AI智算发展对高速光模块的应用需求研究

Study on Application Demand of High-speed Optical Modules for AI Intelligent Computing Development

栾昊立[1]，王晓东[1]，杨　锐[1]，郝建宇[2]，赵铭浩[1]，尹祖新[1]，王丽琼[1]（1. 中国联通研究院，北京 100048；2. 中国联合网络通信集团有限公司，北京 100033）

Luan Haoli[1], Wang Xiaodong[1], Yang Rui[1], Hao Jianyu[2], Zhao Minghao[1], Yin Zuxin[1], Wang Liqiong[1] (1. China Unicom Research Institute, Beijing 100048, China; 2. China United Network Communications Group Co., Ltd., Beijing 100033, China)

摘　要：

AI智算技术的高速发展驱动高速光模块需求量激增，大规模数据处理、大模型训练和推理等任务对高速光模块提出了前所未有的高要求。通过分析大模型训练的分布式并行计算需求，建立通信模型，并以GPT-3为例定量分析大模型通信量，由于通信量巨大，完成大模型训练的数据通信时间远高于并行计算时间。因此，在不降低计算性能的前提下，降低通信时间成为AI智算对通信网络的核心诉求，而采用更高速率的光模块互联、提升有效带宽是解决问题的主要途径。AI智算对高速光模块技术的需求将主要体现在更高速率、更大规模、高集约化、低功耗、高稳定性以及可管可控等方面。

关键词：

AI；智算；大模型；GPT-3；高速光模块

doi：10.12045/j.issn.1007-3043.2024.06.002

文章编号：1007-3043（2024）06-0007-05

中图分类号：TN913.7

文献标识码：A

开放科学（资源服务）标识码（OSID）：

Abstract：

The rapid development of AI smart computing technology drives the surge in demand for high-speed optical modules, and tasks such as large-scale data processing, large model training and reasoning put forward unprecedentedly high requirements for high-speed optical modules. A communication model is established by analyzing the distributed parallel computing requirements for large model training, and quantitatively analyze the large model communication volume by taking GPT-3 as an example. Due to the huge communication volume, the data communication time for completing large model training is much higher than the parallel computing time. Therefore, under the premise of not losing computing performance, reducing communication time becomes the core demand of AI computing on communication network, and adopting higher speed optical module interconnection and improving effective bandwidth is the main way to solve the problem. The demand of AI computing on high-speed optical module technology will be mainly embodied in the aspects of higher speed, larger scale, high intensification, low power consumption, high stability, and manageable and controllable.

Keywords：

AI; Smart computing; Big model; GPT-3; High speed optical module

引用格式：栾昊立，王晓东，杨锐，等. AI智算发展对高速光模块的应用需求研究[J]. 邮电设计技术，2024(6)：7-11.

1　概述

随着人工智能技术的飞速发展，全球智算数据中心建设加速，而光模块是智算数据中心内保障数据传输的关键组件，其市场需求也随之激增。根据Light-Counting 2023年7月发布的《超级数据中心光学报告》预测，用于AI集群的以太网光模块总销售额将持续增长，2028年达到176亿美金，占所有以太网光模块市场的38%，市场占有率较2023年提升13%，其中800G光模块将在2024年规模应用，1.6T光模块将在2026年规模应用。AI智算驱动高速光模块发展是业界共识，但以大模型训练为代表的AI应用是如何影响智算数据

收稿日期：2024-04-29

中心内的通信需求，进而驱动光模块的速率提升和用量的激增，业内却鲜有定量论述。本文对此展开研究，以AI大模型并行计算需求为源，以智算数据中心内高速光模块发展趋势为宿，层层剖析和关联，建立通信模型，进行定量测算。

