

Deep Reinforcement Learning for Adaptive Bandwidth Allocation in Cloud Optical Networks

Cacilia Metz

Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.

caciliawork@ku.edu

Daire Landy

Department of Computer Science, University of North Texas, Denton, TX, USA.

dairel@unt.edu

Wiktor Baszak

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

wiktor1984@unh.edu

Annelie Reising

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

contactannelie@uab.edu

Abstract

Cloud optical networks are becoming a foundational substrate for distributed computing, artificial intelligence services, and latency-sensitive applications. Adaptive bandwidth allocation in such environments must reconcile high capacity, deterministic performance, energy efficiency, and multi-tenant fairness. This paper examines how deep reinforcement learning can be integrated into cloud optical network control planes to support adaptive bandwidth allocation. It frames the problem as a socio-technical systems challenge rather than a purely algorithmic one, emphasizing architecture, governance, deployment, robustness, sustainability, and policy. The analysis begins with the structural characteristics of optical data center fabrics and the limitations of static or rule-based provisioning. It then explains how deep reinforcement learning can learn allocation policies from telemetry, traffic dynamics, and service-level objectives, while also introducing risks related to non-stationarity, reward misspecification, fragility under distribution shift, and opaque decision-making. The paper evaluates architectural choices including centralized, hierarchical, and federated control; interactions with software-defined networking and network function virtualization; and the role of optical circuit switching, elastic spectrum assignment, and frequency-referenced multiplexing. Particular attention is given to resilience under failure, fairness across tenants, energy-aware operation, observability, and human oversight. Deployment pathways are discussed in terms of simulation-to-production transfer, safety envelopes, staged rollouts, and regulatory accountability. The conclusion argues that deep reinforcement learning can improve bandwidth allocation only when embedded within robust governance and infrastructure constraints. The most credible future systems will combine learning-based adaptation with verifiable fallback mechanisms, transparent performance accounting, and institutional rules that protect fairness and public-interest values.

Keywords

cloud optical networks, deep reinforcement learning, bandwidth allocation, optical data center networks, network control, fairness, sustainability, governance.

1. Introduction

Cloud optical networks have evolved from simple high-capacity transport fabrics into programmable infrastructures that support computation, storage, machine learning training, financial transactions, public services, and real-time communication. As demand for bandwidth-intensive and latency-sensitive workloads grows, operators face a persistent tension between efficiency and determinism. Static wavelength assignment, conservative overprovisioning, and manually tuned traffic engineering can protect service quality, but they often waste spectral resources and fail to respond quickly to shifting demand. Conversely, aggressive dynamic allocation can improve utilization while increasing the risk of congestion, instability, and unacceptable variance for premium tenants. This tension makes adaptive bandwidth allocation a central systems problem rather than a narrow optimization task [1], [2].

The architectural context matters because optical networks are not merely faster versions of electronic networks. They combine circuit-like and packet-like behavior, involve physical-layer constraints such as wavelength continuity and signal quality, and often depend on long-lived provisioning actions that cannot be undone instantly without disrupting active flows [3], [4]. In cloud environments, these optical resources must be coordinated with compute scheduling, storage placement, virtual network functions, and service-level agreements. Consequently, bandwidth allocation is entangled with infrastructure design, operations, and policy. A decision that improves aggregate throughput may harm fairness, increase energy consumption, or weaken failure isolation. A decision that protects one tenant may reduce the resilience of the shared substrate.

Deep reinforcement learning has attracted attention because it can learn sequential decision policies from experience rather than relying entirely on hand-crafted rules. In principle, an agent can observe network telemetry, infer latent demand patterns, select allocation actions, and improve over time according to a reward signal that reflects throughput, latency, packet loss, energy use, or service compliance [5], [6], [7]. However, the movement from simulation success to production reliability is difficult. Production networks are non-stationary, partially observable, adversarial in some traffic patterns, and governed by contractual and regulatory constraints. This paper therefore treats deep reinforcement learning for adaptive bandwidth allocation as a socio-technical capability that must be evaluated alongside architecture, governance, and deployment strategy [8], [9].

