DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Reference herein to any social initiative (including but not limited to Diversity, Equity, and Inclusion (DEI); Community Benefits Plans (CBP); Justice 40; etc.) is made by the Author independent of any current requirement by the United States Government and does not constitute or imply endorsement, recommendation, or support by the United States Government or any agency thereof.

University of California, San Diego

Final Scientific/Technical Report

LEED: A Lightwave Energy-Efficient Datacenter

DE-FOA-0001566

Award:	DE-AR0000845
Lead Recipient:	University of California, San Diego
Project Title:	LEED: A Lightwave Energy-Efficient Datacenter
Program Director:	Laurent Pilon
Principal Investigator:	Dr. George Porter
Contract Administrator:	Thomas Bernal
Date of Report:	June 1, 2024
Reporting Period:	August 7, 2017 – March 31, 2024

Public Executive Summary

The Lightwave Energy-Efficient Datacenter (LEED) program is a disruptive "green-field" approach that provides a quantum leap in the energy efficiency of datacenters. LEED's fundamental value proposition is that a novel and re-architected optical network—RotorNet—can deliver "more bandwidth per buck" as well as unique system-level attributes that significantly improve overall datacenter energy efficiency and performance. LEED has developed three system-level testbeds. The first testbed uses calibrated hardware and software power measurements to determine server energy efficiency as a function of network bandwidth and workload. These measurements have shown that increasing network communications bandwidth dramatically increases server energy efficiency providing a realistic path to the overall ENLITENED program goal of doubling the number of transactions per joule. The second testbed demonstrates key hardware: a prototype low-loss, high-port count optical "selector switch". This switch was fabricated, racked, and tested. Measured switch characteristics include loss, bandwidth, crosstalk, switch time, system-level switch time (including the transceivers), and bit error rate. The third testbed demonstrates a fully working and manufactured pinwheel design which dramatically lowers the cost of design, while delivering high switch radix and low reconfiguration times.

The LEED project has tied these three novel photonic switch prototypes together with production servers and software through the development of a novel FPGA-based NIC platform called Corundum. Corundum ensures that the packet-switched protocols supported by commodity operating systems and devices can interface with the Rotor switch design. The LEED group has used this combined hardware and software prototype to characterize applications running at a commercially relevant scale. The project has used a combination of enhanced optical modulation amplitude (OMA) modulators, broadband multiplexers and demultiplexers, avalanche photodiodes, and a novel burst-mode receivers to enable the insertion of LEED-developed optical switches without the need for expensive optical amplification. Our modeling has shown that measured LEED-developed device characteristics can achieve link characteristics of 2 pJ/bit including both transceivers and the Rotor switch.

In summary, the LEED program has demonstrated a credible and practical path, through novel hardware and software, to realize the program objectives of ENLITENED. The net result will ensure that the United States maintains its strength in the crucial sector of Information Technology, which is vital to both our economic security and our national security.

Acknowledgements

We would like to acknowledge support from ARPA-E via the ENLITENED program (DE-FOA-0001566). Support for this project was provided via cost-share by the California Energy Commission ("LEED – a Lightwave Energy Efficient Data Center", EPC-17-051).

Table of Contents

Table of Figures/Tables

Accomplishments and Objectives.

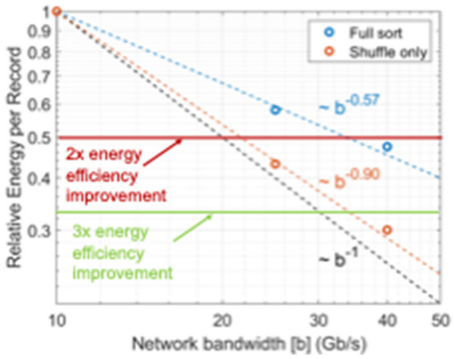
This award allowed the University of California, San Diego to demonstrate several key objectives. The focus of the project was on building a novel phonic interconnect for datacenters and high-performance computing clusters that can deliver data with high bandwidth, low-latency, and a 2x increase in application-level operations per Joule of energy.

A few tasks and milestones were laid out in Attachment 3, the Technical Milestones and Deliverables, at the beginning of the project. The actual performance against the stated milestones is summarized here:

Table 1: *Key Milestones and Deliverables. Tasks are identified as "P1" for Phase 1 and "P2" for Phase 2 of the project.*

Tasks	Milestones and Deliverables
Task P1.1: Development of work plan	*Q1: Development of interim/lower-level tasks and milestones.* **Actual Performance:** (Completed: Q1) The workplan documented in the Gantt chart in the Phase 1 proposal was finalized and approved. The Phase 1 SOPO was also approved.
Task P1.2: Technology to Market 2.1 Techno-Economic Analysis (TEA) 2.2 Pursue Next-Stage Development Opportunities	*(Note: Because of the start date of the contract, there is an approximate six week offset between the reporting cycle and the technical milestone cycle for Phase 1.)* *Q2: Framework of TEA accepted by ARPA-E.* **Actual Performance:** (Completed: Q3) TEA framework document was completed and accepted by ARPA-E *Q3: Analysis of primary market.* **Actual Performance:** (Completed: Q4) Analysis of the primary market was documented in the Technology to Market (T2M) Plan, which was completed and submitted for review on January 22, 2018. *Q4: Review partnership opportunities.* **Actual Performance:** (Completed: Q9)

