

Energy-Efficient Optical Interconnect Architectures for AI Data Centers

Enzio Gulotta

Department of Computer Science, University of Houston, Houston, TX, USA.
enzio858@uh.edu

Brady Gamble

Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
bradymail@uab.edu

Harsaya Prastuti

Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
harsayaprastuti86@unr.edu

Abstract

The rapid expansion of artificial intelligence data centers has transformed interconnect design from a supporting engineering concern into a central determinant of energy efficiency, scalability, and service fairness. As large models scale in parameters, training parallelism, and inference concurrency, electrical packet-switched networks face growing pressure from bandwidth-distance limitations, thermal density, and power delivery constraints. Optical interconnect architectures offer a promising alternative through wavelength parallelism, low propagation loss, and reduced switching energy per delivered bit, but their system-level benefits depend on topology, switching granularity, co-packaging strategies, control-plane maturity, and deployment governance. This paper examines energy-efficient optical interconnect architectures for AI data centers from an interdisciplinary systems perspective. It analyzes silicon photonics, wavelength-switched fabrics, optical circuit switching, hollow core fiber, and co-packaged optics as elements of a broader socio-technical infrastructure. The discussion emphasizes structural trade-offs among bandwidth, latency, radix, reliability, reconfiguration overhead, and sustainability. Particular attention is given to governance mechanisms that shape energy accounting, multi-tenant fairness, carbon-aware scheduling, and infrastructure retrofitting. The paper argues that optical interconnects should be evaluated not merely by device-level efficiency metrics but by lifecycle energy, operational resilience, equity of access, and policy accountability. It concludes that AI data centers require co-designed optical architectures, open interfaces, and transparent energy governance to realize sustainable performance at scale.

Keywords

optical interconnects; AI data centers; energy efficiency; silicon photonics; optical switching; data center architecture; sustainability; governance.

1. Introduction

Artificial intelligence data centers are increasingly best understood as computational infrastructures rather than conventional server farms. The scale of modern model training and

inference has created sustained demand for high-bandwidth, low-latency communication among accelerators, memory pools, and storage tiers. Electrical interconnect architectures, which have historically scaled through incremental improvements in signaling rate and packaging density, now encounter fundamental limits in reach, power, and thermal management [1]. Optical interconnects have therefore emerged as a critical enabling technology because they can transport high-bandwidth signals over longer distances with lower loss and greater wavelength parallelism than copper-based alternatives [2]. However, the transition from electrical to optical fabrics is not a simple substitution of one physical medium for another. It reshapes topology, switching, fault tolerance, cooling, software control, and even the political economy of data center deployment.

Early research on optical interconnects focused on device requirements, modulation formats, and integration pathways for silicon photonics. These studies established that optical links can reduce energy per bit when link distances and aggregate bandwidths exceed certain thresholds [3]. Hybrid electrical and optical switch architectures demonstrated that circuit switching can complement packet switching by providing high-radix connectivity for elephant flows while preserving electrical packet switching for fine-grained traffic [4]. High-radix switch designs further suggested that optics can alleviate the pin-bandwidth and backplane bottlenecks that constrain conventional switch chassis [5]. Silicon photonics then advanced the prospect of energy-efficient communication by integrating photonic components with complementary metal-oxide semiconductor manufacturing [6]. Demonstrations of microprocessors communicating directly through light illustrated the long-term possibility of chip-level optical input and output [7]. More recent work on attojoule optoelectronics has clarified the device-level energy boundaries and the engineering conditions under which optical links become preferable to electrical links [8].

Despite these advances, the literature often treats energy efficiency as a property of transceivers, modulators, or lasers rather than as an emergent property of the entire AI data center. Reviews of optical technologies for data centers have catalogued component progress but have also noted that system integration, packaging, and control remain decisive [9]. Research on optical interconnects for high-performance computing has shown that topology and workload communication patterns strongly affect realized efficiency [10]. For AI data centers, these factors are amplified because training workloads are synchronized, bursty, and sensitive to tail latency. A fabric that is energy-efficient in average throughput may still waste power during synchronization barriers or idle periods. Consequently, this paper examines energy-efficient optical interconnect architectures through a system-level lens that includes architecture, deployment, governance, sustainability, robustness, and fairness.

