

工信部推动“轻量化5G”普及 2025年将实现全覆盖

本报记者 谭伦 北京报道

在我国5G进入第四个商用年头后,产业布局初期留下的技术与市场空白正在获得进一步的完善。

8月29日,工业和信息化部在官网发布《关于推进5G轻量化(RedCap)技术演进和应用创新发展的通知(征求意见稿)》(以下简称《征求意见稿》),旨在推进5G轻量化(RedCap)技术演进、产品研发及产业化,大力推动5G应用规模化发展。

《中国经营报》记者了解到,5G RedCap的全称为5G Reduced Capability,直译为“5G降低能力(版本)”,由于其出自全球5G标准制定组织3GPP定义的一种新型

填补5G中速应用空白

目前仍有40%~50%的5G速率场景中缺少匹配的产品,而这或也成为5G RedCap在政策层面加速获得推动的原因。

对于在目前的时间点决意加快我国5G RedCap发展的目的,多位业内分析人士向记者表示,我国5G发展所处的新阶段是此次文件出台重要的背景,而5G RedCap可以极大填补我国5G在当前应用领域需求的空白。

据Omdia电信战略高级首席分析师杨光介绍,根据3GPP初期定义,5G主要包括高带宽(eMBB)、广连接(mMTC)和超高可靠低时延(URLLC)三大特性。其中,mMTC主要是指物联网,后者可分为高速、中速和低速三大连接应用场景,而高速和低速场景都有相应支持技术,唯独中速场景存有技术空白。

在此背景下,3GPP从5G第三个版本R17开始,对主要服务中速物联网场景的RedCap开始制定标准。截至2022年6月,R17标准正式冻结,5G RedCap的定义也就此明晰,除了对于低成本、大容量高效共存、高集成、多功能给出具体指标外,也明确将主要面向可穿戴设备、工业传感器和视频监控三类业务。

而在日前举行的5G RedCap行业论坛上,中兴通讯无线产品规划总监郝瑞显则更为具体地指出,5G RedCap可通过终端能力裁剪,达到

技术标准,旨在通过减少终端带宽、收发天线数量、降低调制阶数等方式,降低终端成本和功耗,以便更利于5G的普及应用,因此,5G RedCap也在业内被广泛理解为“轻量化5G”。

《征求意见稿》明确提出,到2025年,5G RedCap产业综合能力显著提升,新产品、新模式不断涌现,融合应用规模上量,安全能力同步增强。其中,全国县级以上城市实现5G RedCap规模覆盖,5G RedCap连接数实现千万级增长。

围绕此目标,《征求意见稿》也进一步从技术标准、产业体系、能力升级、应用创新、示范标杆、生态环境、安全保障等多个维度,明确了当前的主要任务,并提出了保障措施。

大幅降低成本、功耗与尺寸的效果,同时继承5G的优质能力,在充分满足应用场景需求的同时,实现5G网络性能与成本平衡,进一步推动5G规模应用,扩展5G应用场景。

值得注意的是,早在2021年3月,中国通信标准化协会(CCSA)便启动了RedCap的相关研究和立项工作,并联合业界在行业会议上通过了一项《征求意见稿》,为相关行业标准制定奠定了基础。

与此同时,2021年7月,工业和信息化部发布了《5G应用“扬帆”行动计划(2021—2023年)》。计划提出,加快弥补产业短板弱项,加快轻量化5G芯片模组的研究及产业化,进一步提升终端模组性价比。IMT 2020(5G)推进组对RedCap寄予厚望,全力推动产业成熟。这也被视为是5G RedCap政策出台的前奏。

此外,有运营商人士告诉记者,虽在标准层面拥有了主要框架,但由于在终端层面仍存在功耗大、成本高等问题未能解决,目前仍有40%~50%的5G速率场景中缺少匹配的产品,这部分阻碍了我国移动互联网的推进,而这或也成为5G RedCap在政策层面加速获得推动的原因。

芯片与模组先行

中国信通院技术与标准研究所副总工程师李侠宇认为,目前5G RedCap产业形成了模组终端双向发力的局面。

标准建立后,产业链上下游的协同发展成为政策关注的重点。记者注意到,此次《征求意见稿》明确提出,推进5G RedCap芯片、模组、终端、网络、仪表等产品研发和产业化,加快RedCap与网络切片、高精度定位、5G LAN(局域网)等5G增强功能结合,满足不同行业场景应用需求。

