

Optical Interconnect Design for Distributed GPU and AI Accelerator Clusters

Mandy Begum

Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV,
USA.

mandy1983@unr.edu

Kacper Woloszyk

Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL,
USA.

kacper740@uab.edu

Abstract

The rapid expansion of distributed GPU and artificial intelligence accelerator clusters has turned interconnect design from a supporting infrastructure concern into a primary determinant of system capability. As model training and inference workloads become increasingly communication-intensive, electrical links face growing pressure in bandwidth density, energy per transmitted bit, reach, and synchronization precision. Optical interconnects offer a promising path toward higher radix, lower latency, and more flexible resource pooling, but they also introduce new architectural, operational, and governance challenges. This paper examines optical interconnect design for distributed accelerator clusters from a system-level perspective. It analyzes workload and traffic characteristics, including collective communication, parameter synchronization, bursty training phases, and latency-sensitive inference. It then evaluates architectural options such as wavelength division multiplexing, optical circuit switching, hybrid electrical and optical fabrics, co-packaged optics, and hollow-core fiber links, with attention to topology, switching granularity, control-plane complexity, and failure containment. The discussion extends to deployment models, composable infrastructure, multi-tenant fairness, robustness, sustainability, and policy implications. The central argument is that optical interconnect design cannot be reduced to a photonic component problem. It is a socio-technical design problem in which performance, economics, energy, reliability, and institutional governance interact. Successful clusters will require co-design across accelerators, network interfaces, optical fabrics, orchestration software, procurement practices, and energy policy. The paper concludes with forward-looking directions for research and deployment.

Keywords

optical interconnects; distributed AI; GPU clusters; co-packaged optics; wavelength switching; data center networks; sustainability; governance.

1. Introduction

Distributed GPU and AI accelerator clusters have become the computational backbone of modern machine learning. The scale of these systems has grown from a few interconnected servers to warehouse-scale environments containing thousands of accelerators, high-bandwidth memory modules, and specialized communication fabrics. This growth has been driven by the end of easy gains from conventional semiconductor scaling, which has pushed

architects toward domain-specific accelerators and massively parallel systems [1]. In such systems, the interconnect is no longer a passive conduit. It shapes what workloads can run, how quickly they converge, how efficiently resources are used, and how failures propagate. The familiar problem of tail latency in large-scale services [2] reappears in AI clusters as straggler effects, synchronization delays, and collective communication stalls. Accelerator performance alone does not determine cluster performance; the fabric that binds accelerators together often does.

The rise of tensor processing units and similar accelerators has demonstrated that specialized hardware can deliver substantial gains for inference and training when data movement is managed carefully [3]. Yet distributed deep learning introduces concurrency patterns that are far more complex than single-device execution [4]. Pipeline parallelism, data parallelism, model parallelism, and hybrid schemes all impose different communication signatures on the network [5]. Some require frequent all-reduce operations, others require point-to-point activation transfers, and still others require parameter server interactions or sharded updates. These patterns create bursty, heterogeneous, and often synchronized traffic. Optical interconnects are attractive because they can provide high bandwidth over longer reaches with lower signal degradation than electrical links, and because wavelength division multiplexing allows many logical channels to share a single physical medium [7]. However, optical fabrics also raise difficult questions about switching latency, circuit setup, buffering, control overhead, and failure recovery [8].

This paper argues that optical interconnect design for distributed GPU and AI accelerator clusters must be treated as a system-level and socio-technical problem. The relevant design space includes photonic devices, transceiver packaging, fiber media, switching architectures, network topologies, collective communication libraries, orchestration systems, energy procurement, and multi-tenant governance. Decisions in one layer constrain or enable options in another. For example, co-packaged optics can reduce electrical SerDes power and improve bandwidth density, but it complicates serviceability and thermal management. Optical circuit switching can reduce energy per bit for large flows, but it may suffer under fine-grained, latency-sensitive traffic unless hybrid control planes are carefully designed. Hollow-core fiber can reduce propagation delay, but its deployment depends on cost, manufacturing maturity, and standards. A narrow focus on any single photonic metric, such as bandwidth per fiber or insertion loss, is therefore insufficient. The central research and engineering challenge is to align optical interconnect architecture with the temporal and spatial structure of AI workloads while preserving robustness, fairness, and sustainability.

