A Scalable, High-Speed Optical Rotor Switch

William M. Mellette[1], Ilya Agurok[2], Alex Forencich[2], Spencer Chang[2], George Papen[2], and Joseph E. Ford[1,2]
[1]*inFocus Networks, San Diego, California*
[2]*UC San Diego, La Jolla, California*
max@infocusnetworks.com

Abstract: Rotary optical switching enables low-loss microsecond-scale reconfiguration between pre-programmed interconnects with thousands of ports, supporting high-bandwidth and low-latency Rotornet datacenter architectures. We describe a 7 μs 128 × 128 port rotor switch with 4 dB fiber-to-fiber insertion loss and a 1-dB spectral bandwidth of 120 nm. © 2023 The Authors

1. Introduction

Optical circuit switches deployed by Google have shown significant capital, energy, and operational savings for large scale datacenters and machine learning clusters [1]. Today's systems use optical switches with 100s of ports and millisecond reconfiguration speeds, but scaling switches to 1,000s of ports and microsecond speeds while maintaining signal integrity presents additional opportunities and potential benefits [1, 2, 3].

One path forward could be to scale the MEMS-based crossbar switches already in production. Thousand-port MEMS switches have been developed [4], but increasing the speed by 1,000× would likely require a reduction in port count, even when using speed-optimized MEMS actuators with small feature sizes and challenging fabrication processes [5]. Another option is integrated waveguide-based switches, which offer fast reconfiguration, but the insertion loss due to cascaded switching elements and fiber-waveguide coupling losses limit port scalability [6].

Regardless of device technology, simply speeding up the optical switch still leaves open the system-level challenge of implementing a correspondingly fast control plane needed to compute a real-time schedule of switch configurations. Rather than trying to make traffic-aware optical network designs faster, another option is to employ so-called "traffic-oblivious" network control, which removes the need for centralized scheduling entirely.

Rotornet is a traffic-oblivious network design that can improve overall throughput/cost ratio [2], improve energy efficiency, and even support traffic delivery latencies faster than the reconfiguration speed of the optical switch [3]. Besides simplifying network control, Rotornet's system-level design simplifies optical switch hardware: rather than requiring an arbitrarily configurable crossbar switch, each switch only needs to cycle between a small number of pre-defined connection patterns. We refer to this specialized type of optical switch as a "rotor" switch.

Here we report on the design, fabrication, and measurement of a microsecond scale optical rotor switch supporting 128 × 128 ports and four interconnection mappings. Unlike previous designs, the new prototype selects between interconnects encoded as free-space optical transforms. The switch architecture is readily scalable to 1,000s of ports and additional mappings with straightforward scale-out of each subsystem. Critically, because scaling does not require cascading additional optical elements, low loss and high signal integrity can be maintained.

2. Design

In traditional switch architectures, individual switching elements acting on individual optical signals perform both switching and spatial routing functions. The rotor switch decouples switching and routing, instead using a common element to switch an array of signals between a set of pre-defined routing structures that spatially rearrange those signals. This switching model allows the use of simple, low-cost, and scalable devices.

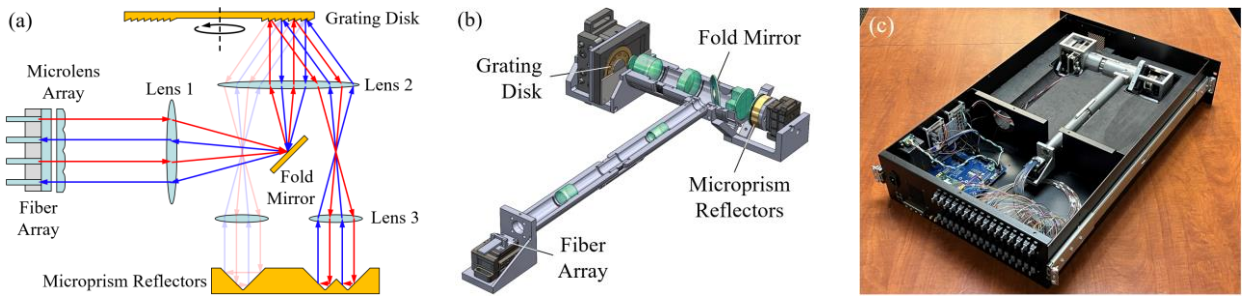

Fig. 1. (a) Cross-sectional diagram of rotor switch architecture with red and blue arrows indicating forward and reverse signal paths, (b) optomechanical CAD with cutouts to show optics, (c) photograph of packaged switch.

