2.3 随着社会智能化进程的演进，对算力水平的需求呈指数级增长

前沿技术推动社会智能化进程的演进，并对算力水平在芯片性能和网络承载等方面提出更高的要求。基于我们设计的算力评估体系，我们对算力未来不同阶段门槛进行预测，发现发展阶段的人均算力需达 10,000 GFLOPS 以上，成熟阶段的人均算力达 29,000 GFLOPS 以上。

2.3.1 智能社会不同阶段的定义及其需要的算力水平估算



资料来源：埃森哲，波士顿咨询，国际机器人学联合会，麦肯锡，Signify Research，Synergy Research，案头研究，罗兰贝格

图 2.12 智能社会不同阶段定义⁵²与智能场景⁵³下算力水平存差异

⁵² 根据全球知名调研机构 Gartner 所述，随着智能渗透率的提升，行业逐步从新兴期发展到成长期、成熟期：新兴期的行业的技术能力、方法、基础设施和生态系统开始发展，增速最快，同时技术不确定性也较高，渗透率一般超过 1%；成长期的行业的技术得到验证，在大多数环境中其价值可相对预测，业务增长和盈利均较好，实现 1% - 5% 的渗透率；成熟期的行业的技术经过验证，行业价值主张得到良好理解，技术可实现大规模商业化应用，智能渗透率一般超过 20%。从新兴期到成熟期的时间间隔在 10 年以上。考量到绝大多数国家需要进入发展 / 成熟阶段，智能渗透率高于 Gartner 所定义渗透率 (5% 和 20%)，即发展阶段平均渗透率为 8%，成熟阶段平均渗透率为 30%。主要国家于 2035 年进入发展阶段，于 2045 年进入成熟阶段

⁵³ 通过选取各主要行业的典型场景，如金融业的智能投顾场景、制造业的工业机器人场景、交通物流业的自动化驾驶场景、医疗业的智能读片场景等，可估算起步阶段、发展阶段、成熟阶段的智能渗透率

算力高低标志着智能社会不同阶段，并随着智能社会的发展而显著增高。算力的发展同时受产业覆盖率、产业渗透率、算力性能自然增长三个因素所影响：针对产业覆盖，智能化场景（即采用人工智能、物联网、大数据、区块链、AR/VR 等创新前沿技术的场景）在各国各产业的覆盖率，如智能驾驶在交通行业的覆盖率，计算公式为应用智能行业数量 / 总行业数量；针对产业渗透，智能化设备在某一特定产业的渗透率，如智能化驾驶设备在驾驶设备的渗透率，计算公式为智能应用场景行业内部智能产品出货量 / 该行业总出货量⁵⁴。针对算力性能自然增长，目前我们看到芯片制程的摩尔定律虽放缓，但仍在继续，需要在算力预估中考量单位成本下算力性能随技术发展自然增长的因素；如无重大技术突破，预计在 10 年后摩尔定律开始失效，故约 2029 年后算力性能自然增长因素将不考量⁵⁵。

基于行业发展不同阶段对应渗透率，智能社会可区分为三个阶段，即起始阶段、发展阶段及成熟阶段（如图 2.13）。

⁵⁴ 针对行业内智能渗透率，等于智能产业产值 / 该行业总产值

⁵⁵ 进一步地，我们考虑到摩尔定律的存在，并在未来十年内（2019-2029）放缓的可能性，推测单位成本下的芯片计算性能可能会随着智能社会的发展而降低。以用于台式电脑的英特尔 i7 处理器为例，通过在 2009-2019 年间的各年度选取当年出售的最顶级的芯片的单位成本计算性能，发现该自然增长系数随时间呈现下降状态，年份和增长倍数的耦合函数为反函数。2029 年后除非出现重大技术突破，将不再考虑摩尔定律，即算力自然增长倍数较上一年为 1

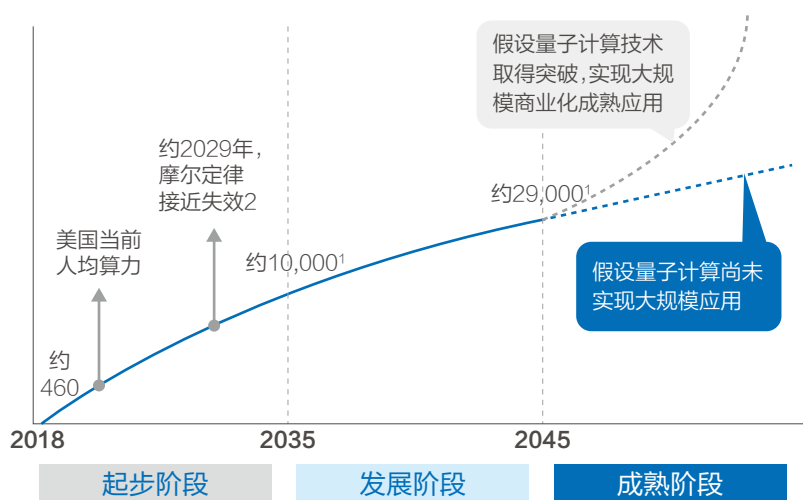
智能社会三大发展阶段算力水平



主要国家/地区人均算力增长趋势 [单位：人均GFLOPS]



不同阶段算力水平预估



- ✓ **发展阶段**：预计2035年到达发展阶段，主要国家/地区的智能化渗透率约8%¹，该阶段人均算力约为10,000 GFLOPS以上
- ✓ **成熟阶段**：预计2045年到达成熟阶段，主要国家/地区的智能化渗透率约30%¹，该阶段的人均算力约为29,000 GFLOPS以上
- ✓ 此外，如量子计算提前实现大规模商用，则阶段实现时间可能提前

1) 从渗透率的角度，5% 和 20% 是发展阶段和成熟阶段的临界点，我们考虑到国家 / 地区间的发展差异，故延迟约 5 年左右，选择了 8% 及 30% 作为实际主要国家 / 地区迈入两个阶段的标志

2) 目前摩尔定律已经开始放缓，未来如无重大技术突破，预计将于 2029 年接近失效

资料来源：案头研究；罗兰贝格

图 2.13 伴随智能化渗透率提升，智能社会三大发展阶段人均算力也不断升高

1) 起步阶段：基于算力估算模型，世界各国均处于该阶段，人均算力均低于 3,000 GFLOPS，其中美国人均算力为 2,522 GFLOPS，新加坡人均算力为 2,132 GFLOPS（如图 2.13）。各产业的智能场景覆盖率较低，覆盖上呈点状分布，同时在特定的智能场景中，智能设备产品的渗透率较低，主要国家 / 地区各行业智能渗透率不到 8%。在此阶段下，尽管人工智能等基于智能计算的前沿技术在各个行业的部分环节与流程中已经有了一定程度的应用，但由于各国对智能技术发展的投入有限，产业链较为复杂、各环节机械化程度水平参差不齐等因素影响，整体智能化程度处于较低水平。在当下 4G 为主导的⁵⁶时代，移动网络⁵⁷带宽实现峰值传输速率约 150Mbps，网络时延约 50 毫秒，每平方公里可支持约 10 万台设备接入；而在起步阶段后半段时，网络技术已全面迈入 5G 时代，网络带宽⁵⁸大幅提升至约 10Gbps，网络时延低于 10 毫秒（将能够完全满足 VR 交互业务（10 毫秒 - 30 毫秒）和智慧家庭（20 毫秒 - 40 毫秒）的延要求⁵⁹），每平方公里可支撑 100 万台 5G 设备接入（包含 NB-IoT 连接，基本满足海量低功耗物联网设备接入需求）。

⁵⁶ 以网络代际演变差异度更高的移动网络为主

⁵⁷ 在起步阶段初期，宽带带宽约 500Mbps

⁵⁸ 在起步阶段后半段，宽带与移动网络的带宽无显著差异，故不进行区分

⁵⁹ 中国信息通信研究院《千兆宽带网络商业应用场景白皮书》2019

- 2) 发展阶段：基于我们的测算，预计对于主要国家 / 地区，人均算力达到约 10,000 GFLOPS 以上时，该地区将步入智能社会的发展阶段。智能场景在制造业、智能驾驶行业、政务行业等产业已实现一定数量的垂直智能行业覆盖，主要国家 / 地区大部分行业的智能渗透率在 8~30%。在该阶段下，芯片性能需求已经突破集成度的限制，同时决策智能、人工智能云服务、智能机器人等技术相对发展成熟，通过人工智能等基于智能计算的前沿技术全面赋能各大产业领域。在此阶段下，技术形态获得大幅度提升，网络技术顺利迈入 6G 时期并逐步成熟，网络带宽提升至 100Gbps–1Tbps⁶⁰ 之间，时延大幅下降至 <0.1 毫秒（可充分保障车辆控制信号端到端的传输⁶¹，并轻松满足在 VR/AR 场景中沉浸式体验所要求 <8 毫秒⁶² 时延要求），每平方公里可支持最多 1 亿台 6G 设备接入。同时，网络技术的进步将会促进包括 3D 即时式生产、智能研发测试、高自动化驾驶等智能场景的出现，显著优化生产周期和效率并降低人工成本，达到安全、环保、高效的目的，从而引领产业生态及商业模式的全面升级与重塑。
- 3) 成熟阶段：基于我们的测算，预计对于主要国家 / 地区，人均算力达到约 29,000 GFLOPS 以上时，该地区将步入智能社会的成熟阶段。智能场景已实现所有行业的全方位覆盖，主要国家 / 地区大部分行业内的智能渗透率均已达到或接近 30% 以上的水平，例如，在交通业等行业中，智能渗透率达到 90% 以上，人类从此进入高等智能社会。比 6G 更为先进的网络标准将实现商用，网络带宽将实现远超 1Tbps 级别的水平，网络时延则可忽略不计，每平方公里设备接入数无明显限制。在该阶段下，芯片性能上可以实现量子计算，同时脑机接口、全息投影等前沿技术发展成熟，智能计算全面赋能各大产业领域，实现包括研发、生产、制造、售后等各产业全场景智能化覆盖。

⁶⁰ University of Oulu 《Key drivers and research challenges for 6G ubiquitous wireless intelligence》2019

⁶¹ GIV 《千兆宽带网络商业应用场景白皮书》2019

⁶² 中国信息通信研究院《千兆宽带网络商业应用场景白皮书》2019

未来发展趋势研判

1 供给端

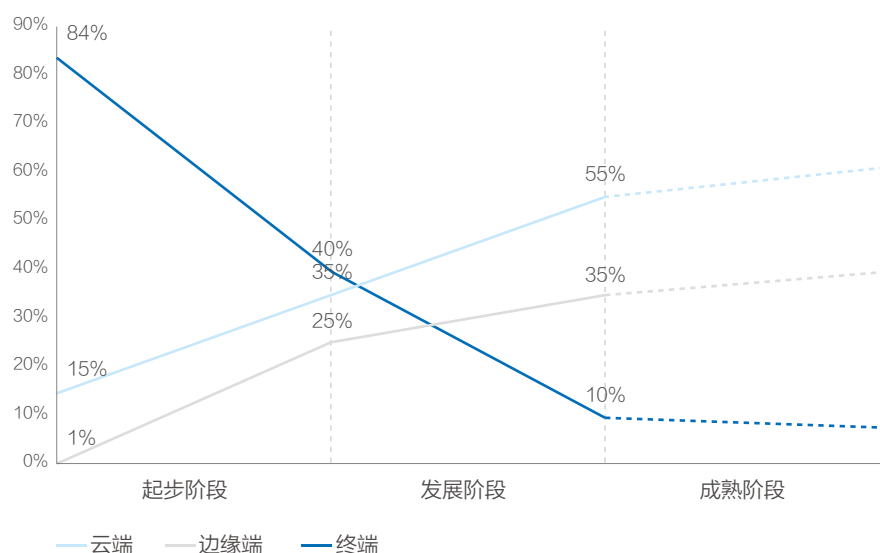
- 未来网络将会突破当前的发展限制,以更高的效率承载更多的数据传输

2 需求端

- 电子终端的需求逐步向便捷化、一体化发展,需要采用更加轻薄的材质设计,意味着其承载的算力将会逐渐降低



云边端算力占比发展趋势 [单位: %]



资料来源: IDC《数字化世界: 从边缘到核心》, 案头研究, 罗兰贝格

图 2.14 云端和边缘端将会在未来实现更为显著的增长

此外, 我们还看到, 随着智能社会阶段的演进, 对泛在算力部署架构的比重也产生了需求变化: 云端和边缘端将会在未来实现更为显著的增长, 从而在整体算力中占据更高的比例。基于我们的测算, 在 2045 年达到成熟阶段时, 终端算力仅占据 10%, 而云端和边缘端将分别占据 55% 和 35%⁶³。目前大部分数据都在终端创建, 但是未来越来越多的数据将会在云端及边缘端进行复制、传输或备份⁶⁴。现实生活中已经有越来越多的智能场景应用来验证上述发展趋势的实现。例如 Google 已经正式发布了云游戏平台 Stadia, 其在服务器上运行游戏, 再把游戏画面从云端传输到用户的设备, 从而为包括笔记本电脑、平板、手机、电视盒子在内的终端设备产品的算力加载减负。我们可以从技术供给端和用户需求端两个层面对此发展趋势做出解释:

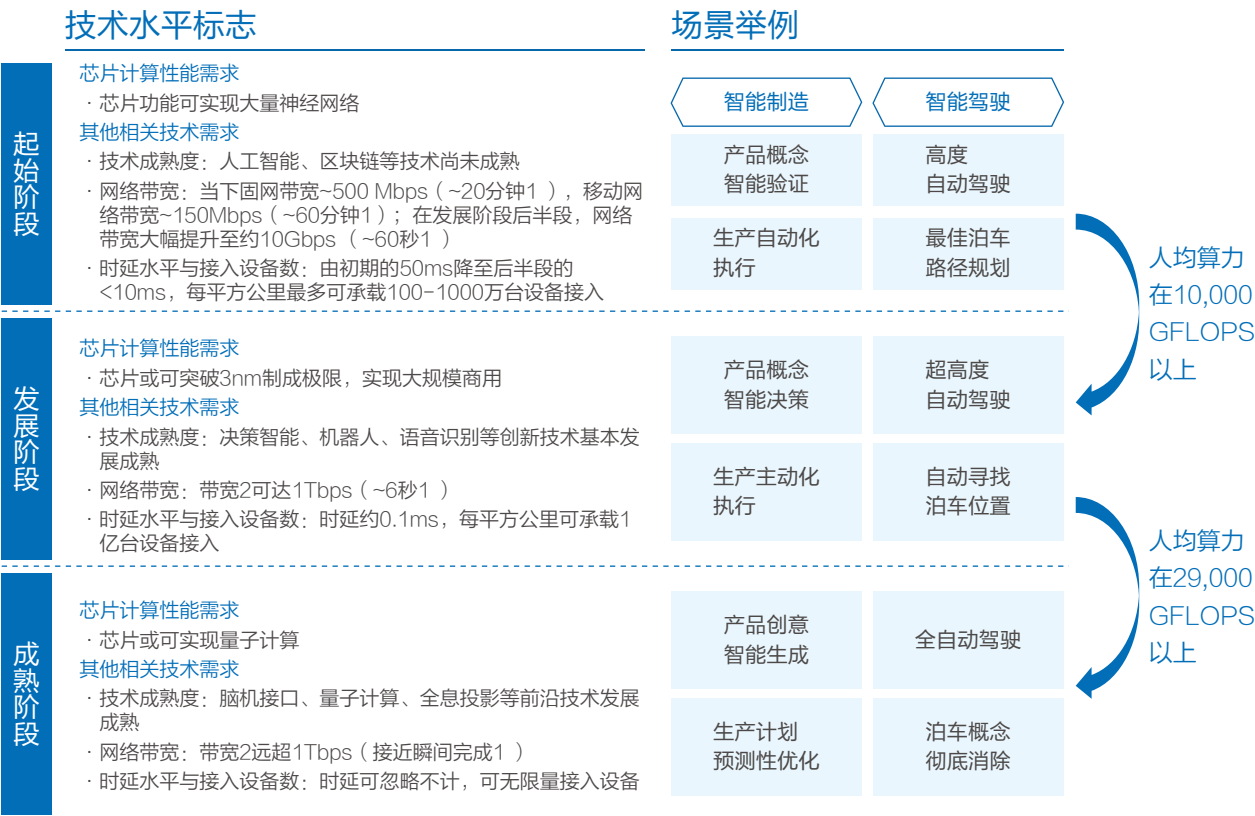
- 1) 从需求端的角度来看, 电子终端的需求逐步向便捷化、一体化发展, 智能电子设备需要采用更加轻薄的材质设计, 意味着终端设备产品承载的算力必将呈现出逐渐减少的趋势
- 2) 从供给端的角度分析, 未来网络将会突破当前的发展限制, 以更高的效率承载更多的数据传输, 因此必将承载更多的算力;

⁶³ IDC《数字化世界, 由边缘到核心》2018

⁶⁴ IDC《Cloud IT Infrastructure Revenues Continue to Expand Despite Slow Down in Spending in 2019, According to IDC》2019

2.3.2 智能社会不同阶段对于算力的具体需求及典型场景展望

智能社会不同阶段技术水平标志与场景举例



1) 以下载一部时长约2小时、4K分辨率、蓝光无损格式的电影所需要的时间为例，文件大小按~60GB计算；2) 此估算基于在发展阶段及之后，宽带与移动网络的带宽无显著差异

资料来源：案头研究，罗兰贝格

图 2.15 随着不同阶段技术水平标志的提升，人均算力水平显著提升

通过一些行业在不同智能阶段的表现形态，阐述算力在其中的重要作用（如图 2.15）。

— 智能制造行业步入智能社会不同阶段的技术形态及场景概述



资料来源：案头研究，罗兰贝格

图 2.16 随着智能制造行业步入智能社会不同阶段，对应场景的智能化程度不断提高

- 1) 起步阶段的场景概述：由于技术形态上的限制导致该阶段的制造业智能化程度有限，只能实现例如模式化智能生产的初级智能化场景（如图 2.16）：具体来看，该阶段的芯片性能可实现大量神经网络，因此可完成大数据量收集后的算法训练，因此可实现对设备的动态调整，大幅提升流水线、装配线等不同工序设备的配合熟练度，从而提升整体制造效率。伴随着人工智能云服务等技术成熟，智能工厂可将大量数据上传至云数据中心进行存储及分析，为数据的测算打下基础，从而实现智能工厂芯片算力的大量输出，通过数据分析赋能智能生产流程优化及生产效率提升。1Gbps 以上的带宽水平和 ~10 毫秒的延迟水平可以满足数据的同步传输，从而可以实现对生产模块中混线生产的定制半成品、成品进行实时追踪、组合配置与调度，实现工厂的柔性化生产，并解决定制化生产带来的不确定性、多样性和复杂性问题。
- 2) 发展阶段的场景概述：技术形态上的成熟导致了快速研发、自主创新、人机协作等制造业智能化场景的出现，制造业从此进入中等智能化时代。芯片集成度上的突破导致芯片性能的显著提高，芯片的神经网络展现出强大的信息提取能力，可以达到芯片结构简洁、计算快速的性能，从而可实现智能研发场景下的快速研发能力。从技术上看，随着智能决策等技术的发展成熟，智能系统可以主动形成 prototype 方案，从而为产品概念的开发主动提供决策建议。网络承载水平的提高可满足工业操作中对实时性体验的需求，智能系统可脱离人为监控和指导，以高效率完成生产订单的执行，从而独立进行高质量产品的制造和生产。

3) 成熟阶段的场景概述：技术形态上的持续提高，导致产品创意智能生成、需求计划预测性优化等场景的出现，制造业从此进入全面智能化时代。芯片性能已实现量子计算，因此智能系统可通过复杂算法对宏观经济发展、科学发展规律、社会政治事件、流行病概率等因素进行预测，自行判断行业外部及企业本身的未来发展趋势，并根据分析结果对需求计划进行优化。从技术成熟度和网络承载上看，智能机器人技术相对成熟，且时延水平可忽略不计，可以通过云技术机器人可将大量运算功能和数据存储功能移至云端，大大降低机器人本身的硬件成本和功耗。

智能驾驶行业步入智能社会不同阶段的技术形态及场景概述



资料来源：案头研究，罗兰贝格

图 2.17 随着智能驾驶行业步入智能社会不同阶段，对应场景的智能化程度亦不断提高

- 1) 起步阶段的场景概述：由于技术形态的限制，只能实现例如感知预警、三维图像模拟的初级智能化场景（如图 2.17）。由于芯片性能受限，算力的神经网络仅可实现自动驾驶情境中的视觉控制。同时，L3 自动驾驶等智能化技术发展成熟，可实现初级的辅助功能，比如盲点监测、车道保持、自动巡航、刹车辅助等，但在特殊条件下期望人类驾驶员能正确响应请求并接管操控。移动网络带宽迈向 10Gbps，中央处理器可根据汽车与目标停车位的相对位置等数据，得出汽车的当前位置、目标位置及周围的环境参数，据此规划计算出最佳泊车路径和策略
- 2) 发展阶段的场景概述：随着技术形态的提高，基本实现高自动化驾驶、最佳泊车路径规划等智能化场景，智能驾驶从此进入中等智能化时代。随着芯片在集成度和性能上的突破，芯片的算力获得极大提升，以自动泊车场景为例，智能系统开发出基于深度学习的系统，可以从视频信息中自动检测车辆及其边界，从而可以准确地检测到开放的停车位。同时，随着决策技能、深度神经学习等技术的完善，自动驾驶系统可以在特定路

况条件下完成全部驾驶操作，即便人类驾驶员无法正确响应请求并接管操控。网络带宽向 1Tbps 靠拢，可实现车联网信息的存储，提高通信概率，极大地增加车载之间的通信机会，实现车与车之间消息的高效传递。

- 3) 成熟阶段的场景概述：随着技术形态的提高，可实现云端车辆控制等智能化场景，智能驾驶从此进入高等智能化时代。由于芯片性能可实现量子计算，以云端车辆控制场景为例，车载中央处理器可根据周围环境以及路况采集信息迅速进行计算处理，识别紧急场景并做出响应。脑机接口等技术达到高度成熟，市民通过脑中植入芯片就可轻松探测并使用周围可以共享的空闲车辆，到达目的地后可即时将车辆共享给其他人使用，无需为寻找停车位置而烦恼，泊车概念完全消失。传输时延降可忽略不计，驾驶权上传至云端后，云端可实时掌控车辆与周围环境数据，实现对车辆高效操控，避免因信息之后而产生的意外。

3

第三章 算力对各国家 / 区域的效益分析

第三章的主要内容

主要内容

3.1 对算力的投入可以带来可观的直接性经济回报

- 作为具有高附加值的高科技行业，针对云、边、端物理承载形态的算力进行投入，可为各国带来直接性经济回报，反映在智能计算产业产值的增长
- 以公有云云服务为例，伴随着1美元的投入，可以带动约4.7美元的直接产业产值增长

3.2 对算力的投入可对众多行业带来极大杠杆效应的延伸性经济增长

- 由于算力投入带来的数字化智能技术加持，可为各国的制造、交通、医疗、能源等多个行业带来延伸性经济增长，体现在产业产值增长、生产效率提升、商业模式创新、用户体验优化等方面
- 以智慧工厂为例，制造业中每1美元的投入可带来约10美元的延伸产业产值增长

3.3 此外，对算力的投入亦能从创新、公共服务、民生三方面提升国家整体竞争力

- 此外，算力投入可提升国家科研创新水平、帮助政府提升政务效率、改善民生方面的体验，极大提升国民的满意度和幸福感，提升国家整体竞争力

资料来源：罗兰贝格

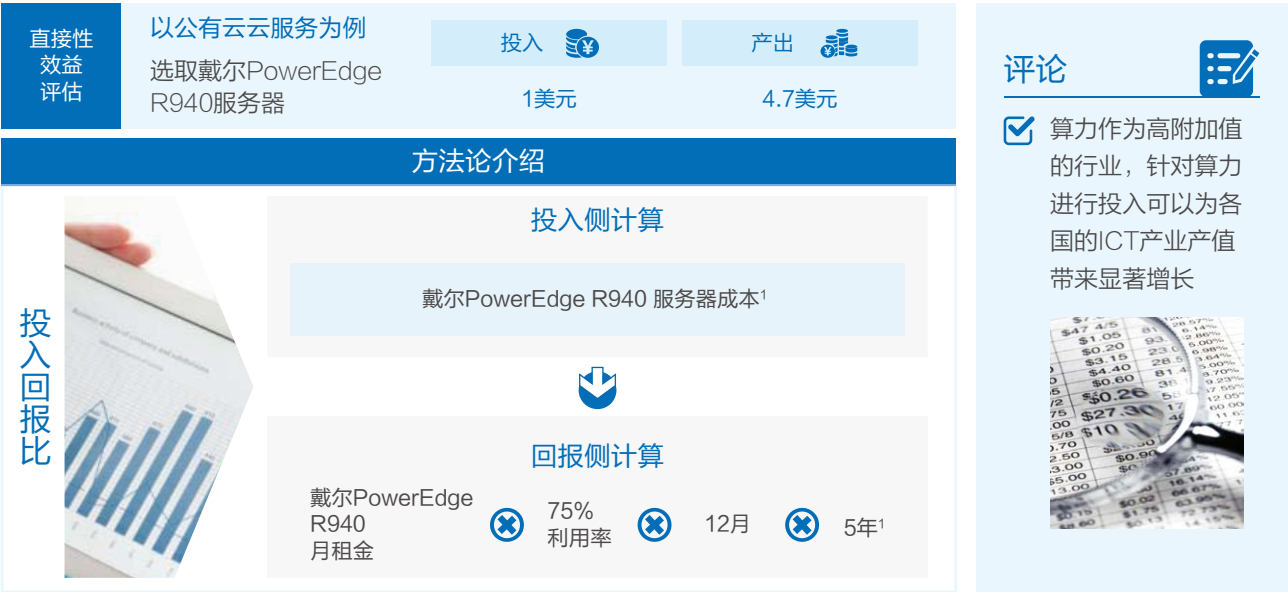
图 3.1 第三章的主要观点介绍

3.1 对算力的投入可以带来可观的直接性经济回报

作为具有高附加值的高科技产业，针对云、边、端物理承载形态的算力进行投入⁶⁵，可以为各国带来直接性经济增长，反映在云、边、端智能计算产业产值的增长，即包括数据中心、服务器、CDN、智能终端等产品形态的销售收入、租赁收入及服务收入。举例来看，云端产业产值主要来源于 IaaS（设备即服务）、PaaS（平台即服务）和 SaaS（软件即服务），各组织单位可按需使用以上各类云服务，则使用服务带来的付费收入即可算作云端产值之一。从实际数据来看，在芯片、服务器、智能终端、数据中心等方面对算力进行投入，可直接带来云、边、端智能计算产业对国家 GDP 增长的直接贡献。

⁶⁵ 对于算力的投入可以从两个维度进行理解：1）一方面，与算力有关的投入领域主要包含服务器、数据中心、边缘服务器、智能终端、芯片、高速网络等；2）另一方面，从投入方式来看，主要包含研发投入、制造投入、人才培养投入、部署实施投入等。以此，对以上任一领域采取的任一投入抓手皆可被认为是对于算力开展的直接投入

直接性经济增长的结果与方法论介绍



1) 主要统计公有云服务器的硬件成本

资料来源：中关村在线，AWS官网，案头研究，罗兰贝格

图 3.2 直接性经济回报估算方法介绍

以公有云云服务场景中的云服务器为例（如图 3.2.），通过计算服务器厂商的硬件设备成本和云服务解决方案提供商的租金收入，可推算出该场景下的投入回报比：从投入来看，戴尔 PowerEdge R940 服务器的出厂价约为 25,100 美元；从产出来看，假设云服务器的使用周期为 5 年，云服务器实现 75% 利用率，则租用戴尔 PowerEdge R940 服务器的租金收入约为 118,586 美元。因此从经济效益上来讲，伴随着 1 美元的投入，可以在云计算行业带动约 4.7 美元的直接产业产值增长⁶⁶。

3.2 对算力的投入可以对众多行业带来极大的杠杆效应，创造巨大的延伸性经济增长

由于算力投入带来的数字化智能技术加持，算力亦可以为各国带来延伸性经济增长，可以为包括制造、交通、医疗、能源在内的多个行业带来显著效益，具体表现在产业产值增长、生产效率提升、商业模式创新、用户体验优化等方面。举例来看，随着智能机器人等基于智能计算的新兴技术在智能制造领域的不断进步，预计未来

