

Joint Optimization of Optical Network Latency and GPU Communication Efficiency in Distributed AI Training

Melchor Tello

Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
melchor553@buffalo.edu

Ditmar Striebitz

Department of Computer Science, University of New Hampshire, Durham, NH, USA.
ditmarstriebitz83@unh.edu

Kauko Tirkkonen

Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
ktirkkonen@colostate.edu

Timothee Sauvage

Department of Electrical Engineering and Computer Science, University of Missouri,
Columbia, MO, USA.
sauvage1985@missouri.edu

Abstract

Distributed artificial intelligence training has become a foundational workload for modern computing infrastructure, yet its performance is increasingly constrained not only by accelerator capacity but also by the latency and efficiency of communication across optical networks and graphics processing unit interconnects. This paper examines the joint optimization of optical network latency and graphics processing unit communication efficiency in distributed artificial intelligence training from a system-level perspective. It argues that latency and communication efficiency cannot be treated as independent engineering variables because choices in optical switching, topology, synchronization, collective communication, congestion control, and workload placement mutually shape one another. The analysis considers architectural trade-offs among optical circuit switching, wavelength switching, packet-based transport, and hybrid electrical-optical fabrics, with attention to clock synchronization, hollow core fiber, in-network aggregation, and collective communication scheduling. It further explores governance, fairness, sustainability, and policy implications, including energy proportionality, carbon accounting, equitable access to advanced network infrastructure, and resilience against optical and synchronization failures. Through cross-domain comparisons with high-performance computing, cloud data centers, and telecommunications, the paper identifies structural tensions between low-latency specialization and general-purpose flexibility. The conclusion emphasizes that joint optimization requires cross-layer telemetry, adaptive control, transparent benchmarking, and institutional coordination. Without such integrated approaches, distributed training systems risk over-investing in raw bandwidth while underutilizing accelerators, exacerbating energy use, and reproducing infrastructural inequities.

Keywords

optical networks; distributed AI training; GPU communication; latency optimization; collective communication; data center architecture; sustainability; governance.

1. Introduction

Distributed artificial intelligence training has moved from a niche parallel computing problem to a central workload of contemporary computational infrastructure. Early large-scale neural network training demonstrated that model quality and training time could be improved by distributing computation across many workers, but it also exposed communication as a persistent bottleneck [1]. Subsequent systems such as TensorFlow, pipeline-parallel training frameworks, and memory-optimized data-parallel methods have expanded the scale and complexity of distributed training, yet they continue to confront the same structural issue: accelerators compute quickly, while the movement of gradients, parameters, activations, and synchronization messages remains comparatively expensive [2], [3], [4]. Communication efficiency is therefore not a secondary concern but a primary determinant of achievable scale.

At the same time, optical networking has become indispensable to modern data centers because electrical interconnects face distance, bandwidth-density, and energy limitations at increasing scale. Optical interconnects offer high bandwidth, low loss, and potential for wavelength-based parallelism, but they also introduce switching latency, clock synchronization complexity, and control-plane overhead [10]. The interaction between optical behavior and graphics processing unit communication is particularly consequential in distributed training, where collective operations such as all-reduce, all-gather, and reduce-scatter must complete within tight windows to prevent accelerator idle time [5]. A network that is merely high-bandwidth may still perform poorly if its latency, jitter, or synchronization behavior disrupts collective scheduling.

This paper addresses the joint optimization of optical network latency and graphics processing unit communication efficiency in distributed artificial intelligence training. The central claim is that these two domains should be optimized together rather than sequentially. Optical path provisioning affects collective communication algorithms, which in turn affect congestion control, energy consumption, and workload placement. Conversely, the communication patterns of distributed training determine which optical switching regimes, topologies, and synchronization mechanisms are viable. A system-level approach must therefore consider architecture, governance, deployment, sustainability, robustness, fairness, and policy as interlocking dimensions rather than separate concerns.

2. Background and System Context

Distributed training systems typically combine several forms of parallelism. Data parallelism replicates the model across workers and exchanges gradients, while model parallelism partitions layers or tensors across devices. Pipeline parallelism assigns different stages of the model to different workers and streams micro-batches through them [3]. Hybrid approaches combine these strategies to balance memory, computation, and communication. The resulting traffic is not uniform: some phases require high-volume bulk transfer, others require low-latency synchronization, and still others require many-to-many collective exchange [4], [5]. Optical networks must support this heterogeneity without forcing every workload into a single traffic profile.