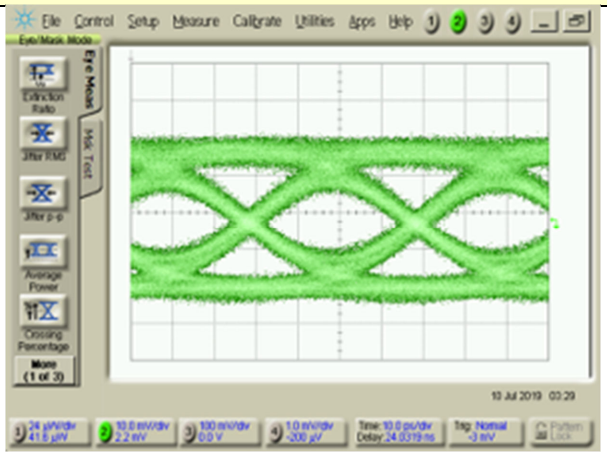
Tasks	Milestones and Deliverables
	Our review of the partnership opportunities was documented in our Technology to Market (T2M) Plan, which was completed and submitted on January 22, 2021.
Task P1.3: LEED Architecture 3.1 Simulation of LEED 3.2 Implementation of Valiant Load Balancing 3.3 Alternative architectures	*Q4: Initial energy efficiency* **Actual Performance:** (Completed: Q5) The initial estimates of the energy efficiency started in Q3 and were completed by Q5. As an example, we conducted energy efficiency measurements on the Comet supercomputer cluster at the NSF-sponsored San Diego Supercomputer center. These measurements are described in detail in the Q3 Phase 1 Progress report. Two important conclusions about the nature of server energy efficiency can be drawn from energy efficiency results. The first is that there is about a 30% variation in the energy-efficiency of skewed load compared to a uniform load. This variation is why several canonical workloads and hardware platforms may be required for meaningful energy-efficiency characterization. The second important conclusion is that CPU utilization is time-varying even over a single application. Moreover, even with careful optimization, the CPU utilization did not exceed 60% for our experiments. This result showed that if a cost-comparable system based on optical switching that can lead to more "bandwidth per buck" can be developed, it can improve the server efficiency and thus the overall energy efficiency of a data center. *Q5: Down select of architectures.* **Actual Performance:** (Completed: Q9) We considered two versions of optically switched architectures for LEED. One version connected the optical circuit switch to the top of rack (ToR) switches, which is representative of common data center networks. The second version connected the optical circuit switch

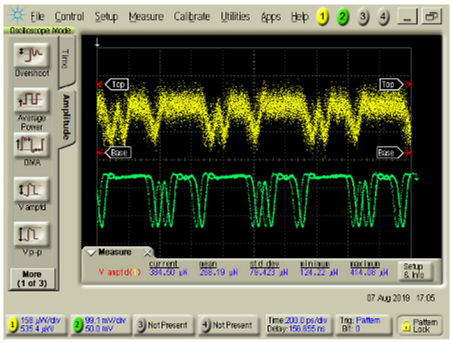
Tasks	Milestones and Deliverables
	directly to the servers. Consideration of our ability to synchronize the packets within a ToR led us to down select an architecture with the optical circuit switch connected directly to the servers. This was the architecture used throughout the program. *Q8: Validation of LEED architecture in testbed.* **Actual Performance:** (95% completion in Phase 1: Q12 – completed in Phase 2) The improvement in server energy efficiency that can be realized by RotorNet was assessed using an open-source measurement methodology and a combination of calibrated hardware power measurements and software tools on the Phase-1 testbed configured with a standard packet switch. The results are shown in the figure below using a sort workload and a shuffle workload between end hosts using a baseline network of 10 Gb/s. The black dashed line is an ideal network for which the energy expended in each transaction (or record for the case of a sort application) is inversely proportional to the bandwidth. This simply means that doubling the bandwidth halves the energy expended per record.  **Figure 1:** *Measured energy efficiency for several applications using a standard packet switch.* *Figure 1 shows* the measured energy efficiency for several applications using a standard packet switch. This data shows that actual measurements of applications have "energy-efficiency slopes" that are not as steep as the ideal black dashed line. For

Tasks	Milestones and Deliverables
	example, the shuffle application as a slope of $b^{-0.9}$ instead of b^{-1}. This flatter slope means that it takes slightly more than twice the bandwidth to double the energy efficiency. For the Sort application, it takes about 3.5x the bandwidth to double the energy efficiency. Even at the low end of this range, the datacenter energy-efficiency improvement is over twice the energy expended in the entire communications network under the reasonable assumption that existing networks expend about 10% of the total energy in a datacenter. These measured results confirm the fundamental working premise of LEED that greater (cost-comparable) network bandwidth leads directly to higher energy efficiency.
Task P1.4: Switch The objective of this project element is to develop high-speed, low-loss switches for use in LEED.	*Q2: Critical design review of upgraded selector switch* **Actual Performance:** (Completed: Q3) As described in the Phase 1 Q2 report and presented in the LEED Review at UCSD on February 28, 2018, LEED made a strategic choice to investigate a modification of the original selector switch from a MEMS tilt-mirror to rotating blazed diffraction grating ``pinwheel''. Refinements to the network architecture from random access to sequential access switching. The new “Rotornet" architecture imposed no network penalty but enabled the simplification of the switch actuator. Details of the modified design were presented at the February 28, 2018 review. *Q4: Selector switch component fabrication and test* **Actual Performance:** (Completed: Q5) The original ``pinwheel” design was based on a rectilinear grating structure with the grating features being invariant in one direction. In the new design, we switched to using a radially symmetric grating. This structure is a conformally-mapped rectilinear grating structure that produces a low loss over a larger sector area compared to the previous rectilinear grating design. Compared to a free-space MEMs switch, the use of pinwheel switch increases the maximum angle of deflection and leads to a 7x reduction in the optical track. The resulting design can be fit within a standard rack of a datacenter. The

