



上海证券
SHANGHAI SECURITIES

增持（维持）

行业： 通信
日期： 2023年06月02日

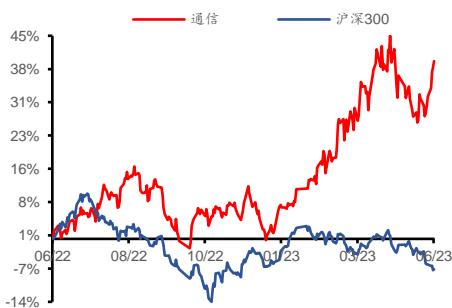
分析师： 刘京昭

Tel:

E-mail:

SAC 编号： S0870523040005

最近一年行业指数与沪深 300 比较



相关报告：

《市场行情疲软，通信行业细分板块出现分化行情》

——2023年05月23日

《财报披露和节日后两天行情疲软，不改长期投资逻辑》

——2023年05月11日

《稳健成长，紧抓数字经济新周期 ——通信行业 2022 及 2023Q1 财报总结》

——2023年05月09日

英伟达发布 DGX GH200 超级计算机，提升通信带宽利好高速光模块

主要观点

今年年初以来，以ChatGPT为代表的生成式人工智能模型持续落地和商业化，引发了业内对人工智能赛道的关注。人工智能技术的不断演进，也对以算力和通信为代表的人工智能基础设施提出了更高的要求。

近日，NVIDIA推出了新型AI超级计算机——DGX GH200。在算力和存储方面，DGX GH200组合了Arm架构的Grace CPU和Hopper架构的GPU，配备了96GB HBM高速显存、576GB显存和144TB共享内存。在通信方面，DGX GH200在GPU间可实现高达900GB/s的数据吞吐速率（带宽），是PCIe Gen 5.0通道传输速率的7倍。相较于此前推出的DGX H100 SuperPOD，DGX GH200更适合生成式人工智能模型训练场景。

我们认为：除算力之外，为了适应生成式人工智能模型的持续迭代，AI数据中心需要更高的带宽和更大的内存。而GH200配置的高内存和大带宽，正是这一发展趋势的代表，这将有利于800G光模块在未来的导入与放量，利好在800G验证和量产方面进度领先的光模块行业头部公司。

投资建议

建议关注：

中际旭创：中高端数通市场龙头，2021年全球排名第一的光模块供应商。公司管理团队激励机制有望得到改善。根据iFinD机构一致预期，截至2023年5月29日，公司2023/2024年的预测PE分别为51/39倍，位于近五年的81%分位。

新易盛：光模块领域龙头，成本管控优秀，具备切入增量云计算/AI客户的能力。根据iFinD机构一致预期，截至2023年5月29日，公司2023/2024年的预测PE分别为39/28倍，位于近五年的46%分位。

天孚通信：光器件整体解决方案提供商。根据iFinD机构一致预期，截至2023年5月29日，公司2023/2024年的预测PE分别为59/46倍，位于近五年的100%分位。

源杰科技：技术成熟的国产激光器厂商。根据iFinD机构一致预期，截至2023年5月29日，公司2023/2024年的预测PE分别为104/76倍，位于近五年的90%分位。

风险提示

海外头部云厂商需求不及预期；人工智能技术落地和商业化不及预期；产业政策转变；宏观经济不及预期等。

目录

1 人工智能基础设施更进一步，英伟达新型超级计算机发布	3
1.1 DGX GH200 在共享内存和 GPU 间通信方面有提升	3
1.2 GPU 间通信带宽增加，提振高速光模块增长空间	5
2 风险提示	6

图

图 1: DGX GH200 组合了 Grace CPU 和 Hopper 架构 GPU	3
图 2: Grace Hopper 架构相对 x86+Hopper 在模型训练和 高性能计算上有比较明显的计算能力优势	4
图 3: DGX GH200 内存大幅增加 (单位: GB)	4
图 4: DGX GH200 连接 NVLink 的拓扑结构	4
图 5: DGX GH200 大内存 AI 工作负载性能大幅优于 DGX H100	4
图 6: GPU-GPU 通过 NVLink4 连接	5

表

表 1: 人工智能领域相关公司对比 (截至 5 月 29 日, 单位: 亿元)	6
--	---

1 人工智能基础设施更进一步，英伟达新型超级计算机发布

1.1 DGX GH200 在共享内存和 GPU 间通信方面有提升

今年年初以来，以 ChatGPT 为代表的生成式人工智能模型持续落地和商业化，引发了业内对人工智能赛道的关注。人工智能技术的不断演进，也对以算力和通信为代表的人工智能基础设施提出了更高的要求。

