61 Port 1×6 Selector Switch for Data Center Networks

William M. Mellette[1], Glenn M. Schuster[1], George Porter[2], and Joseph E. Ford[1]
[1]Department of Electrical & Computer Engineering, and [2]Department of Computer Science & Engineering
University of California San Diego, 9500 Gilman Drive, La Jolla, California, 92093
wmellett@ucsd.edu

Abstract: We present design and preliminary characterization of a scalable MEMS-based "selector switch" for high performance computing networks. The 170 μs, 61-port prototype uses relay image steering to route all 61 SMF channels through one of six pre-structured interconnects.
OCIS codes: (230.4685) Optical microelectromechanical devices; (060.6718) Circuit Switching; (200.6715) Switching

1. Introduction - Selector Switch Architecture

There is interest in exploring optical switching technology as a cost-effective means to increase the aggregate bandwidth of data center networks [1]. In networks where optical amplification or high sensitivity transceivers are a significant cost constraint on growth, it is desirable to maintain low switch insertion loss and crosstalk while scaling to large radix and high speed. Here we investigate a novel and potentially useful switch architecture that can be embodied with multiple switching technologies, but in particular can enable freespace microelectromechanical systems (MEMS) switching to scale in speed and radix without compromising optical performance.

A non-blocking, fully connected $N\times N$ port switch requires a node degree of N. Using 1×2 or 2×2 switching elements requires a multistage switch architecture (e.g. Benes, crossbar, etc.) to satisfy this node degree for $N > 2$. The switch radix is typically limited by accumulated loss and crosstalk. Freespace MEMS switches can scale to high radix with low loss because each $1\times N$ switching element (micromirror) can accommodate $N > 1000$ ports [2], requiring only two stages of switching elements. However, in scaling up N, the micromirror aperture and mechanical tilt range must increase, inevitably reducing the mechanical response speed [3]. These limitations can potentially be overcome using multistage switch architectures, but accumulated insertion loss may still limit scalability. There are currently no cost effective solutions to interconnect data communication networks with thousands of nodes while providing microsecond reconfiguration times. We may need a more holistic, system-level design approach that reconsiders fundamental assumptions about the interplay between network and switch requirements.

We take one such approach by modifying an underlying assumption about the necessary connectivity of a switch, namely, we consider a switch architecture with a node degree M which is less than the number of ports N. The switch selects between a small number of passive (or slowly reconfigurable) interconnections, which make up a subset of all $N!$ possible interconnections of an $N\times N$ cross connect. This "selector switch" architecture does not fully connect all ports in a single pass through the switch; however, if $M \geq \log_2 N$, fixed interconnections can be chosen so any node can be reached in at most M hops through the switch [4]. Figure 1(b) shows an example set of $M = 2$ interconnects which guarantee that $N = 4$ ports are connected with at most two hops. This sacrifice in single-hop connectivity requires memory at each node for temporary data storage, but may be permissible for the data traffic structure of some applications in the network. Applied to freespace MEMS switches, the selector switch architecture's reduction in node degree allows orders-of-magnitude scaling in speed and port count with similar insertion loss compared to a fully connected cross connect architecture.

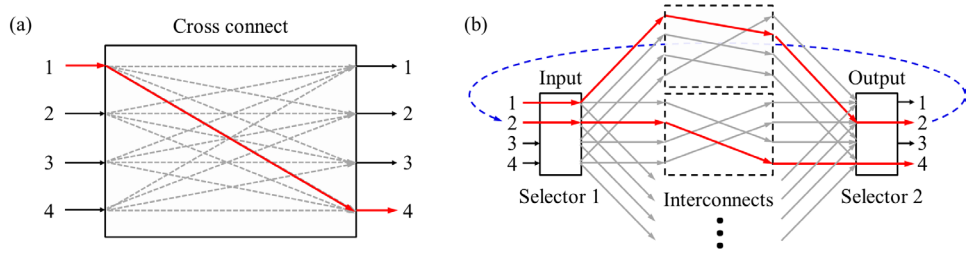

Fig. 1. (a) An $N\times N$ port cross connect can connect any two ports in one pass through the switch. (b) Two selector switches with fixed interconnections are not fully connected in a single pass. However, nodes 1 and 4 are indirectly connected through an intermediate node, 2.

