

Energy-Efficient Optical Switching for Large-Scale AI Computing Clusters Using Adaptive Wavelength Allocation

Puck Jansse

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

puck2001@unr.edu

Chaitanya Ravel

Department of Computer Science, University of Central Florida, Orlando, FL, USA.

chaitanyaravel686@ucf.edu

Vittoria Rosselli

Department of Computer Science, University of New Hampshire, Durham, NH, USA.

vittoria.rosselli@unh.edu

Kurtis Kyne

Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.

kurtis.work@missouri.edu

Abstract

Large-scale artificial intelligence computing clusters are increasingly constrained by the energy and bandwidth demands of dense accelerator-to-accelerator communication. Optical switching offers a promising path beyond electrical packet switching, but its benefits depend on how wavelengths are allocated across tenants, workloads, and time. This paper examines energy-efficient optical switching for large-scale AI clusters through adaptive wavelength allocation, treating wavelength assignment as a system-level control problem rather than a purely physical-layer concern. It argues that adaptive allocation can reduce energy per useful computation by aligning optical resources with traffic dynamics, yet it also introduces structural trade-offs involving latency, fairness, robustness, infrastructure, and governance. The paper analyzes architectural implications for reconfigurable optical fabrics, wavelength-selective switching, clock synchronization, and hollow-core fiber, and it connects these technical choices to sustainability metrics, rebound effects, and policy accountability. It further considers multi-tenant fairness, failure modes, carbon-aware scheduling, and deployment pathways informed by power grids, spectrum governance, and interconnection network design. The central conclusion is that adaptive wavelength allocation can improve the energy efficiency of AI clusters only when paired with transparent telemetry, arbitration rules, lifecycle accounting, and resilient control mechanisms. Without such socio-technical safeguards, efficiency gains may be captured by incremental demand, hidden in operational complexity, or distributed inequitably across users and regions.

Keywords

optical switching, adaptive wavelength allocation, AI computing clusters, energy efficiency, data center infrastructure, sustainability, governance, multi-tenant fairness.

1. Introduction

Large-scale artificial intelligence computing clusters have become a critical form of computational infrastructure, combining thousands of accelerators, high-bandwidth memory, distributed storage, and tightly coupled communication fabrics. As model sizes, training corpora, and inference volumes grow, the limiting factors are no longer only arithmetic throughput and memory capacity but also the energy and latency costs of moving data between compute elements [1]. The end of straightforward semiconductor scaling has made specialized acceleration and interconnection central to continued performance gains, because advances in logic density alone can no longer deliver the historical rate of improvement [2]. In this setting, optical switching has attracted attention as a way to provide high-bandwidth, low-latency, and potentially more energy-efficient connectivity than purely electrical switching architectures [3]. However, the energy efficiency of optical switching is not automatic. It depends on how wavelengths, ports, circuits, and time slots are assigned across a dynamic and heterogeneous workload mix.

Adaptive wavelength allocation is the practice of adjusting wavelength assignments in response to traffic demand, workload phase, thermal conditions, energy prices, carbon intensity, or service-level objectives. In large AI clusters, such adaptation is attractive because training and inference workloads produce highly variable communication patterns. Collective operations, parameter synchronization, gradient exchange, checkpointing, and inference batching can create bursty flows that alternate between periods of intense east-west traffic and periods of relative quiescence. Static optical provisioning may therefore waste energy during low-demand intervals or create contention during peak intervals. Adaptive allocation promises to match optical resources to demand more closely, but it also raises difficult questions about who controls the wavelengths, how quickly reconfiguration can occur, what happens when multiple tenants compete, and how efficiency gains are measured. This paper treats adaptive wavelength allocation as a socio-technical control problem that spans photonics, computer architecture, cluster scheduling, energy accounting, and governance.

The paper proceeds by first situating AI cluster communication within broader data center and interconnection trends. It then conceptualizes adaptive wavelength allocation as a control and resource management problem, with attention to latency, throughput, energy, and fairness trade-offs. Subsequent sections examine architectural and infrastructure implications, energy and sustainability considerations, robustness and governance, deployment pathways, and open research challenges. The analysis emphasizes structural trade-offs rather than narrow component optimization, because the energy efficiency of optical switching depends on the entire system in which it is embedded. The conclusion argues that adaptive wavelength allocation can contribute to more sustainable AI computing only if it is designed with transparent arbitration, lifecycle metrics, and resilient operational practices.

