

# 算力

孙凝晖 张云泉 刘宇航  
中国科学院计算技术研究所

关键词：科学教育 学科核心概念 跨学科概念

中文名：算力

英文名：Computility

学科：计算机系统结构

定义：算力是体现为用户实际效用的计算性能

实质：“算力”的本义是表示某个设备或系

统的计算性能，是计算性能的口语化表达。随着智能时代的到来，智能计算的三要素——算力、算法、数据，逐渐成为信息基础设施的重要组成部分，“算力”的内涵进一步扩大到用户能获得的体现为用户实际效用的计算性能。

“算力”是国内最近几年才频繁使用的概念，目前还没有严格的学术定义，国外也没有准确的英文词汇与之对应。我们新构造了一个英文词“computility”，该词是 comp(ute) 和 utility（效用）的聚合，旨在表达“算力”一词除直观的计算性能(performance)外的新内涵。

## 为何近年来“算力”成为热词？

近年来“算力”逐渐成为学术界、产业界、公众的热词，算力、算力网络、算力指数、算力经济、东数西算、东数西存、东数西训等新名词不断涌现。

“算力”原来只是超级计算领域关注的对象，代表着人们对“算得快”极致能力的不懈追求。深度学习技术的兴起带来了人工智能的热潮，标志着信息化从数字化、网络化走向智能化的新阶段，智能计算时代即将到来。人们把智能计算背后的三个驱动力归结为智能三要素——算力、算法和数据。

因此，“算力”一词开始备受关注。

随着人工智能与物联网技术的广泛应用，“计算无所不在”成为趋势。一方面，越来越多的物件因为嵌入了芯片或微型计算系统而具备了计算能力，例如智能水杯、插座、台灯等，人们乐于用“算力”来衡量物件的计算能力。另一方面，云计算的发展使越来越多的计算任务在云上完成，人们倾向于使用“算力”评估云计算中心的整体计算能力。

国家发改委为了推动“新基建”战略的实施，提出了“东数西算”和国家大数据枢纽节点等重大建设工程；中国移动、中国电信等运营商从云网融合的角度提出了“算力网络”；在学术界，中科院计算所从高通量计算的角度提出了“信息高铁”算力网[1]，鹏城实验室从人工智能发展的角度提出了“中国智能算力网”等。“算力”一词变得越来越热。

## “算力”的本义

“算力”从字面上理解就是计算能力，是从表示能力的电力、运力借鉴而来的。计算机学术界对“算力”这个概念有一个大致的共识，即表示某个设备或系统的计算性能，或者说“算力”是计算性能的口语化表达，类似于“电脑”是计算机的口语化表达。这里的计算设备包括个人电脑、手机、嵌入式设备等，计算系统包括片上系统(SoC)、超级计算机、高性能计算机、云计算系统、数据中心系统、分布式计算系统等。

人们一般使用运算速度作为表征计算性能的指

标,包括每秒百万次指令数(MIPS)、每秒浮点操作数(FLOPS)、每秒定点操作数(OPS)等,表示定点、半精度/单精度/双精度浮点、人工智能常用的8 bit 整数操作等方面的性能。人们一般用这些性能指标测度和表征设备或系统的算力,比如,一颗人工智能计算芯片的算力是256 TOPS,一台超级计算机的算力是1 EFLOPS。

在计算机学术界,为了准确刻画各种计算负载、体系结构对性能的影响,性能指标与基准测试(benchmark)的概念有很多。性能指标包括单位周期执行的指令数(instructions/s)、每秒执行的哈希数(hashes/s)、每秒执行的事务数(transactions/s)、尾延迟(tail latency)等;基准测试包括:面向单线程应用的SPEC CPU,面向数据库应用的TPC,面向超算应用的HPL/HPCG,面向云计算应用的SPEC Cloud,面向多线程应用的PARSEC,面向大数据应用的BigdataBench,面向图计算应用的Graph500,面向人工智能应用的MLperf、AIPerf等。公众知名度最高的可能就是超算TOP500排行榜了,它依据的其实就是使用HPL测度的计算机性能。

## “算力”的泛化

“算力”成为热门词最早出现在比特币挖矿领域。挖矿是比特币区块链的共识机制,也被称作工作量证明(proof of work),是指比特币区块链的所有共识节点通过算力竞争获得记账权的过程。具体而言,就是比特币节点把比特币网络中的未确认交易打包成候选区块,并填充区块中的随机数字段形成完整区块内容,然后按照统一的哈希算法(SHA256)算出一个哈希值。比较哈希值与一个特定目标值(难度值)的大小,如果哈希值小于特定目标值,则该节点就找到了符合要求的区块,取得了记账权,并获得系统奖励和交易手续费。特定目标值通常会根据全网的总算力定期调整难度,从而使得比特币区块链的出块速度维持在一个较稳定的范围内(每10分钟左右出一个区块)。对于单个共识节点而言,要算出这个符合要求的区块是非常困