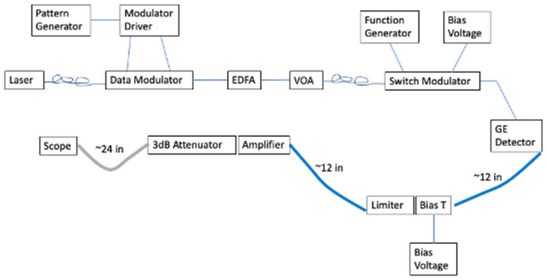
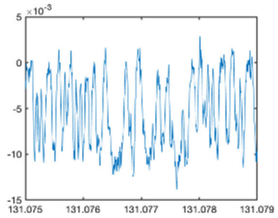
Tasks	Milestones and Deliverables
	total modeled loss through one pass of the switch is about 2.5 dB and is flat over a bandwidth of 100 nm. *Q6: Selector switch full integration.* **Actual Performance:** (Completed: Q9) The assembly of the first Phase 1 Rotor Switch started in December 2018. Initially, we had estimated that all the custom components should arrive by August 2018. The approximately two-month delay was due to several issues associated with the vendors. The vendor responsible for lens fabrication originally quoted a 12-week delivery time, but the actual time will be closer to 24 weeks. The original vendor for the grating pinwheel fabrication quoted a 4-week lead time but ended up not being able to deliver the part after 16 weeks of effort. We had to identify and change to a new vendor, which delayed delivery of the part. The integration of Rotor switch with passive optics in portable breadboard system was achieved in early 2019. This breadboard version of the switch was then "ruggedized" to enable racking into a standard 6U enclosure. The operation of the switch requires two passes through the switch. The two-pass loss for the second-generation pinwheel improved by about 3 dB across the four network configurations and ranged from 5-8 dB. This means that the worst performing channel of the second-generation pinwheel has about the same performance as the best performing channel of the first-generation pinwheel. The spectral response and crosstalk for one channel of the second-generation pinwheel is flat over 100 nm with less than −30 dB of crosstalk over the entire range. These specifications satisfy the 35 nm requirement of the milestone with very low crosstalk. The measured switching speed of the Rotor switch of 25 μs satisfies the milestone specification of 75 μs.

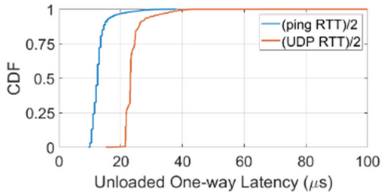
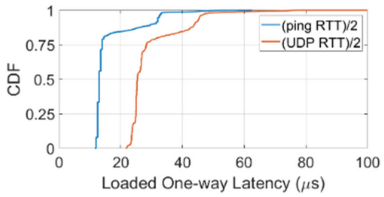
Tasks	Milestones and Deliverables
Task P1.5: Interconnects 5.1 Burst mode 5.2 Integrated APD 5.3 Broadband MUX/DEMUX 5.4 25 Gb/s Modular array w/ closed-loop control 5.5 Packaging of components	*Q3: CMOS tapeout* **Actual Performance:** (Completed: Q5) On July 18, 2018 (which is during the administrative reporting period of Q5), Axalume taped out a place-holder design of a burst-mode receiver which was reviewed by TSMC for pdk and metal stack compatibility prior to the final submission on Aug. 12, 2018. On Sept 15, 2018, masks were ordered and the layout is Axalume's top-level metal layout is designed in collaboration with Sandia Labs to be flip-chip-compatible with PIN and APD arrays designed by Sandia team. The top-level metal layout also supports bond-wire test and verification. These tasks completed the milestone. *Q6: APD demonstration* **Actual Performance:** (Completed: Q9) The key purpose using an APD is to increase the link margin to enable the insertion of an optical switch in the link path, enabling the 2X system energy efficiency goal of the program. To this end, two different avalanche photodiode structures were investigated. These two structures are referred to as lateral and vertical APD structures and initially discussed in the Q2 report. Both structures are based on separate absorption and carrier multiplication (SACM) regions. Initial current-voltage (I-V) curves taken at the wafer-level indicated the expected behavior for both horizontal and vertical APDs. In early July 2019, the first eye-pattern at 25 Gb/s was obtained at a responsivity of 3 A/W sustainably meeting the milestone. This eye pattern is shown in Figure 2.