2. Architectural Foundations of Optical Interconnects in AI Data Centers

The architectural foundation of any data center network is its topology. Conventional cloud data centers often use Clos or leaf-spine topologies because they provide predictable bisection bandwidth, multipath routing, and incremental scalability [11]. The cost of such architectures is not only capital expenditure but also the energy consumed by numerous electrical switches, transceivers, and cooling systems [12]. Large-scale operators have refined these topologies through centralized control and standardized building blocks, as seen in Jupiter-style fabrics that support many thousands of servers [13]. AI data centers inherit these principles but place greater emphasis on accelerator-to-accelerator communication, collective operations, and memory disaggregation. As a result, the network becomes part of the computational critical path rather than a best-effort service layer.

Optical interconnects can be inserted at multiple levels of this architecture. They may replace rack-to-rack links, provide spine connectivity, or eventually connect chips within a package. The choice of insertion point determines the energy and reliability trade-offs. Pluggable optics offer modularity and maintainability but incur electrical SerDes and digital signal processing overhead. Co-packaged optics reduce electrical trace loss and improve bandwidth density but complicate thermal management and serviceability. Silicon photonics enables dense wavelength division multiplexing, but wavelength stability, thermal tuning, and laser reliability remain operational concerns [6]. These trade-offs are not merely technical; they influence vendor ecosystems, maintenance contracts, and the ability of operators to upgrade components independently.

Switching granularity is equally important. Optical circuit switching can provide high-bandwidth paths with low per-bit switching energy, but it requires reconfiguration time and a control plane capable of predicting or reacting to traffic shifts [4]. Optical packet switching promises finer granularity but demands optical buffering, fast switching, and synchronization that remain difficult at scale. Wavelength switching occupies a middle ground by assigning wavelengths to destinations or flows, enabling reconfigurable connectivity without full packet processing. Recent work on low-latency wavelength-switched and clock-synchronized intra-data center interconnects using hollow core fiber illustrates how propagation delay and synchronization can be addressed together [16]. Such architectures are especially relevant for AI workloads that require tightly coupled accelerators and predictable collective communication.

AI workloads also impose distinctive traffic patterns. Distributed training often uses all-reduce, all-gather, and reduce-scatter operations, while inference may involve mixture-of-experts routing or retrieval-augmented generation. These patterns create synchronized bursts that can congest electrical fabrics and cause power spikes. Optical interconnects can reduce congestion by providing reconfigurable high-bandwidth paths, but they also introduce control-plane delay and potential underutilization if reconfiguration is mismatched to workload phases. RotorNet and Sirius demonstrate that optical data center networks can achieve scalability and low complexity through scheduled or flat switching designs [14][15]. The architectural question is therefore not whether optics can outperform electrical links in principle, but how optical resources should be allocated, scheduled, and governed in the presence of dynamic AI workloads.

3. Energy Efficiency and Power Distribution Trade-offs

Energy efficiency in AI data centers must be assessed across computing, cooling, power conversion, and networking. The network share varies by workload and topology, but it is significant enough that interconnect choices affect total facility power. Optical links can reduce energy per bit over distance, yet they add electro-optic conversion, laser bias, thermal tuning, and control overhead. Device-level metrics such as picojoules per bit are useful but insufficient because they exclude cooling, power delivery losses, and embodied carbon [8]. A system-level accounting must therefore consider the entire pathway from accelerator output to remote accelerator input, including electrical-optical conversion, propagation, switching, retiming, and error correction.

Silicon photonics has improved the prospects for dense, manufacturable optical input and output, but energy efficiency depends on how aggressively photonic components are integrated with electronics [6]. Monolithic or hybrid integration can reduce parasitic capacitance and improve bandwidth density, while co-packaged optics can shorten electrical

traces. However, co-packaging concentrates heat near logic dies and complicates thermal design. Lasers remain a particular challenge because they degrade over time, require temperature control, and contribute to failure modes that may be difficult to isolate. Hollow core fiber can reduce latency and nonlinear impairments, which may lower buffering and retransmission energy, but it does not eliminate the need for efficient transceivers and switches [16].

At the system level, energy efficiency is determined by oversubscription, link utilization, and power state management. Electrical networks often keep links active at fixed rates even when traffic is low. Optical circuit switching can save energy by activating paths only when needed, but reconfiguration latency may force overprovisioning or burst buffering. Adaptive link rates, sleep states, and traffic-aware topology reconfiguration can reduce idle power, yet they require telemetry and control loops that are robust to failures. AI training jobs often run for hours or days, which creates opportunities for scheduled optical paths, but inference traffic can be highly variable and latency-sensitive. A hybrid architecture that combines electrical packet switching for fine-grained traffic with optical circuit switching for large flows may offer the best energy profile, provided the control plane avoids oscillation and excessive signaling.

