

Adaptive Optical Network Scheduling for Bursty Cloud Application Traffic

Gero Anders

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

geroa@unr.edu

Abstract

Cloud applications generate highly variable traffic that alternates between long periods of modest demand and intense bursts caused by data-parallel computation, user-facing request surges, storage synchronization, and distributed machine learning. Conventional electrical packet-switched data center networks handle this variability through overprovisioning, buffering, and rapid statistical multiplexing, but these strategies increasingly encounter energy, latency, and scaling limits. Optical networking offers high capacity and low loss, yet static optical provisioning cannot respond efficiently to bursty cloud workloads. This paper examines adaptive optical network scheduling as a socio-technical infrastructure problem rather than a purely physical-layer optimization. It develops an architectural framework that combines optical circuit and packet domains, telemetry-driven control, demand prediction, cross-layer coordination, and policy-aware resource governance. The analysis emphasizes structural trade-offs among reconfiguration latency, spectral efficiency, fairness, robustness, and sustainability. It also considers deployment pathways in brownfield data centers, the role of software-defined control, and the governance challenges that arise when optical capacity becomes a shared and dynamically allocated resource. Case illustrations from hyperscale data centers, edge clouds, high-performance computing, and latency-sensitive services show that no single scheduling model dominates across all contexts. Instead, adaptive optical scheduling must be designed as a modular capability that can shift between circuit-like predictability, packet-like elasticity, and hybrid modes according to workload class, service-level objectives, and operational constraints. The paper concludes that successful adoption depends as much on institutional coordination, transparent policy, and energy accounting as on advances in photonic switching and control algorithms.

Keywords

adaptive optical networks; cloud data centers; bursty traffic; scheduling; optical circuit switching; resource governance; sustainability; robustness.

1. Introduction

Cloud computing has become the computational substrate for a wide range of services, including web search, social media, streaming, scientific simulation, financial analysis, and artificial intelligence. These services depend on data center networks that must move enormous volumes of data among servers, storage systems, accelerators, and external users. The traffic they produce is not smooth. It is bursty, heterogeneous, and often synchronized across many machines, because a single user request can trigger hundreds or thousands of internal operations. Traditional data center designs have responded with hierarchical electrical switches and rich path diversity, allowing statistical multiplexing to absorb short-term fluctuations. However, the continued growth of cloud demand places pressure on this model.

Electrical switching consumes substantial energy, introduces queuing delay, and becomes difficult to scale in both port count and bandwidth density. Optical interconnects have therefore attracted attention as a means of providing high-capacity links and potentially lower energy per bit [1]. At the same time, measurements of production data center traffic show that demand is highly uneven across time, rack, service, and application class [2]. This unevenness creates both a problem and an opportunity for scheduling. If optical resources are allocated statically, they will often be idle in one part of the network while another part experiences congestion. If they are reconfigured too aggressively, control overhead and transient disruption may outweigh the capacity gains [3].

Adaptive optical network scheduling addresses this tension by treating optical capacity as a dynamic resource that can be assigned, reassigned, and released according to observed and predicted demand. The concept spans several layers. At the physical layer, it involves tunable transceivers, wavelength routing, spectrum assignment, and fast optical switching. At the network layer, it requires mechanisms for path computation, topology reconfiguration, and coordination between optical and electrical domains. At the service layer, it intersects with workload placement, virtual machine migration, storage replication, and application-level admission control. At the governance layer, it raises questions about fairness among tenants, accountability for service-level violations, and the environmental consequences of capacity allocation. This paper argues that adaptive optical scheduling for bursty cloud traffic cannot be understood solely as an engineering optimization. It is a systems and policy problem in which technical choices shape organizational behavior and social outcomes.