2. Workload and Traffic Characteristics of Distributed AI Clusters

Understanding optical interconnect requirements begins with the workload. Distributed training typically alternates between computation-intensive phases and communication-intensive phases. During forward and backward passes, accelerators exchange activations, gradients, and model updates. During synchronization, they may perform all-reduce, all-gather, reduce-scatter, or broadcast operations. The duration and frequency of these operations depend on batch size, model architecture, parallelism strategy, and optimizer. Data-parallel training generates large collective communication bursts at the end of each iteration. Pipeline parallelism creates sequential dependencies and point-to-point transfers between stages. Tensor parallelism fragments computations across devices and often requires frequent small transfers. Mixture-of-experts models introduce routing decisions that can create uneven traffic matrices. Inference workloads, by contrast, may be less collective but

more latency-sensitive, especially for interactive services and real-time decision systems. The interconnect must therefore support both throughput-oriented and latency-oriented behavior, sometimes simultaneously.

The temporal structure of this traffic is critical. Many AI workloads are not continuously bandwidth-bound. They alternate between periods of intense local computation and periods of intense communication. If the fabric is provisioned for peak communication demand, it may be underutilized during computation phases. If it is provisioned for average demand, training may stall during synchronization bursts. Optical circuit switching and wavelength switching can help by reallocating bandwidth dynamically, but they require control-plane mechanisms that anticipate or react to traffic shifts quickly. Recent work on low-latency wavelength-switched clock-synchronized intra-data center interconnects enabled by hollow core fiber illustrates how physical-layer choices can reduce latency and improve synchronization precision for tightly coupled workloads [6]. Such advances are relevant because collective communication is sensitive not only to raw bandwidth but also to jitter, skew, and tail latency. A fabric that delivers high average throughput but unpredictable delays can be worse for distributed training than a lower-bandwidth but more deterministic fabric.

Traffic locality and fan-in also matter. Data center measurements have long shown that traffic patterns are often far from uniform, with a mixture of small flows, large flows, and incast-like behavior [10]. AI clusters intensify these patterns. Gradient synchronization can create many-to-one and all-to-all exchanges. Parameter servers can become hotspots. Expert routing can concentrate traffic on a subset of links. Optical interconnects can mitigate hotspots through reconfigurable wavelengths and spatial paths, but only if the control plane can identify and respond to congestion. This creates a tension between optical reconfiguration speed and the fine-grained burstiness of accelerator communication. Slow optical switching may be well suited to large elephant flows but poorly suited to short synchronization messages. Hybrid architectures that combine fast electrical packet switching with slower optical circuit switching are therefore attractive, because they allow latency-sensitive traffic to use electrical paths while bandwidth-heavy flows use optical paths.

Workload diversity also implies that no single optical interconnect design will be optimal for all clusters. A research cluster training large language models may prioritize high-radix all-reduce bandwidth and deterministic latency. An inference cluster serving many small models may prioritize low-latency request routing, isolation, and rapid reconfiguration. A federated or geo-distributed cluster may prioritize reach, security, and policy compliance rather than extreme bandwidth density. Even within a single data center, different tenants may have incompatible communication patterns. The interconnect design must therefore support segmentation, prioritization, and observability. Without these capabilities, optical reconfiguration can amplify unfairness by granting disproportionate bandwidth to tenants whose traffic patterns trigger optical path setup most effectively. The workload analysis thus leads directly to questions of fairness, governance, and multi-tenant resource allocation.