图 1: DGX GH200 组合了 Grace CPU 和 Hopper 架构 GPU

DGX GH200 Technical Specifications	
CPU and GPU	256x NVIDIA Grace Hopper Superchips
CPU Cores	18,432 Arm® Neoverse V2 Cores with SVE2 4X 128b
GPU Memory	144TB
Performance	1 exaFLOPS
Networking	256x OSFP single-port NVIDIA ConnectX®-7 VPI with 400Gb/s InfiniBand 256x dual-port NVIDIA BlueField®-3 VPI with 200Gb/s InfiniBand and Ethernet 24x NVIDIA Quantum-2 QM9700 InfiniBand Switches 20x NVIDIA Spectrum™ SN2201 Ethernet Switches 22x NVIDIA Spectrum SN3700 Ethernet Switches
NVIDIA NVLink Switch System	96x L1 NVIDIA NVLink Switches 36x L2 NVIDIA NVLink Switches
Management Network	Host baseboard management controller (BMC) with RJ45
Software	NVIDIA AI Enterprise (optimized AI software) NVIDIA Base Command (orchestration, scheduling, and cluster management) DGX OS / Ubuntu / Red Hat Enterprise Linux / Rocky (operating system)
Support	Comes with three-year business-standard hardware and software support

资料来源: NVIDIA DGX GH200 技术文档, 上海证券研究所

5月29日, NVIDIA 宣布推出 AI 超级计算机 DGX GH200。相比于 2021 年推出的 DGX H100 SuperPOD 架构, DGX GH200 超级计算机主要的改进有:

(1) 使用了全新的 Grace Hopper 超级芯片。Grace Hopper 超级芯片整合了 Arm 架构的 Grace CPU 和 Hopper 架构的 GPU, 而 DGX H100 则使用单独的 H100 GPU。Grace Hopper 超级芯片相比于传统的 X86+Hopper 方式, 在并行计算上优势明显, 因此 DGX GH200 在模型训练场景下会有更高的计算效率;

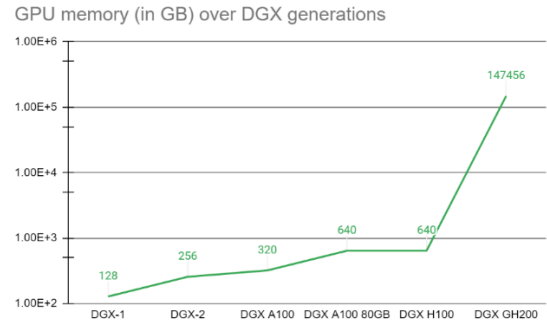
(2) DGX GH200 拥有 144TB 共享内存, 单块 GPU 最高配置 96GB 高速显存, 相较 DGX H100 增加 20%, 因此 DGX GH200 在进行模型训练时, 可以高效地使用更大的训练数据集;

图 2: Grace Hopper 架构相对 x86+Hopper 在模型训练和高效能计算上有比较明显的计算能力优势



资料来源: NVIDIA Grace Hopper 架构白皮书, 上海证券研究所

图 3: DGX GH200 内存大幅增加 (单位: GB)

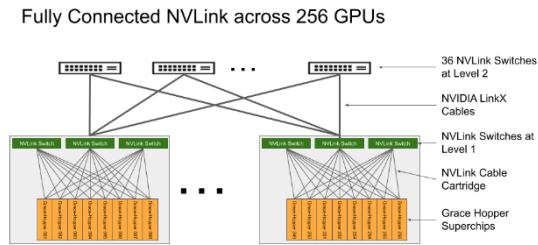


资料来源: NVIDIA 开发者社区, 上海证券研究所

(3) DGX GH200 的 GPU 与 GPU 间通信使用 NVLink4, 可以在 GPU 与 GPU 间实现 900GB/s 的数据吞吐速率 (带宽), 是传统 PCIe Gen 5.0 通道传输速率的 7 倍, 同样有助于提高模型训练效率。

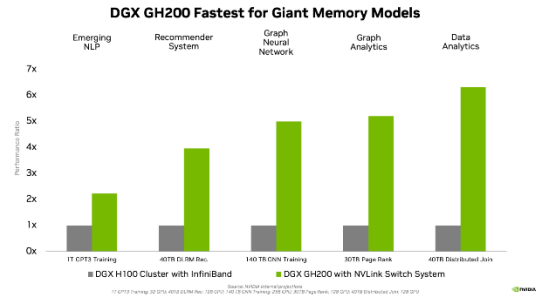
综上所述, 相比于此前的 DGX H100 SuperPOD, DGX GH200 将更适合受到较大关注的生成式人工智能模型训练。