2 AI大模型的分布式并行计算需求

2.1 算力需求激增与单卡训练瓶颈

AI大模型的快速发展和应用场景的拓展驱动智能算力的规模与需求持续提升。截至2023年6月底，我国算力总规模达到197 EFLOPS，其中智能算力规模占比已超过25%并不断提升，增速同比增长45%，较算力规模整体增速高15个百分点[1]。AI大模型具有庞大的参数规模、复杂的网络结构、海量的训练数据以及高精度的计算要求等特点，对算力资源的需求呈指数级增长，从2013年AlexNet的问世到2023年ChatGPT的爆火，算力需求增长了数十万倍，如今GPT系列的大语言模型参数规模通常以数十亿甚至上百亿计，训练数据量达到了数百TB。

AI大模型训练中海量的参数对显存和算力都带来了巨大的挑战，传统模型中单张显卡解决全量计算需求已经不太现实。首先，显存容量是主要限制因素，AI大模型通常包含数以百万计甚至亿计的参数，这需要巨大的内存空间，而单块显卡的显存容量有限，这可能导致无法加载完整的模型，从而限制模型的规模和复杂度。例如，GPT-3在运算时的峰值显存达到了2.8 TB，相当于40张A100 GPU（单卡显存70 GB），而日常应用与游戏的显存需求一般不超过10 GB。其次，AI大模型的训练涉及大量的矩阵运算和深度学习算法，单块显卡无法满足训练时的算力需求，从而导致训练速度缓慢，甚至无法完成训练任务。单块GPU需要32年才能完成总算力为10 000 PFLOP/s-day的GPT-3的训练[2]。

2.2 典型的并行计算技术

AI大模型训练需要大规模的集群算力处理，单个计算设备已经不能满足模型训练的需求。随着数据并行和模型并行技术的不断完善和提升，分布式训练中可以使用千卡或万卡规模的GPU集群来缩短整体训练时长。目前常见的并行技术主要包括数据并行、张量并行和流水线并行。

数据并行（Data parallelism，DP）是指将训练数据划分到多个训练单元，多个训练单元之间按照一定规则定期同步模型参数实现并行训练的一种方式。DP是最早被提出，也是目前应用最广泛的并行智能训练方法[3]。DP将模型复制多份，每一份模型参数放置在1个GPU或多个GPU组成的一个计算单元上。在训练过程中将同时输入多个小批次数据，并让每个GPU负责一个小批次数据的计算，利用大量计算设备提高模型训练吞吐量，减少模型训练时间。

张量并行（Tensor parallelism，TP）是基于算子内切分组织的训练过程，是模型单层所需存储空间大于设备存储空间时训练模型的有效解决方案。TP以张量为计算单位，结合张量的特性以及训练模型所用GPU的特性进行切分，将张量分配给多个GPU进行处理，降低设备的计算存储负载[4]。

流水线并行（Pipeline parallelism，PP）是一种将复杂任务分解为多个子任务，并通过时序控制将这些子任务分别交给不同的处理单元执行的技术。PP的核心思想是将模型按层分割成若干块，每块都交给一个设备。在前向传递过程中，每个设备开始计算并将结果张量传递给下一阶段。在后向传递过程中，每个设备将输入张量的梯度回传给前一个流水线阶段[5]。

3种并行技术对应了不同的数据信息通信和处理方式。DP是在不同服务器中不同的GPU上运行同一批数据的不同子集再进行聚合；TP是将单个数学运算拆分到一个服务器内不同GPU进行运算；PP则是将模型不同层拆分到不同服务器的GPU上运算并进行聚合，完成点对点通信。

3 AI大模型训练的通信需求模型

3.1 AI大模型训练的通信数据流

本节以Transformer架构为例对AI大模型训练的数据流进行分析。Transformer架构是由Google Brain团队在2017年提出的一种用于序列学习的模型[6]，也是现今GPT架构的基础，它以注意力机制为核心，摒弃了传统循环神经网络中的复杂结构，大大简化了模型并提高了并行化能力（见图1），Transformer架构主要由4个部分组成：输入、输出、编码器和解码器。

Transformer架构数据流可以概括为：在输入端将文本指令进行文字向量化处理分割成一个个单词后映射进高维向量空间，再进行位置编码。随后，输入的单词序列进入编码器的自注意力层（图1中红色方框部分），在这里每个单词都会关注到输入句子中的其他单词，并计算出一个新的单词。随后这些文字向