难的,通常需要进行海量(2的几十次方)的哈希碰撞,因而一个共识节点(通常使用专业的矿机)每秒能做多少次哈希计算就成为它在比特币网络中能否取得优势的关键。所有节点的算力总和代表了比特币区块链网络的整体计算能力。矿机计算哈希函数输出的速度,就是“算力”的度量单位。例如,一个矿机达到10T h/s的哈希率时,意味着它每秒可以进行10万亿次计算。

“算力”进一步用在了数字经济领域。华为等企业和政府宏观经济部门提出了“算力指数”,来评估一个企业、一个区域乃至一个国家能提供的总计算能力,用终端和数据中心中安装的处理芯片、CPU核的数量等评估总计算能力,并且研究算力指数的增加与数字经济增加值之间的比例关系,衡量信息新基建的投资效率。“算力经济”也被提出<sup>[2]</sup>,用来表示超算产业以及超算赋能的各行各业的GDP增加值,济南市还率先建设了发展算力经济的超算产业园。

“算力网络”的提出让算力进入了基础设施范畴。一般产业界常用“算力网络”,学术界常用“算力网”。过去几年,国家和地方政府建设了很多超算中心,企业构建了很多云计算中心,高算力设备被密集地部署在这些超算中心,用于提供公共服务,算力的基础设施属性被不断强化。通信网络与互联网是全球性信息基础设施,是信息社会的底座,随着网络功能虚拟化(NFV)、软件定义网络(SDN)、确定性网络等技术的发展,网络的智能测调能力越来越强,自然地,算力与网络在基础设施层面的融合被提出来了。首先,谷歌等国际互联网企业在全局布局了很多数据中心,这些数据中心之间自建了互联的网络,自定义了互联的协议;随后,国内的互联网企业、云计算企业也在全国布局了多个数据中心,通过虚拟专网互联成分布式云;最后,电信运营商本来就具有网络上的优势,近年来也不断加强在公有云、边缘云的投资力度,云网融合成为必然<sup>[3]</sup>。由此,“算力网络”应运而生。

算力网的思想可以追溯到20世纪60年代美国学者约翰·麦卡锡(John McCarthy)提出的“utility

computing”，即“算力应该像电话系统一样，成为一个公共服务，用户可以随用随取，按用付费”；在超算领域，伊安·福斯特（Ian Foster）和卡尔·凯瑟曼（Carl Kesselman）提出的“网格计算”（grid computing）也曾风靡一时。我国高性能计算科技专项很多年都在支持这方面的研究，创业公司“并行科技”利用这类技术服务大量超算用户，对我国超算的普及起到很好的作用。云计算也是这一思想发展的产物，颠覆了企业数据中心市场，为互联网应用提供了可弹性伸缩、可靠性高的算力底座。当前提出的算力网不应该是云计算的简单优化，而应从全球信息基础设施的视角思考它的发展方向。

## 从消费侧扩大“算力”内涵

总体来说，“算力”的上述内涵都是从供给侧出发描述的。例如，在一个公有云里，申请“算力”往往表现为获得多少个虚拟机、多大内存容量等资源。

从人类发展的历史来看，经济的高速发展依赖基础设施的普及与应用深化。农业经济的发展依赖于交通基础设施和水利基础设施，工业经济的发展依赖于能源与电力基础设施、全球货运基础设施、全球金融基础设施和全球通信基础设施，数字经济的发展依赖于互联网、万维网等全球信息基础设施。数字经济进一步向智能化发展，依靠的是算力、算法和数据的全面基础设施化，包括算力基础设施、算法基础设施和数据基础设施。

一个能力的基础设施化的实质是这个能力在消费侧的内涵被充分挖掘。在算力领域，与美欧等发达国家和地区比较，我国的算力集中在终端，云端相对较弱；我国的消费互联网应用强，产业互联网应用弱；我国的算力资源为游戏、视频等信息流服务多，为信息深加工、模型精炼等智能流服务少。缩小这种差距的途径不是一味地增加供给侧的算力资源数量，或者比拼算力供给的价格，而是在消费侧发力，进行原理突破、技术创新、商业模式创新，让算力用户的“获得感”提升，即在一定的资源下及时完成问题求解。此外，利用我国信息基础设施

上的相对优势，也许是缓解我国在集成电路领域长期存在的单芯片性能不足的一个办法。

所以，我们把算力的定义修正为体现用户实际效用的计算性能，就是要强化算力用户的获得感。

为了表达“算力”的新内涵，英文我们没有采用维基百科上的 computing power，而是新构造了一个英文单词“computility”<sup>[4]</sup>，是 comp(ute) + utility 的聚合，体现了需要考虑算力作为公共服务品的一些特征，尤其是需要从用户消费公共服务的角度来考量。在这个理念下，用户只关心完成计算任务需要多少“算力”，即“真正让用户有获得感的操作的速率”（Operations of Interest Per Second, OIPS），类比用于物质加工的电力的功率“焦耳/秒”，computility 是用于信息加工的算力的“功率”；至于算力由谁提供，任务向哪里输送，算力提供方之间如何调度等，用户无需感知。算力网（可翻译成 computility grid）正是传输和调度这种算力任务的网络。

