1. **TCCL: Discovering Better Communication Paths for PCIe GPU Clusters**

这篇论文介绍了TCCL，这是一个专门为通过PCIe通信的GPU集群设计的集体通信库。TCCL包含三个组件：一个用于多传输性能测量的分析器，一个用于发现最佳通信路径的路径发现器，以及一个修改过的NCCL运行时来利用这些路径。论文的评估结果显示，TCCL在三个不同的PCIe依赖GPU集群上的表现超过了现有的最先进的通信库NCCL和MSCCL（高达2.07倍）。

**主要内容总结：**

TCCL（Topology-aware Collective Communication Library）是为通过PCIe通信的GPU集群设计的集体通信库。它包含三个主要组件：性能分析器、路径发现器和修改版的NCCL运行时。TCCL专注于基于环的集体通信算法，这些算法适用于深度学习中的流行通信操作，如AllReduce和AllGather。TCCL的目标是在PCIe依赖的GPU集群中最大化集体通信性能，通过改变通信路径来避免拥塞。

TCCL的性能分析器可以同时执行多个传输并测量所有传输完成所需的时间。路径发现器模块负责搜索通信路径，它通过性能分析器检查每条通信路径的性能，以避免拥塞。由于可能的通信路径数量随着集群中GPU数量的增加而呈指数级增长，TCCL使用动态规划和Dijkstra算法来减少搜索空间，并考虑NUMA对称性。TCCL的运行时是NCCL的一个修改版，它允许使用路径发现器找到的路径替换NCCL中的路径，同时保持与使用NCCL的现有软件的透明接口兼容性。

在三个不同的PCIe依赖GPU集群上的评估结果显示，TCCL的性能超过了现有的最先进的通信库NCCL和MSCCL（高达2.07倍）。此外，TCCL还与具有各种并行类型组合的深度学习训练工作负载一起进行了评估。

**创新点：**

1. **针对PCIe依赖系统的优化：** TCCL是首个为通过PCIe通信的GPU集群设计的集体通信库，这些系统在以往的研究中受到的关注较少。

2. **自动发现最佳通信路径：** TCCL通过路径发现器自动寻找最佳的通信路径，而不是依赖于理论带宽，这有助于避免由于NUMA架构和CPU内存反弹缓冲区放置不当引起的拥塞。

3. **基于环算法的集体通信优化：** TCCL专注于优化基于环的集体通信算法，这些算法在深度学习中非常常见，如AllReduce和AllGather。

4. **与现有软件的兼容性：**由于TCCL是基于NCCL实现的，因此它可以轻松地与现有的使用NCCL的软件（如PyTorch）一起使用，无需进行代码更改。

5. **减少搜索空间的策略：** TCCL使用动态规划和Dijkstra算法来减少可能的通信路径的搜索空间，并利用NUMA对称性进一步优化搜索过程。

6. **性能提升：** 在三个不同的PCIe依赖GPU集群上的评估结果显示，TCCL的性能显著超过了现有的最先进的通信库NCCL和MSCCL（高达2.07倍）。

1. **The Landscape of GPU-Centric Communication**

这篇论文提供了GPU中心通信的全景，重点关注了供应商机制和用户级库支持。它旨在阐明该领域的复杂性和多样化选项，定义术语，并在节点内和跨节点分类现有方法。论文讨论了供应商提供的多GPU执行中的通信和内存管理机制，并回顾了主要的通信库，它们的优势、挑战和性能洞察。然后，它探索了关键的研究范式、未来的展望和开放的研究问题。