Rotor switching can be implemented in a variety of device technologies, including integrated photonics, but we focus on a free-space design targeting scalability and low insertion loss. Figure 1(a) shows a cross-sectional diagram of our rotor switch architecture. A two-dimensional array of single-mode fibers serves as both the input and output with I/O signals spatially interleaved. A microlens array collimates the signals and a telecentric 4-f relay (lens 1 and lens 2 in Fig. 1(a)) images them with demagnification onto a rotating disk. The disk is spatially patterned with conformally mapped blazed diffraction gratings so that as it rotates, the diffraction angle is stable within each annular grating sector [7]. Reconfiguration occurs when the boundary between diffractive sectors is scanned across the spatial extent of the signal array, favoring a rectangular signal array and large disk radius for faster switching speeds. Our prototype uses an 8 × 32 signal array imaged to a 200 × 800 μm area on the grating 20 mm from the center of the disk, which spins at 15,000 RPM to achieve a worst-case reconfiguration delay of 7 μs. This is approximately 1,000× faster than the MEMS tilt mirrors used in commercial optical switches. More ports can be supported at the same or faster speeds by increasing the disk radius, RPM, demagnification ratio, and/or signal array aspect ratio. For example, 7 μs reconfiguration of 1,024 × 1,024 ports can be supported with the same size disk spinning at 30,000 RPM and using a 16 × 128 signal array.

After signals are diffracted from the rotary switching engine, they pass through a second 4-f relay (lens 2 and lens 3 in Fig. 1(a)) which is spatially segmented via aperture division to image the signal array onto one of several locations. At each location, a micro-optic structure spatially rearranges the incoming signal array before reflecting signals to pass back through the switch and to the output fibers. Various types of micro-optics can be employed to provide specific connection patterns. We used reflective prismatic structures with 90° corners and various pitches to achieve connection patterns corresponding to the set of so-called “crossover” mappings [8]. By permuting the input and/or output fiber cabling between the fiber array and bulkheads, the effective connection patterns can be made completely orthogonal between switches (despite each switch having the same internal micro-optic structures), supporting the topology requirements of the overall Rotornet network [2, 3].

Figure 1(b) shows a CAD model of the 128 × 128 port rotor switch with custom lenses and optomechanics. The lens barrel assembly acts as a common mechanical element for integrating all switch subsystems, and we integrated off-the-shelf micro-positioning stages to align the fiber array, grating disk, and microprisms to the barrels. Figure 1(c) shows the packaged switch prototype, including optomechanical assembly, control electronics, management interfaces, and 32 8-fiber MPO bulkheads inside a 3 RU rack-mount enclosure. The entire optomechanical assembly floats on a machined foam insert within the enclosure for mechanical isolation from shock and vibration.

3. Fabrication and Measurement

The optomechanics (Fig. 2(a)) were manufactured using standard machining techniques from a low-CTE alloy to minimize thermal effects. The lenses were ground and polished from commercially available glass blanks, anti-reflection (AR) coated, and bonded into multi-element assemblies using UV-cured optical adhesive. After mounting and aligning all lens elements in the lens barrel assembly, we measured the fiber-to-fiber loss contribution of the lenses to be 1 dB, including glass absorption, surface reflections, lens aberrations, and alignment tolerances.

The fiber array (Fig. 2(b)) was manufactured using an etched silicon plate to position 256 single-mode fibers on a 250 μm pitch grid. An AR coated glass block was bonded to the array face with index matched epoxy to minimize back reflections. An AR coated silicon microlens array (MLA) fabricated with greyscale lithography was positioned above the fiber array to collimate the signals, with the Gaussian beam waist located at the MLA surface. Including fiber positioning and tilt errors along with MLA radius of curvature (ROC) variability, the worst-case insertion loss of the 256-element collimated fiber array assembly was 2 dB (measured by return coupling).

Figure 2(c) shows the diffractive switching element, which was fabricated by nano-imprint replication of a greyscale-lithographically produced master onto a 57 mm diameter substrate. The diffraction efficiency of the 1st order blazed grating was -0.8 dB with 0.3 dB polarization-dependent loss. The 25 mm diameter central hole allows

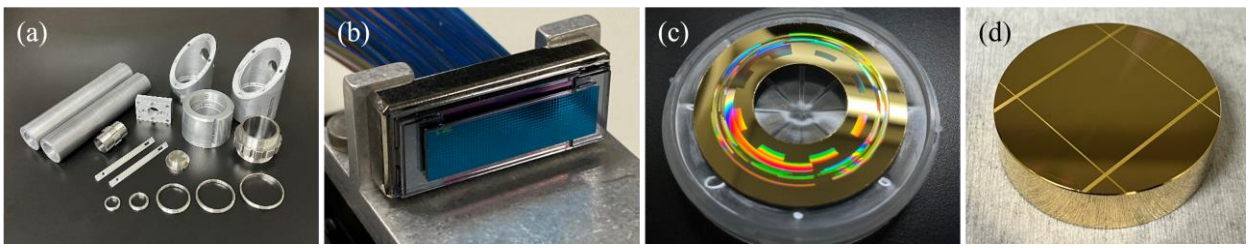

Fig. 2. Switch subsystems: (a) custom optomechanics, (b) 8×32 collimated fiber array, (c) ∅ 50 mm disk with conformally mapped diffraction gratings, (d) microprism reflector arrays on ∅ 40 mm substrate.

the disk to be mounted to a commercially available 15,000 RPM brushless DC motor spindle that consumes ~ 5 W.