Energy efficiency also intersects with AI model design and scheduling. Training large neural networks consumes substantial energy, and the carbon intensity of that energy depends on location, time, and grid mix [17]. Large-scale training runs have drawn attention to the relationship between model size, hardware utilization, and emissions [18]. Global data center energy estimates have been recalibrated to account for efficiency improvements and shifting demand, yet the growth of AI workloads introduces new uncertainty [19]. Climate-aligned AI research calls for transparent reporting, demand shifting, and hardware-software co-design [20]. Optical interconnects contribute to this agenda by enabling higher utilization of accelerators, reducing network energy, and supporting disaggregated architectures that can be powered down or migrated. However, without governance, efficiency gains may be reinvested in larger models or higher request volumes, a rebound effect that undermines sustainability goals.

4. Switching, Topology, and Latency Considerations for AI Workloads

AI workloads are unusually sensitive to latency and jitter because distributed training relies on frequent synchronization. A single slow link can stall an entire collective operation, leaving expensive accelerators idle. Optical interconnects can reduce latency by avoiding deep electrical switch hierarchies and by using low-loss fiber for long reaches. Hollow core fiber further reduces propagation delay, which is important for clock synchronization and tightly coupled computation [16]. Wavelength-switched architectures can provide reconfigurable direct paths while maintaining synchronization across nodes. These properties make optical fabrics attractive for AI clusters that span multiple racks or rows within a data center.

Topology design must balance bisection bandwidth, diameter, radix, and cost. Fat-tree and Clos topologies provide high bisection bandwidth but require many switches and links [11]. Dragonfly and other low-diameter topologies reduce cost but can create contention. Optical circuit switches can create a reconfigurable flat network, as demonstrated by RotorNet and Sirius [14][15]. A flat optical fabric can reduce the number of electrical hops and potentially lower energy per bit, but it requires a scheduler that can match traffic demands to circuit configurations. If the scheduler is centralized, it becomes a scalability and failure concern. If

it is distributed, it may suffer from suboptimal global decisions. These trade-offs are central to energy-efficient design because underutilized circuits waste optical power, while overcommitted circuits cause queuing and retransmission.

Latency and energy are not independent. Longer latency often requires larger buffers, deeper queues, and more aggressive retransmission, all of which consume energy. Optical interconnects can reduce latency by shortening paths and avoiding store-and-forward delays, but they may introduce reconfiguration latency. For AI training, collective communication libraries can be optimized to overlap computation with communication, but overlap depends on predictable network performance. Optical fabrics that provide predictable circuits may improve overlap, while packet-switched optical fabrics may introduce variability. The scheduling of optical resources should therefore be co-designed with collective algorithms, job placement, and checkpointing. This co-design is a system-level challenge that spans hardware, firmware, middleware, and application libraries.

5. Robustness, Reliability, and Failure Governance

Optical interconnects introduce new failure modes that differ from electrical networks. Lasers can degrade, modulators can drift, photodetectors can darken, connectors can be contaminated, and fiber can break. Thermal cycling can misalign components, while power fluctuations can affect wavelength stability. These failures may be gradual or intermittent, making them difficult to detect with simple link-up tests. Robust architectures require redundant paths, optical power monitoring, bit error rate telemetry, and automated fault isolation [1][6]. Co-packaged optics may reduce some connector failures but create new challenges for field replacement and thermal management.

Failure governance is as important as failure detection. AI data centers often operate under strict service level objectives for job completion, checkpoint durability, and inference latency. When an optical link fails, the control plane must decide whether to reroute traffic, reconfigure circuits, or migrate jobs. These decisions have energy consequences because rerouting may increase path length, while migration may require data movement. Multi-tenant environments add fairness concerns: a failure in one tenant's optical path should not cause disproportionate degradation for others. Governance mechanisms should define fault domains, isolation guarantees, and priority policies that are transparent to tenants and regulators. Without such mechanisms, optical efficiency gains may be offset by conservative overprovisioning and redundant power states.