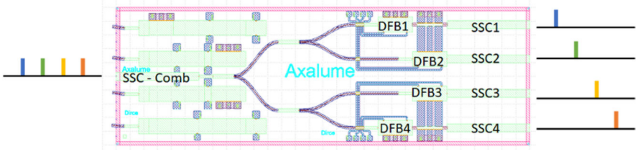
Tasks	Milestones and Deliverables
	 **Figure 2:** *Eye pattern at 25 Gb/s with a responsivity of 3 A/W.* *Q6: Measurement of loss* **Actual Performance:** (Completed: Q9) In support of Milestone M5.3, we fabricated a mux/demux that has successfully met the target specifications determined by our link-level modeling to achieve a 2 pJ/bit link. This mux/demux design is based on a 4-port apodized Bragg filter design. We conducted measurements of loss for the first-generation mux/demux channels and found it to be nearly the 1 dB set target for the milestone. Moreover, to achieve 2 pJ/bit requires less than -18 dB of crosstalk between channels. The design achieves this goal within the O band as 20 dB to 25 dB channel rejection. *Q8: Demonstration of eight channel modulator array* **Actual Performance:** (Completed: Q9) The goal of this work was to design and fabricate a fully modulated 8-channel modulator array with closed-loop control on all eight channels. Our design uses disc modulators and the vertical p-n junction used in depletion mode, which is a competency of the fabrication process at Sandia's Silicon Photonics platform. Conventional disc (resonant) modulators perform better at lower optical powers, since two-photon absorption and free-carrier absorption limit the amount of input power that can be handled by the micro-resonator. Measurements in this project have confirmed and quantified the optical modulation

Tasks	Milestones and Deliverables
	amplitude (OMA) closure at increasing power levels from baseline (unoptimized) test structures, measured at 1550 nm. Test structures were fabricated with 8-disc modulators coupled to a bus waveguide, which are suitable for developing the 8-channel controller. The targeted performance corresponds to a factor of 2x improvement over the highest-reported OMA of disc modulators in the fabrication process we used. The response to a 25 Gb/s data pattern is shown in Figure 3.  **Figure 3:** *The response of the modulator array at 25 Gb/s.* *Q8: Packaged, high-speed burst mode receiver* **Actual Performance:** (Completed: Q9) As discussed in the Q7-Q9 reports, testing of the burst-mode receiver (BMR) started in Q7 and was completed by Q9. Critical blocks were validated experimentally and/or with post-layout simulation to Phase 1 specifications. These parameters include a DC acquisition time of less than 25ns, clock-recovery: time of less than 75 ns, clock speed: 6.25 GHz (suitable for 25 Gbps data input with a 4:1 demux) and an injection locking range greater than 100 ppm. The CMOS receiver was proven functional with power (~1pJ/bit) , bandwidth (25Gbps), and lock-times (40ns) meeting or exceeding Phase 1 specifications. A US patent was issued to Axalume for the optical receiver.

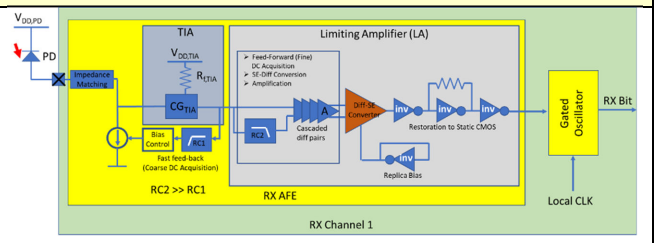
Tasks	Milestones and Deliverables
Task P2: Phase 2 P2.1: A unique energy-efficient and scalable optical circuit-switched architecture P2.2: A manufacturable low-loss, scalable, rate-agnostic optical switch P2.3: The development and demonstration of a photonic integrated chip for a transmitter (Tx) and a photonic integrated receiver chip (Rx)	*Q13: Go-No/Go Milestone* **Actual Performance:** (Completed: Q12) Tasks and milestones for the work plan were refined to satisfy the milestone. *Q14: Down select of switch architecture* **Actual Performance:** (Completed: Q13) Initial characterization of single mode vs. multimode transceivers were conducted along with preliminary analysis of a single mode vs. a multimode switch. Based on these results, we down selected the single mode switch design for the Phase 2 Rotor switch. *Q14: Critical design review of Integrated TX* **Actual Performance:** (Completed: Q18) This design review included several potential approaches for the four WDM laser sources and compatible modulator array based on Axalume's previous Phase 1 work. Axalume's integrated transmitter and modulator tape-outs to support this milestone included the directly modulated laser array (Dirce) described in this report, the Admiral integrated comb source and modulator array also described in this report and two other chips that leveraged LEED funds that were delivered to cloud customers as part of a separate photonic integrated circuit contract. The review of these designs completed the milestone. Details of the two chips (Dirce and Admiral) funded by LEED are provided in other sections of this report. *Q14: Initial benchmarking methodology established.* **Actual Performance:** (Completed: Q15) The original goal of this milestone was mostly incorporated into the Phase 2 work of the STEAM team. To support their effort and complete the initial benchmarking methodology milestone, we worked with the STEAM team and provided simulation code to model the performance of our proposed network architecture. *Q15: Burst-mode operation of avalanche photodiode (APD)*