The central claim is that effective adaptive optical scheduling must be modular, observability-driven, and policy-aware. Modularity allows optical circuit, optical packet, and electrical packet domains to coexist without requiring a complete replacement of existing infrastructure. Observability provides the telemetry needed to distinguish genuine burstiness from measurement noise and to anticipate demand shifts. Policy awareness ensures that allocation decisions reflect service priorities, tenant fairness, energy constraints, and resilience requirements. The remainder of the paper develops this argument through a review of related work, an architectural framework, an analysis of cloud traffic dynamics, a discussion of scheduling mechanisms and cross-layer coordination, an examination of governance and fairness, and an assessment of deployment, sustainability, and robustness. It closes with case illustrations, future directions, and conclusions.

2. Background and Related Work

Research on data center networking has long recognized that traffic patterns are shaped by application architecture. Early measurement studies found that data center traffic is not simply a scaled version of wide-area Internet traffic; it exhibits high locality, incast behavior, and periodic synchronization [3]. The leaf-spine and fat-tree topologies were developed to provide uniform bandwidth between servers and to support the any-to-any communication patterns of distributed applications [4]. These electrical architectures remain dominant because they are flexible, mature, and compatible with packet-switched protocols. Nevertheless, their scalability is constrained by the power and cost of high-radix electrical switches, as well as by the growing mismatch between link capacity and switching capacity.

Hybrid electrical and optical architectures attempt to combine the flexibility of packet switching with the capacity and energy advantages of optics. Helios, for example, explored a hybrid electrical and optical switch design for modular data centers, using optical circuits for large flows and electrical packet switching for fine-grained traffic [5]. OSA proposed an

optical switching architecture that could reconfigure network topology to match traffic demand, thereby improving flexibility without requiring a full optical packet-switched data plane [6]. These designs demonstrate that optical resources can be valuable even when they are not used for every packet. The challenge lies in deciding when a flow or group of flows should be moved to an optical path, how long the path should remain provisioned, and how to manage the transition without disrupting latency-sensitive services.

Recent work on frequency-referenced multiplexing has shown that communications can achieve guaranteed bandwidth and low latency through coordinated optical frequency references, which suggests new possibilities for deterministic optical services in cloud environments [7]. This development is relevant because it points toward a tighter coupling between physical-layer stability and network-level scheduling. If optical channels can provide predictable latency and bandwidth, then scheduling algorithms can treat them as quasi-reserved resources rather than best-effort links. However, deterministic optical capacity is not free. It requires coordination, synchronization, and careful spectrum management, and it may not be suitable for all traffic classes.

Software-defined networking has also shaped the evolution of adaptive infrastructure. In wide-area networks, centralized traffic engineering has been used to improve utilization and to respond to changing demand patterns [8]. Production deployments such as B4 demonstrated that software-defined control can manage heterogeneous network resources at scale, although the approach depends on accurate telemetry and careful failure handling [9]. Data center networks have adopted similar principles, but their timescales are shorter and their traffic patterns more volatile. The cost of cloud data center networking has been analyzed as a first-order design concern, with energy, cabling, and switch complexity all contributing to the need for more efficient architectures [10]. Warehouse-scale computers further illustrate that the data center is best understood as a single machine whose components must be co-designed for efficiency and reliability [11]. These studies provide the background for adaptive optical scheduling, but they also show that optical capacity alone does not solve the broader problem of coordinating resources under bursty demand.

3. Architectural Framework for Adaptive Optical Scheduling

An adaptive optical scheduling architecture can be organized into four planes. The optical data plane consists of reconfigurable transceivers, wavelength-selective switches, optical circuit switches, and possibly optical packet or burst switching elements. The electrical data plane continues to handle fine-grained packet forwarding, buffering, and short-term burst absorption. The control plane computes allocation decisions, maintains topology state, and coordinates handoffs between optical and electrical paths. The management and policy plane translates service-level objectives, tenant agreements, energy constraints, and security requirements into rules that constrain the control plane. This separation is analytical rather than physical; in practice, the planes may be implemented through distributed controllers, programmable switches, and telemetry agents.