2. System Design

The choice of device technology to implement the selector switch architecture depends on the network requirements on reconfiguration time, port count, and insertion loss. An early demonstration of this concept used wavelength switching between fixed interconnects recorded as volume holograms [5], but here we consider freespace MEMS.

Figure 2(a) shows a cross section of the switch layout, based on a 4-f imaging relay with a prism array located near the Fourier plane. Light from a two dimensional (2-D) array of input fibers is imaged onto a single MEMS mirror, which tilts in 2-D to direct reflected light through discrete prism apertures which refract light to shift the output image position so that it couples into one of the output fiber arrays. Switching occurs by aperture selection in the Fourier plane, as opposed to spatial scanning across the fiber array, making the system relatively tolerant to angular misalignments of the mirror. This lowers the necessary drive electronics precision and the effect of underdamped ringing of the mirror. If the single mirror is replaced by a mirror array of smaller and potentially faster mirrors, the input(s) imaged onto each mirror can be routed to an independently-selected interconnection network.

Figure 2(b) shows the Zemax model of a prototype switch with $N = 61$ input ports and $M = 6$ 61-port outputs. For the input and output fiber arrays, we used Chiral Photonics' 2-D pitch reducing optical fiber arrays (PROFAs) [6] for high channel density (10 μm mode field diameter with 37 μm pitch) in 61-core hexagonal arrays. For the micromirror, we used Mirrorcle Technologies' USB-powered A7M8.1 2-axis comb drive device, which is a single 800 μm diameter mirror surrounded by large-area high-force actuators to provide a ±4° mechanical tilt range with sub-millisecond settling. We chose 25 mm and 60 mm focal length lenses for the 4-f relay so that the magnified image of all 61 channels from the input fiber array would be contained within the mirror aperture while minimizing the required tilt range for beam separation at the Fourier plane. The prism apertures were chosen to be 1.3× the beam width. This yields 3% power clipping but keeps the F/# of the focusing lens above F/3, helping to minimize lens aberrations. The fused silica prism apex angle was chosen to be 5° to maximize the F/# of the collimating lens while minimizing chromatic dispersion, which ultimately limits the spectral bandwidth of the switch.

We conducted physical optics propagation in Zemax with a single mode fiber coupling calculation and found that the dispersion contributed to 1 dB excess loss at a bandwidth of ±30 nm. The loss due to lens aberrations at the center of the band (nominally 1560 nm) was predicted to be 0.6 dB. The current system can be tuned to operate in different wavebands by small translation of the fiber arrays. To increase the bandwidth of the switch further or to scale the system to more output apertures, larger tilt angle prisms can be achromatized by combination with a small negative diffractive element, or replaced with reflective facets in a redesigned layout.

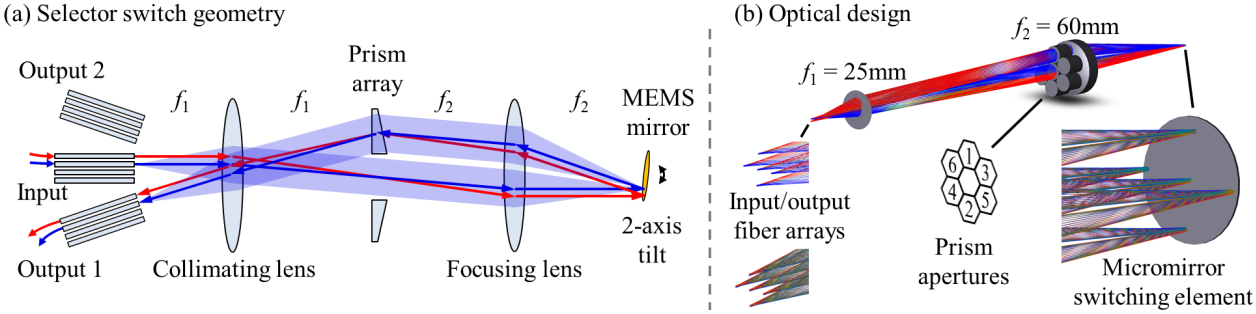

Fig. 2. (a) Cross sectional schematic of selector switch. An input fiber array is relay imaged to couple to an output array by a tilting micromirror and prism array. (b) Zemax model of the prototype system. Gaussian beams from the center and edge channels of the fiber array at different wavelengths are propagated through the system and a single mode coupling calculation quantifies the transmission spectrum.