Tasks	Milestones and Deliverables
	Actual Performance: (Completed: Q16) The experimental setup to demonstrate the burst-mode operation of the APD is shown in Figure 4.  **Figure 4:** *Test set-up for the measurements of the APD.* Measurements were conducted using a real-time scope that could generate 256 Gsamples per second, which is 10 samples per bit interval to evaluate the recovery time of the APD. This proved to be useful for finding problems with the set up. The other advantage of the real time scope is that it allows us to investigate a suite of variable decision threshold techniques and equalization by postprocessing the data. Sample data at 25 Gb/s is shown in Figure 5. We post-processed the data from the sampling scope to estimate the bit error rate (BER) completing the milestone. More details are in the Phase 2 Q2 report.  **Figure 5:** *Data pattern from APD at 25 Gb/s.* *Q16: Implementation of Opera in Phase 1 testbed* **Actual Performance:** (Completed: Q22)

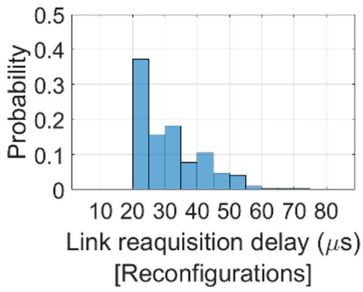
Tasks	Milestones and Deliverables
	The LEED Opera protocol consists of two main components. The first supports low-latency packet transmission, and the second supports bulk data transfers. We have completed the implementation of the low-latency protocol and have tied that into the Linux network stack via our FPGA-based NIC. Within the LEED network interface, the Opera protocol is implemented as a series of logical components that carry out the required tasks of the datapath. We have evaluated the latency of packet transmissions with and without the presence of cross traffic. Figure 6 shows the measured latency on an unloaded network. Figure 7 shows the latency on a loaded network. The "kink" is caused by the presence of cross traffic on the loaded network.  **Figure 6:** *Opera latency on an unloaded network.*  **Figure 7:** *Opera latency on a loaded network with bulk traffic overlapping by 15%.* *Q16: Critical design review of new rotor switch* **Actual Performance:** (Completed: Q15) Most of the full optical design of the Phase II rotor switch was completed in Phase 2 Q2, satisfying all design performance targets in

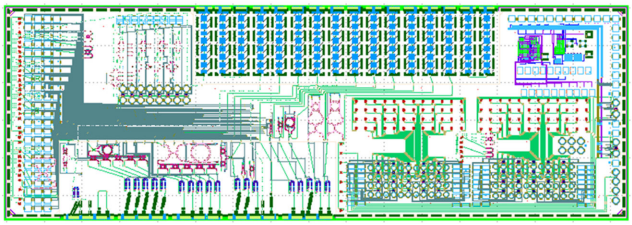
Tasks	Milestones and Deliverables
	simulation. In Q3, we worked with the various vendors to secure quotes for all custom components and with UCSD to issue purchase orders to vendors. *Q16: Tape-our source array* **Actual Performance:** (Completed: Q18) Axalume taped out the first comb laser array designed for C-band operation, codenamed "Dirce". The Dirce is a III/V 4 x DFB covering the wavelength range of 1535.82nm to 1545.32nm with spacing of 3.16nm or ~400GHz with a tunability of 3nm per DFB. The total expected output power is 10mW at a bias current of 140mA. Figure 8 shows the layout of the Dirce comb laser. The outputs of the four DFBs are multiplexed to a single output, designated as SSC -Comb at the left side of the die, to provide 4 wavelengths. The right side of the die provide four individual outputs, SSC1 to SSC4 to characterize and monitor the output of each laser. Dirce has returned from fab to Axalume and preparation for its testing are being made.  **Figure 8:** *Layout of the Dirce 4xDFB comb laser.* *Q16: Commercialization readiness review* **Actual Performance:** (Completed: Q18) The list of three potential products has been down selected from all technologies within the LEED program. **<u>Product 1</u>. LEED Optical Network** 1.A. Optical Switch

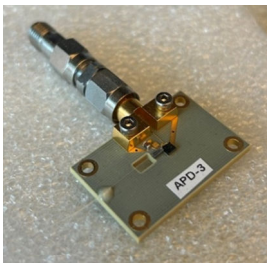
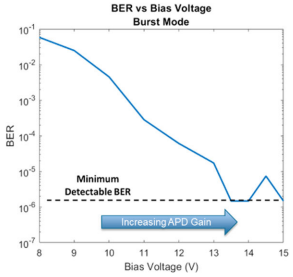
Tasks	Milestones and Deliverables
	1.B. Transport Protocols 1.C. Hardware code for Network Interface Controllers 1.D. Optical Links **Product 2. Upgraded Corundum Platform for Other Applications (independent of LEED network)** **Product 3. Components for Optical Links - Axalume** **Summary** The required commercialization readiness review was completed achieving this milestone. *Q17: Demonstrate multi-rotor synchronization* **Actual Performance:** (Completed: Q17) Practical systems require multiple rotor switches, and these switches needed to be synchronized to each other and to the optical switch. To achieve this milestone, we developed our own custom control board. For global synchronization, we ran Precision Time Protocol (PTP) over the built-in GbE server management ports. This realizes sub-microsecond clock synchronization across the entire network. The performance of the board to maintain a disk phase error for the rotor switch was ± 5 ms, which is 1/1000 of the period of revolution. This performance satisfies the milestone. *Q17: Design of burst-mode receiver (Rx) front end* **Actual Performance:** (Completed: Q14) we completed the simulation, design and verification of the entire single-channel receiver (Figure 9), including TIA, DC acquisition, process sensor as well as the limiting amplifier, as well as the clock and data recovery and bias circuits. The TIA is based on a regulated common-gate amplifier with a resistive load in series with shunt inductors