The most important structural trade-off is between reconfiguration latency and resource utilization. Optical circuit switching can provide very high bandwidth with low per-bit energy, but establishing a circuit takes time. If the circuit is held too long, capacity is wasted during periods of low demand. If it is released too quickly, the network may oscillate, causing repeated setup and tear-down overhead. Adaptive scheduling must therefore estimate the duration and value of a burst before committing optical resources. For long-lived transfers, such as storage replication, model checkpointing, and large-scale data shuffles, optical circuits

are attractive. For short request-response interactions, electrical packet switching is often more appropriate. The architecture should not force all traffic into one mode; it should classify demand and select the mode that offers the best balance of latency, throughput, energy, and control cost.

Telemetry is the foundation of this classification. The architecture requires visibility into link utilization, queue depth, flow size distribution, application identity, tenant identity, and service-level objectives. It also benefits from predictive signals, such as scheduled jobs, batch windows, and application-level deadlines. Machine learning can help identify recurring patterns and forecast demand, but it must be used with caution. Opaque models can produce unstable allocations, and they may encode biases against tenants whose workloads are less predictable. A robust architecture should combine statistical forecasting with explicit policy rules and human-auditable thresholds. The goal is not to replace network operators with automated agents, but to give operators tools that make the consequences of allocation decisions visible and contestable.

Another trade-off concerns centralization and distribution. A logically centralized controller can optimize global objectives, such as minimizing maximum link utilization or maximizing renewable energy use. However, centralization introduces scalability and failure risks. A distributed or hierarchical control model may be more resilient, with local agents handling fast decisions and a global controller managing long-term policy and inter-domain coordination. The choice depends on the size of the network, the criticality of the services, and the regulatory environment. In multi-tenant clouds, the control plane must also prevent one tenant from monopolizing optical capacity or inferring another tenant's activity through side channels. This requires isolation, accounting, and auditability in addition to performance optimization.

4. Traffic Dynamics and Burstiness in Cloud Applications

Burstiness in cloud applications arises from several sources. Data-parallel jobs create synchronized communication phases, such as map-shuffle-reduce patterns, in which many workers exchange data at once. Distributed storage systems generate replication and repair traffic that can spike after failures or during maintenance. Machine learning training creates periodic checkpointing and all-reduce operations that stress interconnects. User-facing services produce diurnal patterns and flash crowds, while microservice architectures amplify a single external request into many internal calls. The tail-at-scale phenomenon shows that even rare slow components can dominate end-to-end latency when requests fan out across many servers [12]. These dynamics mean that average utilization is a poor guide for capacity planning. A network may appear lightly loaded on average while experiencing severe local congestion during bursts.

Cloud computing research has emphasized elasticity, resource pooling, and measured service as defining characteristics [13]. These principles are usually applied to computation and storage, but they apply equally to networking. Bursty traffic requires the network to allocate capacity where it is needed, when it is needed, and for as long as it is valuable. Static wavelength assignments cannot meet this requirement efficiently because they bind capacity to physical paths regardless of demand. Fully dynamic optical packet switching could, in principle, provide fine-grained elasticity, but it remains challenging in terms of buffering, synchronization, and cost. Hybrid approaches therefore dominate practical interest. They allow large, predictable flows to use optical circuits while preserving electrical packet switching for small, latency-sensitive, or unpredictable flows.

The Jupiter deployment illustrates how optical circuit switching can be integrated into a production data center network to increase capacity and flexibility under changing traffic demands [14]. It also shows that the benefits depend on software-defined control, careful topology management, and an understanding of which flows are suitable for optical transport. In edge and access environments, traffic burstiness is compounded by mobility, wireless variability, and the limited capacity of backhaul links. Edge computing moves computation closer to users, but it also creates new coordination demands among edge sites, regional clouds, and centralized data centers [17]. Adaptive optical scheduling must therefore operate across multiple geographic and administrative domains, not only within a single homogeneous data center.