2. System Context and Motivation

Modern AI clusters are warehouse-scale systems in which communication fabrics determine achievable scale and utilization [1]. Distributed training requires repeated synchronization across accelerators, often through collective communication patterns that stress bisection bandwidth and tail latency. Inference systems, meanwhile, may serve many small requests with strict response-time targets, creating traffic that is less synchronized but still sensitive to

congestion. Electrical packet switching has improved substantially through high-radix topologies and advanced scheduling, yet it faces physical constraints in reach, bandwidth density, and energy per bit as link speeds increase [3]. Optical interconnects can carry high bandwidth over longer distances with lower loss, and optical circuit or wavelength switching can potentially reduce the number of active electrical switch stages in the path.

Reconfigurable optical architectures have been explored for data centers for more than a decade. Hybrid electrical and optical designs, such as Helios, demonstrated the feasibility of combining packet switching for fine-grained traffic with optical circuit switching for large flows [4]. Projector and related systems further advanced agile reconfigurable interconnects by using optical switching to create high-bandwidth circuits that can be rearranged according to demand [5]. Microsecond circuit switching research showed that optical reconfiguration could be integrated into data center operations when control planes and buffering are carefully designed [6]. These efforts established an important principle: optical switching is most valuable when it is embedded in a broader traffic engineering and scheduling framework, not when it is treated as a drop-in replacement for electrical switching.

The motivation for adaptive wavelength allocation in AI clusters follows from this principle. AI workloads are not monolithic. A training job may alternate between compute-intensive phases and communication-intensive phases, while a cluster may simultaneously host multiple jobs with different priorities, deadlines, and tolerances for delay. Wavelength division multiplexing allows many independent channels to share a single fiber, but assigning fixed wavelengths to fixed endpoints can lead to stranded capacity. Adaptive allocation can reassign wavelengths to hot paths, consolidate traffic onto fewer active transceivers during low utilization, or shift traffic to links with favorable thermal or energy conditions. At the same time, adaptive allocation requires fast and reliable control, because frequent reconfiguration can consume energy, disrupt synchronization, and create instability. The design problem is therefore not simply how to switch wavelengths but when, for whom, and under what constraints.

Interconnection network research provides a useful vocabulary for these trade-offs. High-radix switch design, topology, routing, and flow control all influence how effectively optical resources can be used [7]. The classical principles of interconnection networks emphasize balance between latency, throughput, cost, and fault tolerance [8]. Adaptive wavelength allocation adds a new dimension: the physical medium itself can be reconfigured, but only within constraints imposed by transceiver tuning ranges, switching speed, crosstalk, and synchronization. For AI clusters, these constraints interact with collective communication libraries, job schedulers, and tenant isolation. A wavelength plan that optimizes aggregate throughput may harm a latency-sensitive inference service, while a plan that guarantees fairness may leave bandwidth underutilized. The remainder of this paper examines how these tensions can be managed at system level.

3. Adaptive Wavelength Allocation as a Control and Resource Management Problem

Adaptive wavelength allocation should be understood as a closed-loop control problem. The system observes traffic demand, queue depths, link utilization, job priorities, energy prices, and possibly carbon intensity. It then decides which wavelengths to assign to which endpoints or flows, how long to hold those assignments, and when to reconfigure. This decision process occurs across multiple timescales. At the fastest timescale, wavelength switching may occur in microseconds or milliseconds to respond to bursty traffic. At intermediate timescales, circuits may persist for seconds or minutes to support training phases. At slower timescales,

capacity planning, topology design, and procurement determine what optical resources exist in the first place. A coherent system must coordinate these layers rather than letting each optimize independently.

The control problem is complicated by the fact that wavelength is a shared and finite resource. A fiber can support many wavelength channels, but not an unlimited number. Optical amplifiers, transceivers, wavelength-selective switches, and spectrum assignment policies impose bounds [16]. When multiple AI jobs or tenants compete, the allocator must decide whose traffic receives a dedicated wavelength, whose traffic is groomed onto shared channels, and whose traffic is delayed. These decisions have energy consequences. Dedicated wavelengths can reduce contention and retransmission, but they may also leave transceivers active when demand is low. Aggressive consolidation can save energy, but it may increase latency and reduce fault tolerance. The allocator must therefore balance energy efficiency against service quality, resilience, and fairness.