Classical parameter-server architectures and ring-based collective communication illustrate the trade-offs. Parameter servers centralize communication and can simplify consistency management, but they may become hotspots and introduce queuing delays [6]. Ring all-

reduce distributes bandwidth more evenly but can suffer from latency accumulation across many steps [5]. Compression and quantization reduce bandwidth demand but may add computation and affect convergence [7], [8]. These techniques are often evaluated in isolation, yet their effectiveness depends on the underlying network. A low-latency optical fabric can make synchronization-heavy patterns more attractive, while a high-latency or jittery fabric may favor compression, larger batch sizes, or topology-aware scheduling.

Data center optical networking provides both opportunity and constraint. Hybrid electrical-optical switch architectures such as Helios demonstrated the potential of reconfigurable optics for modular data centers, but they also revealed the complexity of integrating circuit switching with packet switching [11]. Surveys of optical interconnects for data centers emphasize that optical technologies can alleviate bandwidth and energy bottlenecks, yet practical deployment depends on switching speed, reliability, cost, and management [10]. The distributed training context adds another requirement: the network must not only deliver capacity but also preserve the timing relationships on which collective communication depends.

3. Latency in Optical Networks for Distributed AI Training

Latency in optical networks for distributed training arises from multiple sources. Propagation delay depends on fiber length and material properties, switching delay depends on the reconfiguration time of optical components, and synchronization delay depends on clock distribution and timing recovery. Queuing and congestion delay can occur even in high-bandwidth optical fabrics when traffic is bursty or when control-plane decisions lag behind workload changes [12]. In distributed training, these delays compound because collective operations are often synchronized: a slow participant can stall an entire group, and a brief optical reconfiguration can ripple into accelerator idle time.

Recent work on low-latency wavelength-switched clock-synchronized intra-data center interconnects enabled by hollow core fiber shows that physical-layer choices can materially reduce synchronization overhead and switching latency [14]. Hollow core fiber offers lower propagation delay than conventional silica fiber, while wavelength switching can provide reconfigurable connectivity without full packet-processing overhead. Clock synchronization is especially important because distributed training frequently relies on precise timing for collective operations, gradient aggregation, and pipeline scheduling. If optical paths introduce variable delay, the training framework may compensate with larger buffers or looser synchronization, which reduces effective parallelism.

The implications are architectural. A network designed primarily for average throughput may tolerate moderate latency and jitter, but a distributed training fabric must consider worst-case latency and tail behavior. Optical circuit switching can provide dedicated high-capacity paths, yet reconfiguration time may be too slow for fine-grained collective operations. Wavelength switching can improve flexibility, but it requires tunable transceivers and careful wavelength management. Packet-based optical or hybrid approaches can handle bursty traffic but may reintroduce electrical buffering and congestion. The joint optimization problem is therefore not simply about minimizing one latency metric; it is about matching latency distributions to communication phases.

4. GPU Communication Efficiency and Collective Operations

Graphics processing unit communication efficiency depends on the interplay of intra-node and inter-node interconnects. Within a node, high-speed links such as NVLink allow

accelerators to exchange data quickly, while inter-node communication relies on network interface cards, remote direct memory access, and collective communication libraries [5]. The efficiency of collective operations is determined by algorithm choice, message size, topology, and the ability to overlap communication with computation. If optical latency is low but collective algorithms are poorly mapped to the optical topology, efficiency can remain low. Conversely, if optical latency is high, even well-optimized collectives may leave accelerators waiting.

Communication efficiency is also shaped by compression, aggregation, and scheduling. Quantization and gradient compression reduce the volume of data exchanged, but they introduce encoding and decoding costs and can affect model convergence [7], [8]. In-network aggregation can reduce host and accelerator involvement, but it requires programmable switches or smart network interfaces that may not be available in all optical architectures. Rack-scale parameter servers attempt to exploit locality and high-bandwidth interconnects to improve aggregation [15]. These approaches suggest that GPU communication efficiency is not solely a device-level property; it is a property of the entire communication path, including optical transport, switching, memory movement, and software orchestration.

The fairness dimension is often overlooked. If only some jobs can access low-latency optical paths or dedicated wavelengths, then communication efficiency becomes a source of performance inequality. Large-scale training jobs may monopolize reconfigurable optical resources, while smaller or exploratory workloads experience higher latency and lower accelerator utilization. A joint optimization framework must therefore include allocation policies that balance aggregate throughput against per-job latency and fairness. Otherwise, optical and GPU resources may be optimized for a narrow class of very large models while degrading service for broader research communities.