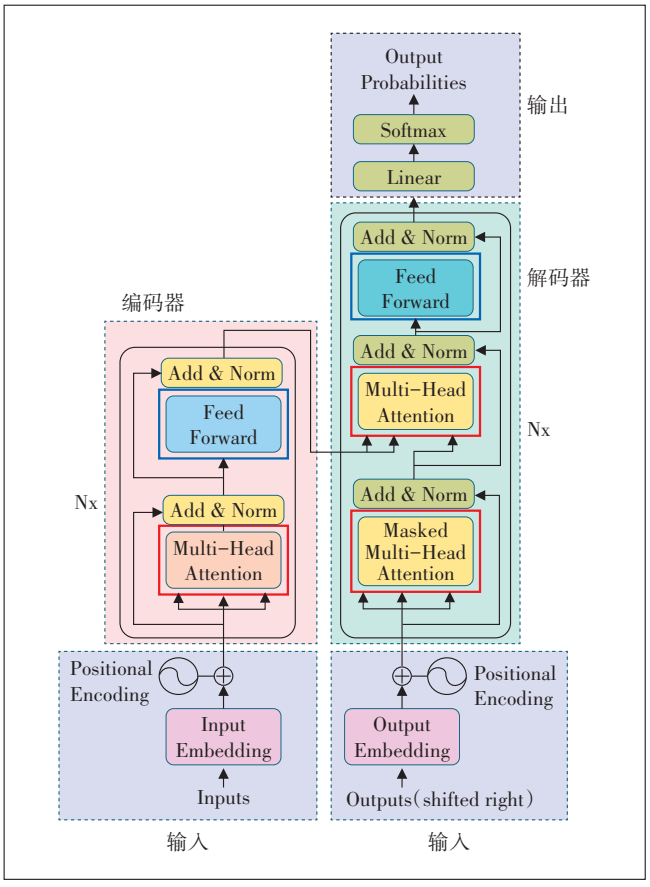

图1　Transformer架构

量会进入多层感知器层(图1中蓝色方框部分)完成非线性变换，提取更高级别的特征。这些过程会在编码器中重复多次，形成多个编码器层，每一层都会进一步提炼输入数据的特征[7]。同样，解码器首先会对其自身的输入进行自注意力处理，并且解码器还有一个额外的交叉注意力层，分析考虑整个输入序列的上下文语境优化解码器的输出。最后，解码器的输出会经过线性变换和Softmax操作[8]，得到每个词的概率分布，模型会选择概率最高的词作为预测结果。这个过程会不断重复，直到生成完整的输出序列。

这里以Transformer架构为基础来量化GPT-3中的参数规模。GPT-3模型中有96个解码器，每个解码器都包含自注意力层和多层感知器层，它们对应的参数规模分别为$4h^2$和$8h^2$，其中h=12 288。因此，GPT-3的参数规模为1.74×10^{11}。

在实际应用中，面对GPT-3这样庞大的参数规模的模型时，DP、TP和PP 3种并行计算方式通常会结合使用情况去提高模型的训练效率和可扩展性。

3.2 GPT-3通信需求量化

本节尝试构建出使用Transformer架构进行大模型训练时，单张GPU卡在一轮迭代中每种并行计算方式的通信量模型，并以GPT-3为例进行量化。一轮迭代通信的数据量可以由式(1)得到：

单次点对点通信数据量×集合通信内通信次数×

单次集合通信量×一轮迭代中集合通信次数　(1)

DP使用了通信集合库(Collective Communication Library)中的AllReduce这一通信操作来传输大模型中各个计算节点的数据。它通过从多个节点获取数据到所有节点并作规约运算，高效地聚合和同步梯度或其他数据。DP的集合通信次数为单张GPU卡上的层数，根据式(1)可以得到DP单卡一轮迭代总通信量(单位为B)为：

$$\frac{48\ (D-1)N_{\text{decoder}}h^2}{P\times T\times D} \tag{2}$$

其中，D、P、T代表了Data、Pipeline、Tensor 3种并行维度，N_{decoder}为模型解码器数量等同于模型层数，h为模型的隐藏层宽度，决定了模型的大小。