3. Prototype Fabrication and Preliminary Characterization

We assembled a prototype system following the design in Section 2, and have populated the input fiber array and two of the six available output arrays (Fig. 3(a)). 4-axis translation stages were used to position the fiber arrays with one rotational and three linear degrees of freedom. We fabricated an antireflection coated prism array by dicing coated fused silica wedges and bonding them together in a hexagonal pattern with optical UV-curing adhesive (Fig. 3(b)). The entire prototype, including cables, optics, and MEMS driver is contained on an 18×24 inch breadboard.

Light from a broadband erbium ASE source was sent through the switch to measure the transmission spectrum. The fibers were aligned to maximize coupling at the 1560 nm gain peak of source. To assess the on-axis performance of the lenses, the micromirror was left un-tilted and light from the center core of the input PROFA was coupled back into itself, and then traveled through a circulator to an optical spectrum analyzer. Figure 3(d) shows good agreement between the modeled and measured transmission for self-coupling (the measurement was band-limited by the bandwidth of the circulator). Fresnel reflections from uncoated surfaces account for 2.1 dB loss (with an additional 0.1 dB from the gold mirror) leaving 0.3 dB excess loss from on-axis lens aberrations. Next, the mirror was tilted and light was coupled into the center core of a second device. The measured signal had 1 dB excess loss compared to the model, which we attribute to off-axis lens aberrations, giving 3.6 dB loss at the center wavelength (Fig. 3(d)). The measured dispersion introduced by the prism agrees with the model, giving an insertion loss of 4.6

dB over a 60 nm bandwidth. In the switch architecture shown in Fig. 1(b), the total insertion loss for the 2-stage switch is twice the single pass loss, giving 9.2 dB loss over 60 nm. By antireflection coating all uncoated surfaces with standard commercial coatings, the loss may be reduced to 2.6 dB per pass. With custom lenses, it may be possible to reduce the loss further by reducing off-axis lens aberrations.

Figure 3(e) shows the *x* and *y* axis drive waveforms applied to the mirror and the corresponding measured optical switching response. The drive waveforms were digitally filtered using the inverse transfer function of the mirror to give a fast response time while suppressing ringing. The optical power traces in Fig. 3(e) show light exiting the center core of one PROFA and coupling to the center core of another with negligible ringing. In this case the two prism apertures were adjacent, and we observe a brief period where light is weakly coupled into both cores. The switching time (90-90% power) was measured to be 170 μs. By reshaping the drive waveform to allow a small amount of mirror ringing (within 10% of the settled power), we expect the switching time can be reduced further.