Security and physical robustness also matter. Optical fibers are difficult to tap without detection, but they are not immune to interception or denial-of-service through physical damage. Data center operators must secure cable trays, patch panels, and switching rooms. They must also protect control planes that reconfigure optical paths. A compromised scheduler could create persistent congestion, eavesdrop on traffic patterns, or disable redundancy. These risks suggest that optical interconnect governance should include authentication, authorization, audit logging, and anomaly detection. The same telemetry used for energy optimization can support security monitoring, provided privacy and tenant confidentiality are preserved.

6. Deployment Pathways and Infrastructure Retrofitting

Deploying optical interconnects in AI data centers is not only a greenfield design problem. Many operators must retrofit existing facilities with limited power, cooling, and cable capacity. Retrofitting optical links can relieve bandwidth bottlenecks without replacing entire

switch hierarchies, but it may require new patch panels, power shelves, and thermal management. Pluggable optics offer an incremental path because they can be inserted into existing switch ports. Co-packaged optics offer greater density and efficiency but require new switch designs and operational procedures. The choice depends on facility age, service contracts, and the pace of AI hardware refresh [9][10].

Standards and interoperability shape deployment. Open interfaces allow operators to mix vendors, avoid lock-in, and extend the life of optical components. Proprietary interfaces may offer better performance in the short term but can slow innovation and increase costs. Governance bodies and consortia can accelerate adoption by defining power, thermal, and management specifications. Training and certification for technicians are also necessary because optical handling requires cleanliness, precision, and diagnostic skills that differ from electrical cabling. These human infrastructure factors are often overlooked in technical papers but are decisive in real deployments.

Operational models must evolve as well. Software-defined networking can abstract optical circuits and packet paths, but it must expose energy metrics and failure states to orchestration systems. AI schedulers can use this information to place jobs where network energy is lowest or where renewable energy is abundant. However, such scheduling requires accurate telemetry and standardized interfaces. If energy data is incomplete or inconsistent, optimization may shift load to locations with hidden carbon costs. Deployment governance should therefore require transparent measurement, verifiable reporting, and continuous auditing.

7. Sustainability, Fairness, and Policy Implications

Sustainability in AI data centers encompasses operational carbon, embodied carbon, water use, and electronic waste. Optical interconnects can reduce operational energy by lowering network power and improving accelerator utilization, but they also have manufacturing impacts. Silicon photonics, lasers, and hollow core fiber require specialized materials and processes that may have significant embodied carbon [20]. A lifecycle assessment should compare optical and electrical architectures over their full service life, including manufacturing, transportation, operation, maintenance, and disposal. Without such assessment, claims of energy efficiency may be incomplete or misleading.

Fairness is also at stake. Access to efficient optical infrastructure may be uneven across regions, institutions, and firms. Large cloud providers can invest in custom silicon photonics and co-packaged optics, while smaller research groups may rely on leased capacity or older electrical fabrics. This disparity affects not only cost but also the ability to conduct cutting-edge AI research. Policy instruments such as efficiency standards, public procurement requirements, and shared research facilities can broaden access. Carbon-aware scheduling and transparent energy pricing can also align incentives, but they must be designed to avoid shifting environmental burdens to disadvantaged communities [17][19].

Policy and governance should address measurement, accountability, and innovation. Standard metrics for energy per useful AI computation, network energy intensity, and carbon intensity per job would allow comparison across architectures and operators. Reporting requirements could include both operational and embodied emissions. Governance frameworks could encourage open interfaces, circular economy practices, and responsible decommissioning. The challenge is to avoid rigid rules that stifle innovation while ensuring that efficiency gains are real, verifiable, and equitably distributed. Optical interconnect architectures are not merely

technical choices; they are infrastructure decisions with social and environmental consequences [20].

8. Cross-Domain Comparisons and Future Research Directions

Comparing AI data centers with high-performance computing, telecommunications, and cloud computing reveals both parallels and differences. High-performance computing has long used optical interconnects for scale-out fabrics, and its experiences with topology, routing, and resilience can inform AI clusters [10]. Telecommunications has developed coherent optics, dense wavelength division multiplexing, and operational practices for long-haul reliability that may be adapted to intra-data center contexts. Cloud computing has perfected Clos topologies, centralized control, and commodity switching [13]. AI data centers combine these traditions but add tight coupling between computation and communication, making latency, synchronization, and collective operations more central.