**主要内容总结：**

"The Landscape of GPU-Centric Communication" 这篇论文提供了GPU中心通信的全景，重点关注了供应商机制和用户级库支持。它旨在阐明该领域的复杂性和多样化选项，定义术语，并在节点内和跨节点分类现有方法。论文讨论了供应商提供的多GPU执行中的通信和内存管理机制，并回顾了主要的通信库，包括CUDA-aware MPI, NCCL/RCCL, NVSHMEM, 和 ROC\_SHMEM，它们的优势、挑战和性能洞察。然后，它探索了关键的研究范式，如无CPU网络、通信调试工具，并讨论了未来的发展方向和未解决的研究问题。通过广泛描述GPU中心通信技术在软件和硬件堆栈中的实现，论文为研究人员、程序员、工程师和库设计者提供了如何充分利用多GPU系统的见解。

**创新点：**

1. **GPU中心通信的全面概述：** 论文提供了一个全面的视角来审视GPU中心通信，这是随着GPU在HPC和机器学习应用中变得越来越重要而出现的一个新兴领域。
2. **供应商机制和用户级库的支持：** 论文详细讨论了供应商提供的机制，如NVIDIA的NCCL和AMD的RCCL，以及用户级库如何支持多GPU通信。
3. **跨节点和节点内的通信方法分类：** 论文不仅关注了节点内的GPU通信，还涵盖了跨节点的通信方法，为读者提供了一个全面的了解。
4. **关键研究范式的探索：** 论文探讨了如无CPU网络等关键研究范式，这些范式可能定义了未来GPU中心通信的发展方向。
5. **未来展望和开放性问题：** 论文不仅回顾了当前的技术状态，还提出了未来可能的发展方向和当前领域中尚未解决的问题，为未来的研究提供了思路。
6. **调试和分析工具的讨论：** 论文强调了调试和分析工具在多GPU编程中的重要性，并讨论了现有工具的局限性，指出了这一领域未来可能的改进方向。
7. **GPU Cluster Scheduling for Network-Sensitive Deep Learning**

这篇论文提出了一种新的GPU集群调度器，用于分布式深度学习（DDL）工作负载，它可以根据DDL作业对预期通信网络延迟的敏感性，实现基于邻近性的GPU资源合并。调度器包括三个主要组件：(i) 一个经典的延迟调度算法，以促进作业放置和合并；(ii) 一个网络敏感的作业抢占策略；(iii) 一个“自动调谐器”机制，用于优化延迟计时器，以有效进行延迟调度。

**主要内容总结：**

"GPU Cluster Scheduling for Network-Sensitive Deep Learning" 这篇论文提出了一个名为Dally的新型GPU集群调度器，专门为分布式深度学习（DDL）工作负载设计，以减少与网络延迟相关的通信开销。Dally调度器包括三个主要组件：

1. **延迟调度算法：**用于促进作业的放置和整合，以优化GPU资源的整合。
2. **网络感知的作业抢占策略：**根据DDL作业对预期通信网络延迟的敏感性来优先处理对网络条件更敏感的作业。
3. **自动调谐器机制：**优化延迟计时器，以实现有效的延迟调度。

此外，为了支持大规模实验的成本效益方法，研究者开发了一个数据驱动的DDL集群模拟平台ArtISt-sim。通过模拟平台，Dally与现实工作负载跟踪的几个最新替代方案进行了比较，展示了其设计的优势。Dally调度器能够在训练所有作业的端到端Makespan上提供高达69%的改进，同时在网络拥塞条件下将平均作业完成时间减少高达83%，并将通信开销最小化高达98%。

**创新点：**

1. **网络感知的作业整合：**Dally调度器根据DDL作业对网络延迟的敏感性来整合GPU资源，这是对现有调度器通常不考虑基于邻近性的作业整合的改进。
2. **自动调谐器：**Dally引入了一个自动调谐器，能够根据系统的网络使用情况动态调整延迟计时器，这是对传统手工调整延迟计时器的创新。
3. **数据驱动的模拟平台：**研究者开发了ArtISt-sim模拟平台，它能够模拟现代网络硬件设备，如NVIDIA Quantum和Spectrum交换机，这对于研究现代DDL工作负载至关重要。
4. **性能提升：**通过广泛的模拟评估，Dally显示出在减少作业完成时间和通信开销方面的显著改进，特别是在网络拥塞条件下。
5. **成本效益：**Dally的设计和ArtISt-sim模拟平台的开发，使得研究者能够在不需要昂贵的大规模实验设置的情况下，开发和评估新的DDL调度技术。
6. **Exploring GPU-to-GPU Communication: Insights into …**