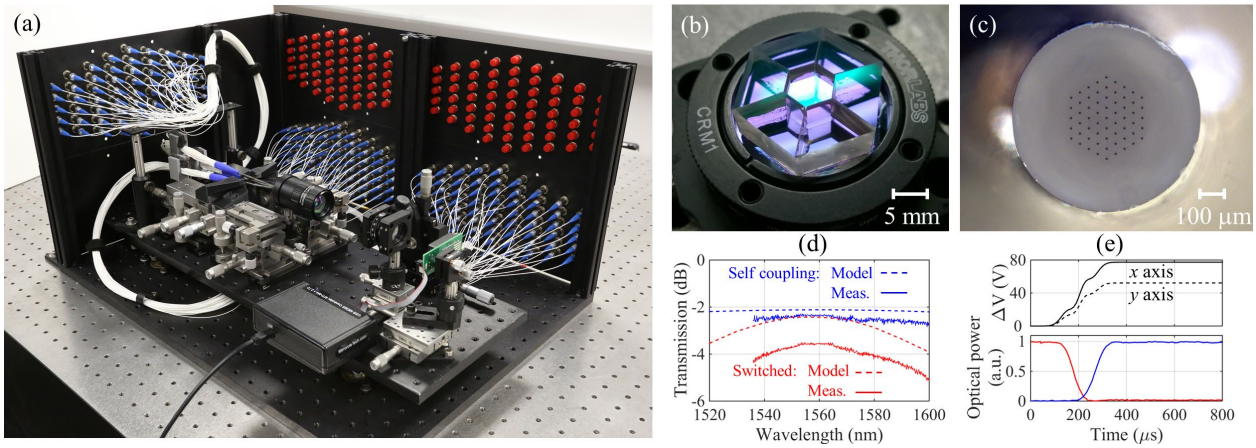

Fig. 3. (a) Prototype system with three of seven 61-core fiber arrays mounted. (b) Custom fabricated antireflection coated prism array. (c) Microscope image of 61-core fiber array. (d) Modeled and measured transmission spectrum, aligned for maximum coupling at 1560nm. (e) Drive waveforms for *x* and *y* mirror axes and measured switching response between the center cores of two fiber arrays.

We are in the final stages of assembly and characterization of the prototype system, which involves replacing the temporary 3D-printed PROFA mounts with aluminum mounts to improve stability, and attaching the remaining output fiber arrays. We will measure and report on array-to-array coupling between all 61 input and output cores.

4. Conclusion

The switch design presented here demonstrates a novel optical switch architecture which leverages a reduced node degree for improved scalability. Our prototype was designed for 1×6 selection of 61-port outputs, but with a different commercially available MEMS mirror array the system can scale to 1×18 (2 rings of hexagonally packed apertures) ×61 ports and potentially larger with a fully custom MEMS array. With M = 18 output apertures, the system can connect 2^{18} = 262,144 nodes in at most 18 hops with the appropriate fixed interconnections. The number of ports integrated into a single 1×M ×N switch will be driven by the cost and yield of the fiber and MEMS arrays. To scale to high port counts, a number of selector switches can be interconnected together in parallel using fiber optics to form a 2-stage switching network. Because it incorporates relay imaging, this switch geometry is also compatible with space-division-multiplexed signals in either multimode or multicore fibers. By replacing the fiber optic interconnections with freespace relays built from structured micro-optics, freespace interconnections can be implemented which further extend the scalability of the switch and eliminate cabling complexity.

This work is supported by the National Science Foundation under grant NSF CNS-1314921.

5. References

[1] N. Farrington, G. Porter, S. Radhakrishnan, H. Bazzaz, V. Subramanya, Y. Fainman, G. Papen, and A. Vahdat, "Helios: A Hybrid Electrical/Optical Switch Architecture for Modular Data Centers," in Proc. ACM SIGCOMM, 2010, 339-350.

[2] J. Kim *et. al.*, "1100 x 1100 Port MEMS-Based Optical Crossconnect With 4-dB Maximum Loss," IEEE Photonics Technology Letters, **15**(11), 1537-1539 (2003).

[3] W. Mellette and J. Ford, "Scaling Limits of MEMS Beam-Steering Switches for Data Center Networks," JLT **33**(15), 3308-3318 (2015).

[4] I. Stoica, R. Morris, D. Karger, M. Kaashoek, and H. Balakrishnan, "Chord: A scalable peer-to-peer lookup service for internet applications," in Proc. ACM SIGCOMM, 2001, 149-160.

[5] J. Ford, Y. Fainman, S. Lee. "Reconfigurable array interconnection by photorefractive correlation," Appl. Opt. **33**(23), 5363-5377 (1994).

[6] V. Kopp, J. Park, M. Wlodawski, J. Singer, D. Neugroschl, and A. Genack, "Chiral Fibers: Microformed Optical Waveguides for Polarization Control, Sensing, Coupling, Amplification, and Switching," JLT **32**(4), 605-613 (2014).