3. Architectural Foundations of Optical Interconnect Design

At the physical layer, optical interconnect design involves choices among transceivers, modulators, detectors, waveguides, fibers, and packaging strategies. Silicon photonics has emerged as a promising platform because it allows optical components to be fabricated using semiconductor manufacturing techniques, potentially reducing cost and improving integration [13]. Device requirements for optical interconnects to silicon chips include low energy per bit, high temperature stability, dense integration, and compatibility with existing packaging [14].

In practice, these requirements are difficult to satisfy simultaneously. Lasers may be external or integrated, modulation may be direct or external, and multiplexing may be coarse or dense. Each choice affects link budget, power, reliability, and thermal behavior. For distributed accelerator clusters, the relevant unit of design is not a single link but a fabric composed of many links, switches, and control elements. The fabric must scale in radix, maintain signal integrity, and provide predictable latency under load.

Wavelength division multiplexing is a foundational technique because it allows multiple optical channels to share a single fiber. This increases bandwidth density and reduces the number of physical fibers required. However, it also introduces complexity in wavelength management, crosstalk, amplification, and filtering. Optical circuit switching can establish dedicated light paths between endpoints, which is attractive for large transfers because it avoids per-packet switching overhead and can reduce energy per bit. Hybrid electrical and optical switch architectures such as Helios showed that optical circuits can complement electrical packet switching in modular data centers [9]. More recent work on high-radix switch design explores how optics can enable switches with many ports while managing buffering and contention [16]. The central trade-off is between reconfiguration flexibility and switching speed. Electrical packet switches offer fine-grained multiplexing but consume significant power at high radix. Optical circuit switches offer low power per bit and high capacity but require setup time and may waste capacity if traffic is bursty.

A second architectural axis is the degree of integration. Pluggable optics dominate current deployments because they are serviceable and standards-based. On-board optics and co-packaged optics move optical engines closer to switch application-specific integrated circuits or accelerator packages. This reduces electrical trace length and SerDes power, which is increasingly important as bandwidth density rises [15]. However, co-packaged optics complicate thermal management, repair, and upgrade cycles. If an optical engine fails, replacing it may require replacing the switch package or board. This shifts reliability engineering from field-replaceable units to integrated subsystems. For AI clusters, where accelerator downtime is extremely expensive, serviceability must be considered alongside performance. A fabric that offers excellent bandwidth but cannot be repaired quickly may reduce overall cluster availability. Deployment models therefore influence whether a photonic technology is economically and operationally viable.

A third axis is the fiber medium. Standard single-mode fiber offers low loss and mature ecosystem support. Multimode fiber can be cost-effective for short reaches but has bandwidth-distance limitations. Hollow-core fiber is notable because light propagates largely in air, which can reduce latency and nonlinear effects. The required reference in this paper illustrates how hollow-core fiber can support low-latency, wavelength-switched, clock-synchronized intra-data center interconnects [6]. This matters for AI clusters because synchronization overhead accumulates across thousands of accelerators. Reducing propagation delay and jitter can improve collective communication efficiency. Yet hollow-core fiber is not a universal replacement. It requires careful handling, splicing, and connectorization, and its long-term reliability and cost profile continue to evolve. Optical interconnect design must therefore match fiber choice to reach, latency sensitivity, and deployment constraints.

4. Topology, Switching, and Resource Allocation Trade-offs

Network topology determines how accelerators are connected and how traffic flows. Traditional data center topologies such as fat-tree and leaf-spine provide high bisection

bandwidth and predictable routing but can be expensive at extreme radix. Dragonfly and torus topologies reduce cabling and switch count but create longer paths and more complex routing. Optical interconnects can enable new topologies by decoupling logical connectivity from physical cabling. For example, an optical circuit switch can reconfigure logical links between racks or pods, allowing the network to adapt to changing communication patterns. This is appealing for AI workloads because different parallelism strategies create different traffic matrices. However, reconfigurable topologies introduce control-plane challenges. The system must decide when to reconfigure, which paths to establish, how long to hold them, and how to handle conflicts. If reconfiguration is too frequent, control overhead and transient disruption may outweigh benefits. If it is too slow, the fabric may not respond to workload changes.