2. Cloud Optical Networks and Bandwidth Allocation

Cloud optical networks provide high-capacity connectivity among data centers, availability zones, edge facilities, and wide-area transport points. Their physical layer supports dense wavelength division multiplexing, reconfigurable optical add-drop multiplexers, optical circuit switching, and increasingly flexible spectrum assignment. These technologies allow operators to create virtual topologies that match traffic demands more closely than fixed physical topologies [2], [17]. Yet flexibility introduces control complexity. The control plane must decide when to establish or tear down lightpaths, how to route them, how much spectrum to assign, and how to protect critical services when failures occur. These decisions interact with higher-level orchestration systems that manage virtual machines, containers, and network functions [18].

Bandwidth allocation in this setting can be understood as the continuous matching of finite optical capacity to heterogeneous service requirements. Some applications require high throughput but tolerate moderate delay, while others require deterministic latency, low jitter, or strict isolation. Cloud providers often express these needs through service classes, quotas, and service-level objectives. A purely aggregate approach that maximizes total utilization may violate these commitments because it treats all bits as equivalent. A purely conservative approach may preserve quality but strand capacity and increase cost. Adaptive allocation must therefore operate across multiple objectives and time horizons [10], [15].

Recent optical research has explored mechanisms that provide guaranteed bandwidth and low latency through frequency-referenced multiplexing, illustrating that physical-layer design can directly shape what control policies are feasible [15]. When the substrate can offer stronger performance guarantees, learning-based allocation can focus on coordinating demand rather than compensating for unstable capacity. Conversely, when the physical layer provides only best-effort capacity, the learning system must internalize uncertainty and apply larger safety margins. This relationship between optical capability and algorithmic ambition is often overlooked. Better photonics does not eliminate the need for intelligent control, but it changes the control problem by making some constraints less severe and others more visible.

Elastic optical networks add another layer by allowing spectrum to be assigned in flexible blocks rather than fixed wavelength grids. This improves spectral efficiency but creates fragmentation and assignment constraints that evolve over time. Routing, modulation, and spectrum assignment become tightly coupled, and a poor decision can block future requests even if current demand is satisfied [14]. Deep reinforcement learning is appealing in this context because it can learn to anticipate fragmentation, reserve capacity for predicted bursts, and adapt to changing traffic mixes. At the same time, the action space is large, delayed consequences are common, and safety violations may affect multiple tenants. These features make the problem suitable for learning but also demand strong guardrails.

3. Deep Reinforcement Learning as Adaptive Control

Deep reinforcement learning combines reinforcement learning with deep neural networks to approximate policies or value functions in complex environments. The agent interacts with an environment by observing a state, choosing an action, receiving a reward, and updating its policy to improve long-term performance [8]. In network control, the state may include link utilization, queue depths, flow statistics, failure indicators, and service-level measurements. Actions may include allocating spectrum, rerouting traffic, adjusting priorities, or reconfiguring optical paths. Rewards may combine throughput, latency, packet loss, energy consumption, and fairness penalties [9]. The appeal is that such agents can discover non-obvious policies that outperform static heuristics in dynamic conditions [5], [6].

Several families of methods are relevant. Value-based methods learn to estimate the long-term value of actions and have been used in resource management and traffic optimization [10], [13]. Policy-gradient methods directly optimize a policy and are often better suited to continuous or high-dimensional actions, such as adjustable bandwidth shares or spectrum assignments [7]. Actor-critic designs combine both ideas and can stabilize learning in complex systems. Model-based approaches learn or exploit a model of network dynamics to plan ahead, which can improve sample efficiency but introduces model error risks. In optical networks, the choice among these approaches depends on the observability of the environment, the cost of exploration, the acceptable level of risk, and the availability of high-fidelity simulation [16].

Deep reinforcement learning also inherits several well-known difficulties. Reward misspecification can produce policies that optimize a proxy while harming the actual service. For example, an agent rewarded only for throughput may starve latency-sensitive traffic or create unfair allocations across tenants. Non-stationarity can cause policies to become obsolete as traffic patterns, applications, and infrastructure change. Partial observability means the agent may not see the true bottleneck or the intentions of other controllers. Distribution shift can make a policy that performed well in simulation behave unpredictably in production. Finally, exploration itself can be dangerous in an operational network because trying unfamiliar actions may disrupt live traffic. These challenges do not invalidate deep reinforcement learning, but they require a deployment philosophy based on constraints, simulation, staged rollout, and continuous oversight [11], [12].