国信证券研报指出,运营商、设备商、模组厂商及芯片厂商是当前布局RedCap的产业代表,各方目前在终端测试、标准制定、技术研发等方面持续进展,有望于今年开始迈入商用,推动物联网产业发展。

运营商方面,记者从中国移动了解到,目前,中移物联已推出了RedCap模组——MR880A系列产品,同时,已与主流网络设备厂商、芯片厂商紧密协作,主导开展面向商用的5G RedCap端到端测试验证,并已完成5家主设备厂商5G RedCap组网试验,正在进一步开展终端芯片和模组的技术试验。

中国电信方面则透露,其已

规模商用面临三大挑战

当前窄带物联网等蜂窝物联网技术已趋于成熟,想让垂直行业接受RedCap新技术,需全产业链共同努力打造产业生态。

虽然5G RedCap产业已初具商用能力,但业内认为,在迈入大规模商用前,仍然需要面临来自成本、网络覆盖及生态方面的挑战。

记者从一家模组企业人士处了解到,商用初期5G RedCap模组的价格约在200元左右,而目前占据主流的4G LTE Cat.4模组在60~80元之间,Cat.1模组成本更是低至40元以内。“4G模组经过了这么多年商用,出货量达到亿级左右,成本才降到这个水平,5G RedCap肯定也需要经历较长时间的出货量爬坡,



工业和信息化部日前在官网发布《关于推进5G轻量化(RedCap)技术演进和应用创新发展的通知(征求意见稿)》,旨在推进5G轻量化(RedCap)技术演进、产品研发及产业化,大力推动5G应用规模化发展。

对接RedCap行业终端超过50款,验证了超20个行业应用场景。此外,中国联通围绕RedCap技术研究、测试验证等方面开展了一系列工作,在测试认证方面,中国联通在业内建立了“仪表+模拟网+实网+场景试验场”四位一体的测试认证体系。

设备商方面,中兴通讯相关负责人表示,已与IMT-2020 5G推进组及中国四大运营商完成国内5G全频段RedCap功能及性能

才能摊薄成本。”该人士向记者表示。

而与成本因素同源,4G网络的成熟度也提升了当前5G RedCap的推进难度。在杨光看来,4G商用距今不到10年,也是目前覆盖最完善、传输最稳定、技术最成熟的通信网络,这意味着以4G网络为基础的模组产品仍有较长的生命周期,在成本优势与4G网络释放明确退网信号前,产业链替代意愿会比较低。

同时,4G网络建立的优势也表现在其覆盖深度。工信部数据显示,截至2021年年底,我

国已建成并开通了590万4G基站,占基站总数的比例近60%,即使是偏远地区和农村地区,也实现了4G的覆盖。对此,C114通信网主编周桂军向记者表示,虽然我国已建成全球规模最大的5G网络,但与覆盖更加完善的4G网络相比,5G网络在深度覆盖方面仍然有待提升,尤其是在室内等特殊场景的深度覆盖亟须加强,这将加大RedCap的普及难度。

此外,5G RedCap生态建立的挑战也是目前行业共识。在今年年初举行的第三届全球6G技

术大会上,5G物联网产业联盟秘书长解洲指出,当前窄带物联网等蜂窝物联网技术已趋于成熟,想让垂直行业接受RedCap新技术,需全产业链共同努力打造产业生态。

“物联网是个长尾市场,包括运营商、设备商、芯片模组、应用等多个方面,一项新技术的落地,需要产业链各方协作磨合,产业生态的建立更需要时间。”杨光表示,这意味5G RedCap的繁荣仍是一个较为漫长的过程,目前政策方面释放了积极的信号,而产业链需要努力和耐心。

AI算力不止GPU 国产大模型有望“摆脱”英伟达?

本报记者 秦鼎 北京报道

自ChatGPT爆发之后,AI大模型的研发层出不穷,而在这场“百模大战”激战正酣之际,美国芯片公司NVIDIA(英伟达)却凭借其GPU在大模型计算中的出色发挥赚得盆满钵满。

就在近日,科大讯飞创始人、董事长刘庆峰表示:“华为的GPU(图形处理器)能力现在已经跟英伟达A100一样了,现在已经做到对标英伟达的A100。”