Resource allocation in optical fabrics involves both bandwidth and wavelength assignment. Wavelengths are finite, especially in cost-sensitive systems with limited transceiver tuning ranges. Assigning wavelengths poorly can lead to blocking, where a connection cannot be established even though capacity exists on other wavelengths. This is analogous to spectrum allocation in wireless networks and requires scheduling algorithms that account for traffic demand, priority, and fairness. In multi-tenant AI clusters, fairness is particularly important. A tenant running a large training job may legitimately need high bandwidth for collective communication, but it should not starve inference tenants with latency-critical requests. Admission control, quotas, and preemption policies may be necessary. These policies must be expressed in terms that optical switches can enforce. Otherwise, fairness remains an aspiration in orchestration software that the fabric cannot realize.

Switching granularity is another key trade-off. Packet switching provides fine-grained multiplexing and is well suited to bursty, short messages. Circuit switching provides high throughput and low energy per bit for long-lived flows. Wavelength switching sits between these extremes, offering optical path setup without full spatial circuit switching. A hybrid fabric can combine all three: electrical packet switching for control and small flows, wavelength switching for medium flows, and spatial circuit switching for large transfers. The challenge is to classify traffic correctly and route it to the appropriate plane. Misclassification can cause latency spikes or underutilization. Machine learning techniques may help predict traffic patterns and optimize reconfiguration, but they also introduce dependency on training data and model robustness. A control plane that is opaque or unpredictable may be unacceptable in production clusters. Explainability and fail-safe behavior are therefore important design goals.

Contention and congestion control must also be reconsidered. Electrical data center networks rely on congestion control protocols such as Data Center TCP to manage incast and buffer pressure [12]. Optical circuits change the nature of contention because dedicated paths bypass shared queues. This can reduce congestion but also reduce statistical multiplexing. If a circuit is underutilized, its capacity may be stranded. If traffic exceeds the circuit capacity, there is no easy way to buffer at the optical layer. Hybrid architectures must therefore coordinate electrical congestion control with optical path management. The control plane must know when to keep traffic on the electrical plane, when to establish an optical circuit, and when to tear it down. This coordination is a systems problem that spans network interface cards, switches, collective communication libraries, and job schedulers. Without end-to-end coordination, optical reconfiguration may simply move bottlenecks rather than eliminate them.

5. Optical Integration and Deployment Models

Deployment models for optical interconnects range from rack-level upgrades to full fabric replacement. A gradual approach introduces optical links where electrical links are most constrained, such as inter-rack or inter-pod connections. This preserves existing electrical switching and software while gaining reach and bandwidth benefits. A more aggressive approach deploys an optical fabric throughout the cluster, including accelerator-to-switch and switch-to-switch links. This can enable greater reconfigurability and resource disaggregation, but it requires new management tools, training for operators, and revised failure recovery procedures. Co-packaged optics may be most attractive at the switch and accelerator package level, where electrical link power dominates. However, integration reduces field replaceability. Operators may need to adopt new sparing strategies, such as spare line cards or whole-package replacements. The economic calculus depends on failure rates, repair costs, and downtime costs, which vary by organization and workload criticality [20].

Disaggregated and composable infrastructure is a natural beneficiary of optical interconnects. In a composable cluster, GPUs, memory pools, storage, and network accelerators can be dynamically assigned to jobs. Optical circuits can provide high-bandwidth, low-latency connections between disaggregated components. This improves resource utilization and allows heterogeneous accelerators to be matched to workloads. However, disaggregation increases the blast radius of failures. If an optical switch or control plane fails, multiple jobs may lose access to their resources. The system must therefore include strong isolation, rapid failure detection, and graceful degradation. It must also define clear ownership and accountability when resources are shared across administrative domains. In multi-tenant environments, a tenant may not know which optical paths its job uses or how failures are being handled. Transparency and service-level agreements become part of the interconnect design.