5. Joint Optimization Framework

A joint optimization framework for optical latency and GPU communication efficiency requires cross-layer visibility. Telemetry must span optical power, wavelength occupancy, switch reconfiguration time, link latency, congestion signals, collective operation progress, and accelerator utilization [9]. Without such visibility, оптимизация becomes reactive: schedulers tune one layer while creating bottlenecks in another. For example, a scheduler might assign a communication-heavy job to a path with spare bandwidth, unaware that the path has high jitter that will disrupt synchronization. Conversely, a latency-sensitive job might be placed on a low-latency path that lacks sufficient bandwidth for bulk gradient exchange.

The framework should treat latency, bandwidth, energy, and fairness as simultaneous objectives. In practice, this means using multi-objective control policies that can adapt to workload phase. During gradient exchange, bandwidth and congestion control may dominate; during synchronization or pipeline handoff, latency and jitter may dominate [12]. During periods of low utilization, energy proportionality may justify consolidating traffic onto fewer optical paths and powering down unused components. The control plane must coordinate these transitions without causing oscillations or instability. This is a governance as well as an engineering problem because different tenants, operators, and institutions may have different priorities.

A joint framework also needs benchmarking that reflects real distributed training patterns rather than synthetic point-to-point measurements. Scientific benchmarking of parallel

systems has long emphasized reproducible, meaningful performance reporting [9]. For optical-AI systems, benchmarks should include collective operation latency, tail latency, reconfiguration impact, accelerator idle time, energy per useful training step, and fairness across job classes. Such benchmarks would allow operators to compare architectural choices and policymakers to assess infrastructure investments. They would also reduce the risk of optimizing for misleading aggregate metrics that hide poor behavior under realistic synchronization and failure conditions.

6. Architecture and Deployment Considerations

Deployment choices determine whether joint optimization is feasible. A data center may adopt a fully optical fabric, a hybrid electrical-optical topology, or an overlay of optical circuit switches for elephant flows. Each choice affects the granularity of control and the types of collective communication that perform well. Hybrid architectures can preserve packet-switched flexibility while using optical paths for long-lived, high-volume transfers [11]. However, hybrid designs require careful traffic classification and fast switching between domains. If optical reconfiguration is slow, it may only benefit large transfers and fail to help latency-sensitive synchronization.

Physical infrastructure matters as well. The use of advanced fiber and wavelength switching can reduce latency, but it may require new transceivers, amplifiers, and monitoring equipment [13], [14]. Data center operators must consider backwards compatibility, standardization, maintenance, and fault management. Disaggregated architectures, in which compute, memory, and accelerators are separated, create new opportunities for optical networking but also new dependencies. If a training job spans disaggregated resources, the optical path becomes part of the critical path for every memory access and collective operation. This increases the importance of deterministic latency and robust synchronization.

Governance and policy enter through procurement, standards, and access. National research clouds, university clusters, and private hyperscale facilities may have different missions, yet they increasingly compete for similar optical and accelerator technologies. If low-latency optical infrastructure is concentrated in a few commercial providers, public-interest research may face higher costs and limited experimentation. Deployment should therefore include open interfaces, shared testbeds, and documentation of performance under realistic AI workloads. Sustainability also affects deployment: energy-efficient optical components can reduce operational carbon, but the embodied carbon of specialized hardware and the energy cost of cooling and reconfiguration must be accounted for [20].

7. Governance, Fairness, Sustainability, and Policy

The joint optimization of optical latency and GPU communication efficiency has significant governance implications. Infrastructure decisions determine who can train large models, how quickly, and at what environmental cost. Energy and policy considerations for deep learning have shown that training can have substantial carbon footprints and that efficiency improvements do not automatically reduce total impact if usage grows [16]. Carbon emissions studies of large neural network training further highlight the importance of location, hardware, and scheduling in determining environmental outcomes [17]. Optical networks can improve energy efficiency by reducing electrical switching and transmission losses, but they can also enable larger and more frequent training runs, creating rebound effects.

Sustainability requires metrics that go beyond peak throughput. Green AI research argues for measuring energy and carbon alongside accuracy and performance [18]. For optical-AI

systems, this means tracking energy per collective operation, energy per gradient exchange, and carbon intensity of the power supply during training. Workload scheduling can shift flexible training jobs to times and locations with cleaner energy, but doing so may increase latency or reduce GPU utilization. Joint optimization must therefore expose these trade-offs to operators and policymakers rather than hiding them in aggregate efficiency claims.

Fairness is equally important. Access to low-latency optical paths can become a competitive advantage for well-funded laboratories and firms. If reconfigurable optical resources are allocated purely by bid or job size, smaller institutions may be systematically disadvantaged. Governance mechanisms could include reserved capacity for public research, transparent latency and utilization reporting, and tiered service classes that balance large-scale training with broad access. Ethical concerns about large models, including concentration of power and environmental externalities, are relevant here [19]. Optical and GPU communication efficiency should not be treated as neutral technical goals; they shape the distribution of scientific and economic capability.