Future research should pursue co-design across devices, packaging, topology, control, and applications. Monolithic integration of photonics and electronics could reduce energy and improve density, but it requires advances in materials, thermal management, and testing. Hollow core fiber could reduce latency and support new synchronization schemes [16]. Optical memory and in-network computing could reduce data movement, but they raise new governance questions about data locality and tenant isolation, particularly when computation is pushed into the network and data crosses administrative boundaries. Future research should therefore investigate how optical resources can be abstracted, scheduled, and audited without compromising security or fairness. Another promising direction is the integration of optical interconnects with renewable-aware scheduling and dynamic voltage and frequency scaling across the entire AI cluster. If network energy is exposed as a first-class signal to job schedulers, then training and inference tasks can be placed in ways that minimize carbon intensity while respecting latency constraints. This requires cross-layer telemetry, open application programming interfaces, and governance policies that reward verifiable efficiency rather than nominal peak bandwidth.

Comparative studies across domains also suggest that the optical community can learn from telecommunications in the areas of fault management, wavelength planning, and service-level agreements. Telecommunications operators have decades of experience in managing optical links with strict availability targets, and their practices for monitoring optical signal-to-noise ratio, chromatic dispersion, and polarization mode dispersion may be adapted to intra-data center fabrics. At the same time, AI data centers have unique requirements for collective communication and synchronization that are not typically found in telecom networks. A fruitful research agenda would combine telecom reliability engineering with AI workload characterization to produce optical architectures that are both resilient and efficient.

From a methodological perspective, future work should move beyond single-metric evaluations. Energy per bit, latency, throughput, and reliability must be assessed together, and the results should be reported with sensitivity analyses that account for workload variability, climate conditions, and hardware refresh cycles. Simulation frameworks should incorporate realistic AI traffic traces, including all-reduce and mixture-of-experts patterns, and they should model power delivery losses, cooling overhead, and embodied carbon. Experimental testbeds with programmable optical switches and co-packaged optics can validate these models and reveal operational challenges that are invisible in simulation. Such testbeds should be shared across institutions to promote reproducibility and broaden participation, which is itself a matter of fairness in infrastructure research.

Finally, the governance of optical interconnect deployment should be treated as a continuous process rather than a one-time compliance exercise. As AI models evolve and data center workloads shift, the assumptions behind initial architecture decisions may become obsolete. Operators should therefore establish review cycles that evaluate energy performance, failure modes, equity impacts, and environmental externalities. Regulators and standards bodies can support this process by defining metrics, encouraging transparency, and facilitating coordination among vendors, operators, and researchers. The goal is not to slow innovation but to ensure that innovation is directed toward architectures that are sustainable, robust, and just.

9. Conclusion

This paper has examined energy-efficient optical interconnect architectures for AI data centers from a system-level and interdisciplinary perspective. The analysis shows that optical interconnects are not merely a faster physical layer but a reconfiguration of the data center's topology, control plane, energy profile, and governance structure. Silicon photonics, wavelength switching, optical circuit switching, hollow core fiber, and co-packaged optics each offer distinct advantages, yet their benefits depend on workload characteristics, deployment pathways, and operational practices. Device-level efficiency metrics such as picojoules per bit are necessary but insufficient; lifecycle energy, cooling overhead, reliability, and fairness must also be considered.

Architecturally, AI data centers require fabrics that can handle synchronized collective communication, bursty inference traffic, and memory disaggregation. Hybrid electrical and optical designs that use packet switching for fine-grained flows and circuit switching for large transfers appear promising, but they demand sophisticated schedulers that can avoid oscillation and excessive signaling. Flat optical fabrics and wavelength-switched architectures can reduce hop count and latency, while hollow core fiber can further reduce propagation delay. However, these advantages are realized only when the control plane is robust, the packaging is thermally managed, and the failure modes are well understood.

Sustainability and fairness are inseparable from technical design. Optical interconnects can reduce network energy and improve accelerator utilization, but they also have embodied carbon and manufacturing impacts. Without transparent measurement and carbon-aware scheduling, efficiency gains may be offset by rebound effects or shifted to vulnerable communities. Policy frameworks should require verifiable energy reporting, support open interfaces, and promote equitable access to advanced infrastructure. Governance should also address multi-tenant fairness, fault isolation, and security of reconfigurable optical paths.