⁶⁶ AWS《AWS Storage Gateway》2019

将为美国制造业带来 1,700 亿美元的额外经济产值提升⁶⁷。生产产值的提升可以体现在宏观和微观层面：从宏观来看，算力投入可以带动整个智能产业规模的提升，如智能驾驶系统的投入可以带动整个交通物流行业规模增长；从微观来看，算力投入可以带动特定智能场景的产值提升，如数字化硬件设备的投入，可以提升 Amazon Go 无人便利店的营业额提升。

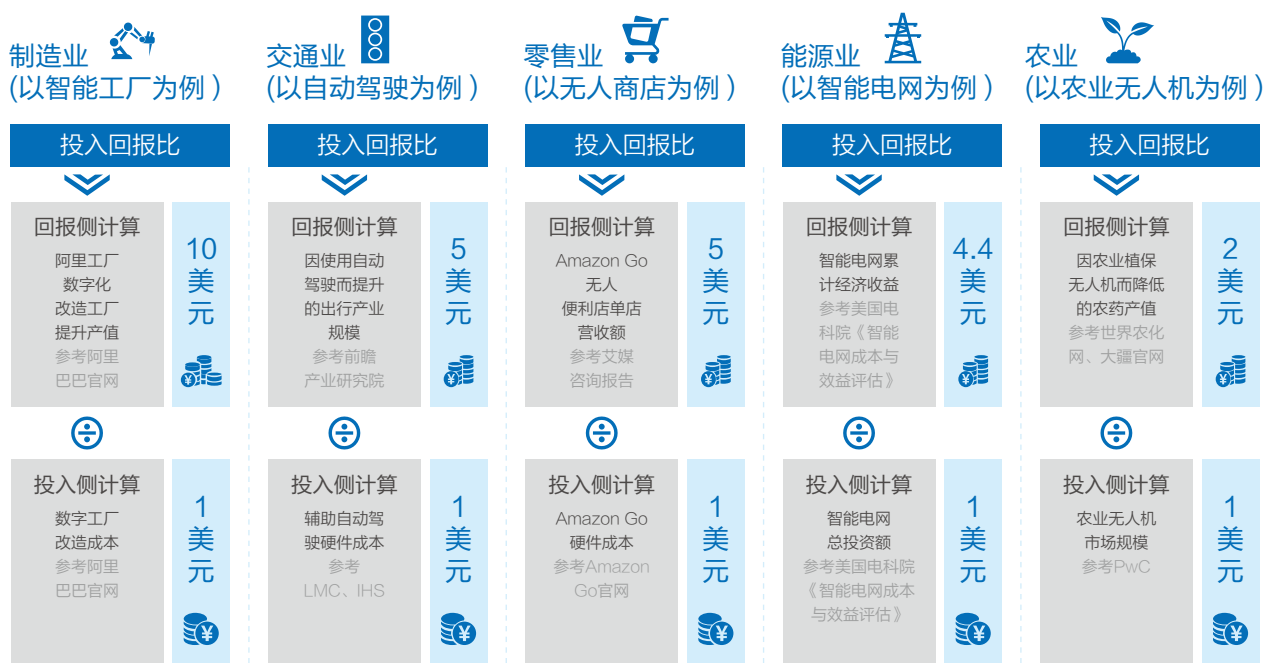


图 3.3 算力投入对大部分行业产值都有较为显著的延伸带动性

具体来看，算力投入对包括制造业、交通物流业、零售业在内的大部分行业产值都具有较为显著的带动性。当前各国在能源、文娱等行业均对算力进行积极投入，并实现了显著的经济效益，综合提升国家实力（如图 3.3）。

1) 对于制造业而言，选取阿里巴巴改造的首间数字化工厂——“杭州点石工厂”为例，50,000 元的工厂数字化改造成本（包括服务器、传感器、机器人等硬件成本，不包括软件、人力、网络部署等成本）作为算力投入；改造后工厂年产值 160 万元，数字化提升产效 6%，假设投入硬件的使用周期为 5 年，则对应的 96,000 元的工厂数字化带动产值为算力回报。从经济效益上来看，这意味着每 1 美元的投入，可以带动 10 美元的智能制造产业产值提升⁶⁸。

2) 对于交通物流业而言，以自动驾驶汽车为例，2018 年中国乘用车销量为 2370 万辆，基于 ADAS 等系统

⁶⁷ 美国劳工部

⁶⁸ 阿里巴巴《淘工厂数字化：五分钟产 2000 件衣服，以前是同样的，今后要不同》2018

而实现 L3 级别以下辅助驾驶的汽车渗透率为 19%，单车自动驾驶硬件成本（包括计算平台、摄像头、激光雷达等）为 23,725 美元，则价值 7,585 亿元的辅助自动驾驶硬件成本作为算力投入；中国交通出行市场规模为 5936 亿元，自动驾驶提升交通出行效率 60%，假设使用周期为 10 年，则对应的 35,616 亿元的自动驾驶带动交通出行行业产值规模为算力回报。从经济效益上来看，这意味着每 1 美元的投入，可以带来 5 美元的交通物流业产值提升⁶⁹，说明算力投入可以明显带动交通物流业产值的增长。

- 3) 对于零售业而言，以亚马逊的无人便利店——“Amazon Go”为例，价值 147 万美元的单店硬件成本（包括 RFID 投入、电脑、摄像头等）作为算力投入；Amazon Go 的单店年营业额为 150 万美元，假设硬件的使用周期为 5 年，则价值 750 万美元的单店五年营业额为算力回报。从经济效益上来看，这意味着每 1 美元的投入，可以带来 5 美元的零售业产值提升⁷⁰，说明算力投入可以明显带动零售业产值的增长。
- 4) 对于能源业而言，以智能电网为例，根据美国电力研究院的研究表明，如在美国建设未来现代化电力输送系统，预估投资额应在 3380-4760 亿美元之间，预计收益将达到 1.294-2.028 万亿美元，从经济效益上来看，每 1 美元的投入，可以带来约 4.4 美元的能源业产值提升⁷¹，可见算力投入可以明显带动能源业产值的增长。从具体的案例来看，南方电网深圳供电局通过与华为⁷²合作，利用人工智能物联网（AIoT）技术，通过部署输电视频监控终端进行视频巡检，改革传统的人工巡检，使得巡检周期从 20 天缩短至 2 小时，巡检效率提高约 80 倍。具体在南方电网深圳供电局与华为的合作案例中，深圳市电力局当下需要针对 110kV 及以上架空线路共计 511 回 3900 余公里的电路进行例行的巡检，而这些工作量在传统的人工巡检方式下需要 20 天（约 160 小时）完成，而基于人工智能物联网的视频巡检可将工作效率由 24 公里 / 小时提升至 195 公里 / 小时，同时识别准确率大大提高，能及时发现传统人工地面巡视不易发现的隐患点。提升核心手段在于 1）建立云边端 AI 协同架构，包括在云端部署云计算中心和云资源池作为人工智能平台、在边侧部署 Atlas 500 智能小站或轻量云实现厂站侧的 AI 升级、在端侧的智能摄像头上部署华为 Atlas 200 AI 加速模块；2）建立有针对性的输配电节点巡检方案，如针对架空线路进行自动视频分析和无人机视频分析和针对变电站、配电房、输电隧道管廊使用轻量云 AI 和 1 拖 N 智能摄像头等方案。在提升巡检效率之外，针对电力行业数据采集难和数据无法使用的问题，华为的物联网采集技术结合电网实际业务统一了数据规范，提升了通道能力，端侧数据采集量比过去提升了 30 倍，提升电网运营的稳定性与安全性。

⁶⁹ 高工智能产业研究院，MarkLines，HIS，蔚来资本，英飞凌，三菱日联摩根士丹利

⁷⁰ Amazon，加拿大皇家银行资本市场，thebestreview，艾媒咨询

⁷¹ 美国电科院

⁷² 南方电网深圳供电局官网

- 5) 对于农业而言，以农业无人机为例，价值 4 亿美元的中国农业无人机市场规模作为算力投入；使用农业无人机可减少用药量 50%，中国农药市场规模为 65 亿美元，植保无人机的作业面积占全国耕地面积 5%，假设无人机的使用周期为 5 年，则对应的 8 亿元的无人机减少用药量产值作为算力回报。从经济效益上来看，这意味着每 1 美元的投入，可以带来 2 美元的农业产值提升⁷³，说明算力投入可以明显带动农业产值的增长。
- 6) 对于文娱行业而言，美国电影企业皮克斯通过投入高性能计算机，实现了 3D 电影画质和清晰度的飞跃，从而提高用户体验。随着电影观众对于电影特效要求的提升，电影厂商需要带来更加逼真的特效体验，而为此需所搭配的图形渲染算法的演进（如对于毛发、纹理和光线）对于计算能力提出了更高的要求。以全球动画电影巨头皮克斯影业为例，该厂投入大量资金建造的自有高性能计算中心能够进入全球超算前 25 名。据估计，仅在 2013 年皮克斯就投资拥有 2,000 台高性能服务器和 24,000 颗计算核心，这降低了大量的动画与特效电影的特效渲染时长，并提高了电影画质。动画及特效电影对于算力的要求来源于每帧画面的素材量和画质要求。以目前行业顶尖水准的 IMAX 3D 电影为例，该类型电影画面需要普通工作站运行 50 个小时才能渲染出一帧，而在一部电影长度 1 个半小时，每秒钟 24 帧为准，一台普通的工作站需要 740 年才能计算完成，无法满足电影制作的时效性要求。据专家透露，皮克斯的高性能计算机中心，在理想情况下仅需要约 1 周的时间就能完成一部同类电影渲染工作，降低渲染时长，使得电影画质由原来的 1080P 快速升级到 4K IMAX 3D，打造了《阿凡达》、《猩球帝国》、《复仇者联盟》等耳熟能详的电影。

尽管对于包括人工智能在内的智能计算技术进行投入，可以对国家的经济增长带来显著的带动效应，智能计算本身也存在“边际效应”，即随着时间的推移，智能计算对于经济增长的带动效应可能会在发展及成熟阶段逐渐降低。以工业机器人自动化为例，伦敦政治经济学院和瑞典知名高校乌普萨拉大学的研究人员通过研究 17 个国家的行业数据发现，工业机器人可以显著提高劳动生产率和增加值，但随着机器人密集度的增加，额外每增加一单位机器人带来的工业产业附加值却在减少，该现象被解释为由于机器人的过度使用而导致的“拥挤效应”，即机器密度增多，边际效应下降。因此，包括工业机器人在内的智能计算行业，可能会因为未来智能设备产品密度的增加，导致单位算力对经济增长带来的边际效应的下降。

3.3 此外，对算力的投入亦能从创新、公共服务、民生等三方面提升国家整体竞争力

针对科研领域进行算力投入，可进一步提高国家高新技术研发能力，在重大科研领域实现定点突破，提升国家整体竞争力。针对公共服务业进行算力投入，可进一步提高政府工作效率，提高各级政府公共服务能力，进而创建平安和谐的社会环境。针对民生进行算力投入，在教育、科研、医疗、就业、社区服务等领域不断加大建设数字化基础设施和生态系统，可有效提高居民的生活质量，提升居民的幸福感和获得感（如图 3.4）。

⁷³ 赛迪顾问、第一财经、Phillips McDougall、太平洋证券

	创新领域（科研）	政务领域（公共服务）	民生领域（医疗）
案例	<ul style="list-style-type: none"> · 中国国防科技大学投入建设“天河一号”超级计算机推动医学进步 	<ul style="list-style-type: none"> · 爱沙尼亚政府投入数字化政府建设服务居民 	<ul style="list-style-type: none"> · 欧盟和 CERN 搭建螺旋星云云计算中心服务科学研究 
算力投入	<ul style="list-style-type: none"> · 耗资 6 亿元建设“天河一号”超级计算机（2010 年全球超级计算机 Top 500 第 1 名） 	<ul style="list-style-type: none"> · 云计算中心 · 互联网络 · 终端电脑 	<ul style="list-style-type: none"> · 投入约 530 万欧元自建云平台
算力回报	<ul style="list-style-type: none"> · 助力新药筛选、药物毒性筛查、药物重开发等环节，助力医疗科研突破 · 每天服务 8000+ 科研计算任务，推动国家整体竞争力提升 	<ul style="list-style-type: none"> · 46.7% 的爱沙尼亚居民通过互联网投票 · 98% 的居民拥有电子身份证，实现绝大多数政务可通过网络办理，最大化的节省居民办事时间成本和企业的合规成本 	<ul style="list-style-type: none"> · 帮助 Pan-Cancer 项目筛选癌症基因学，通过提高早筛成功率显著降低了当地的癌症患病率，延长当地民众预期寿命

资料来源：案头研究，罗兰贝格

图 3.4 各国投入算力对科学创新、政务效率和民生幸福起到了积极的推动作用，提升国家竞争力

算力投入可拉动效率驱动型、客户中心型、工程技术型和科学研究型四类创新，有效推动国家整体竞争力提升。效率驱动型创新指有助于减少成本、缩短生产时间、提高产品质量的制造业改善，其主要涉及大宗化学品、纺织、电气设备、工程机械等行业。企业级计算中心、终端（如智慧工厂内物联网设备）为代表的算力投入可以推动企业通过大数据等创新模式对企业成本去向进行分析并提供解决方案，有效降低运营成本。客户中心型创新是指通过产品、服务或业务模式上的进步来解决消费者的问题，其相关行业包括互联网软件与服务、家用电器和家居产品。以云算中心和智能终端（如智能手机、电子货签等）的算力投入推动了以快消品为代表的 2C 行业向消费者驱动型的新产品研发模式演变，从而更好的为消费者推出更优质的产品。工程技术型创新是指设计开发全新的产品，主要相关行业为飞机制造、汽车制造、电信设备等。算力的投入可以帮助设备制造商在生产、维护、质量监控领域利用大数据与 AI 等技术进行生产工艺流程优化、设备预测性维护、生产质量全自动检测。科学研究型创新是指通过基础学科研究成果商业化来开发新的产品，其主要涉及的领域包括品牌药物、生物技术、半导体设计和专用化学品等。以超级计算机和超大型计算中心为代表的云端算力设施建设投入可以帮助研究机构、企业进行药物分子发现和新材料研发，推动初步研究成果到最终商业化的进程加快。在这四类创新中，我们以科学研究型创新中药物创新为例，中国搭建以“天河一号”为代表的超级计算机，解决传统药物研发中的所遇到的高昂的筛选成本、毒性测试成本和极高的失败率等主要痛点。中国自进入 2000 年以来，持续在超算领域推进研究、投入和建设，其中具有里程碑意义的事件是 2010 年 11 月，由国防科学技术大学与天津滨海新区耗资 6 亿元合作研制的超级计算机“天河一号”投入使用，其峰值性能达到 4700 万亿次每秒，在 2010 年世界超级计算机 Top 500 中排名第一。天河一号及其他大量超算中心的问世为生物医药研发提供了大