8. Robustness, Failure Modes, and Resilience

Distributed training systems are sensitive to failures and stragglers. An optical link failure, wavelength misconfiguration, or clock synchronization loss can disrupt collective operations across many accelerators. Recovery may require checkpointing, job restart, or dynamic rerouting. If the network control plane is not integrated with the training scheduler, recovery can be slow and expensive [6]. Resilience should therefore be designed into the joint optimization framework from the beginning. This includes redundant optical paths, graceful degradation to electrical switching, adaptive collective algorithms, and health monitoring that can distinguish transient congestion from persistent faults.

Tail latency is a particular concern. Even if average optical latency is low, rare high-latency events can stall synchronous training. In large jobs, the probability of such events increases because many participants must coordinate. Congestion control mechanisms designed for data centers, such as delay-based approaches, can help but may interact poorly with optical reconfiguration [12]. A robust system must anticipate these interactions and provide fallback strategies. For example, if an optical path becomes unstable, the scheduler might switch to a compressed communication mode or reduce the degree of synchronization. These adaptations should be automated and governed by policies that consider fairness and cost.

Security is another dimension. Optical networks can be difficult to monitor at fine granularity, and compromised control planes could misroute traffic or disrupt synchronization. Distributed training also involves valuable models and data, making confidentiality and integrity important. Joint optimization should therefore include observability, authentication, and policy enforcement across optical and compute domains. Resilience is not only about avoiding downtime; it is about maintaining trustworthy and equitable service under stress.

9. Case Illustrations and Cross-Domain Comparisons

Hyperscale AI clusters illustrate the benefits and limits of joint optimization. These facilities can deploy custom optical fabrics, high-bandwidth accelerators, and sophisticated schedulers. They may achieve high GPU utilization for large jobs, but they also concentrate resources and energy use. National research clouds and university consortia face different constraints: they may have heterogeneous hardware, shared governance, and limited ability to customize optical infrastructure. Joint optimization in these settings must prioritize interoperability, fairness, and reproducible benchmarking over peak performance alone [4], [9].

High-performance computing offers a useful comparison. Supercomputing systems have long used low-latency interconnects and topology-aware collective algorithms. However, AI training workloads differ in their communication patterns, failure tolerance, and software stacks. Techniques from HPC can inform optical-AI design, but they must be adapted to the scale and variability of AI workloads [3], [10]. Cloud data centers provide another comparison. They emphasize multi-tenancy, elasticity, and cost efficiency, which can conflict with the dedicated low-latency paths needed for tightly synchronized training [11], [13]. Telecommunications networks contribute expertise in wavelength switching, synchronization, and fault management, but they operate under different traffic and regulatory regimes.

These comparisons suggest that no single architecture is universally optimal. A research cluster might benefit from hybrid optical circuit switching and packet switching with strong fairness policies. A hyperscale provider might use wavelength-switched paths for bulk transfers and electrical switching for latency-critical control. A federated or edge-based training system might prioritize robustness and privacy over raw latency. Joint optimization must therefore be context-sensitive rather than prescriptive.

10. Future Research Directions

Future research should develop cross-layer controllers that can learn from telemetry and adapt in real time. Machine learning for optical control is promising, but it must be explainable and safe because misconfiguration can disrupt large training jobs. Digital twins of optical-AI infrastructure could allow operators to test scheduling and reconfiguration policies before deployment. Standardized interfaces between training frameworks and optical control planes would reduce integration costs and improve portability. Benchmarking should include tail latency, fairness, and carbon metrics, not only bandwidth and throughput [9], [18].

Another direction is the co-design of collective communication algorithms with optical topology. If the network can reconfigure quickly, collective algorithms might change their structure dynamically. If reconfiguration is slow, algorithms should be designed to minimize switching. Compression and aggregation techniques could be integrated with optical path selection [7], [8]. Sustainability research should examine rebound effects and develop policies that align efficiency gains with absolute reductions in energy and carbon [17], [20]. Finally, governance research should explore how to allocate low-latency optical resources fairly across public and private stakeholders, ensuring that advances in distributed AI training do not deepen existing inequalities [19].

11. Conclusion

The joint optimization of optical network latency and graphics processing unit communication efficiency is essential for the next generation of distributed artificial intelligence training. Optical networks offer the bandwidth and reconfigurability needed to connect thousands of accelerators, but their latency, synchronization, and control-plane behavior directly affect collective communication and accelerator utilization. Conversely, the communication patterns of distributed training determine which optical architectures are practical. Treating these domains separately leads to fragmented optimization, underutilized hardware, excessive energy consumption, and inequitable access to advanced infrastructure.