Deployment also involves standards and interoperability. Optical interconnects must interoperate with Ethernet, InfiniBand, Compute Express Link, and proprietary accelerator fabrics. Different standards emphasize different features: some prioritize packet semantics, others prioritize memory semantics, and others prioritize collective operations. An optical fabric that only supports one protocol may limit deployment options. Multi-protocol support can increase flexibility but also complexity. Network operating systems must expose optical path state to orchestration software, and telemetry must capture optical-specific metrics such as wavelength utilization, amplifier noise, and transceiver temperature. Without such telemetry, operators cannot diagnose performance problems or plan capacity. The observability stack is therefore as important as the photonic hardware. Optical interconnect design must include data models, APIs, and dashboards that make the optical layer legible to both human operators and automated control systems.

6. Robustness, Federation, and Governance

Robustness in optical AI clusters encompasses physical failures, control-plane failures, and workload-induced instability. Optical components can degrade over time due to laser aging, connector contamination, fiber stress, or thermal cycling. A single degraded link may cause retransmissions, collective communication slowdowns, or job failures. The system must detect degradation before it becomes catastrophic. This requires continuous monitoring and predictive maintenance. It also requires redundancy at multiple layers. Spare wavelengths, redundant fibers, and backup electrical paths can provide resilience, but they add cost and complexity. The design must determine which failures are tolerable and which require immediate failover. In large training jobs, even a brief interruption can force checkpoint

recovery, which may waste hours of computation. The value of redundancy must therefore be assessed against the cost of lost training time.

Federation and geo-distribution introduce additional robustness and governance challenges. Some AI workloads are distributed across multiple data centers for scale, data sovereignty, or energy reasons. Optical interconnects can link these sites, but long-haul optics face different constraints than intra-data center optics, including amplification, dispersion, and latency. Governance becomes more complex when data and computation cross jurisdictional boundaries. Policies may require data to remain within a region, or may restrict the use of certain accelerator technologies. The interconnect must support policy enforcement, such as path selection based on data residency or encryption requirements. It must also provide auditability. If an optical path is shared across tenants or agencies, the system must be able to demonstrate that isolation and compliance requirements were met. This is not merely a technical matter; it involves contracts, regulations, and institutional trust.

Security is closely related to robustness. Optical fibers are often considered difficult to tap without detection, but they are not immune to interception, jamming, or denial-of-service attacks. An attacker with physical access could bend a fiber, introduce a splitter, or inject high-power light. The control plane could also be attacked to misroute traffic or exhaust wavelengths. Multi-tenant clusters must therefore authenticate endpoints, encrypt traffic, and isolate control channels. They must also protect against side-channel leakage through timing or power analysis. Optical switching may reduce some electrical crosstalk risks, but it introduces new ones, such as wavelength collision and amplifier transients. Security must be designed into the optical layer, not added as an afterthought. This is especially important for clusters used in healthcare, finance, defense, and scientific research, where data sensitivity is high.

Fairness and governance also concern access to the interconnect itself. In a shared cluster, the optical fabric is a common-pool resource. If one tenant can monopolize wavelengths or reconfiguration opportunities, others suffer. Fair allocation may require quotas, reservations, priorities, and preemption. It may also require transparency about how resources are allocated. Tenants should be able to understand why their job experienced latency or bandwidth limitations. Operators should be able to audit decisions and adjust policies. This suggests that optical interconnect design should include policy interfaces that connect technical resource allocation to organizational rules. Without such interfaces, fairness remains informal and difficult to enforce. The design of the fabric thus intersects with institutional design, not only engineering.