数据处理平台，提升了在医疗领域的科学研究型创新。中国大量科学单位和企业利用这一平台开展了抗癫痫、艾滋病、癌症等行业研究并取得了大量科研成果。从底层逻辑上看，超级计算机主要能够在三方面帮助药物研发：一是新药筛选。通过对数以百万计的化合物筛选，确定该药物是否可结合在与疾病相关的靶蛋白上；二是药物毒性筛查。通过测试该药物是否会与人类大约一千个蛋白进行潜在结合，形成毒性，避免毒性药物进入最昂贵的临床试验阶段；三是帮助药物重开发。尝试匹配现有药物到不同的靶定蛋白，重新利用现有药物，缩短研发周期。据统计，“天河一号”每天服务 8000 余个生物医药、石油资源开发、装备制造等行业的科研计算任务，累计支持国家科技重大专项和研发计划等 1500 余项，已经成为不可替代的科技创新“国之重器”。

算力投入可极大提升政府政务办公效率，促进社会公平高效，创造平安和谐的社会环境。以公共服务业中的数字化政府领域为例，爱沙尼亚大量投入云端数据中心，建造全国电子数据库，实现绝大多数政务可通过网络办理，最大化的节省居民办事时间成本和企业的合规成本。爱沙尼亚在从苏联独立后，基于重构其政府公共服务网络的需求和提升政府服务效率的目标，投入云计算中心、互联网络、终端电脑等设施建设，大力推行数字化政府，实现电子化存储该国公民从出生记录、选举名册、房契到银行凭证等信息。以该国数字化税收系统为例，该国居民仅需 2 分钟就可以完成报税的流程工作，显著降低合规成本。在该税收系统中，大型云计算中心扮演数据处理、分析、和存储等功能；完善互联网络帮助居民从任意数据；大量的终端电脑作为报税信息输入节点，辅以高度易用的交互软件，降低电子报税门槛。数据显示，46.7% 的爱沙尼亚居民通过互联网进行民主投票，98% 的该国居民拥有电子身份证，99% 的政府服务已经可以通过网络处理。统计数据显示，数字化进程每年可帮助该国直接节省相当于 2%GDP 的社会运行成本。

算力投入对促进民生方面具有显著效益，可显著提升民众的生活质量，从而有效增强民众的满意度和幸福感。以科研行业为例，欧盟政府致力于将研究机构、数据供应商、公开信息化基础设施、商业化云服务供应商聚拢，实现商业化运营科学云服务。通过投入螺旋星云（The Helix Nebula）科学云服务，投入 530 万欧元建立欧洲混合云平台，以支持用于科学研究的高性能计算和大数据功能的部署，从具体的使用案例来看，螺旋星云在诸多科学领域产生显著作用，代表性案例有三个：第一，帮助 ESRF 的 3DIX 项目通过 X 射线生成纳米级的 3D 成像，以研究物体在纳米级水准的特征并广泛应用于化学、生命科学、材料结构等科学领域；第二，帮助 HADDOCK 建立平台，超过 1 万名用户安全地将已识别的或预测的蛋白质界面的信息编码写入模糊的交互约束（AIRs），以驱动对接过程。第三，协助 Pan-Cancer 项目每月从 5000+ 肿瘤样本中识别 12 种肿瘤类型。PanCancer 是目前研究癌症基因学的最大项目，其大量的数据（约 1PB）和有限的时间窗口（每月）能够充分利用螺旋星云大数据处理能力。其研究结果通过提高癌症的早筛成功率，通过尽早的治疗及预防介入，进而显著降低当地的癌症患者的死亡率，延长当地民众预期寿命，有效提升民众的生活幸福感和满意度。

4

第四章 算力发展面临的挑战与政府的建议举措

第四章主要内容

主要内容

4.1

算力发展过程中将面临功耗激增、网络水平不足、算力生态等三大挑战，及基础设施建设和安全两大问题

- 算力功耗增长挑战智能社会可持续发展，算力功耗总量随算力需求逐年增长
- 网络发展水平不均影响网络算力部署形式，需提升网络质量和网络覆盖以实现算力全量输出
- 多样化的智能场景需要多样化的算力生态
- 算力基础设施建设及安全问题持续面临，将影响多样化的智能场景高效实现

4.2

各国政府可采用规划倡议、公共设施/服务提供、财税支持、法律监管四类政策抓手应对算力发展挑战与安全问题

- 针对基础设施建设的必然要求，积极布局算力基础设施、推动云计算和AI等技术发展
- 针对算力功耗挑战，需：1）引行业设节能目标；2）直接投资新能源；3）以绿色设备购置抵税等手段推广低耗设备；4）建立算力碳排放交易机制
- 针对网络质量挑战，需：1）形成矩阵网络建设监控体系；2）以公私合营带动社会资本投入网络建设
- 针对生态多样性挑战方面，需：1）规划算力生态多样性目标、推动移动化场景发展，拉动多样化架构市占率；2）建立多样化生态联盟；3）专项资金支持各类架构研发企业；4）放宽生态企业准入门槛
- 针对算力安全问题，需：1）制定国际安全标准；2）以公共财政资金支持第三方安全机构运营；3）制定完善的究责机制，促相关从业者重视并切实遵守算力的相关安全规则

资料来源：罗兰贝格

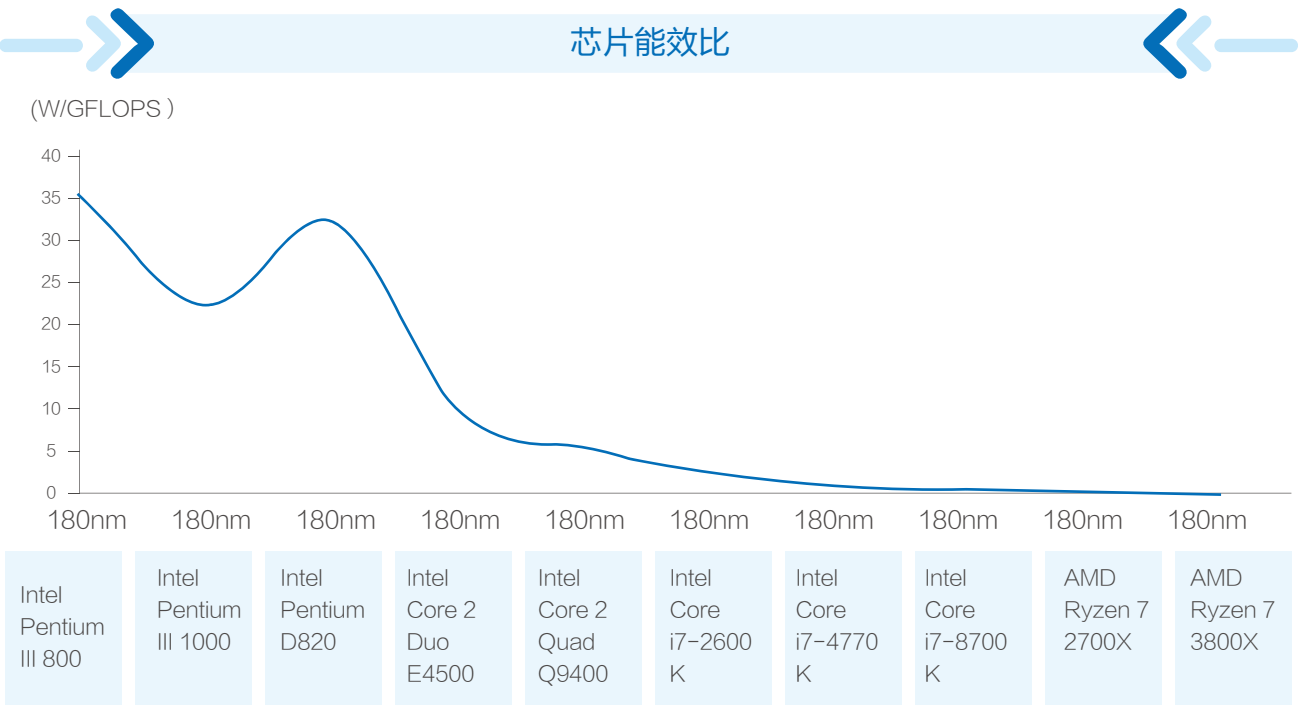
图 4.1 第四章的主要观点介绍

4.1 算力在发展的过程中面临来自于功耗、网络和生态三方面的现实挑战，并需重视基础设施建设及安全问题

算力发展对一国各行各业的提振，乃至未来智能社会的实现都是不可或缺的，但是其未来发展必然不会是一帆风顺的。一方面随着算力的发展，某些并发性问题将会逐渐显现（如算力功耗高速膨胀等），需要各国同心协力克服；另一方面，算力发展的某些必要条件远远滞后于理想情境（如网络质量的提升、生态配套的完善等），需要各国认真对待，积极推进。因此，本文从算力自身及其支撑配套角度出发，定义出了当前 3 大算力核心挑战，分别为：算力功耗增长挑战智能社会可持续发展、网络发展水平不均影响算力部署形式、算力生态单一影响多样化智能场景的实现。此外，作为 ICT 产业，基础设施的积极建设及安全问题亦为重点面临问题。

1) 在算力功耗所带来的可持续发展挑战方面：

功耗是影响算力水平发展的三大因素之一，在单位算力功耗呈现出逐年递减，且正向着极限逼近的态势下，算力的整体量级却仍然随着其广泛的应用而持续大规模增长。而在这背后凸显出的问题是，一方面人类必须依托算力的大规模使用以实现智能社会的演进，而另一方面这一进程又必将伴随着碳排放的提升而导致全球气候的不可预测性显著增加，这一双手互搏式的矛盾综合体是各国皆需关注的重中之重。



资料来源：Intel、AMD官方数据；罗兰贝格

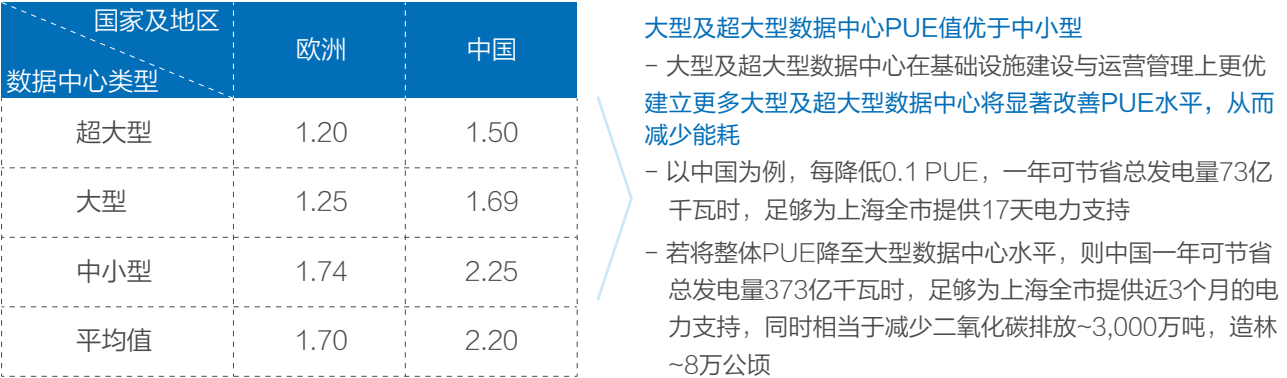
图 4.2 单算力功耗随着纳米制程的提升显著下降

整体来看，单位算力功耗在过去 10 多年间经历了显著下降，但随着硅基芯片工艺制程提升的难度凸显，其进一步下探幅度有限。从图 4.2 的趋势分析中可以看出，随着工艺制程的进步，输出单位算力所需要的功耗水平（W/GFLOPS）从 90 纳米制程开始呈现出了指数级下降的趋势（如图 4.2），发展到当下 7 纳米制程的先进处理器芯片时已开始趋近于 0 W/GFLOPS，下探难度愈发增大。从原理来看，由于功耗水平下降的核心驱动力来自于芯片制造工艺的低制程化，而这一趋势正受到硅基芯片上晶体管栅长缩减引发的量子隧穿效应挑战，一旦到达临界制程可能造成晶体管闭合功能的全面无效化。因此在没有新型材料替代的条件下，单位算力功耗难以进一步下降。

而更为艰巨的是，在芯片单位算力功耗下降幅度有限的背景下，智能社会应用的广泛普及无可避免，势必导致算力功耗总量的进一步提升。以中国为例，随着人工智能、物联网、区块链等技术的发展，中国数据中心的总

用电量连续 8 年增速超过 12%。至 2018 年，中国数据中心总用电量为 1,608.89 亿千瓦时⁷⁴，占中国全社会用电量的 2.35%，已经显著超过上海市 2018 年全社会用电量（1,567 亿千瓦时）。面向未来，数据中心的算力将在信息技术产业的大力发展下继续增长，2023 年将较 2019 年增长 66%，年均增长率将达到 10.64%，功耗总量将进一步提高。因此，为保障未来智能社会的可持续发展，优化高能耗的数据中心的 PUE 水平将成为算力发展一大核心挑战。