TP同样使用了AllReduce操作计算节点间的交互，但与数据并行不同的是，它是在一台服务器的多张GPU上进行AllReduce。其中TP的集合通信次数为正向计算、反向计算、重计算3轮，并且在模型的每一层中的自注意力层都与多层感知器层通信一次。综上可以得到TP单卡一轮迭代总通信量(单位为B)为：

$$\frac{24\ (T-1)\times N_{\text{decoder}}B\times s\times h}{P\times T\times D} \tag{3}$$

其中，B表示单次传递给模型用以训练的数据个数，s为输入序列的长度。

PP不同于前2种使用AllReduce操作的计算方式，它是在相邻服务器间点对点传递计算结果，通信次数为正向计算、反向计算、重计算共3次，由此可得到PP单卡一轮迭代通信总量(单位为B)为：

$$\frac{6B\times s\times h}{D} \tag{4}$$

表1所示为GPT-3参数数值。根据表1，将千卡GPT-3的参数值带入到式(2)~(4)中得到数据并行、张量并行、流水线并行3种并行模式一轮迭代总通信量分别为：9.5 GB、567 GB、13.5 GB，由此一个千卡配置的GPT-3一轮的通信量接近600 GB。

3.3 AI大模型训练对通信网络的核心需求

表1　GPT-3参数数值

参数	h	s	N_{decoder}	B	D	T	P
GPT-3数值	12 288	2 048	96	1 536	16	8	8

通过3.2节的测算可知，完成一轮迭代的通信量巨大，所以训练时间的长短十分重要，将直接影响业务性能。大模型的训练时间包括计算时间和通信时间，而随着并行计算技术的应用和GPU性能的不断增强，计算时间大幅缩减，通信时间则成为影响训练时间的关键因素。表2以GPT-3数据并行的方式处理9.5 GB通信量为例，展示了对应的通信时间。

表2　GPT-3数据并行通信时间

GPT-3	端口带宽/(Gbit/s)	网络跳数	单设备转发时延/μs	单设备数据传输时间/μs	光纤传输时间/μs	总时间/μs
数据并行9.5 GB	200	3	2.4	376 239.4	10	376 256.6

通信时间=主机内内存拷贝和协议栈处理时延+数据传输时间+交换机转发时延。数据传输时间=通信数据量/有效带宽，时间为秒级，对通信时间的影响最大，占比超过90%。转发时延=单设备转发时延×跳数，它与内存拷贝和协议栈处理所用时间均为毫秒级及以下。

如果通信时间过长不仅会拖慢训练进度，影响训练效率，甚至可能导致数据不一致或训练失败。因此，优化通信网络、减少通信时间，对于保障大模型进行高效训练至关重要。首先，对于优化微秒级的交换机转发时延来减少通信时间的效果可以忽略不计。其次，减少主机内的内存拷贝和协议栈处理时延主要依靠远程直接内存访问（Remote Direct Memory Access）技术将数据直接从一台计算机的内存传输到另一台计算机的内存，绕过了操作系统的介入来实现[9]，但是整体通信时间的提升效果并不显著。因此，在AI模型通信数据量不变的前提下，提升有效接入带宽来减少数据传输时间是实现通信时间优化的关键，实现的方式主要有如下2种。

a）使用400G/800G/1.6T等更高速光模块来增大接入端口容量。

b）采用Fat-tree组网结构[10]，上行带宽∶下行带宽=1∶1，网络流量不收敛，但同样需要更高的上层光口带宽。

4 AI智算对光模块的需求

4.1 更高速率、更大规模

在自动驾驶、实时语音识别等重要AI应用场景中，对响应时间有着极高的要求，由第3章分析可知，只有提升AI智算的网络有效带宽，减少数据传输的延迟，才能满足高实时性的性能要求。通过提升光模块速率来提高单通道通信能力，通过增大光模块互联规模来提供更多的数据传输通道，进而在AI算法中处理大量的数据和计算任务时，获得更高的数据吞吐能力和更强的并行处理能力。