## 算力网的六个挑战

在智能时代，算力面临的技术挑战、商业挑战、生态挑战很多，需要新的算力部件和高效的算力站，算力基础设施成型后还需要与算法基础设施、数据基础设施进行深度耦合。从消费侧出发，我们认为算力网包括了六个技术挑战：多云统一、站网分离、变租为用、单一计量、网程抽象、质量测度。

多云统一是将“七国八制”的企业云计算中心、超级计算中心统一成“planet scale”级的算力基础设施，有统一的应用 API、云账号、云间算力拆借，做到应用跨云迁移无感。

站网分离是将算力资源站与算力网运营商分离，在技术手段上支撑无算力站的虚拟算力运营商的诞生。

变租为用是算力用户改变使用算力的方式。相对传统的个人计算，云计算是变买资源为租资源（变买为租）；相对云计算，算力网是变租资源为买效用（变租为用），有助于解决云计算存在的用户占大量资源而不用的问题。

单一计量是要有电力那样的单一计量方式，即用功率（瓦）和千瓦时（度）同时刻画消费侧与供给侧的行为。消费侧的设备都有功率标识，并且可以通过电表来度量；供给侧的发电装备产生的电能有很多参数，但对外提供服务时都简化为功率和千瓦时两个指标。

网程抽象是要给算力网用户提供表达任务的高级抽象，类似于万维网中的网页、算力站中的进程。网程就是算力网的网页加进程；进程通过线程、全局/局部数据、消息、文件、管道/套接字、信号量等少数概念支撑了繁多的单机任务、并行任务、分布式任务、网络任务的高效表达与指派（assignment）。

质量测度是定义算力消费质量的指标及其测量方法，比如通量（goodput），可同时体现吞吐率（throughput）与尾延迟，收费依据不能仅仅是算力资源数量的消耗与占有时长。

## 哲学思考——计算的概念簇

算力概念是与时俱进的时代产物，我们把时空尺度拉得更大一点，思考往科学哲学上靠一点，可以形成一个计算的概念簇<sup>[5]</sup>：算术、算法、算礼、算力、算势，其中算术、算法、算礼刻画了计算方法，算力、算势刻画了计算能力。

算术是关于数值的计算方法，是狭义的算法，也是最基本、最顾名思义的算法，强调计算的技巧，如中国乘法和印度乘法就使用了不同的技巧。

算法是广义的计算方法，包括数值算法和非数值算法，强调计算的机械的规则，尽管算法是由人脑设计，但人脑本身善于巧思，不善于执行机械的规则；大量的问题往往通过机械的规则容易解决，使用人工智能技术设计算法应是可行的。

人脑与计算机之间存在着紧密联系，又存在着内在机制和语义沟通上的鸿沟。算礼是关于计算如何在计算机系统上进行的规范<sup>[6]</sup>，比算法更接近人脑，更关注计算机系统的整体，强调计算的系统可被人脑直接评估。算礼是不成文的，相对算法而言是软性的，但其褒贬意义上的影响力不可替代。

算力是应用程序能实际获得的计算能力，其世界是算势。算势强调计算速度的潜能，不同数量级的算势能求解的问题复杂度也有数量级的差异。每一个量级的算势对应一个可求解的问题域，随着算势增大，可解域也在增大。算势是社会生产力的一个重要指标。

算力来源于算势，受限於算势。如何弥合两者之间的鸿沟，实现从算势到算力的高效转化，是包括计算机系统结构和系统软件在内的整个计算机科学需要研究解决的核心问题。

从科学哲学的角度来看，算力本质上是一种“功率”，算力网络是算势的载体，本质上是算力场（computility field）。 ■

### 作者介绍

#### 孙凝晖

CCF 会士、常务理事。中国工程院院士，中国科学院计算技术研究所研究员。主要研究方向为计算机系统结构、高性能计算机。

#### 张云泉

CCF 杰出会员、理事、高性能计算专委会秘书长。中国科学院计算技术研究所研究员。主要研究方向为并行算法、并行软件。zyq@ict.ac.cn

#### 刘宇航

CCF 高级会员、CCCF 特邀专栏作家。中国科学院计算技术研究所副研究员。主要研究方向为计算机系统结构、存储系统。liuyuhang@ict.ac.cn

## 参考文献

- [1] 徐志伟, 李国杰, 孙凝晖. 一种新型信息基础设施: 高通量低熵算力网(信息高铁)[J]. 中国科学院院刊, 2022, 37(1): 46-52.
- [2] 张云泉. 算力经济登上历史舞台[OL]. 2020-10-11. baijiahao.baidu.com/s?id=1680185917944441133&wfr=spider&for=pc.
- [3] 李正茂等. 云网融合——算力时代的数字信息基础设施[M]. 中信出版社, 2022.
- [4] 孙凝晖, 张云泉, 张福波. 算力的英文如何翻译? [J] 中国计算机学会通讯, 2022, 18(9): 87-88.

更多参考文献: <http://dl.ccf.org.cn/cccf/list>

(本文由 CCF 术语工委执行委员林俊宇推荐)