欧洲和中国不同类型数据中心PUE水平



资料来源：工信部、哥本哈根经济、公开资料整理，罗兰贝格

图 4.3 大型数据中心的 PUE 显著高效

而数据中心 PUE 优化的一大核心方向在于建立更多超大型（>10,000 个标准机架可共置至少 100,000 台服务器）和大型数据中心（>3,000 个标准机架可共置放至少 30,000 台服务器）⁷⁵，从而带动整体数据中心 PUE 值优化。从中国和欧洲的数据中心 PUE 值来看，我们发现数据中心的规模越大，其 PUE 值的表现亦会更优（如图 4.3）。在中国，超大型数据中心的 PUE 水平（1.5）较整体平均值（2.2）高出 31%，欧洲超大型数据中心的 PUE 水平（1.2）较整体平均值（1.7）高出 29%。究其原因，超大型 / 大型数据中心由于投资金额庞大、能耗量高等因素，通常较中小型数据中心更会从基础设施和管理系统层面来进行优化，降低功耗。通过测算，当前中国数据中心每降低 0.1 PUE，可节省总发电量 73 亿千瓦时，足够为上海市提供 17 天电力支持；倘若可将数据中心 PUE 整体降至大型数据中心水平，则可节省 373 亿千瓦时，足够为上海市提供近 3 个月的电力支持，同时相当于减少二氧化碳排放量 ~3,000 万吨，造林 ~8 万公顷。

在基础设施建设层面，超大型 / 大型数据中心会采用能耗比表现更优的服务器设备，同时大力投入节能型供电、冷却配套硬件解决方案，从而实现 PUE 值优化。以百度云计算（阳泉）中心为例，其设计容纳 16 万台服务器，

⁷⁴ 绿色和平 & 华北电力大学《中国数据中心能耗与可再生能源使用潜力报告》2019

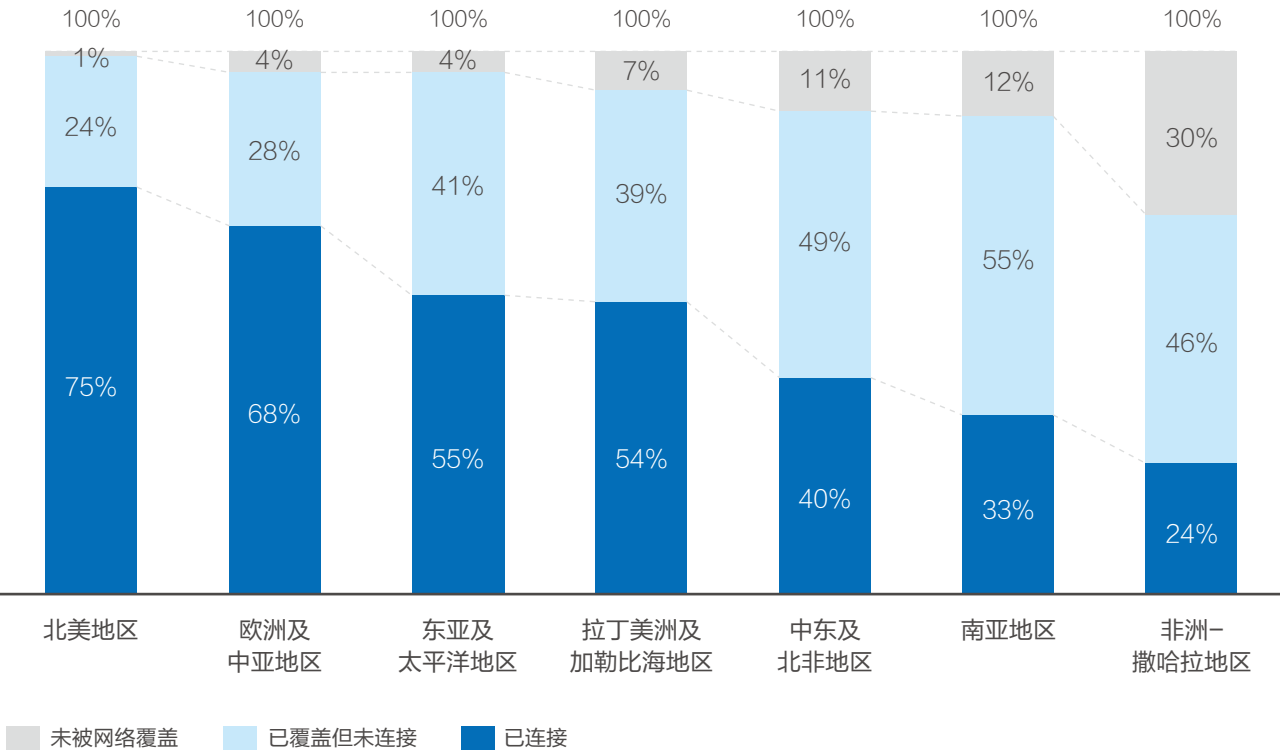
⁷⁵ 工业和信息化部、发展改革委、国土资源部、电监会、能源局《关于数据中心建设布局的指导意见》2013

因此采用了先进的整机柜服务器，以模块化设计将供电、散热系统与服务器置于同一机柜内，针对每个机柜服务器的实际运行及功耗进行个体化调节，从而实现年功耗节省 20%。同时辅以高压直流 + 市电直供的供电模式，实现传统 UPS 供电无法达到的 96% 能源效率，共同实现了其年均 PUE 降至 1.23 的出色功耗控制表现。在数据中心运营管理软件层面，大型数据中心会采用基于 AI 的智能化管理系统对供电、冷却系统的运营状态进行自适应的动态化调整。如华为廊坊数据中心所采用的 iCooling 数据中心能效解决方案，通过 AI 对历史运营数据分析，寻找供电、通风、冷却等影响能耗的不同关键因素，建立预测模型并对相应系统进行动态调整，降低 PUE，实现能耗优化。

2) 在网络发展水平不均所带来的算力性能制约挑战方面：

基于前文论述，网络管道的延时及带宽限制将显著影响算力在云、边、端各个层面的协同性能发挥。目前各国面临的现实是，网络盲区依旧存在，即使是对于已拥有网络覆盖的地区而言，其网络质量（网络带宽及时延决定）各有千秋。

全球各地区人口网络覆盖情况



资料来源：GSMA；罗兰贝格

图 4.4 全球各地区人口网络覆盖存在明显差距

首先，世界各个地区网络覆盖情况存在明显差距。研究表明⁷⁶，当前网络连接情况最好的当属北美地区，仅 1% 的人口尚未被基础网络设施覆盖，但即使是连接情况最好的地区仍旧有 24% 的人口尚未实现与互联网的连接。而与之形成鲜明对比的是，南亚、非洲等发展中国家集中的地区大部分人群依旧未实现网络连接，在非洲地区甚至有 30% 的人尚未被网络基础设施所覆盖（如图 4.4）。各国之间巨大的网络发展水平差异是比导致其未来智慧化社会进程受阻，所以互联网网络覆盖及连接水平的进一步提升是全球各国家及地区仍需继续加强的环节，从而为算力的全量性能输出创造必要基础条件。

全球各国家及地区网络质量情况（部分）

全球网络TOP 5国家及地区				全球网络Bottom 5国家及地区		
平均带宽	国家及地区	固网 [Mbps]	移动网络 [Mbps]	国家及地区	固网 [Mbps]	移动网络 [Mbps]
	新加坡	191.93	50.43	黎巴嫩	7.11	35.49
	韩国	156.18	97.44	莫桑比克	6.17	17.16
	中国台湾	151.75	43.66	苏丹	6.09	10.77
	中国香港	151.02	31.24	阿尔及利亚	4.18	7.23
	罗马尼亚	130.79	36.68	委内瑞拉	3.61	6.79
平均时延	国家及地区	移动网络 [Mbps]		国家及地区	移动网络 [Mbps]	
	新加坡	30.7		加纳	157.1	
	匈牙利	33.4		科特迪瓦	156.5	
	捷克	33.7		伊拉克	139.4	
	瑞士	34.0		乌兹别克斯坦	120.5	
	中国香港	34.5		塞内加尔	115.3	

核心发现

- 不同国家间的网络带宽与时延差距显著
- 经济发展水平较好的国家网络发展水平较高

资料来源：案头研究、华为，罗兰贝格

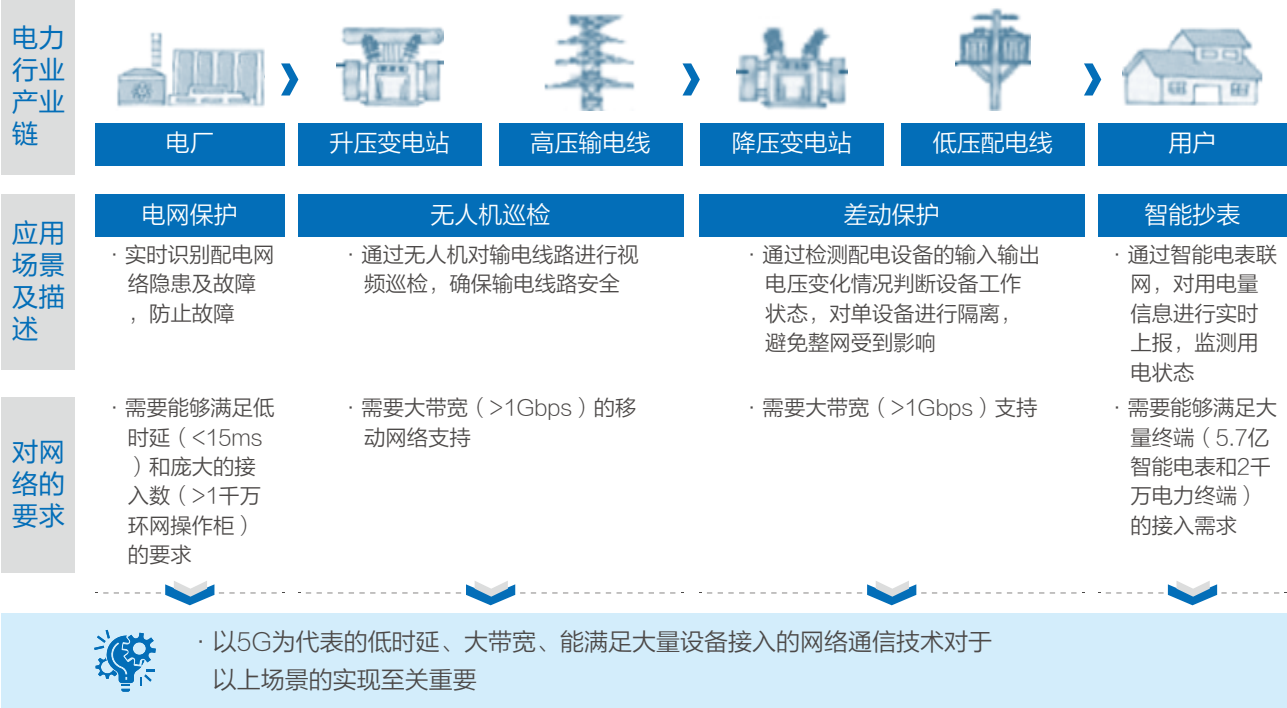
图 4.5 全球各国及地区网络质量情况差异显著

其次，在网络基础设施已经覆盖的国家中，不同国家之间的网络速率及时延差距依旧显著。在网络速率方面，以当前 Speedtest 专业测试软件录得的网络下行速率数据为例（如图 4.5），得益于长期的网络基础设施投入，新加坡、韩国等国家以近 200 Mbps 的固网带宽、以及平均 50 Mbps 的移动网络带宽名列全球各国前列。但是排名末尾的国家（如委内瑞拉、阿尔及利亚）与前列国家相差超 50 倍，网络条件最差的委内瑞拉固网速率仅录得 3.61 Mbps，仅可支持普通 WEB 网页的浏览，无法支撑算力的发挥。而在网络时延方面，全球不同国家之间的平均水平仍旧差距较大。从数据统计来看，当前表现最好的新加坡与表现最为堪忧的加纳相比网络延时水平差距在 5 倍以上，其中加纳 150 毫秒以上的移动网络延迟连最为基础的汽车应急呼叫功能都无法顺畅实现。

⁷⁶ 全球移动通信系统联盟（GSMA）《2018 年移动互联网连接现状》2018

网络的速率与时延水平对于算力输出会产生极大的影响，以输电网络为例，因为发电区域和用电区域之间常存在较大的地理间隔，所以需要借助特高压网络实现电力的远距离传输，在这期间，网络自身性能对于电力传输各环节都会产生较大影响，而当下，许多电力领域的应用场景（包括电网保护、无人机巡检、差动保护、智能抄表，涵盖电网领域发、输、配、售四大环节）正受限于网络带宽的限制，而无法得以应用或实现效益的充分发挥。（如图 4.6）

网络在电力领域对于算力输出的影响



资料来源：案头研究、华为，罗兰贝格

图 4.6 网络对于电力领域算力的输出至关重要

我们以无人机巡检为具体案例来看，目前的 4G 网络带宽无法支持无人机电力线巡检传输超高清图像的需求，导致电网检测维修人员无法实时检测输电线路的微小缺陷并进行及时的隐患排除。具体来说，在电力巡检过程中所采用的无人机主要为网联无人机，其与地面的通信主要有三种目的：图传、数传和遥控。其中图传对于无人机网络通信能力的要求是最高的，如果无人机采用目前最先进的 4G LTE 蜂窝通信技术（约 100 Mbps），即使在网络基站覆盖到位的情况下，仅能支持无人机与基站或遥控器之间进行 720P 清晰度的视频传输，无法满足 4K 超高清视频的传输要求。从而导致电网维修人员难以凭借无人机回传的视频图像找到输电线路的微小缺陷并进行及时的隐患排除，影响无人机巡检的使用范围与应用场景，所以以 5G 为代表的高速通信网络对于电力输电环节的无人机巡检应用场景至关重要。在以 5G 为代表的高速网络到来之后，通过无人机现场高清视频传输，工人无需亲自爬上百米高的电力塔亲自检测，从而在提升巡检效率的同时，更是杜绝了人员作业风险，显著提升安全性。

3) 在生态配套带来的多样化挑战方面:

当下算力生态的发展距成熟仍有较大差距。虽然 X86 架构的服务器芯片可满足通用计算场景需求，但是在未来智能社会即将到来的大背景下，场景将极大程度的向多样化演进，这对与不同类型的芯片架构将产生不同的需求。比如以大数据搜索、云游戏等为代表的场景，其提出了高并发性、端云协同性等要求，然而，依托现有的以 X86 为主的架构难以针对各类场景都提供有效支持，因而提出了对于以 ARM 为代表的 RISC 等更多类架构芯片发展、同时实现端到端的适配要求，即算力生态多样化。具体来看，当下算力生态多样性正面临以下三大挑战：a) 多样化的场景应用对不同类型芯片架构的要求；b) 不同类型芯片架构生态对端到端的适配要求；c) 端到端的适配对大量高端人才的要求，以上三者的任意缺失会导致未来不同类型智能化的场景无法全量实现，阻碍人类社会向高阶智能化形态演进。