The microprism reflector arrays (Fig. 2(d)) were fabricated by diamond ruling four orthogonal sets of 70 µm deep grooves into a common metal substrate, which was then coated with a layer of gold. The largest feature was a single 140 µm wide v-groove used to spatially invert the entire signal array. The smallest features were an array of 16 adjacent v-grooves with 32 µm pitch which swapped neighboring signals pairwise within the signal array. A manufacturing defect in the initial microprism device blocked 90% of the signal paths through one of the four interconnection patterns. Refabrication of the microprisms is underway to correct the defect, but the current device provided sufficient connectivity for initial characterization of the switch.

Figure 3(a) shows the measured fiber-to-fiber transmission and crosstalk spectra for a representative path through the switch. The center-band insertion loss is 4 dB, and despite the diffractive switching element, the imaging layout of the switch enables a 1-dB bandwidth of 120 nm. Crosstalk between nearest-neighbor fibers in the array is -50 dB, and crosstalk due to the -1 order of the diffraction grating is -35 dB. The polarization-dependent loss (PDL) is 0.7 dB. The loss, -1 order crosstalk, and PDL can all be improved with optimizations to the diffraction grating design and fabrication process. PDL can also be eliminated entirely by introducing a λ/4 waveplate before lens 3 to rotate polarization by 90° between the first and second bounce off the diffraction grating [9]. Loss can be further reduced by improving the fiber position, fiber tilt, and MLA ROC tolerances in the fiber array assembly.

Figure 3(b) shows the reconfiguration speed for two different switching transitions: 7 µs when the signals are switched between opposite edges of the signal array (worst-case), and < 0.5 µs when the signals are in the same row in the array (best-case). The switch can be phase-locked to an external or internal pulse-per-revolution (PPR) reference synchronization signal. An optical encoder provides PPR feedback on disk position to an MPSoC, which runs our PID control software achieving ± 5 µs phase synchronization between the disk and reference signal at 15,000 RPM (see Fig. 3(c)). Lower phase error may be possible using multi-pulse-per-revolution feedback.

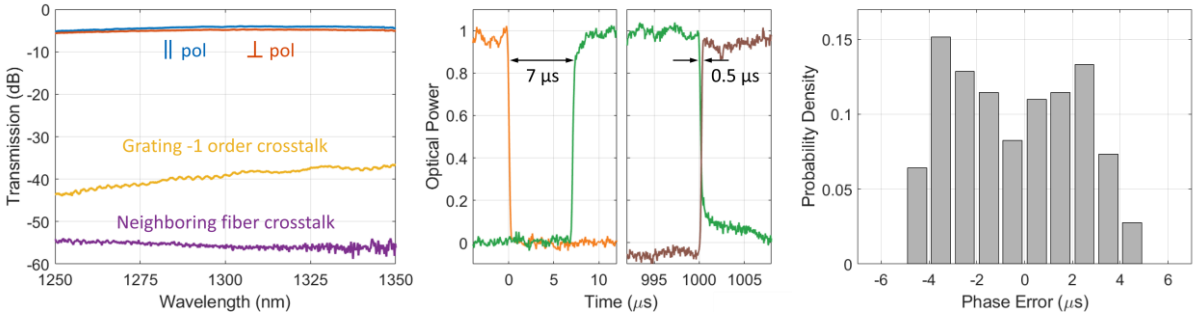

Fig. 3. (left) Crosstalk and transmission spectra for worst- and best-case polarizations, (center) worst- and best-case reconfiguration speeds of 7 µs and < 0.5 µs, (right) ± 5 µs phase stability of rotor under closed-loop control.

4. Conclusion

We have demonstrated a novel optical rotor switch architecture based on image relay, diffractive rotary switching, and free-space micro-optic interconnection. This class of switch can support high-bandwidth, low-latency, traffic-oblivious network architectures such as Rotornet. The prototype demonstrates that microsecond-scale reconfiguration speeds are possible at large port count, low insertion loss, and large spectral bandwidth using scalable, low-cost, and low-power bulk optical devices.

Acknowledgements

This work was funded by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0000845.

References

[1] H. Liu et al., “Lightwave fabrics: at-scale optical circuit switching for datacenter and machine learning systems,” proc. SIGCOMM, 2023.
[2] W. Mellette et al., “Rotornet: a scalable, low-complexity, optical datacenter network,” proc. SIGCOMM, 2017.
[3] W. Mellette et al., “Expanding across time to deliver bandwidth efficiency and low latency,” proc. 17th NSDI, 2020.
[4] J. Kim et al., “1100 x 1100 port MEMS-based optical crossconnect with 4-dB maximum loss,” Photonics Technology Letters 15(11), 2003.
[5] W. Mellette et al., “Scaling limits of MEMS beam-steering switches for data center networks,” Journal of Lightwave Technology 33(15), 2015.
[6] X. Tu et al., “State of the art and perspectives on silicon photonic switches,” Micromachines 10(1), 2019.
[7] L. Wu et al. “Fast quasi-static beam steering via conformally-mapped gratings,” Optics and Photonics Information Processing XIII, 2019.
[8] J. Jahns and J. Murdocca, “Crossover networks and their optical implementation,” Applied Optics 27(15), 1988.
[9] J. Ford et al., “Wavelength add/drop switching using tilting micromirrors,” Journal of Lightwave Technology 17(5), 1999.