Energy-aware optical networking research has shown that the energy consumed by communication systems depends on more than the raw line rate. Transceiver efficiency, switching overhead, cooling, and idle power all matter [9]. In cloud and data center environments, the balance between processing, storage, and transport energy can shift as workloads and architectures change [10]. Adaptive wavelength allocation can reduce energy by putting unused optical components into low-power states, selecting more efficient paths, or aligning communication with periods of low carbon intensity. However, the control overhead itself can be significant. Frequent reconfiguration may require active monitoring, control-plane messaging, clock synchronization, and retraining of scheduling models. The net benefit depends on whether the energy saved by better matching supply and demand exceeds the energy spent on adaptation.

Adaptive allocation also raises questions of stability and predictability. If multiple controllers respond to the same congestion signals, they may oscillate, causing wavelengths to be reassigned before flows can complete. If the allocator is trained on historical workloads, it may perform poorly when new model architectures or traffic patterns emerge. If tenants can influence allocation through their own scheduling decisions, strategic behavior may emerge. These concerns suggest that adaptive wavelength allocation should include dampening mechanisms, hysteresis, explicit service-level objectives, and escape rules for degraded conditions. It should also include audit trails, because energy and fairness claims depend on knowing why a particular allocation occurred. In this sense, the control problem is not only technical but also institutional: it determines how a shared computational infrastructure distributes a scarce resource.

4. Architecture and Infrastructure Implications

The architectural implications of adaptive wavelength allocation extend from the photonic layer to the cluster scheduler. At the photonic layer, the system requires tunable transceivers, wavelength-selective switches, multiplexers, demultiplexers, optical amplifiers, and potentially hollow-core fiber for low-latency transmission. Hollow-core fiber is particularly relevant for AI clusters because it can reduce propagation delay and support tight synchronization across distributed accelerators. Recent work on wavelength-switched clock-synchronized intra-data center interconnects using hollow-core fiber illustrates how physical-layer advances can enable low-latency reconfigurable paths [17]. This line of research is important because it shows that adaptive wavelength allocation is not only a bandwidth strategy but also a latency and synchronization strategy. In AI training, small delays in

collective communication can accumulate across many iterations, so reducing propagation and switching latency can improve both performance and energy efficiency.

At the switch and control layer, adaptive allocation requires a control plane that can translate workload intent into optical configuration. Job schedulers know which accelerators belong to which training job, which communication patterns are expected, and which deadlines are critical. The optical control plane knows which wavelengths are available, which paths are healthy, and which switches can be reconfigured. Bridging these domains requires APIs, telemetry standards, and policy engines. Without such integration, adaptive allocation risks becoming a local optimization that ignores global job performance. For example, assigning a dedicated wavelength to a large all-reduce operation may improve that operation while starving an inference service that needs low tail latency. A system-level architecture must therefore expose enough information for coordinated decisions while preventing any single tenant from monopolizing the optical fabric.

Infrastructure implications also include power distribution, cooling, and physical layout. Optical switches and transceivers have different thermal profiles than electrical switches, and their placement affects cooling requirements and maintenance. If adaptive allocation allows some components to idle, the cooling system must be able to scale down accordingly; otherwise, energy savings may be limited. Conversely, if reconfiguration increases thermal cycling, component reliability may suffer. Data center design must therefore consider not only peak bandwidth but also the dynamic range of optical utilization. This is especially important for AI clusters, which may run at high utilization for long periods and then drop sharply after a training run completes. The infrastructure should support graceful transitions between these states.

Finally, adaptive wavelength allocation has implications for network topology and oversubscription. Optical circuit switching can provide high-bandwidth paths between selected endpoints, but it cannot instantly create connectivity between every pair of accelerators. The topology must balance static electrical connectivity for fine-grained traffic with reconfigurable optical connectivity for large flows [4], [6]. In practice, hybrid designs are likely to dominate. Electrical packet switching may handle control messages, small flows, and load balancing, while optical switching handles bulk data movement and synchronized collectives. The allocator must decide which flows are eligible for optical circuits and how to avoid fragmentation. These decisions affect energy efficiency because every electrical hop that can be bypassed reduces energy per bit, but every reconfiguration adds overhead. The optimal balance will vary by workload, cluster size, and deployment maturity.

5. Energy Efficiency, Sustainability, and Rebound Effects

The sustainability case for adaptive optical switching rests on the premise that moving data with less electrical switching and better resource matching can reduce energy per useful computation. Optical transmission can be more energy-efficient than electrical transmission over certain distances and bandwidths, and wavelength switching can reduce the number of active switch stages [3], [9]. Adaptive allocation can further improve efficiency by consolidating traffic, powering down unused transceivers, and shifting communication to times or locations with lower carbon intensity. In AI clusters, where communication can be a substantial fraction of total energy, these savings could be meaningful. However, the magnitude of the benefit depends on workload characteristics, hardware efficiency, and system boundaries.