RISC架构对未来智能化社会要求更加匹配，以ARM为例：



资料来源：案头研究，罗兰贝格

图 4.7 以 ARM 为代表的芯片架构对于未来智能社会场景有关键支持作用

a) 未来智能社会的应用场景在并发性、兼容性等方面要求各不相同，因此需要不同芯片架构满足需求。典型的，虽然 X86 架构的服务器设备可满足通用计算场景需求，但为了更高效地应对某些场景的技术特性，RISC 服务器的存在往往是更优的选择（如图 4.7）。首先在并发性处理方面，以搜索引擎场景为例，其基于大数据的搜索技术（如 MapReduce）对于算力高并发性能的支撑需求尤其突出，核心原因在于该项技术本身要求将海量数据进行切片，并生成多个任务进行并发处理，以应对在同一时间点内发起的并行式搜索需求。而采用 RISC 架构的芯片可以凭借更多的物理核数，以单个核心应对单个任务请求的逻辑，提高此类搜索应用的并发处理效率，获得更好的性能。从实际效果来看，以拥有同级别主频（2.4 GHz）和功耗（90W）的 E5-2640v4（X86 架构，10 核）和 Hi1616（ARM 架构，32 核）芯片进行对比，采用双路 ARM 架

构的服务器并发效率较 X86 架构可提升近 70%。此外，在大数据存储与分析领域，RISC 架构下的性能表现要比 X86 架构有逾 25% 的性能表现优势。其次，在兼容性优势方面，以逐渐普及的云游戏场景为例，在对游戏应用进行云化迁移时，RISC 架构服务器天然具备的兼容性优势。核心原因在于当前大部分手机及平板端的游戏应用为代表性 RISC 架构的 ARM 原生应用，倘若将这些游戏应用进行云化部署，以 X86 架构为基础的服务器需要经历指令翻译器、Android 仿真器、VM/Docker 三层处理才能实现应用兼容，而若迁移到以 ARM 架构为基础的服务器上，其仅需要通过 VM/Docker 一层处理，从而显著降低编译难度、提升效率和可操作性，整体提升迁移性能。在实际的迁移案例中，应用 RISC 服务器可降低 40% 由指令翻译所带来的性能损耗。最后，在功耗控制方面，以云计算场景为例，其价值创造的核心方式即在于利用大规模服务器阵列所构成的超大型数据中心，形成虚拟化的计算平台，为用户提供快速灵活的算力服务。而在可持续发展要求下的绿色算力要求下，数据中心的功耗控制便显得尤为重要。从实际数据来看，RISC 架构芯片在相同性能的前提下功耗普遍比 X86 架构芯片低 20% 左右⁷⁷，更为重要的是，服务器每年的能耗费用基本是服务器采购费用的 40%，因此利用 RISC 服务器的部署实现数据中心功耗控制既符合可持续发展的根本需求，更在经济上能够实现运营成本降低。因此 RISC 服务器的应用在云计算场景下亦是必要的（如图 4.8）。

b) 不同的芯片架构需要端到端的适配，从而保障芯片算力的充分利用。端到端的芯片架构适配主要包括三个层面：首先在芯片层面，需要依靠 IP 微核心设计、版图拼接、晶圆生产等玩家配合，实现算力的实体化；其次在软件层面，一方面需要操作系统与底层编译器进行对接，以实现底层资源调用，另一方面需要应用程序实现开发套包对底层命令的识别，实现编程开发；最后在硬件设备层面，需要实现不同部件总线接口带宽与芯片计算能力的配合，从而实现算力的充分利用。足可见无论是芯片、软件还是硬件层面，都需要与底层算力架构进行配合，才能保证算力的真正应用，缺一不可。以目前的 RISC 架构的服务器生态为例，RISC 架构服务器的发展乏力主要原因在于缺少端到端的适配，其中一个主要的问题是系统开发的不适配。为保障与服务器端一致的开发与测试环境，减少不兼容的问题，目前大部分软件开发者都在使用 X86 架构的计算机进行服务器端软件开发。AMD 全球副总裁潘晓明曾介绍，在服务器领域，市占率需在 25% 以上才可以实现足够的供应链利润，维持整体产业链生态技术创新的活力⁷⁸。如 RISC 架构未来服务器将来形成生态，势必要将 RISC 架构的服务器芯片市占率提升到至少 20% 以上，才会有足够的操作系统开发商、软件应用开发商、云计算服务商去适配 RISC 架构的服务器，真正发挥 RISC 架构服务器超高性能功耗比的特点，满足不同场景的算力需求特征。

c) 为了能够从软件和设备层面及场景层面对芯片指令集进行适配，必然需要大量芯片领域的高端人才来提供支撑，扮演“润滑油”角色，从而彻底将芯片算力导入场景，激活算力生态。整个算力生态中，从底层芯片

⁷⁷ ARM, 2016

⁷⁸ 环球网《AMD全球副总裁潘晓明：打破 x86 服务器垄断地位》2015

设计，到设备生产，再到软件开发，每一个环节的发展进步都与高端专业人才密不可分，可以说高端人才就是整个算力生态系统的“润滑油”。无论是操作系统和应用软件的开发者，智能设备的设计者，还是芯片研发过程中的软硬件架构师和工程师，都对整个算力生态系统的构建起到了至关重要的作用。未来若要实现计算能力的进一步提升及智能场景的大规模应用，社会需要大规模、高质量以及多种类的高端人才。

4) 此外，我们也看到基础设施建设及安全问题是算力发展面临的突出问题。例如在云计算设施建设方面，基于前文论述，国家的算力水平与其经济发展水平具有极强的相关性，且算力投入对于各个行业都具有极强的促进作用。如果一个国家的算力基础设施推进速度与经济发展水平形成错配，将会影响到该国各产业发展速度，进而影响到国家的整体竞争力。根据麦肯锡研究数据，2018 年美国企业的上云率为 80%，而中国同期只有 40%，这较大的影响了中国企业的数字化运营效率。此外，富士通数据中心宕机亦是云计算设施建设缺少足够投入的典型案例，其位于硅谷的数据中心因变压器发生故障导致数据中心供电受损，所有的 SaaS 和公有云服务暂时下线，导致所有客户耗时 5 天才恢复正常，严重影响大量客户的正常运营。在算力安全输出保障方面，由于云计算算力输出的过程中信息的传递会经过网络、软件、设备、芯片四个层面的信息传递，其信息安全需要每一个层面都具有完善的安全保障机制。如果缺失任何一层，国家就难以保障其政府、企业、民众的信息安全，从而发生严重的信息泄露，从而引发公共安全问题。以美国国家得宝（Home Depot）的信用卡用户信息暴露问题为例，由于缺少网络和软件层面的完善保障，5600 万用户的信用卡信息被泄露，对该国的公众财产安全形成严峻挑战。

4.2 各国政府可采取规划倡议、设施 / 服务提供、财税支撑、法律监管等抓手，针对性的推动算力发展

应对所需的政策举措



资料来源：案头研究，罗兰贝格

图 4.8 对政府的举措建议总结

为奠定算力发展基础，应对功耗提升、网络质量提升、生态多样性建设三方面的挑战，并综合考量算力生态体系安全等问题，政府需要积极利用各类举措抓手，推动各国算力水平的发展，从而满足未来智能社会应用场景所带来的算力需求。从举措抓手上看，政府需综合利用以规划及倡议、公共设施 / 服务提供、财税支持、法律及监管等不同类型的方式形成由针对性的组合，从而应对算力发展挑战（如图 4.8）。

奠定算力发展基础的政策建议

	目标内容	具体举措
目标一	· 积极布局算力基础设施建设	· 建立底层设施基础：政府可自建自营数据中心、超算中心等基础物理设施，奠定算力发展关键设施基础 · 共享数据与开发应用：政府可推动开放数据共享、开发上层应用等手段，提升数据平台的应用价值
目标二	· 推动云计算、AI等关键应用技术普及与发展	· 规划发展目标：政府应规划技术发展战略、明确分阶段目标 · 设立对口管理机构：政府可在国家层面建设伞状的自上而下的统一算力发展与管理领导部门 · 匹配资金等发展所需资源：以政府直接投资为牵引、带动社会资金投入等形式匹配发展所需资源投入 · 制定法律法规等约束边界：政府制定包括伦理在内的法律法规，明确发展边界

资料来源：案头研究；罗兰贝格

图 4.9 针对拉动算力发展基础的政策建议总结



1) 为奠定算力发展基础，政府应积极布局算力基础设施建设、同时推动云计算、AI 等关键应用技术的普及与发展（如图 4.9）。

在基础建设领域，我们看到包括欧盟、中国、日本在内的世界各国政府都已出台相关政策与举措积极布局，建立底层设施基础、共享数据与开发应用。

在建立底层设施基础方面，2016 年，欧盟正式启动欧盟云计算行动计划，重点关注欧盟大数据基础设施的建设，规划了以下举措：于 2016 年，创建服务于科技界的欧盟科学云基础，整合和强化欧盟虚拟科研基础设施网络平台；2018 年，启动新兴量子技术研发创新行动计划，加速新一代超级计算机开发；2020 年，全面开发和部署大规模高性能计算、大数据存储和高速宽带基础设施，包括建设欧盟大数据处理存储中心、升级科技创新骨干网络、打造新一代超级计算机以跻身世界前三强。在规划的牵引下，欧盟将建设 8 个世界级超级计算机中心，其中 3 台将具备每秒 15 亿亿次浮点运算的能力，另 5 台将具备每秒 4,000 万亿次浮点运算能力，从而推动欧盟朝着成为全球顶级超级计算区域迈进。为此，欧盟从设备采买与人才培养方面皆有所布局：在设备采买

方面，欧盟计划在 2020 年前购买和安装两台近亿亿次计算机与另外 3-4 台千万亿次计算机。在人才培养方面，欧盟建立了玛丽居里项目，资助高校学生在欧盟内部大学间交流以及在数字化企业实习、以推广其数字技能人才理念与实践，从而奠定算力发展所需的人才池基础。除上述政策外，中欧还已出台了诸多政策，旨在推动云计算、AI 的基础建设发展（如图 4.10）。

中国与欧盟在人工智能与云计算领域相关政策

	AI	云计算
 中国	<ul style="list-style-type: none"> · 2018.11《促进新一代人工智能产业发展三年行动计划（2018-2020年）》 - 在核心基础、制造技术装备、支撑体系方面提出要求，规划到2020年全国90%以上地区的宽带速率和时延满足AI行业应用需求 · 2019.8《国家新一代人工智能创新发展试验区建设工作指引》 - 加强网络、大数据、计算机基础设施建设水平，明确到2023年，布局建立20个左右AI创新发展试验区 	<ul style="list-style-type: none"> · 2015.1《国务院关于促进云计算创新发展培育信息产业新业态的意见》 - 规划到2020年，云计算基础设施不断优化，新建大型云计算数据中兴PUE优于1.5；创新能力得到提升，突破艾字节(EB)级云存储系统等关键技术 · 2017.1《云计算发展三年行动计划（2017-2019年）》 - 对云计算技术增强、产业发展、应用促进提出发展规划，计划到2019年突破一批核心技术以优化基础设施，新建数据中心PUE值普遍优于1.4
 欧盟	<ul style="list-style-type: none"> · 2018.4《欧盟人工智能合作宣言》 - 强调应增加投资，升级人工智能基础设施，提出建立世界级AI研究的欧洲跨国研究所，计划到2020年底增加至少200亿欧元投资 · 2018.6《欧洲2021-2027数字化转型》 - 提供25亿欧元重点资助AI在数据及算法存储、AI测试、国家合作方面发展 	<ul style="list-style-type: none"> · 2012.9《云计算战略计划》 - 计划到2020年，对云计算技术研发和基础建设增加450亿欧元直接投入，预计投入直接产生的累计GDP将达9,750亿欧元 · 2016.4《欧盟云计算行动计划》 - 总投入67亿欧元，规划到2017年强化科研基础建设网络平台，到2020年全面部署建设高性能计算中心、大数据存储中心，提升科技创新和超算能力 · 2018.6《欧洲2021-2027数字化转型》 - 提供27亿欧元资助超算中心建设，并于2018年1月提出EuroHPC联合业务，和成员国建设世界级超算基础设施