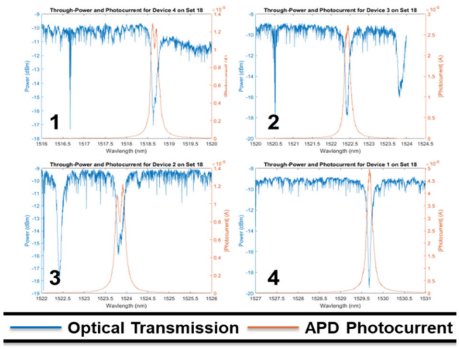
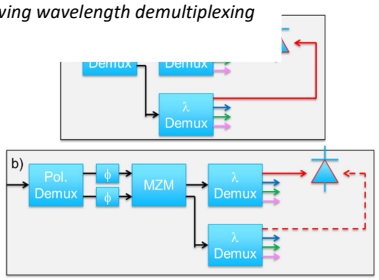
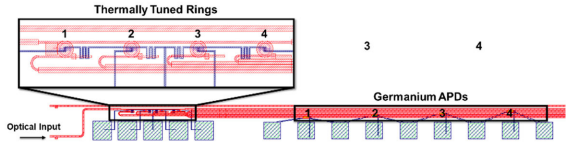
Tasks	Milestones and Deliverables
	 **Figure 9:** *Design of the single channel receiver.* *Q18: Complete fabrication of new rotor switch* **Actual Performance:** (Completed: Q25) The complete fabrication of the switch was completed in late 2024 and was installed into a rack-mount enclosure. We measured the fiber-to-fiber loss, crosstalk, and switching speed for a subset of all 128 switch ports. The results exceed all milestone targets and the switch was used to achieve the other system level milestones for Phase 2. The results were presented as a conference paper at the Optical Fiber Communications Conference (OFC) in March 2024. *Q18: Revised benchmarking* **Actual Performance:** We worked with the STEAM team over several iterations to help define the benchmarking approach. This included reviewing proposed workloads and assessing and providing feedback on the modeling results based on our simulation code and the STEAM team's benchmarking workloads. The generated set of workloads and modeling approaches constituted the final benchmarking methodology. *Q19: Report BER of new rotor switch* **Actual Performance:** (Completed: Q25) We performed a BER test of new rotor switch. We evaluated the full NxN set of port-to-port communication patterns across all time steps. As part of this milestone, we implemented a novel protocol that essentially causes the FPGA-based NICs to measure packets through

Tasks	Milestones and Deliverables
	the rotors to identify when the point is on the rotor disk. Since the rotor is periodic, we can then have the NIC "pause" traffic for 2-3 microsecond granularities to avoid these points. The result worked great—we were able to achieve 98% of link rate using TCP traffic over the rotors. Note that TCP is especially sensitive to packet losses, and since each of these points would have resulted in a lost packet, this would have severely reduced TCP performance. *Q19: Demonstration of transmitter (Tx) controller* **Actual Performance:** (Completed: Q18) A multi-channel, laser driver and tuning controller ("Controller") has been designed, simulated, and built. This controller electronics board is designed to drive, monitor, and tune four channels of the transmitter. The board and enclosure are shown in Figure 10.  **Figure 10:** *Laser Controller/Tuner board and enclosure.* The Controller has been tested in our lab electrically and satisfied all specifications. The Controller has 4 independent channels each accessible via a touch-screen panel. Each channel also has two monitoring port inputs. These monitoring port inputs serve to realize a feedback algorithm to tune the lasers to a desired set of laser frequencies. The controller also has a microprocessor to coordinate all functions and communicate with a PC if necessary. The Controller can be operated stand alone, without the need for a PC to implement our proprietary laser tuning algorithm. This, as well as all the other functions of the Controller are realized via a touch screen on the controller enclosure. The Controller also has the capability to control a Thermoelectric Coupler (TEC) to set/maintain the temperature of the integrated laser die between 15 to 50 degC. *Q19: Commercialization*

Tasks	Milestones and Deliverables
	Actual Performance: (Completed: Q26) During the project LEED-developed technology has been licensed to a spin-out company called inFocus Networks, which has engaged in over 80 meetings with systems vendors, component manufacturers, and system integrators. A key commercialization hypothesis has been that demonstrating the LEED switch in production would reduce technical risk sufficient to pilot a commercial insertion, which at the time of this report has not yet been achieved. The LEED project has also supported Axalume which has been able to pilot the development of the transceiver technology needed to enable LEED-compatible switching in HPC and datacenter applications. *Q20: Practical workloads for datacenter circuit-switching and HPC* **Actual Performance:** (Completed: Q26) We have achieved the major goal of supporting practical workloads—we can support TCP traffic with an unmodified Linux network stack. This means that applications can work with the LEED architecture without modifications. To complete this milestone, our plan has been to evaluate the performance of a set of datacenter workloads "head to head" with what a typical packet switch would support. In addition to showing network-level statistics like latency and throughput, we measured the time to acquire signal using commodity transceivers. Our findings were quite surprising, in a good way, in that we were able to achieve very low lock times with off-the-shelf transceivers with no special modifications.  **Figure 11:** *Measured link reaquisition delays for the LEED switch.*