Data center energy research cautions against simplistic claims. Global data center energy use has grown more slowly than some early forecasts suggested, partly because of efficiency improvements in cooling, power distribution, and computing hardware [11]. Yet the rapid growth of AI workloads has renewed concerns about energy demand, water use, and carbon emissions. Studies of deep learning training have highlighted the energy and policy implications of large models [12], while subsequent work has called for measuring and reporting the carbon footprint of machine learning more systematically [13], [14]. Adaptive wavelength allocation should be evaluated within this broader accounting framework. It is not enough to show that an optical switch uses less power than an electrical switch; the analysis must include transceiver power, control overhead, cooling, embodied carbon, and the effects on job completion time.

Rebound effects are a central risk. If optical switching makes communication cheaper or faster, users may run larger models, more frequent experiments, or more redundant computations. The Jevons paradox and related rebound literature show that efficiency gains can be offset or even exceeded by increased demand [15]. In AI clusters, this could manifest as longer training runs, larger batch sizes, or more parallel experiments that consume the freed capacity. Adaptive wavelength allocation could therefore improve energy per bit while increasing total cluster energy. To avoid this outcome, efficiency metrics must be tied to useful output, such as energy per trained model, per inference, or per scientific result. Carbon-aware scheduling and demand-side policies may also be needed to ensure that efficiency gains are not simply reinvested in additional consumption.

Lifecycle considerations further complicate the picture. Optical transceivers and switches contain sophisticated materials and components whose manufacturing and disposal have environmental impacts. Hollow-core fiber and advanced photonic devices may offer performance benefits, but their production may be more energy-intensive than conventional alternatives. Adaptive allocation can extend hardware lifetime by reducing unnecessary utilization, but frequent reconfiguration may increase wear on tunable components. A sustainable design must therefore consider the full lifecycle, including manufacturing, operation, maintenance, and end-of-life recovery. It should also consider water use and land use, especially in regions where data centers compete for scarce resources. Energy efficiency is a necessary but not sufficient condition for sustainability.

6. Robustness, Fairness, and Governance

Robustness is essential because adaptive wavelength allocation introduces new failure modes. A misconfigured wavelength assignment can create silent data corruption, synchronization errors, or transient outages. A control-plane failure can leave the optical fabric in an inconsistent state. An adversarial tenant might attempt to exhaust wavelength resources or infer the traffic patterns of other tenants. The system must therefore include fault detection, isolation, rollback, and degradation modes. In a large AI cluster, a single training job may span thousands of accelerators, so a localized optical fault can have disproportionate consequences. Resilience requires redundant paths, conservative reconfiguration policies, and clear operational procedures for reverting to static or electrical fallback modes.

Fairness is equally important. AI clusters are often shared by multiple teams, projects, or external customers. Adaptive wavelength allocation determines who gets high-bandwidth circuits, who experiences congestion, and whose jobs are delayed. If allocation is based solely on aggregate throughput, large jobs may dominate at the expense of smaller or more interactive workloads. If allocation is based on priority, lower-priority users may be starved.

A fair system needs explicit policies for sharing, minimum guarantees, burst allowances, and appeal mechanisms. These policies should be transparent to users and auditable by operators. Fairness is not only an ethical concern; it also affects system stability, because users who cannot predict performance may overprovision or behave strategically.

Governance frameworks must address the fact that optical infrastructure is both technical and political. The allocation of wavelengths can shape which research questions are computationally feasible, which regions benefit from AI development, and which communities bear the environmental costs of data centers [20]. Corporate or institutional control over adaptive allocation may concentrate advantage unless accompanied by oversight. Governance mechanisms might include published allocation policies, independent audits of energy and fairness metrics, stakeholder representation in infrastructure planning, and reporting requirements for carbon and water use. In multi-tenant environments, contracts should specify service-level objectives, data protection, and dispute resolution. Adaptive allocation should not be a black box that hides consequential decisions behind technical complexity.

Security is another dimension of governance. Wavelength-level attacks, side-channel leakage, and denial-of-service attempts could target the optical control plane. Because adaptive allocation relies on telemetry about traffic and job behavior, privacy and confidentiality must be protected. Aggregated telemetry may be sufficient for control, but granular data could reveal model architectures, training data characteristics, or business strategies. Operators should therefore apply data minimization, access control, and encryption. They should also test the control plane against adversarial conditions, including sudden traffic shifts, sensor spoofing, and compromised tenants. Robustness and fairness are intertwined: a system that fails unpredictably or unequally will lose legitimacy, even if it is energy-efficient in average conditions.