中国、欧盟政府正从政策层面，积极推动云计算、AI的基础建设发展

资料来源：案头研究；罗兰贝格

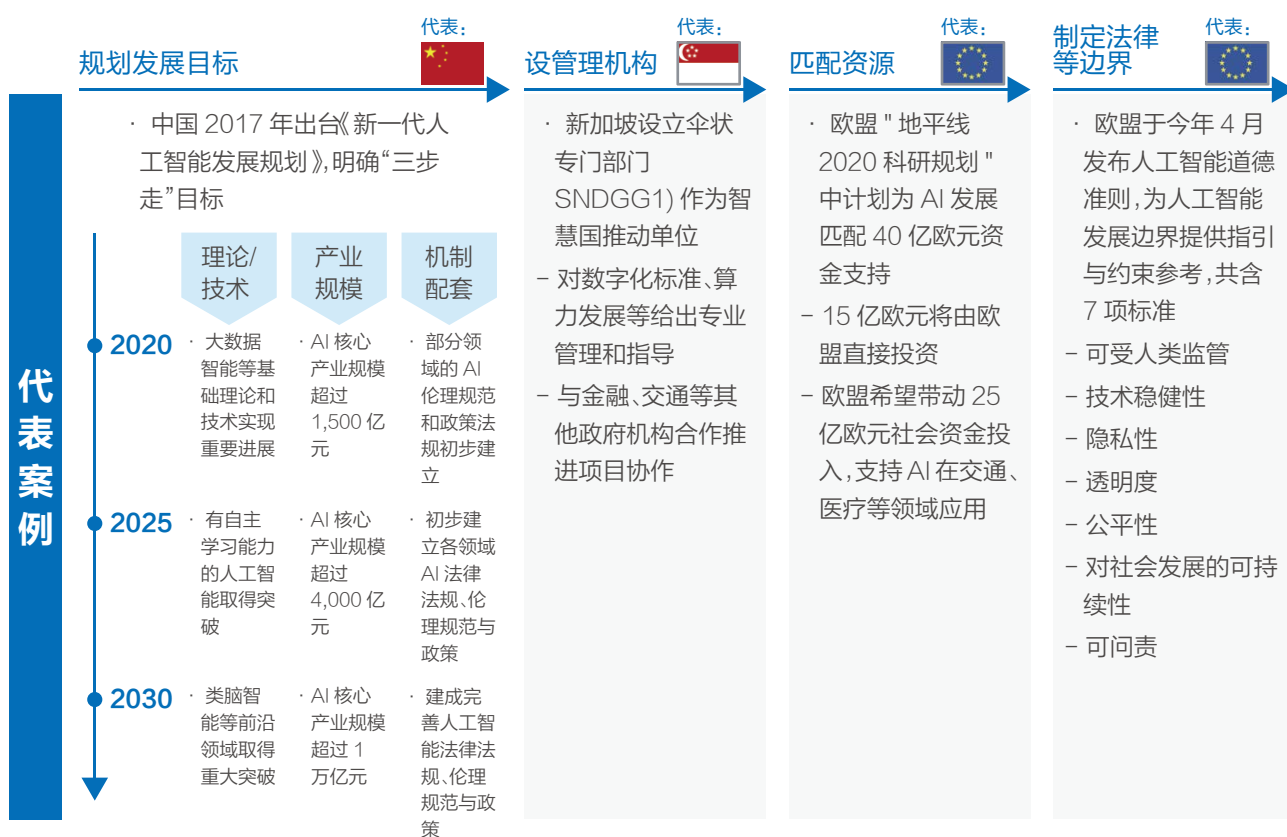
图 4.10 中欧积极出台政策推动 AI、云计算的基础建设

在共享数据与开发应用方面，早在 2012 年，欧盟政府就出台了欧洲云战略（European Cloud Strategy）——其目标是将欧洲建设成一个基于数据和知识的共享经济体，使得欧盟各国跨学科研究人员都可轻松访问所有资源，从而释放跨学科研究潜力、提高对数据密集型研究和数据科学的认识，具体内容包括：1）为 170 万研究人员提供“开放科学云”服务，按照 FAIR 原则（可查、可访问、可互操作、可重用）发布研究/科学成果存储、共享和重新使用跨学科和跨边界的数据，该“开放科学云”的在线门户网站初步版本于去年 11 月 23 日正式发布，成为欧盟科研工作者开展科研活动的关键数据来源；2）通过开放共享欧盟 2020 地平线（H2020）未来所产生的所有科研信息数据，鼓励刺激大量科研数据集的共享与重新再利用，从而为算力基础建设中的数据再添砖加瓦。

与此同时，欧盟、中国等各国政府也在通过规划发展目标、设立对口管理机构、匹配资金等发展所需资源、制定法律法规等约束边界以积极推动云计算、AI 等关键应用技术的普及与发展。

比如，在规划层面，在中国 2017 年出台的《新一代人工智能发展规划》中，国务院则从人工智能理论与技术体系、产业规模、机制配套三大方面提出“三步走”的发展战略，明确了截止 2020 年、2025 年与 2030 年的阶段发展目标，包括：从 2020 年到 2030 年在理论与技术领域要实现由易到难的研究突破、产业规模要实现从 1,500 亿元到破万亿元的提升、机制配套要实现逐步完善的法律法规与伦理规范等，从而实现了先“定目标”的关键步骤。在目标的牵引下，中国近年来人工智能发展态势喜人，据统计，2017 年我国在人工智能领域发表的论文数量已居世界第一⁷⁹，近两年，中国科学院大学、清华大学、北京大学等高校纷纷成立人工智能学院，开展相关前沿技术研究；与此同时，人工智能核心产业规模达近 600 亿，正朝着规划目标不断迈进。

各国推动AI技术发展的关键举措案例



1) 智慧国及数字政府工作团

资料来源：案头研究；罗兰贝格

图 4.11 各国从规划发展目标、设立管理结构等方面推动 AI 技术发展

⁷⁹ 爱思唯尔文献数据库统计结果 2017

进而，立足规划目标，以中国、欧盟、新加坡为代表的各国政府已先后通过设立管理机构、资金支持等手段，切实推动 AI 等应用技术的发展（如图 4.11）。比如：新加坡即设立了伞状的专门部门“智慧国及数字政府工作团 SNDGG”作为智慧国的重要推动单位。SNDGG 作为管理组织，一方面对数字化标准、算力发展、人才培养等问题做出专业的管理和指导；一方面负责与金融、交通、商务等其他政府机构广泛合作，共同推进项目协作、资源共享。此外，在匹配资源与制定法律等边界方面，欧盟也已采取了相应举措，以推动 AI 等技术发展。

2）针对未来算力功耗激增的挑战，政府的核心目标应考量节能型“绿色算力”发展，平衡智能社会发展的切实需求及日益增加的能源消耗所带来的温室气体排放激增（如图 4.12）。未来，打造“绿色算力”意味着算力自身需朝着碳平衡的目标前进，通过算力手段推动其他行业节能减排，这样将有效提升社会节能效率。据国际电联全球电子可持续性举措 (GeSI) 报告，2020 年，ICT 行业的二氧化碳排放量将占全球排放总量的 2.8%，但在 2020 年通过 ICT 行业可减排 78 亿吨的二氧化碳，减排量相当于常态下 2020 年排放量的 15%，是其自身排放量的 5 倍⁸⁰。

“绿色算力” 推动政策建议

		具体举措	代表案例
举措及抓手	规划及倡议	✓ 引行业联盟设节能目标，树节能减排泛化意识	· 欧盟委员会重磅推出《欧洲绿色协议》（European Green Deal），以2050年实现净零排放为目标，其中涉及包括ICT等各个产业，倡议实现可持续性的数字化发展，并将推出相关法律法规 · 在GSMA协会提出移动产业气候行动路线图倡议下，沃达丰等>50家成员企业承诺2050年0温室气体排放
	设施/服务提供	✓ 设新能源投资计划，以直接政府投资带动新能源规模提升	· 印尼政府投入约360亿美元政府资金，5年内新增1080万千瓦可再生能源发电规模
	法律及监管	✓ 建立能源碳排放交易机制，以高额成本换取有限排放配额机制，引导企业主动减排	· 德国通过设立排放配额交易机制，推动碳排放量较1990年减少近30%

资料来源：中国《绿色数据中心白皮书》；案头研究；罗兰贝格

图 4.12 针对推动“绿色算力”的政策建议总结

– 在规划及倡议方面，政府可积极号召云计算服务器、数据中心等领域内领先企业共同设立能耗控制目标，树立节能减排意识，推动“绿色算力”发展。在这方面，欧盟委员会刚刚发布了《欧洲绿色协议》，提出

⁸⁰ 《国际电联全球电子可持续性举措报告》

将推行一系列政策，力求在实现 2050 年净零排放的目标。该协议明确要可持续性地发展数字化的方向，并将拟定相关政策政策，提高“从宽带到数据中心到 ICT 终端设备”的能源效率和循环经济表现。此外，以 GSMA 为首的全球移动通信系统协会已开始在此方面积极践行，其在《巴黎协定》中制定了移动产业气候行动路线图倡议，提出在 2050 年之前实现温室气体净零排放目标。此后，GSMA 下逾 50 家成员单位（包括沃达丰、西班牙电信等）加入倡议，已陆续开始公布年度能源消耗情况，制定每年的能源消耗削减目标及策略计划，从而实现对其数据中心的能源消耗控制。

- 在公共服务及设施提供方面，政府可通过直接政府投资带动新能源规模提升，从而强化新能源作为数据中心等算力核心发源地的能源供给。以风能、潮汐能、地热能等清洁能源作为数据中心等算力的主要能源供给方式可帮助政府在源头上控制算力发展碳排放的影响。目前，各国都陆续出台了自己的新能源计划。其中，印尼政府在 2014 年提出 5 年新能源投资计划，通过投入约 360 亿美金的政府资金，导入水电、地热形式的新能源设备发展，新增 1080 万千瓦可再生能源发电规模，到 2019 年将可再生能源的装机总量从 1070 万千瓦增加至 2150 万千瓦。在 2017 年，印尼政府的投资计划已收效显著，可再生能源发展已经在其国家电力供应中占比已超过 12.62%，远远高于当初设定的计划⁸¹。
- 在法律及监管机制方面，各国可积极推动碳排放交易机制的建立，通过以高额成本换取有限排放配额机制，引导企业主动减排。在碳排放交易机制下，国家可针对各行业、各地区、各领域的信息技术行业设定碳排放总量上限，建立并完善碳排放交易机制和体系，强制要求参与企业通过购买或出售排放配额来满足其排放需求，实现总体节能目标。以德国联邦环境署为例，该署成立了专门负责管理碳排放交易活动的德国碳排放交易管理局，通过分配排放配额、监督配额拍卖、评估年度排放报告等，对碳排放交易活动进行监督管理。德国的碳排放交易机制已在碳排放领域取得了良好的效果，推动 2017 年德国的碳排放量下降到 9.05 亿吨，相较 1990 年减少了 27.7%。在中国，也已有北京、深圳、上海作为碳排放交易机制试点，自 2012 年起，开展碳排放交易方案在区域范围内的落地执行。在这样的大背景下，无疑对于企业降低碳排放提出了更高的要求。对于算力相关企业而言，AI 等新兴技术则是一个可有效降低能耗的手段——其可对数据中心的运营方式进行监控及智能化调整，从而实现高效的绿色升级。以华为华北区云数据中心（廊坊云数据中心）为例，其借助 AI 节能，从“制冷”升级为“智冷”。该数据中心由 AI 系统控制水温优化调节和功率优化调节，完成对于冷机、塔、泵的智能调节控制和水阀、风扇的调节控制。其 AI 系统通过收集风火水电、IT 负载等多维数据，利用 DNN 建模，精确预测 PUE，并采用深度强化学习在线训练，对接控制系统，形成闭环，自动求解最优 PUE 下的参数调节。从结果来看，在无 PUE 优化的状态下，其全年 PUE（预计）约为 1.42，在通过以 AI 系统为主的整体优化后，PUE 降低 0.116 至 1.304，获得显著提升。

⁸¹ 中国经济网《印尼大力发展可再生能源》2018

3) 针对网络质量挑战，政府部门的核心目标应在于利用各种政策手段及资源，提高国家整体网络速率及时延水平，从而为算力性能的发挥打好基础。其核心可用的政策抓手在于规划倡议及财税支持（如图 4.13）：

网络基础设施建设发展推动政策建议

		具体举措	代表案例
举措及抓手	规划及倡议	<ul style="list-style-type: none">形成地理覆盖与网速布局的矩阵式监控体系，持续推动网络建设进程通过调整网络基础设施使用权鼓励竞争，推动网络建设进程	<ul style="list-style-type: none">德国在明确网络建设“四步走”发展目标基础上，常态化开展矩阵式监控，推动网络建设有序开展西班牙通过MARCO法案1调整网络基础设施使用权的方式促进光纤网络建设
	财税支持	<ul style="list-style-type: none">采用公私合营模式（PPP），以少量政府财政投入牵引社会资本投入网络基建	<ul style="list-style-type: none">欧盟委员会和欧洲投资银行注资2.4亿欧元，后引导以德国复兴信贷银行等社会资金加入注资，共同推动欧洲欠发达地区宽带网络基础设施建设

1) MARCO：西班牙政府推出的电信法案，规定所有电信运营商需要无条件开放建筑物内的网络接入基础设施给所有运营商

资料来源：案头研究；罗兰贝格

图 4.13 针对提升网络质量的政策建议总结

- 在国家层级的规划方面，各国政府可在制定网络基础设施规划目标基础上，形成地理覆盖面与网速布局的矩阵式监控体系，持续把控与推动网络建设进程。以德国为例，其在 2016 即发布了“数字化战略 2025”，明确分阶段的网络建设“四步走”发展目标，在后续的网络建设进程中，政府从建设的地理覆盖面与网速两大维度，以矩阵式开展常态进程监控：在地理覆盖方面，政府会监控网络建设是否满足初期规划的由中心家庭到周边工业园区再到全部地理区域的覆盖；在网速布局方面，政府会监控是否按照从 50Mbps 到更高速的光纤、5G 网络的持续发展。在该矩阵式监控的保障下，网络建设得以顺利开展，截止 2017 年底，75.5% 的德国家庭现在能够获得至少每秒 50 兆以上的联网速度。以西班牙为例，西班牙政府通过 MARCO 法案有效推进了光纤网络覆盖率的提升，为超高速网络的发展和未来智能社会应用场景夯实了基础。2009 年，西班牙 CNMC（国家市场竞争委员会）实施 MARCO 法案，要求所有电信运营商无条件允许互相共享网络

接入设施，刺激了行业竞争并将新一代网络（特别是光纤）的部署成本降低了 75%。自此，各大国际电信运营商开始进入西班牙，与西班牙电信（Telefónica）展开竞争，光纤网络建设迅速铺开。到 2014 年，西班牙光纤覆盖率达到欧洲第一⁸²。此后，西班牙政府希望进一步促进运营商向农村等网络建设成本较高的地区部署光纤网络。因此，2016 年，西班牙政府对 MARCO 法案进行更新，规定 2016 年之后新建的光纤基础设施的产权归为运营商私有，不必强制开放。这项政策掀起了网络部署薄弱地区的基础设施建设热潮。到 2018 年，西班牙光纤入户率达到 63%，是欧洲均值的 2.6 倍。

— 另一核心政策抓手即是利用公私合营（PPP）的模式，以少量政府财政投入牵引大体量的社会资本投入，推动网络基础设施建设。欧盟于 2015 年发布的“数字化单一市场 2025”的规划中，提出到 2025 年为学校、医院及大型企业提供 1Gbps、在欧洲境内全部城市和乡镇地区实现最少 100Mbps 的高质量网络全范围覆盖。为满足这一需求，欧盟委员会和欧洲投资银行启动的宽带基础设施基金——即“连接欧洲宽带基金”，致力于对欧洲网络欠发达地区的宽带网络基础设施进行投资。该基金作为欧洲战略投资基金项目下首个投资平台，由欧盟委员会和欧洲投资银行注资 2.4 亿欧元，后引导以德国复兴信贷银行、意大利投资银行等各国银行为代表的社会资金加入注资，共同推动欧洲人口较少地区、欠发达地区的宽带网络基础设施建设。此外，新加坡政府以资讯通信发展管理局（IDA）为核心，出资引导成立负责设备运营的私营公司开放网络（OpenNet）。OpenNet 负责全国宽带网络的建设、运行和维护，而原有运行商如新加坡电信等以租用模式获取网络接入为终端客户提供服务。新加坡政府采用该运行模式，建立了政府牵引投资的引导方式，通过出资 7.5 亿新元，以催化剂的角色拉动了多家企业共多达 30 亿新元的投入，实现了以社会资源服务于社会的目标，同时也加快了新加坡高速宽带网络的建设和发展速度，为未来在线学习、视频会议、远程医疗等智慧化社会场景提供了良好的网络基础，奠定了智能社会的基石。