这篇论文全面描述了三个超级计算机的GPU到GPU通信性能，涵盖了目前可用的HPC内部和节点间互连技术的很大一部分：Alps、Leonardo和LUMI。分析指出了从路由到通信库的几个改进机会。首先，默认的软件配置没有充分利用所有三个系统的潜力，需要付出相当大的调整努力才能在单个节点和更大规模上实现良好的性能。其次，每个通信库都有自己的一系列优化，因此根据场景的不同，它们可能更适合或不太适合。通常，我们发现\*CCL在集体操作上提供了更高的性能，而GPU-Aware MPI在点对点传输上表现更好。

**主要内容总结：**

1. 对Alps、Leonardo和LUMI三个超级计算机的GPU到GPU通信性能进行了全面描述。

评估了节点内和节点间的互连性能，使用了不同的基准测试。

1. 分析了不同软件解决方案的性能，包括显式设备到设备复制、NCCL/RCCL和GPU-Aware MPI。
2. 评估了网络噪声对GPU到GPU数据移动的影响，展示了它如何严重影响工作负载的可扩展性。
3. 提供了八个关键观察结果，为系统架构师、研究人员、实践者和软件开发人员提供了优化大规模多GPU系统中数据移动的宝贵见解。

**创新点：**

1. **全面的GPU到GPU通信性能评估：**论文提供了对三个不同超级计算机的GPU到GPU通信性能的全面描述，这是对现有研究的重要补充，因为以往的研究往往只关注单个系统或特定的通信技术。
2. **多系统和大规模评估：**研究者们不仅评估了单个节点内的GPU通信，还扩展到了跨节点的通信，并且在多达4096个GPU上进行了评估，这为理解大规模多GPU系统的通信性能提供了宝贵的数据。
3. **深入分析和优化建议：**论文不仅描述了性能表现，还深入分析了性能的局限性和优化机会，为实际应用提供了指导。
4. **关键观察结果：**论文提出了八个关键观察结果，这些观察结果涵盖了从网络到软件优化的多个方面，为系统设计者和软件开发者提供了优化数据移动的实用见解。
5. **网络噪声的影响分析：**论文评估了网络噪声对GPU到GPU数据移动的影响，这是对现有研究的重要补充，因为网络噪声是影响大规模多GPU系统性能的关键因素之一。
6. **Evaluating Modern GPU Interconnect: PCIe, NVLink, …**

这篇论文通过彻底描述现代GPU互连，包括PCIe、NVLink版本1、NVLink版本2、NV-SLI、NVSwitch和GPUDirect，填补了多GPU并行编程、执行和性能模型的知识空白。它测量了它们的原始启动延迟、可持续的单/双向带宽、网络拓扑、通信效率、路由和NUMA效应，并基于这两种通信模式：点对点（P2P）和集体（CL）。

**主要内容总结：**

"Evaluating Modern GPU Interconnect: PCIe, NVLink, NV-SLI, NVSwitch and GPUDirect" 这篇论文对现代GPU互联技术进行了深入的性能评估。研究团队测试了六种不同类型的GPU互联技术：PCIe、NVLink-V1、NVLink-V2、NVLink-SLI、NVSwitch以及支持GPUDirect的InfiniBand。这些测试涵盖了从单个节点到多个节点的高性能计算（HPC）平台，包括NVIDIA的P100-DGX-1、V100-DGX-1、DGX-2、RTX2080-SLI系统，以及橡树岭国家实验室的SummitDev和Summit超级计算机。研究的目的是了解这些互联技术在多GPU应用性能上的实际影响，特别是在点对点（P2P）和集体（CL）通信模式下的表现。