英伟达在2023年5月发布了DGX GH200超级计算机，它使用NVLink互连技术与交换机系统，形成了一个两级、无阻塞的胖树（Fat-Tree）NVLink结构，将256个GH200超级芯片组合在一起（见图2）。DGX GH200中的每个GPU都能以900 GB/s访问其他GPU的内存和所有NVIDIA Grace CPU的扩展GPU内存。

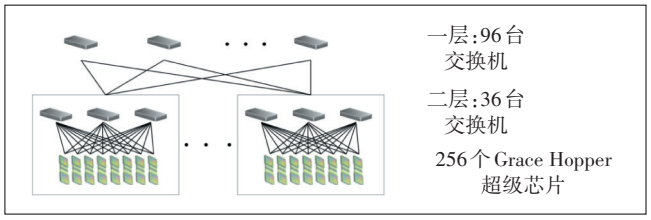

图2　DGX GH200硬件架构

在GH200硬件结构中，第1层、第2层各有96和36台交换机，每台交换机均有32个800G速率的端口。因为距离较近，一层交换机与256个GH200超级芯片通过电缆代替光模块连接，而一二层交换机互联时需要36×32×2=2 304个800G光模块。此外，GH200还配备了24台IB交换机用于IB网络，需768个800G光模块。因此，1台DGX GH200需1 920个800G光模块。

2024年3月，英伟达发布了最新的基于Blackwell架构的GB200，GB200可以灵活扩展至更大的集群规模，如8个机柜配置下的576个GPU集群，这就需要更多高性能光模块实现跨机柜的无阻塞全互联。与GH200相比，GB200不仅提高了光模块与GPU的配比（1∶9），而且光模块的速率也从800 G提升至1.6 T。所以，光模块必须不断提高速率并且增大规模，才能匹配更高计算性能对并行通信量提出的要求。

4.2 高集约化、低功耗

随着光模块规模和速率提升，如何降低功耗成为未来AI智算发展的主要挑战。根据统计，过去10年间全球数据中心的网络交换带宽提升了80倍，与此同时光模块功耗增加了26倍。降低光模块功耗的一条主要路径是推动光模块的封装结构演进，从传统DSP光模块架构到线性驱动可插拔光模块（LPO），再到共封装光学（CPO）[11]。

LPO技术方案取消了传统光模块中高功耗的数字

信号处理（DSP）和时钟数据恢复（CDR）芯片，将相关功能集成到设备侧的交换芯片中。这种设计简化了光模块的结构，实现了高密度集成，并且较传统的光模块功耗降低50%[12]。目前这种方案是演进到CPO的折中选择，最大的挑战是标准化程度不够，互联互通性和支持的模式受到影响。CPO技术通过在同一高速主板上同时封装交换ASIC芯片和硅光引擎[13]，有效地降低了信号衰减、系统功耗和成本，实现了高度集成，并且支持向更高速率演进，是未来光模块架构的发展方向。

CPO架构的基础是硅基光电子技术，下文简称为硅光技术。硅光技术的核心理念是以光补电，它利用硅材料作为光的传输介质，并通过CMOS工艺在硅基上集成光子器件，原本通过电信号传递的数据通过激光束进行承载传输[14]。硅光技术能够将多个光电器件集成到一个微芯片上，大幅提高了集成度，从而减小了光模块的尺寸，功耗也显著降低。

4.3 高稳定性、可管可控

大模型的训练往往需要数天甚至数周的时间，在这期间可能会出现硬件故障、数据损坏或网络中断等问题，并且由于大模型计算训练具有分布式、多并行、强同步等特点，任何一个小训练单元的故障都会影响整个大模型训练的性能甚至导致训练失败[15]。根据实验数据，0.1%的丢包都会造成大模型算力下降50%。光模块作为大模型训练中的关键组件，稳定的工作状态是保证训练顺利完成的关键。实际工程中光模块的失效率在2‰~4‰。以一个万卡集群参数面网络为例，400G和200G光模块数量为几万个，相当于平均不到一周就会出现一个光模块故障。此外，光模块的故障定位和检查处理尚不能实现智能化，主要依靠人工判断和维护，处理的周期长。