图 4: DGX GH200 连接 NVLink 的拓扑结构



资料来源: NVIDIA 开发者社区, 上海证券研究所

图 5: DGX GH200 大内存 AI 工作负载性能大幅优于 DGX H100



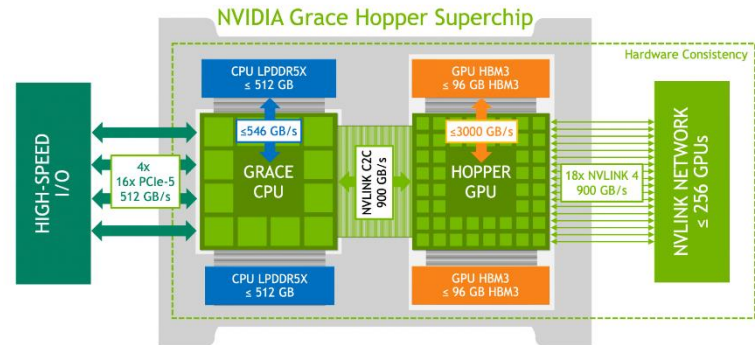
资料来源: NVIDIA 开发者社区, 上海证券研究所

DGX GH200 在 Grace CPU 与 GPU 间使用 NVLink C2C 技术进行通信, 在 GPU 间通过 NVLink 交换机进行通信。所使用的 NVLink 交换机结合了第四代 NVLink 与第三代 NVSwitch 技术。如图 4 所示, DGX GH200 第一级配置 96 台 NVLink 交换机, 第二级配置 36 台 NVLink 交换机, 每 8 个 Grace Hopper 超级芯片为一组与 3 台第一级的 NVLink 交换机互相连接。结合 800G 光模块的封装方式和 NVSwitch 的端口数量, 假设 Grace Hopper 超级芯片与第一级的 NVLink 交换机通过线缆连接, 而第一级的 NVLink

请务必阅读尾页重要声明

交换机与第二级的 NVLink 交换机通过光通信连接，则 DGX GH200 中一块 GPU 对应 9 个 800G 光模块；假设 Grace Hopper 超级芯片与第一级的 NVLink 交换机、第一级的 NVLink 交换机与第二级的 NVLink 交换机均通过光通信连接，DGX GH200 中一块 GPU 对应约 18 个 800G 光模块。因此，DGX GH200 中一块 GPU 对应的 800G 光模块约为 9-18 个。

图 6: GPU-GPU 通过 NVLink4 连接



资料来源: NVIDIA Grace Hopper 架构白皮书, 上海证券研究所

考虑到 NVIDIA 推出的 CUDA 架构在高性能计算领域的重要地位，DGX GH200 超级计算机的设计思路和具体实现，也将对后续面向模型训练场景的 AI 服务器的设计产生较大的影响。

1.2 GPU 间通信带宽增加，提振高速光模块增长空间

光通信是 AI 数据中心协同运转的重要通信手段，光模块则是光通信必不可少的组件。AI 数据中心的网络通信建设，将进一步拓展以 400G 和 800G 光模块为代表的中高速光模块的应用前景，对目前处于 400G 光模块放量、800G 光模块加速导入的周期起点的光模块行业具有重要意义。

我们认为：除算力之外，为了适应生成式人工智能模型的持续迭代，突破生成式人工智能模型训练时间较长的瓶颈，AI 数据中心需要更高的带宽和更大的内存。而 DGX GH200 正是这一发展趋势的代表，这将有利于 800G 光模块在未来的导入与放量，利好在 800G 验证和量产方面进度领先的光模块行业头部公司。

表 1: 人工智能领域相关公司对比 (截至 5 月 29 日, 单位: 亿元)