4. Architectural Integration and Deployment

The integration of deep reinforcement learning into cloud optical networks can take several architectural forms. A centralized agent may control a broad domain and benefit from global visibility, but it introduces scalability limits, communication overhead, and a single point of failure. A hierarchical design may place fast local agents near optical nodes while a slower global agent coordinates long-term topology and capacity planning. A federated or multi-agent design may distribute learning across domains, allowing each domain to adapt locally while sharing limited information. Each architecture embodies different trade-offs among responsiveness, coordination, privacy, and resilience [12], [16]. The choice should follow from operational boundaries, not from algorithmic fashion.

Software-defined networking provides a natural integration point because it separates control from data forwarding and exposes programmable interfaces. Software-defined optical networks extend this logic to optical elements, enabling centralized or logically centralized control of lightpaths and spectrum [17]. Network function virtualization further allows service functions to be placed and migrated dynamically, which changes bandwidth demand patterns and creates opportunities for joint optimization [18]. In such environments, a deep reinforcement learning agent could interact with an orchestrator that enforces policy, validates actions, and resolves conflicts among multiple objectives. The agent should not be granted unconstrained authority over physical resources. Instead, it should propose allocations within safety envelopes defined by capacity, protection, and service-level constraints.

Deployment pathways are likely to be incremental. Early systems may use learning for recommendation rather than direct control, allowing human operators to approve or reject actions. Later systems may automate narrow decisions with clear boundaries, such as adjusting backup capacity or rebalancing best-effort traffic. Mature systems may coordinate multiple domains, but only if observability, rollback, and accountability mechanisms are strong. Simulation-to-production transfer remains a major hurdle because optical networks have long-tailed failure modes and because traffic generated in testbeds may not reflect production behavior. Digital twins, shadow mode deployment, and canary testing can reduce risk by comparing learned policies against existing heuristics before granting them authority [4], [10].

5. Robustness, Resilience, and Failure Governance

Resilience is a first-order requirement in cloud optical networks because failures can affect many tenants and critical services simultaneously. Fiber cuts, amplifier failures, transponder faults, software bugs, and configuration errors can create abrupt changes in topology and

capacity. A learning-based allocation system must detect these changes, avoid unsafe actions, and recover quickly. If the policy has been trained mainly on normal conditions, it may fail precisely when adaptation is most needed. Robustness therefore requires training with stress scenarios, adversarial traffic, degraded capacity, and partial observability. It also requires fallback mechanisms that can revert to conservative rules when uncertainty is high [11], [13].

Failure governance is not only technical. Operators need clear rules about which decisions can be automated, who is accountable when harm occurs, and how affected tenants are notified and compensated. An agent that reroutes traffic to protect a high-priority service may degrade another tenant without explicit human intent. If the reward function encodes only technical metrics, the resulting policy may be legally or contractually problematic. Governance must define priority classes, isolation guarantees, restoration obligations, and transparency requirements. It must also require logging of state, action, and outcome so that incidents can be reconstructed [20]. Without such governance, deep reinforcement learning becomes an opaque layer of risk rather than a tool for resilience.

6. Fairness, Multi-Tenancy, and Policy Implications

Cloud optical networks are usually multi-tenant infrastructures. Bandwidth allocation affects not only efficiency but also distributive fairness. A policy that maximizes aggregate utility may systematically disadvantage tenants with bursty workloads, smaller contracts, or less ability to adapt. Fairness can be defined in multiple ways: equal shares, proportional shares, minimum guarantees, envy-free allocations, or priority according to social importance. Each definition implies different technical mechanisms and different policy commitments. Deep reinforcement learning can incorporate fairness penalties or constraints, but the choice of fairness objective is ultimately normative, not purely technical [19]. Engineers and policymakers must therefore collaborate on what fairness means in a given deployment.

Policy implications extend beyond individual networks. If cloud optical infrastructure becomes a critical dependency for public services, financial systems, and emergency response, then bandwidth allocation is a matter of public interest. Regulators may require transparency about service classes, outage reporting, and non-discriminatory access. Competition policy may examine whether dominant providers can use adaptive allocation to favor their own services or disadvantage rivals. Data protection rules may apply when telemetry is used for learning, especially if traffic patterns reveal sensitive information about users or organizations. These concerns argue for governance frameworks that treat learned allocation policies as auditable institutional decisions rather than private implementation details [19], [20].