Tasks	Milestones and Deliverables
	Figure 11 shows that using off-the-shelf transceivers, we can acquire/reacquire the link in 10s of microseconds, which meets the requirements of the milestone. We also evaluated several HPC workloads against the LEED switch, including distributed sort, select, map/reduce, and kernels of AI and machine learning algorithms. Because of the point-defect masking approach we developed, we have been able to support TCP traffic at speeds within 2% of the ideal link rate, and our "request/response" latency which is critical for HPC applications is equivalent to a state-of-the-art packet switch. *Q20: Demonstration of integrated transmitter chip* **Actual Performance:** (Completed: Q20) We completed the design, verification, and tapeout of the *Admiral* silicon photonic Tx chip. The chip dimensions were 8.2mm x 2.8mm and was taped out for fab in the GF 45SPCLO process. The process includes electronic and photonic integration in a 45nm node. The Admiral chip, shown in Figure 12 includes the Tx demo circuit which is a 1x8 high-speed ring modulator array with grating-coupler optical I/O whose design was described earlier. In addition, an o-band 4 (+1) channel laser was included in the layout for hybrid integration with an edge-coupled o-band gain chip. Designs for high-speed ring modulators suitable for electrical and optical probing were included in the Figure below. Microring modulator bias-control test circuits designed in the 45nm electronic process were placed in the top right corner of the chip with wire-bond pad access. Optical I/O to the chip includes laser couplers (for the gain chip), fiber attach structures, and grating couplers. High-speed electrical probe pads as well as DC wire-bond pads were included for electrical I/O. Circuits are accessible through a serial peripheral interface.  **Figure 12:** *Final layout of the Admiral integrated transmitter chip.* *Q20: Demonstration of integrated APD Rx*

Tasks	Milestones and Deliverables
	The tape out, fabrication, and characterization of the integrated APD Rx is complete. Under phase 1, Sandia demonstrated burst mode operation of a Ge APD receiver which was fabricated at Sandia's MESA facility, utilizing germanium deposited by an external vendor. This demonstration showed no errors in 1e8 bits at 10 Gb/s but required equalization to achieve operation at 25 Gb/s due to reflections in the cables used to probe the devices. During phase 2 of the program, we were able to package these devices (**Error! Reference source not found.**) with proper impedance matching which reduced reflections and allowed for high bit rate operation. We were able to demonstrate error free operation in burst mode up to 23 Gb/s (Figure 14), which was the upper limit of bit rate that we could achieve with available equipment at the time.  **Figure 13:** *Packaged APD with proper impedance matching for high data rate operation* To realize 4 wavelength, polarization independent operation, we designed and fabricated a new series of devices. For wavelength division multiplexing of the 4 wavelengths, we explored two different options: 1) a series of tunable ring resonator devices and 2) arrayed waveguide gratings. For polarization independence, we also explored two options. In the first option, illustrated in Figure 15a, light is split into TE and TM polarization before passing to two independent wavelength demultiplexing circuits. The TE and TM polarizations of a single wavelength are then recombined incoherently at a single APD. In the second option, shown in Figure 15b, light is split into TE and TM polarizations, but then the TM mode is rotated to TE polarization. The two optical paths are then  **Figure 14:** *Burst mode bit error rate measurements of packaged APD device at 23 Gbps.*

Tasks	Milestones and Deliverables
	 **Figure 17:** *Optical transmission and APD photocurrent, showing wavelength demultiplexing operation of rings.*  **Figure 15:** *Two approaches to polarization independent operation* coherently recombined before passing to the wavelength demultiplexing circuit. A wide range of these devices were designed and taped out during Phase 2. Unfortunately, due to challenges in fabrication which impacted the germanium APDs, we were unable to  **Figure 16:** *Mask layout of 4-channel ring resonator based demultiplexing circuit.* fully demonstrate these devices. Initial work was done to demonstrate the polarization control and demultiplexing. For example, shown in Figure 16, a circuit with 4 independently tunable ring resonators. In Figure 17 we see the optical transmission (blue) and APD photocurrent (red) for each of the four ring resonators, clearly showing the ability to select out four wavelength channels. Frequency response

Tasks	Milestones and Deliverables
	measurements of the APDs through the rings was also collected. However, we only observed bandwidths of 8-9 GHz, which was not high enough to achieve 25 Gb/s data rates. The cause of this poor frequency response was determined to be due to material quality issues with the germanium and the electrical contacts to the germanium. Extensive failure analysis was performed on APD devices and we observed a number of unexpected features. These features included voids in the germanium around the contacts and germanium segregation in the silicon buffer layer below the device. As a result of this failure analysis, we worked with the external vendor to implement process changes to correct these issues. Unfortunately, these process changes were unsuccessful, and we continued to observe poor frequency response from germanium APDs. In parallel to our efforts to improve quality from the external vendor, we began efforts (funded internally) to develop our own capability to deposit germanium. These initial germanium detectors completed fabrication near the end of Phase 2 and showed excellent responsivity and frequency response, comparable to what we have historically seen. Several wafers which included the 4 wavelength, polarization independent receiver circuits were redirected for in-house deposition of germanium, however, they were unable to complete fabrication before the end of the Phase 2.