细分板块	股票简称	22 营业收入	22 归母净利润	23E 营业收入	23E 归母净利润	23E 估值	24E 营业收入	24E 归母净利润	24E 估值	近五年 PE 分位数 (%)
算力	寒武纪	7	-13	11	-8	--	16	-5	--	--
	紫光国微	71	26	92	34	23	119	45	17	0
	澜起科技	37	13	50	15	54	73	21	36	30
	北京君正	54	8	59	9	49	70	12	36	27
	芯原股份	27	1	34	2	295	43	2	189	--
	国科微	36	2	53	4	51	69	6	37	43
	瑞芯微	20	3	25	4	77	32	6	57	97
服务器	中科曙光	130	15	152	20	36	180	26	28	36
	紫光股份	741	22	852	27	32	984	33	26	74
	浪潮信息	695	21	823	26	22	959	32	18	44
光模块	中际旭创	96	12	112	15	51	145	20	39	81
	新易盛	33	9	38	9	39	50	13	28	46
	博创科技	15	2	20	3	31	25	3	25	40
	剑桥科技	38	2	--	--	--	--	--	--	58
	华工科技	120	9	151	12	25	188	16	20	26
	光迅科技	69	6	77	7	29	87	8	25	44
光芯片	天孚通信	12	4	16	5	59	22	6	46	100
	源杰科技	3	1	4	2	104	5	2	76	90
	仕佳光子	9	1	11	1	61	13	1	42	47

资料来源: iFinD, 上海证券研究所

*盈利预测来自 iFinD 机构一致预期; 仅列举各板块部分标的

2 风险提示

海外头部云厂商需求不及预期; 人工智能技术落地和商业化不及预期; 产业政策转变; 宏观经济不及预期等。

分析师声明

作者具有中国证券业协会授予的证券投资咨询资格或相当的专业胜任能力，以勤勉尽责的职业态度，独立、客观地出具本报告，并保证报告采用的信息均来自合规渠道，力求清晰、准确地反映作者的研究观点，结论不受任何第三方的授意或影响。此外，作者薪酬的任何部分不与本报告中的具体推荐意见或观点直接或间接相关。

公司业务资格说明

本公司具备证券投资咨询业务资格。

投资评级体系与评级定义

股票投资评级：	分析师给出下列评级中的其中一项代表其根据公司基本面及（或）估值预期以报告日起 6 个月内公司股价相对于同期市场基准指数表现的看法。
买入	股价表现将强于基准指数 20%以上
增持	股价表现将强于基准指数 5-20%
中性	股价表现将介于基准指数±5%之间
减持	股价表现将弱于基准指数 5%以上
无评级	由于我们无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使我们无法给出明确的投资评级
行业投资评级：	分析师给出下列评级中的其中一项代表其根据行业历史基本面及（或）估值对所研究行业以报告日起 12 个月内的基本面和行业指数相对于同期市场基准指数表现的看法。
增持	行业基本面看好，相对表现优于同期基准指数
中性	行业基本面稳定，相对表现与同期基准指数持平
减持	行业基本面看淡，相对表现弱于同期基准指数
相关证券市场基准指数说明：A 股市场以沪深 300 指数为基准；港股市场以恒生指数为基准；美股市场以标普 500 或纳斯达克综合指数为基准。	

投资评级说明：

不同证券研究机构采用不同的评级术语及评级标准，投资者应区分不同机构在相同评级名称下的定义差异。本评级体系采用的是相对评级体系。投资者买卖证券的决定取决于个人的实际情况。投资者应阅读整篇报告，以获取比较完整的观点与信息，投资者不应以分析师的投资评级取代个人的分析与判断。

免责声明

。本公司不会因接收人收到本报告而视其为客户。

本报告版权归本公司所有，本公司对本报告保留一切权利。未经书面授权，任何机构和个人均不得对本报告进行任何形式的发布、复制、引用或转载。如经过本公司同意引用、刊发的，须注明出处为上海证券有限责任公司研究所，且不得对本报告进行有悖原意的引用、删节和修改。

在法律许可的情况下，本公司或其关联机构可能会持有报告中涉及的公司所发行的证券或期权并进行交易，也可能为这些公司提供或争取提供多种金融服务。

本报告的信息来源于已公开的资料，本公司对该等信息的准确性、完整性或可靠性不作任何保证。本报告所载的资料、意见和推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值或投资收入可升可跌。过往表现不应作为日后的表现依据。在不同时期，本公司可发出与本报告所载资料、意见或推测不一致的报告。本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本报告中的内容和意见仅供参考，并不构成客户私人咨询建议。在任何情况下，本公司、本公司员工或关联机构不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告中的任何内容所引致的任何损失负责，投资者据此做出的任何投资决策与本公司、本公司员工或关联机构无关。

市场有风险，投资需谨慎。投资者不应将本报告作为投资决策的唯一参考因素，也不应当认为本报告可以取代自己的判断。