7. Sustainability, Economics, and Policy

Sustainability is becoming a central concern for large-scale AI infrastructure. Training and serving advanced models consume significant electricity, water, and embodied carbon in hardware. Optical interconnects can improve energy efficiency by reducing electrical SerDes power and enabling longer reaches with lower loss [14]. They can also support higher utilization through disaggregation and dynamic reconfiguration, which reduces the number of idle accelerators. However, optical components have their own environmental footprint. Manufacturing lasers, modulators, and specialized fibers requires energy and materials. Co-packaged optics may increase integration complexity and reduce repairability, potentially increasing electronic waste. A full sustainability assessment must therefore consider the lifecycle of the interconnect, not only operational power. It must compare the energy saved during operation with the energy and materials used in production and disposal.

Economics shape deployment as much as physics. Data center network cost includes capital expenditure for switches, transceivers, fibers, and installation, as well as operational expenditure for power, cooling, maintenance, and downtime [11]. Optical interconnects can reduce some costs while increasing others. For example, they may lower power per bit but require more expensive transceivers and specialized technicians. They may reduce cabling complexity but increase control-plane complexity. The business case depends on workload scale, utilization, and energy prices. In regions with high electricity costs or strict carbon limits, optical efficiency may be more attractive. In regions with cheap power and low labor costs, electrical solutions may remain competitive longer. Policy can influence this calculus through carbon pricing, efficiency standards, and procurement rules. Governments and large enterprises can accelerate adoption by requiring lifecycle assessments and energy reporting for network equipment.

Policy also affects supply chains and resilience. Optical interconnects depend on specialized components that may be concentrated in a few geographic regions. Export controls, trade disputes, and natural disasters can disrupt supply. Organizations may need to diversify suppliers, maintain strategic reserves, or design for interoperability to avoid vendor lock-in. Standards bodies play a role in ensuring that optical interfaces are compatible across vendors. However, standards development can be slow relative to AI innovation. There is a tension between rapid proprietary deployment and slower but more open standardization. The right balance depends on the organization's risk tolerance and time horizon. For national laboratories and large cloud providers, proprietary solutions may be acceptable if they provide performance advantages. For universities and smaller enterprises, standards and interoperability may be more important for affordability and collaboration.

Sustainability and policy also intersect with fairness. If advanced optical interconnects are only affordable to large technology companies, they may concentrate AI capabilities in a few institutions. This raises concerns about equitable access to research infrastructure, economic opportunity, and scientific participation. Public funding agencies may need to support shared optical testbeds and open standards. Universities may need to collaborate on regional AI clusters that pool resources. Policy makers may need to consider incentives for energy-efficient design and transparency about environmental impact. The interconnect is not just a technical artifact; it is an infrastructure choice with distributional consequences. Designing it well requires attention to who benefits, who bears the costs, and how decisions are made.

8. Future Directions

Several promising directions could shape the next generation of optical interconnects for AI clusters. Co-packaged optics is likely to become more common as electrical link power becomes prohibitive. This will require advances in thermal management, testing, and repair. Silicon photonics will continue to mature, potentially enabling denser integration and lower cost [13]. Hollow-core fiber may find niche deployment in latency-sensitive intra-data center links, especially for synchronization-heavy workloads [6]. Wavelength switching and optical circuit switching will likely be integrated with electrical packet switching in hybrid fabrics. The control plane will need to become more autonomous, using telemetry and machine learning to predict traffic and reconfigure paths. However, autonomy must be bounded by safety and explainability requirements. An optical fabric that reconfigures itself unpredictably could destabilize training jobs and complicate debugging.

Another direction is the co-design of interconnect and collective communication. Today, collective libraries often treat the network as a black box. Future systems could expose optical

path information to the library, allowing it to schedule all-reduce operations along low-latency circuits or to align communication with wavelength availability. Similarly, job schedulers could place related accelerators on the same optical domain to reduce path length. Compilers could generate communication patterns that are more compatible with optical reconfiguration. This cross-layer co-design is difficult because it requires collaboration among hardware architects, network researchers, compiler engineers, and machine learning practitioners. Yet it offers substantial potential gains. The alternative is to continue optimizing layers independently, which often leads to suboptimal system performance.