7. Sustainability and Energy-Aware Operations

Energy consumption is an increasingly important constraint in optical and cloud infrastructure. Optical transmission can be more efficient than electronic switching over long distances, but the overall system includes transponders, amplifiers, cooling, and control overhead. Bandwidth allocation affects energy use because underutilized lightpaths may be kept active for resilience or latency reasons, while overprovisioning can waste power. A deep reinforcement learning agent could learn to consolidate traffic, power down unused capacity, or choose routes that balance latency and energy. However, energy-aware policies may conflict with resilience and fairness. Shutting down redundant capacity can save energy but reduce failure tolerance. Consolidating traffic can improve efficiency but create hotspots and increase risk [2], [17].

Sustainable operation therefore requires multi-objective reasoning. The reward function should not collapse all values into a single scalar without scrutiny. Operators may prefer constrained optimization, where energy is minimized subject to hard service guarantees, or Pareto-aware policies that expose trade-offs to human decision-makers. Measurement is also essential. Without accurate accounting of energy per service class, per tenant, or per allocation decision, sustainability claims remain weak. Deep reinforcement learning can support energy-aware bandwidth allocation, but only if the infrastructure exposes fine-grained telemetry and if governance defines which trade-offs are acceptable. Otherwise, efficiency gains may be achieved by shifting costs to resilience, fairness, or future capacity [18].

8. Observability, Human Oversight, and Institutional Governance

Deep reinforcement learning systems depend on observability. If the agent cannot see relevant network conditions, it cannot learn reliable policies. Yet full observability is rarely available in large optical networks because telemetry is sampled, delayed, or partitioned across domains. Improving observability requires investment in monitoring, time synchronization, data pipelines, and semantic models of network state. It also raises privacy and security concerns because detailed traffic telemetry can reveal sensitive behavior. A credible deployment must balance the need for learning with data minimization, access control, and auditability [16]. Observability is not merely a technical prerequisite; it is a governance resource that enables accountability.

Human oversight remains necessary even when automation performs well. Operators need the ability to inspect policies, understand why an action was taken, override decisions, and adjust objectives. Explanations may be approximate, but they must be useful for incident response and regulatory review. Human-in-the-loop control can take the form of approval workflows, safety constraints, periodic policy review, or adversarial testing. The goal is not to keep humans in every loop but to ensure that meaningful control exists where risk is high. Institutional governance should define roles, responsibilities, escalation paths, and redress mechanisms. It should also require periodic reassessment as traffic, infrastructure, and social expectations change [20]. Deep reinforcement learning can adapt quickly, but institutions must adapt deliberately.

9. Evaluation, Benchmarks, and Cross-Domain Lessons

Evaluation of deep reinforcement learning for bandwidth allocation cannot rely solely on simulated reward curves. It must include measures of service-level compliance, tail latency, fairness, resilience under failure, energy use, and operational burden. Benchmarks should represent realistic traffic mixes, multi-tenant behavior, adversarial conditions, and physical-layer constraints. They should also include strong baselines such as rule-based provisioning, optimization-based planning, and human-operated heuristics. A learned policy is only valuable if it improves meaningful outcomes without creating hidden costs [14], [13]. Reproducibility requires open or shared testbeds, documented assumptions, and careful reporting of training conditions.

Cross-domain lessons from data center traffic optimization, congestion control, and resource management suggest that learning works best when the action space is constrained and the reward is aligned with operational goals [10], [11]. In data centers, reinforcement learning has been used to optimize traffic and resource allocation, but production deployment often requires safety layers and fallback policies. In congestion control, learned policies can adapt to diverse network conditions but must avoid unfairness and instability. These lessons transfer

to optical bandwidth allocation: the most successful systems are likely to be hybrid, combining learning with optimization, control theory, and human expertise. Deep reinforcement learning should be treated as one component in a broader control architecture, not as a replacement for engineering judgment.

10. Conclusion

Deep reinforcement learning offers a promising path toward adaptive bandwidth allocation in cloud optical networks, but its value depends on how it is embedded in infrastructure and institutions. The technical promise is substantial: learned policies can respond to dynamic demand, anticipate congestion, manage spectrum fragmentation, and balance multiple objectives more flexibly than static rules. The risks are equally substantial: non-stationarity, reward misspecification, fragile generalization, unfair outcomes, energy trade-offs, and opaque decision-making. These risks cannot be solved by better neural network architectures alone. They require architectural choices that constrain authority, observability systems that support accountability, governance frameworks that define fairness and priority, and deployment practices that protect live services during learning.