华为方面对此并没有回应,不过,多位业内人士对《中国经营报》记者表示,目前,国内在AI芯片领域已经取得了一定的成就,但要与英伟达A100相媲美,还存在一些挑战。

GPU被掣肘

英伟达在去年收到美国政府的“通知”,通知称:“若对中国(含中国香港)和俄罗斯的客户出口两款高端GPU芯片——A100和H100,需要新的出口许可。”不仅如此,该许可证要求还包括未来所有的英伟达高端集成电路,只要其峰值性能和芯片间I/O性能均大于或等于A100的阈值,以及包括这些高端电路的任何系统,都被纳入许可证限制范围。

不过,英伟达针对中国客户推出了替代型号A800,与原有的A100系列计算卡相比,A800系列的规格基本相同,比较大的区别在于NV-Link互连总线的连接速率,A100系列为600GB/s,A800系列则被限制在了400GB/s,综合使用效率只有

弯道超车

“中国有很多GPU的初创公司,不要低估中国在芯片领域的追赶能力。”英伟达CEO黄仁勋曾表示。

虽然距离国内厂商赶超英伟达尚有较大的距离,但AI芯片并非只有GPU。

AI场景需要多核、高并发、高带宽的AI芯片。AI芯片,也被称为AI加速器或计算卡,即专门用于处理人工智能应用中的大量计算任务的模块。当前,AI芯片主要分为GPU、FPGA(现场可编程门阵列),以及TPU、VPU为代表的ASIC(特定应用集成电路)芯片。

FPGA属于具备可编程硬件结构的集成电路,其可编程性和灵活

性可快速适应不同AI算法要求,但存在高功耗的问题。ASIC属于专用芯片,通过算法固化实现更高的算法利用率和能耗比,但开发周期较长、灵活性较弱。

郭涛表示,中国在AI芯片领域的研发和生产方面已经取得了一些重要成果,除GPU芯片外,未来有望在CPU(中央处理器)、FPGA芯片、ASIC芯片等多个方面实现弯道超车。

张孝荣也认为,除了GPU之外,国内厂商还可以研发其他类型的加速器,如FPGA、DSP等。这些加速器可以针对特定的AI计算任务进行优化,提高计算效率和性

能。除此之外,还可以通过优化算法和模型设计,减少对计算资源的依赖,提高计算效率。例如,剪枝、量化等技术可以减少模型的参数量和计算量。将AI计算能力下沉到设备端,减少对云端计算资源的依赖。通过在设备端进行本地计算,可以提高响应速度和数据隐私保护。总之,国内可以通过不断技术创新和技术积累,在AI芯片领域有所突破,不要着眼于不切实际的弯道超车,而要以满足实际需求为主,逐渐缩短GPU领域的差距。

安光勇表示,国内致力于发展本土芯片产业,通过政策和资金支持,促进产业生态的形成,这将有助

于国内企业在芯片领域进行合作和创新。不仅如此,中国在人工智能算法和应用方面也有很强的优势。即使在芯片技术上难以迅速迎头赶上,通过优化算法和软件,仍可以在性能和效率上取得优势。

实际上,国内的厂商也在不断实现突破。“当前中国大模型中有一半是由华为昇腾AI来支撑的。”今年7月,华为轮值董事长胡厚崑对外透露,截至目前,昇腾已认证了30多家硬件伙伴、1200多家软件伙伴,联合孵化了2500多个AI场景方案。在中国,平均每两家AI企业就有一家选择昇腾。



刘庆峰表示,华为在GPU领域取得了重大进展,实现了与英伟达A100 GPU相当的能力和性能。

芯片产业高速发展,在一大批关键核心技术取得新突破,目前已经具备生产媲美英伟达A100芯片的能力,但若想实现大规模出货仍然面临很多挑战。

全联并购公会信用管理委员会专家安光勇也认为,目前,国内

在AI芯片领域已经取得了一定的成就,但要与英伟达A100相媲美,还存在一些挑战。英伟达A100是一款面向高性能计算和人工智能任务的GPU,具备出色的计算和训练能力,但要达到A100的水平仍需要时间。

作为“科创板AI芯片第一股”的寒武纪此前在互动平台回复,公司设计、研发的智能芯片不属于GPU,是面向人工智能领域专门设计的芯片。智能芯片的性能和能效优势主要集中于智能应用,在人工智能领域可以替代GPU芯片,但不适用于人工智能之外的其他领域。

而在“传统”的GPU领域,国内厂商也有建树。摩尔线程全功能GPU显卡可以完成AI模型的训练、推理、精调全流程,目前可以进行ChatGLM和GLM系列模型的推理,并且支持超大大模型的分布式推理和offload推理。