Standards and governance will also evolve. Open APIs for optical network control, common telemetry formats, and interoperability profiles can reduce integration costs. Policy frameworks for multi-tenant fairness, data sovereignty, and environmental reporting can guide deployment. International collaboration may be needed to avoid fragmentation. At the same time, competition and national interests may drive proprietary approaches. The field must navigate this tension carefully. The goal should be a modular ecosystem where optical components, switches, control planes, and orchestration systems can be combined in different ways. Modularity supports innovation, resilience, and fairness. It also allows organizations with different needs and budgets to adopt optical interconnects incrementally. The future of optical AI clusters is therefore not a single architecture but a family of architectures shaped by technical, economic, and institutional choices.

9. Conclusion

Optical interconnect design for distributed GPU and AI accelerator clusters is a foundational systems challenge. As accelerator performance grows, communication increasingly determines scalability, efficiency, and reliability. Optical technologies offer compelling advantages in bandwidth density, reach, and energy per bit, but they also introduce new constraints in switching latency, control complexity, serviceability, and governance. This paper has argued that the design process must be holistic. It must begin with workload traffic patterns, including collective communication, synchronization, and burstiness. It must then consider physical-layer options, switching architectures, topologies, and resource allocation mechanisms. It must extend to deployment models, robustness, fairness, sustainability, and policy. Each layer interacts with the others, and decisions made in one layer can enable or undermine the goals of another.

The path forward requires interdisciplinary collaboration. Photonic engineers, computer architects, network researchers, software developers, operators, economists, and policy makers all have roles to play. The most successful systems will likely be hybrid, adaptive, and transparent. They will combine electrical packet switching with optical circuit and wavelength switching where appropriate. They will expose telemetry and control interfaces that allow orchestration systems to make informed decisions. They will include safeguards for fairness, security, and failure recovery. They will be evaluated not only on peak bandwidth but also on energy, cost, maintainability, and equity. Optical interconnects are not a silver bullet, but they are an essential part of the future AI infrastructure stack. Designing them well requires recognizing that they are simultaneously technical, economic, and social systems.

References

1. Shalf, J. (2020). The future of computing beyond Moore's Law. *Philosophical Transactions of the Royal Society A*, 378(2166), 20190061. <https://doi.org/10.1098/rsta.2019.0061>