7. Deployment Pathways, Policy, and Cross-Domain Lessons

Deployment of adaptive wavelength allocation will likely be incremental rather than revolutionary. Existing data centers have large investments in electrical switching, and many AI workloads can tolerate hybrid architectures. A practical pathway begins with optical links for the highest-volume, most predictable flows, such as inter-rack communication for large training jobs. Over time, as control-plane maturity and workload awareness improve, more flows can be admitted to the optical fabric. This incremental approach allows operators to validate energy savings, reliability, and fairness before expanding scope [1], [4]. It also allows standards bodies and vendors to converge on interfaces for telemetry, wavelength assignment, and fault management. Premature standardization could freeze suboptimal designs, while delayed standardization could fragment the market and raise integration costs.

Cross-domain comparisons offer useful lessons. Electrical power grids use market mechanisms, reserves, and congestion management to balance supply and demand across many generators and users. Spectrum governance uses licenses, sharing rules, and interference protection to allocate a scarce wireless resource. Air traffic control uses centralized separation and scheduling to maintain safety under high density. Adaptive wavelength allocation in AI clusters faces analogous challenges: scarce capacity, competing users, real-time constraints, and safety or reliability requirements. Lessons from these domains suggest the value of layered control, clear property rights, fallback modes, and independent oversight. They also suggest that purely market-based allocation may

underprovide resilience and fairness, while purely centralized allocation may be too slow or inflexible for dynamic AI workloads.

Policy implications extend to procurement, incentives, and international coordination. Governments and large institutions that fund AI infrastructure can require energy and carbon reporting, interoperability, and fair access as conditions of support. They can also promote research into open control planes, standardized telemetry, and reusable components. International coordination is needed because AI supply chains and data center impacts cross borders. At the same time, policy should avoid locking in specific technologies or vendors. The goal should be to create a regulatory environment that rewards measurable efficiency, transparency, and equitable access while allowing innovation in optical switching and adaptive allocation. The required reference on wavelength-switched interconnects illustrates how quickly the physical layer can evolve; policy must be adaptable enough to accommodate such changes without sacrificing accountability [17].

8. Open Challenges and Future Research

Several open challenges remain. First, workload-aware allocation requires better models of AI communication patterns. Current schedulers often treat jobs as black boxes, but adaptive optics could benefit from predictions of collective operations, synchronization frequency, and sensitivity to latency. Second, control-plane scalability must be demonstrated at cluster sizes with thousands of endpoints. Centralized controllers may become bottlenecks, while distributed controllers may struggle with global optimization. Third, energy accounting must become more granular and standardized. Without consistent metrics, claims about energy-efficient optical switching will be difficult to compare across systems [13], [14].

Fourth, security and privacy in optical control planes require further study. Wavelength assignment data could reveal sensitive workload information, and control messages could be targeted by attackers. Fifth, hardware reliability under frequent reconfiguration is not fully understood. Tunable lasers, wavelength-selective switches, and hollow-core fiber may have different aging characteristics than static optical components. Sixth, the interaction between optical switching and advanced interconnection topologies needs more research. High-radix electrical switches, optical circuit switches, and emerging photonic fabrics may complement each other in ways that are not yet fully mapped [7], [8]. Seventh, policy and governance frameworks must keep pace with technical change. Adaptive allocation decisions may have distributional consequences that are not visible in standard performance benchmarks [20].

Future research should therefore combine photonics, computer architecture, scheduling, economics, and policy analysis. Testbeds that emulate realistic AI workloads across multiple tenants would be valuable. Digital twins of optical fabrics could help operators explore allocation policies without disrupting production. Open datasets and benchmarks for energy, latency, fairness, and carbon intensity would improve comparability. Finally, researchers should study how adaptive wavelength allocation interacts with other efficiency strategies, such as model compression, sparsity, quantization, and carbon-aware scheduling. No single technique will solve the energy challenge of large-scale AI. The most promising path is an integrated system design in which optical switching, workload scheduling, and sustainability governance reinforce one another [2], [18].