5. Scheduling Mechanisms and Cross-Layer Coordination

Scheduling mechanisms for adaptive optical networks can be categorized by the resource they allocate and the timescale on which they operate. At the longest timescale, capacity planning determines which links and switches are optical, how much spectrum is available, and where reconfigurable elements are placed. At an intermediate timescale, traffic engineering computes paths and assigns wavelengths or spectrum bands to traffic aggregates. At the shortest timescale, admission control and flow scheduling decide which individual transfers enter the optical domain and when they should be preempted or migrated. These timescales must be coordinated. A long-term decision to deploy optical switches in a particular pod, for example, constrains the short-term options available to the scheduler. Conversely, persistent short-term congestion may justify new optical capacity or a change in topology.

Cross-layer coordination is essential because optical scheduling interacts with transport protocols, application runtimes, and storage systems. Transport protocols that assume stable round-trip times may perform poorly when paths are reconfigured. Applications that are unaware of optical availability may issue transfers at times that create unnecessary contention. Storage systems that replicate data without considering network topology may generate traffic that crosses expensive optical paths. An adaptive architecture can mitigate these problems by exposing network state through well-defined interfaces and by allowing applications to express deadlines, priorities, and tolerance for delay. The emerging optical data center vision emphasizes programmable, software-defined control that can align optical and electrical resources with application requirements [15]. Machine learning can further improve this alignment by predicting demand, detecting anomalies, and tuning parameters, but it must be integrated with explainable policy so that operators can understand why particular allocations occur [16].

Fairness is a cross-layer concern as well. In a multi-tenant cloud, optical capacity is a shared resource. A scheduler that maximizes aggregate throughput may starve tenants with bursty but low-volume workloads. A scheduler that enforces strict equal sharing may waste capacity on tenants that cannot use it. A practical approach is to define fairness in terms of service-level objectives rather than raw bandwidth. For example, each tenant might receive a guaranteed minimum optical capacity during congestion, with additional capacity allocated according to weighted priorities or market-like mechanisms. Such policies require transparent accounting and dispute resolution. They also require robustness against strategic behavior, because tenants may attempt to inflate their reported demand to obtain more capacity.

6. Governance, Fairness, and Policy

Adaptive optical scheduling transforms network capacity into a governed resource. Governance includes the rules, institutions, and practices that determine who gets access, under what conditions, and with what accountability. In a private cloud, governance may be handled through internal service-level agreements and capacity planning committees. In a public cloud, it involves contracts, pricing, regulatory obligations, and consumer protection. In critical infrastructure, it may involve national or sectoral regulators. The technical design of the scheduler shapes these governance arrangements. A scheduler that records detailed flow-level telemetry, for instance, creates privacy risks because traffic patterns can reveal application behavior, tenant relationships, or user activity. A scheduler that relies on opaque machine learning may make it difficult to explain or contest allocation decisions. A scheduler that allows rapid reconfiguration may improve efficiency but complicate security auditing and fault attribution.

Fairness in optical scheduling has multiple dimensions. Distributive fairness concerns how capacity is divided among tenants, applications, and regions. Procedural fairness concerns whether allocation rules are transparent and consistently applied. Temporal fairness concerns whether short bursts are treated differently from sustained demand. Spatial fairness concerns whether some racks or pods receive systematically better service because of their location relative to optical switches. These dimensions can conflict. For example, prioritizing latency-sensitive traffic may improve user experience but reduce the throughput available to batch jobs. Prioritizing geographic balance may increase energy use if renewable energy is concentrated in one region. A policy-aware scheduler should make these trade-offs explicit and allow stakeholders to adjust weights as conditions change.