4）针对生态多样化发展的挑战，各国政府应从规划层面积极推动各类架构芯片、设备与应用开发、牵引移动化场景发展，以应对多样化挑战，同时通过建立多样化生态产品联盟，缔结各类架构上下游资源、提供人才支持；此外，在财税方面，通过项目专项资金的投入，切实支持相关多样化架构等技术研发，并配套相应法律法规支持（如图 4.14），具体来说：

⁸² 西班牙电信，2017

生态多样性发展推动政策建议

		具体举措	代表案例
举措及抓手	规划及倡议	<ul style="list-style-type: none">✓ 规划多样架构、设备与应用的发展方向，包括市占率等具体目标✓ 顺延网络发展，政府可通过推动智慧城市建设等手段，创造更多移动化应用场景	<ul style="list-style-type: none">· 欧盟资助欧洲处理器计划（EPI），打造欧洲自主研发的处理器家族· 伦敦议会发布《智慧伦敦规划》方案，牵引相关生态产业玩家发展
	设施/服务提供	<ul style="list-style-type: none">✓ 可通过建立多样化生态联盟，助力各类架构缔结产业上下游资源、同时为产业发展提供人才支持	<ul style="list-style-type: none">· 中国工信部支持成立绿色计算产业联盟，引入RISC服务器全生态链企业适配RISC· 欧盟资助EPI联盟，引入26家产学研用机构，计划2024年前推出三代多架构欧洲自主微处理器家族
	法律及监管	<ul style="list-style-type: none">✓ 放宽算力相关市场准入门槛，优化相关政务治理流程与效率，切身支持行业发展	<ul style="list-style-type: none">· 新加坡将数字金融专利审批时长从2年缩短至6个月，降低准入门槛

1)《关于建设人工智能上海高地 构建一流创新生态的行动方案（2019-2021年）》

资料来源：案头研究；罗兰贝格

图 4.14 针对推动算力生态多样化发展的政策建议总结

在针对未来的规划及倡议方面，政府一方面可通过规划针对各类架构的芯片、适配设备与应用的研究，应对应用场景丰富所带来的多样化挑战，一方面可通过推动具备移动化特征的场景建设，牵引以 ARM 为代表的 RISC 等算力生态发展。面向未来，在算力应用场景愈发丰富的大背景下，需要不同类型的芯片架构以适应多样化的智能社会场景。因此，政府应首先认知到未来智能化社会所提出的生态多样化需求，进而从规划做起，支持以 RISC 架构为代表的各类架构的芯片开发，设立分阶段的各类架构产品市场占有率等发展目标。为确保最终的应用效果，在规划中也应强化对于设备厂商、应用开发商基于 RISC 等架构的适配工作，从而推动各类架构体系发展形成良性循环，推动可持续发展的生态构成。以欧盟为例，欧盟委员会间接注资欧洲处理器计划（EPI）。欧盟希望通过 EPI 设计并执行欧洲微处理器自主发展的路线图，发展欧洲自主的低功耗微处理器，以适应百万兆级计算、高效运算和智能应用场景（如自动驾驶）等市场需求。计划提出到 2021 年，EPI 将推出的第一代欧洲自主研发的微处理器家族 Reha 将采用多样化的芯片架构。其主力通用芯片将基于 ARM ZEUS 平台，其加速器将基于 RISC-V 架构，并探索 MPPA 众核方案，eFPGA 方案和量子计算方案。另一方面，顺延网络发展的大背景，政府也应关注移动化应用场景的规划与牵引，可通过推动智慧城市建设等方式，为契合移动化场景需求的 RISC 架构的生态企业提供用武之地。以伦敦市为例，伦敦市议会发布了

建设未来智慧城市的《智慧伦敦规划》方案，规划了包括智慧交通、智慧生活等一系列具备移动化特征的应用场景，从而牵引相关生态产业玩家发展。

- 在设施 / 服务提供方面，可通过建立多样化生态联盟的方式，助力各类架构缔结产业上下游资源、同时为产业发展提供人才支持。比如，2016 年 4 月，在中国工业和信息化部的支持下，由多家国际领先服务器与云计算方案提供商以及国内高校、科研机构共同宣布成立“绿色计算产业联盟”，该联盟的出现为 RISC 架构生态的发展带来了显著产业资源与人才支持优势。在产业资源方面，截止 2019 年 10 月底，已有 55 个成员单位，涵盖完整的 RISC 服务器产业生态链，包括：传统服务器厂——商惠普、戴尔、华为、联想，互联网企业——阿里巴巴、百度，芯片厂商——ARM 和飞腾，以及软件公司——中软等等。以今年 10 月刚刚加入绿色计算产业联盟的 Marvell 为例，作为基于 ARM 架构开发服务器处理器的厂商，其已表示将进一步为基于 Arm 架构平台上的大数据、企业和云计算所需的核心应用程序开发提供支持。此外，该类联盟的出现也将为 ARM 等多类架构的发展带来人才支持。比如：绿色计算产业联盟已于 2018 年与上海交通大学建立了 ARM HPC 联合创新项目，通过引入联盟会员单位中专家提供前沿技术讲解、开展 ARM HPC 技术与生态研讨等方式，支持 ARM64 相关科研项目 and 人才培养，从而为 ARM 生态的发展提供重要人才基础支持。再比如，以欧盟为例，欧洲处理器计划（汇集了来自 10 个欧盟国家的 27 家合作伙伴）事实上是欧洲处理器计划联盟（汇集了来自 10 个欧盟国家的 27 家合作伙伴 Consortium）推动的落地计划之一，由欧盟地平线 2020 计划通过框架伙伴协议注资成立。汇集了来自 10 个欧盟国家的 27 家合作伙伴的研发力量依托于汇集了来自 10 个欧盟国家的 27 家合作伙伴联盟，汇集了来自 10 个欧盟国家的 27 家成员单位，覆盖了芯片研发、芯片生产、应用场景等多个产业环节。其中包括：高性能计算研发机构——苏黎世理工大学、E4 Computer Engineering 等，芯片生产——Bull，Extoll 等，汽车平台——宝马、劳斯莱斯等。汇集了来自 10 个欧盟国家的 27 家合作伙伴希望通过产业联盟的形式重组欧洲芯片行业的竞争力，确保高阶芯片设计的核心能力留在欧洲。
- 在法律及监管方面，可针对算力自身（芯片设计生产）及体系配套（设备、软件等）企业适当放宽市场准入门槛，并从提升政务治理流程效率出发，切身支持行业发展。以新加坡在数字金融领域推出的科技快速通道政策（FinTech Fast Track Initiative）为例，其特地加快专利批准的速度，并降低市场准入门槛，为适应金融科技软件的生命周期需求，推动金融数字化的发展。从 2018 年 4 月 26 日起，金融科技项目的专利审批时长将从 2 年缩短至 6 个月。
- 在上述政府的支持下，未来，算力有望实现多样化生态均衡发展（如图 4.15），具体包括：在芯片层，形成 X86、RISC 为代表的其他诸多架构类芯片多足鼎立的态势；在设备和软件层，针对不同架构也应具备良好的对接的设备、系统与应用，从而实现真正端到端的良好适配，最大化不同架构芯片的算力效用。

算力生态多样化理想格局



资料来源：案头研究，罗兰贝格

图 4.15 算力生态多样性旨在实现纵向配套完善、芯片层横向多指令集架构均衡发展的格局

5) 除了针对各项算力生态发展挑战的举措之外，在算力的安全问题上，各国政府需要关注的核心在于提升对于算力风险漏洞的防范能力，在规划及倡议方面，各国政府应重视联合制定国际安全规范；在财税支持方面，政府需要提供资金支持专业团队开展潜在风险识别及预警；在法律及监管层面，政府应建立并完善相关法律法规，鼓励相关组织和个人保护算力的安全（如图 4.16）：

安全应对能力提升推动政策建议

		具体举措	代表案例
举措及抓手	规划及倡议	<ul style="list-style-type: none"> 各国共同制定国际化安全标准，为全球算力发展提供标准参考红线 	<ul style="list-style-type: none"> 芯片层面，ISO与IEC共同制定ISO/IEC 9798，ISO/IEC 11770等标准，对密钥管理和加密算法等提出诸多明确建议，抵抗密码攻击
	财税支持	<ul style="list-style-type: none"> 公共财政提供专项资金，建设安全监测专业机构，且支持第三方安全机构持续监测运营 	<ul style="list-style-type: none"> 德国每年投入~5,000万欧元资助亥姆霍兹信息技术安全研究中心，持续开展针对网络安全与隐私风险的学术研究
	法律及监管	<ul style="list-style-type: none"> 建立完善的究责机制，督促相关从业者重视并切实遵守算力的相关安全规范 	<ul style="list-style-type: none"> 法国电子通信服务提供商若违反欧盟数据保护指令，将受到监禁5年处罚与30万欧元罚款

资料来源：案头研究；罗兰贝格

图 4.16 针对保障算力安全的政策建议总结

在规划及倡议方面，各国政府应联合共同制定国际化的标准安全规范，为全球算力发展提供统一安全底线。算力安全架构主要由网络、软件、设备、芯片四个层面构成（如图 4.17），在每一层面，皆需各国政府联合制定统一的标准安全规范，以确保算力发展的安全底线。比如在芯片层面，ISO（国际标准化组织）与 IEC（国际电工委员会）共同制定的 ISO/IEC 9798，ISO/IEC 11770 等诸多标准对实体鉴别、密钥管理和加密算法等提出了明确建议，比如：ISO-9796 RSA 加密算法是截止 2018 年 9 月最有影响力的公钥加密算法，它可抵抗截止到 2018 年 9 月的所有密码攻击；在设备层面，SHARE（由 IBM 发起的独立信息技术协会组织）制定的 SHARE78 容灾国际标准，涵盖了设备安全、容灾备份、日志审计等 7 大层次的容灾方案，具备极强的适用性。在软件层面，IEEE（电气电子工程师学会）通过 Std 982.1-2005 软件可靠性度量词典定义了软件可靠性度量参数，使用户可了解与评价软件可靠性；在网络层面，3GPP（通信技术的标准化组织）所发布的 SA3-5G SECURITY 标准建议采用基于密钥的技术方案，保证网络切片间隔离安全，从而支持各种应用场景对于安全的差异化需求。另外一方面，全球统一的标准则更有助于算力生态企业实行标准化开发，而无需为不同国家的不同标准开发针对性产品，从而降低研发成本与难度，推动整体算力生态企业发展进程。

算力安全架构

	安全内容				规范与机制	
网络层	接入认证	访问控制	密钥管理	...	1 国际安全规范建立	
软件层	漏洞管理	数据隔离	补丁管理	...	2 常态化的安全漏洞通报	
设备层	设备安全	容灾备份	日志审计	...	3 刑事纠责机制建设	
芯片层	数据加密	随机数自检	安全启动	...	4 第三方监管机制建设	
					5 安全隐私保护条例建立	
					...	

资料来源：案头研究，罗兰贝格

图 4.17 在规划及倡议方面，政府应重视算力安全架构及相应国际安全规范，从源头预防安全问题

在财税支持方面，政府可以通过公共财政提供专项资金，建设安全监测专业机构，且支持第三方机构的持续监测运营，以提前识别及预警潜在安全漏洞风险。以德国为例，每年由德国联邦政府投入 4,500 万欧元、地方政府注资 500 万欧元以支持德国亥姆霍兹国家研究中心联合会的信息技术安全研究中心开展信息安全研

究，以应对不断增加的网络犯罪风险。在该资金的支持下，约有 800 名行业专家在亥姆霍兹信息技术安全研究中心开展针对网络安全与隐私风险的学术研究，包括语义识别技术的认知系统，识别虚假信息，从而鉴别威胁情报。该研究所的科研成果可有效保护自动化系统、自动驾驶系统以及医学数据的信息安全，判定企业交易过程中的潜在风险，从而有望在未来有效扭转网络犯罪经济。

- 在法律法规及监管层面，需要政府建立完善的究责机制，督促相关从业者重视并切实遵守算力的相关安全规则。在德国，根据《联邦数据保护法规定》，对于未经公开访问的大量人数的个人资料，故意而未经授权传输数据给第三方，或出于商业目的使用其他方式使它们可被访问，将受到法律制裁和罚款；此外，《德国网络安全法》明确了“关键基础设施”运营者的法律责任，若网络运营商未在信息安全局提出风险警报后未及时修补漏洞，亦将被政府处以罚款的惩罚。

综上所述，为应对算力发展的未来挑战，政府需要在规划战略、基建服务、财税支持、法律监管等方面做好充足的准备，为一国算力水平的发展做好掌舵手的角色，才有可能在不远的将来迎来智能化社会的实现，再一次建立起国家竞争力的核心优势高地。

华为技术有限公司



深圳龙岗区坂田华为基地

电话: +86 755 28780808

邮编: 518129

www.huawei.com

商标声明

 HUAWEI, HUAWEI,  是华为技术有限公司商标或者注册商标，在本手册中以及本手册描述的产品中，出现的其它商标，产品名称，服务名称以及公司名称，由其各自的所有人拥有。

免责声明

本文档可能含有预测信息，包括但不限于有关未来的财务、运营、产品系列、新技术等信息。由于实践中存在很多不确定因素，可能导致实际结果与预测信息有很大的差别。因此，本文档信息仅供参考，不构成任何要约或承诺，华为不对您在本文档基础上做出的任何行为承担责任。华为可能不经通知修改上述信息，恕不另行通知。

版权所有© 华为技术有限公司 2020。保留一切权利。

非经华为技术有限公司书面同意，任何单位和个人不得擅自摘抄、复制本手册内容的部分或全部，并不得以任何形式传播。