智算网络的健壮性对于保障数据传输的可靠性和完整性、提升网络使用体验和服务质量有着重要的意义，所以高稳定性且易于管控和维护的光模块是未来光互联技术发展的重点之一。

5 结束语

本文针对AI智算技术对高速光模块的应用需求开展研究，逐层剖析、建立关联。通过分析AI大模型训练的分布式并行计算需求，建立通信需求模型，并以GPT-3为例定量分析。通过分析大模型训练的响应时间构成，得到减少数据传输时间是对通信网络的核心需求，而提升有效带宽是最有效的实现方式。光模块将向更高速率、更大规模、更低功耗、更加稳定等应用方向不断创新和优化，以满足AI智算日益增长的网络性能需求。

参考文献：

[1] 中国算力大会. 中国算力白皮书（2023年）[R/OL]. [2024-01-18]. https://www.sgpjbg.com/baogao/93378.html.
[2] 百度智能云. 智算中心网络架构白皮书（2023年）[R/OL]. [2024-01-18]. https://zhuanlan.zhihu.com/p/648137403.
[3] SERGEEV A, BALSO M D. Horovod: fast and easy distributed deep learning in TensorFlow [J/OL]. [2024-01-18]. https://arxiv.org/abs/1802.05799.
[4] 岳宗乾. 基于张量的高效并行计算方法研究与实现[D]. 北京：北京邮电大学，2022.
[5] 卢凯，赖志权，李笙维，等. 并行智能训练技术：挑战与发展[J]. 中国科学：信息科学，2023，53（8）：1441-1468.
[6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need[J]. arXiv, 2017.
[7] 朱炫鹏，姚海东，刘隽，等. 大语言模型算法演进综述[J]. 中兴通讯技术，2024，30（2）：9-20.
[8] 黄光红，林广栋，吴尔杰，等. 深度神经网络Softmax函数定点算法设计[J]. 中国集成电路，2022，31（7）：60-64.
[9] 梁嘉诚，余江，王洪波，等. 基于RDMA的高性能单向数据采集技术研究[J]. 计算机工程，2023，49（10）：31-40.
[10] 李洋. 基于Fat-Tree型数据中心网络的流量调度方法研究与实现[D]. 南京：东南大学，2022.
[11] 卞玲艳，曾艳萍，蔡莹，等. 大数据时代光电共封技术的机遇与挑战[J]. 激光与光电子学进展，2024，61（9）：0900006.
[12] 张平化，王会涛，付志明. 数据中心光模块技术及演进[J]. 中兴通讯技术，2024，30（1）：89-98.
[13] IPEC. IPEC成功立项OIO研究项目，探索下一代数据中心交换芯片技术演进[EB/OL]. [2024-01-15]. https://www.ipec-std.org/zh/5219.html.
[14] 朱振华，顾恩婷，李凯. 硅光技术与测试挑战[J]. 中国集成电路，2020，29（12）：80-84.
[15] 冯杨洋，汪庆，谢旻晖，等. 从BERT到ChatGPT：大模型训练中的存储系统挑战与技术发展[J]. 计算机研究与发展，2024，61（4）：809-823.

作者简介：
栾昊立，毕业于曼彻斯特大学，助理工程师，硕士，主要从事传输网规划、研究工作；王晓东，高级工程师，硕士，主要从事网络规划工作；杨锐，高级工程师，硕士，主要从事传输网络规划、光网络新技术研究、智库研究工作；郝建宇，毕业于北京师范大学，工程师，硕士，主要从事政府客户网络通信及信息化业务服务支撑工作；赵铭浩，毕业于北京邮电大学，高级工程师，硕士，主要从事光通信等传输技术研究工作；尹祖新，毕业于哈尔滨工业大学，教授级高级工程师，长期从事传输网规划、研究等工作；王丽琼，高级工程师，硕士，主要从事光网络规划研究工作。