Regulation and policy also matter because optical networks are part of broader communications and energy infrastructures. Governments may impose requirements for lawful interception, data localization, or critical infrastructure protection. Energy regulators may promote efficiency standards or carbon reporting. Competition authorities may scrutinize whether dominant cloud providers use control over optical infrastructure to disadvantage smaller competitors. The design of adaptive scheduling can either support or undermine these goals. Open interfaces, standardized telemetry, and auditable allocation logs can improve accountability. Proprietary control planes and opaque algorithms can entrench advantage and reduce the ability of regulators or customers to verify claims. International traffic engineering research has long emphasized the importance of coordinated control and policy in large networks [20]. Adaptive optical scheduling extends these concerns to the photonic layer, where capacity is abundant in principle but constrained by physics, spectrum, and switching hardware.

7. Deployment, Sustainability, and Robustness

Deployment of adaptive optical scheduling will occur largely in brownfield environments. Existing data centers contain large investments in electrical switches, fiber plants, and operational practices. A practical strategy is incremental: introduce optical circuit switches for specific traffic classes, measure the benefits, and expand as operational confidence grows. This approach reduces risk but may create fragmentation, with some traffic using optical paths and other traffic remaining on electrical paths. The control plane must therefore manage a mixed topology and provide consistent service-level guarantees across both domains. Deployment also requires skilled staff, updated monitoring tools, and new procedures for fault diagnosis. These organizational factors are often underestimated in technical research.

Sustainability is a central motivation for optical networking. Optical links can carry more data per watt than electrical links over distance, and optical circuit switching can avoid the energy cost of electronic packet processing for large flows. However, the full life-cycle impact includes manufacturing, cooling, and the energy used by control and telemetry systems. Adaptive scheduling can improve sustainability by matching capacity to demand, consolidating traffic onto fewer links, and shifting flexible workloads to times or locations with cleaner energy. It can also worsen sustainability if aggressive reconfiguration increases overhead or if efficiency gains lead to rebound effects, in which lower cost per bit encourages greater consumption. Energy-efficient dynamic traffic offloading and reconfiguration have been studied as ways to reduce data center energy use, but they require careful system-level accounting [19]. A sustainable scheduler should therefore report energy intensity per delivered service, not only per bit, and should consider the embodied energy of optical equipment.

Robustness is equally important. Optical components can fail, fiber can be cut, and control planes can be compromised. An adaptive network that reconfigures frequently may be more vulnerable to transient errors and oscillations. It must include safety mechanisms such as conservative fallback to electrical paths, rate limiting on reconfiguration, and validation of topology changes before they are committed. It must also handle correlated failures, in which a single optical switch or amplifier affects many flows. Resilience planning should consider not only physical redundancy but also control-plane diversity, graceful degradation, and clear operational procedures for reverting to known-good states. In cloud environments, robustness is a service-level expectation, not an optional feature. Adaptive optical scheduling must therefore be designed with failure as a normal condition rather than an exception.

8. Case Illustrations and Cross-Domain Comparisons

Hyperscale cloud providers offer the most immediate context for adaptive optical scheduling. These providers operate large data centers with diverse workloads, centralized control, and significant engineering resources. They can deploy optical circuit switches at scale and use software-defined control to manage traffic across thousands of servers. The Jupiter deployment shows that optical switching can be integrated into a production network to improve capacity and flexibility [14]. Hyperscale environments also have the telemetry and automation needed to support adaptive scheduling, although they must still manage fairness among internal services and external customers. The main challenge is not whether optics can work, but how to govern a dynamic optical fabric without creating new operational risks.

Edge and access networks present a different set of constraints. Edge clouds are distributed across many small sites, often with limited power, cooling, and staff. Traffic is highly variable because it depends on mobile users, local sensors, and wireless conditions. Optical connections between edge sites and regional data centers can provide high-capacity backhaul, but scheduling must account for intermittent demand and diverse ownership. Edge computing research emphasizes low latency, context awareness, and coordination across cloud and edge tiers [17][18]. Adaptive optical scheduling in this context may be less about fine-grained circuit switching and more about dynamic wavelength or spectrum allocation between edge aggregation points and regional hubs. Fairness becomes geographically salient because rural or underserved areas may have fewer optical resources and less political influence.