A system-level approach must integrate architecture, telemetry, scheduling, governance, sustainability, robustness, and fairness. It should use realistic benchmarks, support adaptive control, and account for tail latency and failure modes. It should also recognize that infrastructure choices have policy consequences, shaping who can participate in large-scale

AI research and how environmental costs are distributed. The path forward requires collaboration among computer scientists, optical engineers, data center operators, policymakers, and research communities. By optimizing optical latency and GPU communication efficiency together, distributed AI training can become more performant, resilient, and socially accountable.

References

1. Dean, J., Corrado, G., Monga, R., Chen, K., Devin, M., Mao, M., Ranzato, M., Senior, A., Tucker, P., Yang, K., Le, Q., & Ng, A. Y. (2012). Large scale distributed deep networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 25, pp. 1223–1231). Curran Associates.
2. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., ... Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation* (pp. 265–283). USENIX Association.
3. Narayanan, D., Harlap, A., Phanishayee, A., Seshadri, V., Devanur, N. R., Ganger, G. R., Gibbons, P. B., & Zaharia, M. (2019). PipeDream: Generalized pipeline parallelism for DNN training. In *Proceedings of the 27th ACM Symposium on Operating Systems Principles* (pp. 1–15). ACM. <https://doi.org/10.1145/3341301.3359646>
4. Rajbhandari, S., Rasley, J., Ruwase, O., & He, Y. (2020). ZeRO: Memory optimizations toward training trillion parameter models. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (pp. 1–16). IEEE. <https://doi.org/10.1109/SC41405.2020.00024>
5. Sergeev, A., & Del Balso, M. (2018). Horovod: Fast and easy distributed deep learning in TensorFlow. arXiv. <https://arxiv.org/abs/1802.05799>
6. Li, M., Andersen, D. G., Park, J. W., Smola, A. J., Ahmed, A., Josifovski, V., Long, J., Shekita, E. J., & Su, B.-Y. (2014). Scaling distributed machine learning with the parameter server. In *Proceedings of the 11th USENIX Symposium on Operating Systems Design and Implementation* (pp. 583–598). USENIX Association.
7. Alistarh, D., Grubic, D., Li, J., Tomioka, R., & Vojnovic, M. (2017). QSGD: Communication-efficient SGD via gradient quantization and encoding. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30, pp. 1709–1720). Curran Associates.
8. Lin, Y., Han, S., Mao, H., Wang, Y., & Dally, W. J. (2018). Deep gradient compression: Reducing the communication bandwidth for distributed training. In *International Conference on Learning Representations*. OpenReview.
9. Hoefler, T., & Belli, R. (2015). Scientific benchmarking of parallel computing systems: Twelve ways to tell the masses when reporting performance results. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (pp. 1–12). ACM. <https://doi.org/10.1145/2807591.2807644>

10. 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>
11. 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. In *Proceedings of the ACM SIGCOMM 2010 Conference* (pp. 339–350). ACM. <https://doi.org/10.1145/1851182.1851223>
12. Mittal, R., Lam, V. T., Dukkupati, N., Blem, E., Wassel, H., Ghobadi, M., Vahdat, A., Wang, Y., Wetherall, D., & Zats, D. (2015). TIMELY: RTT-based congestion control for the datacenter. *ACM SIGCOMM Computer Communication Review*, 45(4), 537–550.
<https://doi.org/10.1145/2829988.2787510>
13. Cheng, Q., Bahadori, M., Glick, M., Rumley, S., & Bergman, K. (2018). Recent advances in optical technologies for data centers: A review. *Optica*, 5(11), 1354–1370.
<https://doi.org/10.1364/OPTICA.5.001354>
14. 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.
15. Luo, L., Nelson, J., Ceze, L., Phanishayee, A., & Krishnamurthy, A. (2020). Parameter hub: A rack-scale parameter server for distributed deep neural network training. In *Proceedings of the ACM Symposium on Cloud Computing* (pp. 41–54). ACM.
<https://doi.org/10.1145/3419111.3421280>
16. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3645–3650). ACL. <https://doi.org/10.18653/v1/P19-1355>
17. 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>
18. Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54–63. <https://doi.org/10.1145/3381831>
19. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). ACM.
<https://doi.org/10.1145/3442188.3445922>
20. Masanet, E., Shehabi, A., Lei, N., Smith, S., & Koomey, J. (2020). Recalibrating global data center energy-use estimates. *Science*, 367(6481), 984–986.
<https://doi.org/10.1126/science.aba3758>