Project Activities

This project aims to double the energy-efficiency of datacenter, HPC, and cluster computer systems by improving network bandwidth and latency through the introduction of a novel photonic optically-circuit-switched architecture called LEED. LEED is broken into three main approaches. First, a unique energy-efficient and scalable optical circuit-switched architecture. This architecture uses a realistic control plane for a large port-count optical-circuit switched network and leverages advanced photonic switching and interconnects. This optical network can support both bulk latency-insensitive traffic as well as latency-sensitive traffic. Second, a manufacturable low-loss, scalable, rate-agnostic optical switch technology that enables both the LEED architecture and the development of broadband wavelength-division multiplexed (WDM) interconnect technology. Third, we have developed and demonstrated a photonic integrated chip for a transmitter (Tx) and a photonic integrated receiver chip (Rx). The Tx chip integrates a power-tunable WDM source, a ring-based modulator array and the associated control. The Rx chip integrates a WDM demultiplexer, and a burst-mode compatible APD. Taken together, the composed system validates our working hypothesis which is that optically circuit-switched network architectures can increase the efficiency of compute clusters by executing more work per unit time/energy. The results of this work include a working end-to-end

prototype and testbed insertion that delivers low-latency traffic at latencies equivalent to state-of-the-art packet switches, as well bandwidth for bulk-sensitive traffic that is within 2% of ideal for a packet-switched design under ideal conditions.

During this project, the award timeline changed due to issues caused by the COVID-19 pandemic, including supply chain disruptions which prevented the switch's completion on schedule. While we changed the timeline, we did not change the overall goals of the project.

Project Outputs

A. Journal Articles

J Kelley, A Forencich, G Papen, W Mellette, "Characterization of Burst-Mode Links for Optical Circuit Switching," Journal of Lightwave Technology 40 (9), 2823-2829

W Mellette, A Forencich, J Kelley, J Ford, G Porter, A Snoeren, G Papen, "Optical networking within the lightwave energy-efficient datacenter project," Journal of Optical Communications and Networking 12 (12), 378-389

B. Papers

W Mellette, I Agurok, A Forencich, S Chang, G Papen, and J Ford, “A Scalable, High-Speed Optical Rotor Switch,” 2024 Optical Fiber Communications Conference and Exhibition (OFC)

Shaya Fainman, Joseph Ford, William M Mellette, Shayan Mookherjea George Porter, Alex C Snoeren, George Papen, Saman Saeedi, John Cunningham, Ashok Krishnamoorthy, Michael Gehl, Christopher T DeRose, Paul S Davids, Douglas C Trotter, Andrew L Starbuck, Christina M Dallo, Dana Hood, Andrew Pomerene, Anthony Lentine, "Leed: A lightwave energy-efficient datacenter," 2019 Optical Fiber Communications Conference and Exhibition (OFC)

L Wu, W Mellette, G Schuster, J Ford, "Fast quasi-static beam steering via conformally-mapped gratings," Optics and Photonics for Information Processing XIII 11136, 231-240

W Mellette, R Das, Y Guo, R McGuinness, A Snoeren, G Porter, "Expanding across time to deliver bandwidth efficiency and low latency," 17th USENIX Symposium on Networked Systems Design and Implementation (NSDI 20)
W Mellette, A Forencich, R Athapathu, A Snoeren, G Papen, G Porter, “Realizing RotorNet: Toward Practical Microsecond Scale Optical Networking,” ACM Sigcomm 2024.

C. Status Reports

N/A

D. Media Reports

N/A

E. Invention Disclosures

Commented [MM1]: Check with Axalume and Sandia

See issued patents (F)

F. Patent Applications

US Patent 10,193,636 - DC-Coupled Optical Burst-Mode Receiver
US Patent 11,632,330 - Optimizing connectivity in reconfigurable networks
US Patent 11,368,336 - Method and apparatus for interfacing with a circuit switched network
US Patent 11,550,104 - Selector switch

G. Licensed Technologies

See issued patents (F). Patents owned by UCSD are licensed to inFocus Networks.

H. Networks/Collaborations Fostered

The project has fostered collaborations between UC San Diego, inFocus Networks, Sandia National Labs, and Axalume.

I. Websites Featuring Project Work Results

N/A

J. Other Products (e.g. Databases, Physical Collections, Audio/Video, Software, Models, Educational Aids or Curricula, Equipment or Instruments)

Software and simulation results are available at https://github.com/ucsdsysnet/corundum and https://github.com/TritonNetworking/opera-sim

K. Awards, Prizes, and Recognition

N/A

Follow-On Funding

Additional funding committed or received from other sources (e.g. private investors, government agencies, nonprofits) after effective date of ARPA-E Award.

Table 2. Follow-On Funding Received.

Source	Funds Committed or Received
National Science Foundation – SBIR	$225,000
San Diego Tech Coast Angels	$1,000,000