High-performance computing and financial services provide contrasting cases. High-performance computing workloads are often scheduled in advance, with predictable communication phases and known job requirements. Optical scheduling can be integrated

with job schedulers to reserve capacity for large collective operations. Financial services require extremely low and predictable latency, often for small messages. Optical circuit switching may be less suitable for individual transactions, but it can support bulk data replication between trading centers and analytics clusters. These cases show that the value of adaptive optical scheduling depends on the temporal and spatial structure of the workload. A scheduler designed for hyperscale web traffic may not serve a scientific supercomputer or a latency-sensitive trading system well. Cross-domain comparison therefore argues for modular architectures with configurable policies rather than one-size-fits-all designs.

9. Future Directions

Future research should pursue three interconnected directions. The first is observability and explainability. Adaptive schedulers will need richer telemetry that captures application intent, energy source, topology state, and tenant policy without violating privacy or creating excessive overhead. Explainable machine learning can help operators understand why a particular allocation was chosen and how to adjust it. The second direction is standardized governance. If optical capacity becomes dynamically allocated across organizational boundaries, there will need to be common formats for service-level objectives, accounting, and audit. Standards bodies, cloud providers, and regulators all have roles to play. The third direction is sustainability-oriented scheduling that optimizes for carbon intensity, water use, and embodied energy rather than only bandwidth or latency. This will require integrating network control with energy markets, workload managers, and building management systems.

Another promising direction is the use of digital twins and simulation for safe experimentation. Because optical reconfiguration can disrupt live services, operators need ways to test policies in simulated or shadow environments before deployment. Digital twins can model traffic, failures, and control decisions, allowing researchers to explore trade-offs without risking production traffic. They can also support training for operators and help regulators evaluate the systemic effects of different governance rules. Finally, adaptive optical scheduling should be studied as part of a broader trend toward programmable, software-defined infrastructure. The boundary between network, computation, and storage is blurring, and future cloud platforms may allocate optical paths as naturally as they allocate virtual machines. Achieving that vision will require technical advances, but it will also require institutions that can manage shared infrastructure fairly, sustainably, and robustly.

10. Conclusion

Adaptive optical network scheduling for bursty cloud application traffic is a response to a fundamental mismatch between static optical capacity and dynamic cloud demand. Optical networking offers high bandwidth and energy efficiency, but its benefits depend on the ability to allocate capacity in time, space, and spectrum according to workload needs. This paper has argued that the problem should be approached as a socio-technical systems challenge. The architectural framework must combine optical and electrical domains, central and local control, telemetry and policy, and performance and fairness. Traffic analysis must recognize the diverse sources of burstiness, from data-parallel computation to microservice fan-out and storage repair. Scheduling mechanisms must operate across timescales and coordinate with transport, applications, and storage. Governance must make allocation decisions transparent, contestable, and aligned with service-level objectives and sustainability goals. Deployment must be incremental, robust, and sensitive to organizational capacity. Case illustrations show that no single model fits all environments; hyperscale clouds, edge networks, high-performance computing, and financial services each require different balances of circuit-like

predictability and packet-like elasticity. The path forward lies in modular architectures, standardized governance, explainable automation, and sustainability accounting. If these elements are developed together, adaptive optical scheduling can help cloud infrastructure meet future demand without sacrificing fairness, resilience, or environmental responsibility.