论文中，研究者们测量了这些互联技术的原始启动延迟、可持续的单/双向带宽、网络拓扑、通信效率、路由以及NUMA效应。研究结果揭示了四种新的GPU通信网络NUMA效应，其中三种由NVLink的拓扑结构、连接性和路由触发，一种由PCIe芯片组设计问题引起。这些观察结果表明，对于在多GPU节点上运行的应用程序，选择正确的GPU组合可以显著影响GPU通信效率和应用程序的整体性能。

**创新点：**

1. 全面的GPU互联技术评估：论文提供了对多种现代GPU互联技术的全面评估，包括PCIe、NVLink的两个版本、NV-SLI、NVSwitch以及GPUDirect，这种全面的评估在以往的研究中较少见。
2. 新的NUMA效应观察：研究者们发现了四种新的NUMA效应，这些效应对于理解和优化多GPU系统中的数据移动具有重要意义。
3. 跨多个平台的测试：研究不仅在单个节点上进行，还扩展到了跨节点的测试，涵盖了从高端服务器到HPC平台的多种配置。
4. 实际应用性能影响的分析：论文不仅关注了互联技术的性能指标，还分析了这些技术对于实际多GPU应用程序性能的影响，提供了实际应用中的优化指导。
5. 多GPU性能模型的构建：研究结果有助于构建实用的多GPU性能模型，这对于共享环境中的GPU任务分配、调度和迁移至关重要。
6. 通信模式的深入分析：论文深入分析了P2P和CL两种通信模式，为理解不同通信模式下的性能表现提供了数据支持。
7. **CGX: Adaptive System Support for Communication-Efficient Deep Learning**

这篇论文提出了一个名为CGX的框架，它为ML应用提供了高效的软件支持，用于压缩通信，无论是多GPU单节点训练还是更大规模的多节点训练。CGX基于两项技术进步：在系统级别，它依赖于为ML框架重新开发的通信栈，提供了灵活、高效的压缩通信支持。在应用级别，它提供了与流行框架的无缝、无需参数的集成，以便最终用户无需修改训练配方，也无需大量修改训练代码。

**主要内容总结：**

"CGX: Adaptive System Support for Communication-Efficient Deep Learning" 这篇论文提出了一个名为CGX的框架，旨在通过算法和系统设计替代昂贵的硬件过配置方法，以提供机器学习（ML）应用中的压缩通信的高效软件支持。CGX框架适用于多GPU单节点训练以及更大规模的多节点训练。它基于两项技术进步：

1. **系统级别：**CGX依赖于为ML框架重新开发的通信栈，提供灵活、高效的压缩通信支持。
2. **应用级别：**CGX提供与流行框架的无缝、无需参数的集成，这意味着最终用户无需修改训练配方或大量训练代码。此外，CGX还包括一种层级自适应压缩技术，动态平衡压缩增益与准确性保持。

CGX与流行的ML框架集成，能够在基于普通硬件的多GPU节点上提供高达3倍的速度提升，在多节点设置中提供数量级改进，同时对准确性的影响可以忽略不计。

**创新点：**

1. **替代硬件过配置：**CGX展示了如何通过软件和系统设计替代昂贵的硬件过配置方法，这对于降低成本和提高资源效率具有重要意义。
2. **无缝集成与参数自由：**CGX提供了与现有数据并行训练流程的无缝集成，无需额外的超参数调整，这大大简化了开发者的使用难度。
3. **层级自适应压缩技术：**CGX引入了一种新的层级自适应压缩技术，可以动态平衡压缩增益与模型准确性，这是对现有压缩技术的创新。
4. **跨节点和单节点的性能提升：**CGX在单节点和多节点设置中都显示出显著的性能提升，这表明其技术具有广泛的适用性。
5. **对准确性的最小影响：**尽管CGX提供了显著的速度提升，但它对模型的准确性影响很小，这对于保证ML模型的质量至关重要。