The most credible future systems will be hybrid and layered. Fast local controllers may handle immediate allocation decisions, while slower global agents manage topology and capacity planning. Human operators will set objectives, approve high-risk actions, and audit outcomes. Physical-layer advances such as frequency-referenced multiplexing can strengthen performance guarantees, but they do not remove the need for policy and control. Sustainability, resilience, and fairness must be treated as first-class design constraints rather than afterthoughts. If these conditions are met, deep reinforcement learning can contribute to cloud optical networks that are not only more efficient but also more robust, equitable, and accountable. If they are ignored, the result may be optimized infrastructure that fails the social and operational tests that matter most.

References

1. Al-Fares, M., Loukissas, A., & Vahdat, A. (2008). A scalable, commodity data center network architecture. *ACM SIGCOMM Computer Communication Review*, 38(4), 63–74.
2. Kachris, C., & Tomkos, I. (2012). A survey on optical interconnects for data centers. *IEEE Communications Surveys & Tutorials*, 14(4), 1021–1036.
3. Farrington, N., Porter, G., Radhakrishnan, S., Bazzaz, H. H., Subramanya, V., Fainman, Y., Papen, G., & Ford, J. (2010). Helios: A hybrid electrical/optical switch architecture for modular data centers. *ACM SIGCOMM Computer Communication Review*, 40(4), 339–350.
4. Poutievski, L., Singh, A., Ong, J., Obenshain, D., Tariq, M., Hines, A., ... & Vahdat, A. (2022). Jupiter evolving: Transforming Google's datacenter network via optical circuit switches and software-defined networking. *Proceedings of the ACM SIGCOMM 2022 Conference*, 66–85.
5. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533.

6. Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2015). Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.
7. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.
8. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.
9. Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine, 34(6), 26–38.
10. Mao, H., Alizadeh, M., Menache, I., & Kandula, S. (2016). Resource management with deep reinforcement learning. Proceedings of the 15th ACM Workshop on Hot Topics in Networks, 50–56.
11. Jay, N., Rotman, N., Godfrey, P. B., Schapira, M., & Tamar, A. (2019). A deep reinforcement learning perspective on internet congestion control. Proceedings of the 36th International Conference on Machine Learning, 3050–3059.
12. Xu, Z., Tang, J., Yin, C., Wang, Y., & Xue, G. (2018). Experience-driven networking: A deep reinforcement learning based approach. IEEE INFOCOM 2018 - IEEE Conference on Computer Communications, 1871–1879.
13. Chen, L., Lingys, J., Chen, K., & Liu, F. (2018). AuTO: Scaling deep reinforcement learning for datacenter-scale automatic traffic optimization. Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication, 191–205.
14. Chen, X., Guo, J., Zhu, Z., Proietti, R., Castro, A., & Yoo, S. J. B. (2018). Deep-RMSA: A deep reinforcement learning framework for routing, modulation and spectrum assignment in elastic optical networks. Journal of Lightwave Technology, 36(20), 4895–4906.
15. Zhou, Z., Wei, J., Luo, Y., Clark, K. A., Sillekens, E., Deakin, C., ... & Liu, Z. (2023). Communications with guaranteed bandwidth and low latency using frequency-referenced multiplexing. Nature Electronics, 6(9), 694-702.
16. Boutaba, R., Salahuddin, M. A., Limam, N., Ayoubi, S., Shahriar, N., Estrada-Solano, F., & Caicedo, O. M. (2018). A comprehensive survey on machine learning for networking: Evolution, applications and research opportunities. Journal of Internet Services and Applications, 9(1), 16.
17. Thyagaturu, A. S., Mercian, A., McGarry, M. P., Reisslein, M., & Kellerer, W. (2016). Software defined optical networks (SDONs): A comprehensive survey. IEEE Communications Surveys & Tutorials, 18(4), 2738–2766.
18. Mijumbi, R., Serrat, J., Gorricho, J. L., Bouten, N., De Turck, F., & Boutaba, R. (2016). Network function virtualization: State-of-the-art and research challenges. IEEE Communications Surveys & Tutorials, 18(1), 236–262.
19. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.

20. Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A*, 376(2133), 20180080.