In conclusion, the path toward energy-efficient optical interconnects for AI data centers is a co-design challenge that spans materials, devices, packaging, topology, control, software, and policy. Progress will depend on interdisciplinary collaboration, shared testbeds, and a commitment to evaluating architectures by their total system performance and social impact. If pursued thoughtfully, optical interconnects can help AI data centers scale without proportional increases in energy consumption and environmental harm. If pursued narrowly, they may simply enable larger models and higher consumption without delivering durable sustainability. The choice is not purely technical; it is a matter of governance, values, and long-term infrastructure planning.

References

1. Benner, A. F., Ignatowski, M., Kash, J. A., Kuchta, D. M., & Ritter, M. B. (2005). Exploitation of optical interconnects in future server architectures. *IBM Journal of Research and Development*, 49(4/5), 755-775. <https://doi.org/10.1147/rd.494.0755>
2. 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>
3. 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>
4. 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/10.1145/1851182.1851223>
5. Petracca, M., Bergman, K., & Carloni, L. P. (2008). Photonic networks-on-chip for future generations of chip multiprocessors. *IEEE Transactions on Computers*, 57(9), 1246-1260. <https://doi.org/10.1109/TC.2008.78>
6. Jalali, B., & Fathpour, S. (2006). Silicon photonics. *Journal of Lightwave Technology*, 24(12), 4600-4615. <https://doi.org/10.1109/JLT.2006.885782>
7. Sun, C., Wade, M. T., Lee, Y., Orcutt, J. S., Alloatti, L., Georgas, M. S., ... & Stojanović, V. M. (2015). Single-chip microprocessor that communicates directly using light. *Nature*, 528(7583), 534-538. <https://doi.org/10.1038/nature16454>
8. Miller, D. A. B. (2017). Attojoule optoelectronics for low-energy information processing and communications. *Journal of Lightwave Technology*, 35(3), 346-396. <https://doi.org/10.1109/JLT.2017.2647779>
9. Kachris, C., Kanonakis, K., & Tomkos, I. (2013). Optical interconnection networks in data centers: Recent trends and future challenges. *IEEE Communications Magazine*, 51(9), 39-45. <https://doi.org/10.1109/MCOM.2013.6588648>
10. Ahn, J. H., Binkert, N., Davis, A., McLaren, M., & Schreiber, R. S. (2009). HyperX: Topology, routing, and packaging of efficient large-scale networks. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis* (pp. 1-11). ACM. <https://doi.org/10.1145/1654059.1654113>
11. Al-Fares, M., Loukissas, A., & Vahdat, A. (2008). A scalable, commodity data center network architecture. *ACM SIGCOMM Computer Communication Review*, 38(4), 63-74. <https://doi.org/10.1145/1402958.1402967>
12. Greenberg, A., Hamilton, J., Maltz, D. A., & Patel, P. (2009). 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>
13. Singh, A., Ong, J., Agarwal, A., Anderson, G., Armistead, A., Bannon, R., ... & Vahdat, A. (2015). Jupiter rising: A decade of Clos topologies and centralized control in Google's datacenter network. *ACM SIGCOMM Computer Communication Review*, 45(4), 183-197. <https://doi.org/10.1145/2785956.2787508>

14. Mellette, W. M., McGuinness, R., Roy, A., Papen, G., & Porter, G. (2017). RotorNet: A scalable, low-complexity, optical datacenter network. In Proceedings of the Conference of the ACM Special Interest Group on Data Communication (pp. 267-280). ACM.
<https://doi.org/10.1145/3098822.3098838>
15. Ballani, H., Costa, P., Behrens, C., Cletheroe, D., Haller, I., Jozwik, K., ... & Thomsen, B. (2020). Sirius: A flat datacenter network with nanosecond optical switching. In Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication on the Applications, Technologies, Architectures, and Protocols for Computer Communication (pp. 782-797). ACM.
<https://doi.org/10.1145/3387514.3406591>
16. 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.
17. Bashir, N., Irwin, D., Shenoy, P., & Spreitzer, M. (2011). Enabling sustainable clouds via environmentally opportunistic computing. In 2011 IEEE International Conference on Cloud Computing Technology and Science (pp. 430-437). IEEE.
<https://doi.org/10.1109/CloudCom.2011.65>
18. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 3645-3650). ACL. <https://doi.org/10.18653/v1/P19-1355>
19. Masanet, E., Shehabi, A., Lei, N., Smith, S., & Koomey, J. (2020). Recalibrating global data center energy-use estimates. *Science*, 367(6481), 984-986.
<https://doi.org/10.1126/science.aba3758>
20. Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54-63. <https://doi.org/10.1145/3381831>