9. Conclusion

Adaptive wavelength allocation offers a credible pathway toward more energy-efficient optical switching in large-scale AI computing clusters. By matching wavelength resources to

dynamic traffic demands, operators can reduce idle capacity, bypass unnecessary electrical switching, and improve the energy per useful computation. Optical interconnects, reconfigurable fabrics, and hollow-core fiber make this vision technically plausible, and recent advances in wavelength-switched, clock-synchronized intra-data center interconnects demonstrate the potential for low-latency, high-bandwidth operation [17]. However, the benefits are not automatic. Adaptive allocation introduces complexity, control overhead, fairness dilemmas, and new failure modes. It can also trigger rebound effects if efficiency gains are reinvested in additional demand.

The central argument of this paper is that adaptive wavelength allocation should be treated as a socio-technical system. Technical design must address switching speed, synchronization, topology, telemetry, and resilience. Institutional design must address transparency, arbitration, accountability, and equitable access. Energy and sustainability metrics must cover the full lifecycle and connect efficiency to useful output. Governance must prevent wavelength allocation from becoming an opaque mechanism that concentrates computational advantage or hides environmental costs. If these conditions are met, adaptive optical switching can help large AI clusters become more efficient, robust, and fair. If they are ignored, efficiency gains may be partial, fragile, or socially uneven. The task for researchers and operators is therefore not only to build faster optical switches but to design the rules, metrics, and institutions that make those switches serve sustainable and legitimate computational infrastructure.

References

1. Barroso, L. A., Clidaras, J., & Hölzle, U. (2018). *The datacenter as a computer: Designing warehouse-scale machines* (3rd ed.). Springer.
2. Shalf, J. (2020). The future of computing beyond Moore's Law. *Philosophical Transactions of the Royal Society A*, 378(2166), 20190061.
3. Kachris, C., & Tomkos, I. (2012). A survey on optical interconnects for data centers. *IEEE Communications Surveys & Tutorials*, 14(4), 1021–1036.
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.
5. Ghobadi, M., Mahajan, R., Phanishayee, A., Devanur, N., Kulkarni, J., Ranade, G., Blanche, P.-A., Rastegarfar, H., Glick, M., & Kilper, D. (2016). Projector: Agile reconfigurable data center interconnect. In *Proceedings of the 2016 ACM SIGCOMM Conference* (pp. 216–229). ACM.
6. Porter, G., Strong, R., Farrington, N., Forencich, A., Chen-Sun, P., Rosing, T., Fainman, Y., Papen, G., & Vahdat, A. (2013). Integrating microsecond circuit switching into the data center. *ACM SIGCOMM Computer Communication Review*, 43(4), 447–458.
7. Binkert, N., Davis, A., Jouppi, N. P., McLaren, M., Muralimanohar, N., Schreiber, R., & Ahn, J. H. (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.
8. Dally, W. J., & Towles, B. (2004). *Principles and practices of interconnection networks*. Morgan Kaufmann.

9. Tucker, R. S. (2011). Green optical communications—Part II: Energy limitations in networks. *IEEE Journal of Selected Topics in Quantum Electronics*, 17(2), 261–274.
10. Baliga, J., Ayre, R. W. A., Hinton, K., & Tucker, R. S. (2011). Green cloud computing: Balancing energy in processing, storage, and transport. *Proceedings of the IEEE*, 99(1), 149–167.
11. Masanet, E., Shehabi, A., Lei, N., Smith, S., & Koomey, J. (2020). Recalibrating global data center energy-use estimates. *Science*, 367(6481), 984–986.
12. 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.
13. Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54–63.
14. Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D., Texier, M., & Dean, J. (2021). Carbon emissions and large neural network training. *arXiv*. <https://arxiv.org/abs/2104.10350>
15. Gossart, C. (2015). Rebound effects and ICT: A review of the literature. In L. M. Hilty & B. Aebischer (Eds.), *ICT innovations for sustainability* (pp. 435–448). Springer.
16. Essiambre, R.-J., Kramer, G., Winzer, P. J., Foschini, G. J., & Goebel, B. (2010). Capacity limits of optical fiber networks. *Journal of Lightwave Technology*, 28(4), 662–701.
17. Zhou, Zichuan, et al. "Low-latency wavelength-switched clock-synchronized intra-data center interconnects enabled by hollow core fiber." *Optics Express* 31.15 (2023): 24739-24748.
18. Winzer, P. J., Neilson, D. T., & Chraplyvy, A. R. (2018). Fiber-optic transmission and networking: The previous 20 and the next 20 years. *Optics Express*, 26(18), 24190–24239.
19. Kahn, J. M., & Ho, K.-P. (2001). Spectral efficiency limits and modulation techniques for fiber-optic communication systems. In *Optical Fiber Communication Conference* (pp. TuK1). Optica Publishing Group.
20. Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.