References

1. 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>
2. Benson, T., Akella, A., & Maltz, D. A. (2010). Network traffic characteristics of data centers in the wild. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement* (pp. 267–280). ACM. <https://doi.org/10.1145/1879141.1879175>
3. 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>
4. 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/10.1145/1402946.1402967>
5. 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>
6. Chen, K., Singla, A., Singh, A., Ramachandran, K., Xu, L., Zhang, Y., Wen, X., & Chen, Y. (2014). OSA: An optical switching architecture for data center networks with unprecedented flexibility. *IEEE/ACM Transactions on Networking*, 22(2), 498–511. <https://doi.org/10.1109/TNET.2013.2253128>
7. Zhou, Z., Wei, J., Luo, Y., Clark, K. A., Sillescu, E., Deakin, C., ... & Liu, Z. (2023). Communications with guaranteed bandwidth and low latency using frequency-referenced multiplexing. *Nature Electronics*, 6(9), 694–702.
8. Hong, C.-Y., Kandula, S., Mahajan, R., Zhang, M., Gill, V., Nanduri, M., & Wattenhofer, R. (2013). Achieving high utilization with software-defined WAN. *ACM SIGCOMM Computer Communication Review*, 43(4), 15–26. <https://doi.org/10.1145/2534169.2486012>
9. Jain, S., Kumar, A., Mandal, S., Ong, J., Poutievski, L., Singh, A., Venkata, S., Wanderer, J., Zhou, J., Zhu, M., Zolla, J., Hözl, U., Stuart, S., & Vahdat, A. (2013). B4: Experience with a globally-deployed software defined WAN. *ACM SIGCOMM Computer Communication Review*, 43(4), 3–14. <https://doi.org/10.1145/2534169.2486019>
10. 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>

11. Barroso, L. A., Clidaras, J., & Hölzle, U. (2013). The datacenter as a computer: An introduction to the design of warehouse-scale machines. *Synthesis Lectures on Computer Architecture*, 8(3), 1–154. <https://doi.org/10.2200/S00516ED2V01Y201306CAC024>
12. 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>
13. Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud computing: State-of-the-art and research challenges. *Journal of Internet Services and Applications*, 1(1), 7–18. <https://doi.org/10.1007/s13174-010-0007-6>
14. Poutievski, L., Mashayekhi, O., Chen, S., Chowdhury, M., Sivakumar, R., Mehra, M., Kaushik, R., Pemberton, D., She, Y., Gopal, A., & Vahdat, A. (2022). Jupiter evolving: Transforming Google's datacenter network via optical circuit switches and software-defined networking. In *Proceedings of the ACM SIGCOMM 2022 Conference* (pp. 66–85). ACM. <https://doi.org/10.1145/3544216.3544265>
15. Vahdat, A., Liu, H., Zhao, X., & Johnson, C. (2015). The emerging optical data center. In *Optical Fiber Communication Conference* (pp. OTuH2). Optical Society of America. <https://doi.org/10.1364/OFC.2015.OTuH2>
16. Musumeci, F., Rottondi, C., Nag, A., Macaluso, I., Zibar, D., Ruffini, M., & Tornatore, M. (2019). An overview on application of machine learning techniques in optical networks. *IEEE Communications Surveys & Tutorials*, 21(2), 1383–1408. <https://doi.org/10.1109/COMST.2018.2880039>
17. Taleb, T., Samdanis, K., Mada, B., Flinck, H., Dutta, S., & Sabella, D. (2017). On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration. *IEEE Communications Surveys & Tutorials*, 19(3), 1657–1681. <https://doi.org/10.1109/COMST.2017.2705720>
18. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5), 637–646. <https://doi.org/10.1109/IIOT.2016.2579198>
19. Baccarelli, E., Cordeschi, N., Mei, A., Panella, M., Shojafar, M., & Stefa, J. (2016). Energy-efficient dynamic traffic offloading and reconfiguration of networked data centers for big data stream mobile computing: Review, challenges, and a case study. *IEEE Network*, 30(2), 54–61. <https://doi.org/10.1109/MNET.2016.7437025>
20. Akyildiz, I. F., Lee, A., Wang, P., Luo, M., & Chou, W. (2014). A roadmap for traffic engineering in SDN-OpenFlow networks. *Computer Networks*, 71, 1–30. <https://doi.org/10.1016/j.comnet.2014.06.002>