2. Dean, J., & Barroso, L. A. (2013). The tail at scale. *Communications of the ACM*, 56(2), 74-80. <https://doi.org/10.1145/2408776.2408794>
3. Jouppi, N. P., Young, C., Patil, N., Patterson, D., Agrawal, G., Bajwa, R., Bates, S., Bhatia, S., Boden, N., Borchers, A., Boyle, R., Cantin, P.-L., Chao, C., Clark, C., Coriell, J., Daley, M., Dau, M., Dean, J., Gelb, B., ... Yoon, D. H. (2017). In-datacenter performance analysis of a tensor processing unit. *ACM SIGARCH Computer Architecture News*, 45(2), 1-12. <https://doi.org/10.1145/3079856.3080246>
4. Ben-Nun, T., & Hoefler, T. (2019). Demystifying parallel and distributed deep learning: An in-depth concurrency analysis. *ACM Computing Surveys*, 52(4), 1-43. <https://doi.org/10.1145/3320060>
5. Narayanan, D., Harlap, A., Phanishayee, A., Seshadri, V., Devanur, N. R., Ganger, G. R., Gibbons, P. B., & Zaharia, M. (2019). PipeDream: Generalized pipeline parallelism for DNN training. In *Proceedings of the 27th ACM Symposium on Operating Systems Principles* (pp. 1-15). ACM. <https://doi.org/10.1145/3341301.3359646>
6. Zhou, Z., Dzieciol, H., Clark, K., Luo, Y., Richardson, D., Poletti, F., ... & Liu, Z. (2023). Low-latency wavelength-switched clock-synchronized intra-data center interconnects enabled by hollow core fiber. *Optics Express*, 31(15), 24739-24748.
7. Kachris, C., Kanonakis, K., & Tomkos, I. (2012). Optical interconnection networks in data centers: Recent trends and future challenges. *IEEE Communications Magazine*, 50(9), 39-45. <https://doi.org/10.1109/MCOM.2012.6295708>
8. Rumley, S., Nikolova, D., Hendry, R., Li, Q., Bergman, K., & Carloni, L. P. (2017). Optical interconnects for extreme scale computing systems. *Parallel Computing*, 64, 65-80. <https://doi.org/10.1016/j.parco.2016.10.004>
9. Farrington, N., Porter, G., Radhakrishnan, S., Bazzaz, H. H., Subramanya, V., Fainman, Y., Papen, G., & Vahdat, A. (2010). Helios: A hybrid electrical/optical switch architecture for modular data centers. *ACM SIGCOMM Computer Communication Review*, 40(4), 339-350. <https://doi.org/10.1145/1851182.1851229>
10. Kandula, S., Sengupta, S., Greenberg, A., Patel, P., & Chaiken, R. (2009). The nature of data center traffic: Measurements and analysis. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement* (pp. 202-208). ACM. <https://doi.org/10.1145/1644893.1644918>
11. Greenberg, A., Hamilton, J., Maltz, D. A., & Patel, P. (2008). The cost of a cloud: Research problems in data center networks. *ACM SIGCOMM Computer Communication Review*, 39(1), 68-73. <https://doi.org/10.1145/1496091.1496103>
12. Alizadeh, M., Greenberg, A., Maltz, D. A., Padhye, J., Patel, P., Prabhakar, B., Sengupta, S., & Sridharan, M. (2010). Data center TCP (DCTCP). *ACM SIGCOMM Computer Communication Review*, 40(4), 63-74. <https://doi.org/10.1145/1851182.1851192>
13. Biberman, A., & Bergman, K. (2012). Optical interconnection networks for high-performance computing systems. *Reports on Progress in Physics*, 75(4), 046402. <https://doi.org/10.1088/0034-4885/75/4/046402>
14. Miller, D. A. B. (2009). Device requirements for optical interconnects to silicon chips. *Proceedings of the IEEE*, 97(7), 1166-1185. <https://doi.org/10.1109/JPROC.2009.2014298>

15. Taubenblatt, M. A. (2012). Optical interconnects for high-performance computing. *Journal of Lightwave Technology*, 30(4), 448-457.
<https://doi.org/10.1109/JLT.2011.2172989>
16. Binkert, N., Davis, A., Jouppi, N. P., McLellan, H., Nitta, C., Simon, A., & Vahdat, A. (2011). The role of optics in future high radix switch design. In *Proceedings of the 38th Annual International Symposium on Computer Architecture* (pp. 437-447). ACM.
<https://doi.org/10.1145/2000064.2000113>
17. Shen, Y., Meng, X., Cheng, Q., Rumley, S., Abrams, N., Gazman, A., ... & Bergman, K. (2019). Silicon photonics for extreme scale systems. *Journal of Lightwave Technology*, 37(2), 245-259. <https://doi.org/10.1109/JLT.2018.2877682>
18. Kachris, C., & Tomkos, I. (2012). A survey on optical interconnects for data centers. *IEEE Communications Surveys & Tutorials*, 14(4), 1021-1036.
<https://doi.org/10.1109/SURV.2011.122111.00069>
19. Cisco. (2020). Cisco annual internet report (2018-2023). Cisco Systems.
<https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>
20. Barroso, L. A., Clidaras, J., & Hölzle, U. (2013). *The datacenter as a computer: An introduction to the design of warehouse-scale machines* (2nd ed.). Morgan & Claypool.
<https://doi.org/10.2200/S00516ED2V01Y201306CAC024>