On-chip optical matrix-vector multiplier for parallel computation

Lin Yang, Lei Zhang, and Ruiqiang Ji

A new integrated system enables the miniaturization of digital signal processors, while also improving their performance and stability.

Matrix-vector multiplication is a fundamental operation in modern digital signal processing, applicable to digital images, radar signals, and coherent optical communication. Inspired by the intrinsic parallelism offered by optical computation, many efforts have been made to develop optical devices capable of performing parallelizable operations.\(^1\)–\(^3\) The Stanford multiplier is one of the most notable examples,\(^1\) and generally consists of a light source array, an optical lens, a spatial light modulator (SLM), and a photo-detector (PD) array. However, in most cases this architecture produces systems that are large and consume a lot of power. Moreover, the number of distinct elements used in the design makes these multipliers extremely sensitive to vibration, which limits their application.

To overcome these limitations, we propose an on-chip optical matrix-vector multiplier (MVM).\(^4\) In the standard MVM approach, a combination of lenses is used to split light into several parts, which are then independently processed by a SLM, before subsequently being converged by another combination of lenses. The splitting and converging processes are called fan-out and fan-in. In our approach, they are implemented by an integrated power splitter and wavelength multiplexer. We also replace the traditional bulky and slow SLM with an integrated, high-speed wavelength-selective modulator matrix, reducing the size and complexity of the system and subsequently improving its stability and performance.

We have designed and fabricated a prototype of a system capable of performing a multiplication of a \(M \times N\) matrix \(A\) by a \(N \times 1\) vector \(B\) to give a \(M \times 1\) vector \(C\). The mathematical procedure of MVM can be split into multiplications and additions, which is reflected in our design. Figure 1 shows a schematic of the architecture we propose. The elements of \(B\) are represented by the power of \(N\) modulated optical signals with \(N\) different wavelengths \((\lambda_1, \lambda_2, \ldots, \lambda_N)\), generated by \(N\) modulated laser diodes, either alone or together with \(N\) Mach-Zehnder modulators. These signals are multiplexed, passed through a common waveguide, and then projected onto \(M\) rows of the modulator matrix by a \(1 \times M\) optical splitter. Each element \(a_{ij}\) of matrix \(A\) is represented physically by the transmissivity of the microring modulator located in the \(i^{th}\) row and the \(j^{th}\) column of the modulator matrix. Each modulator in any one row only manipulates an optical signal with a specific wavelength.

The multiplication processes of MVM are carried out when the \(M \times N\) optical pulses pass through the modulator matrix, while

\[ A \cdot B = C \]

Continued on next page
the $i^{th}$ accumulation process (the addition part of MVM) is carried out when the $N$ optical signals (each with a different wavelength) in the $i^{th}$ row are combined by being guided along the common output waveguide to the $i^{th}$ PD. The signal at this PD gives the $i^{th}$ element of the resultant vector $C$.

As a proof of concept, we fabricated a $4 \times 4$ microring modulator matrix on an 8-inch silicon-on-insulator wafer: see Figure 2. Along with off-chip laser diodes, commercial Mach-Zehnder (MZ) modulators, multiplexers, and PDs, this constitutes our prototype. The waveguides have a height of 220nm, a width of 400nm and a slab thickness of 70nm. The radii of the four microrings are 9.97, 10.00, 10.03, and 10.06\(\mu\)m, respectively, in order that the four resonance wavelengths are uniformly distributed throughout the whole free spectral range (FSR).

Figure 3(a) shows our experimental setup. The original resonance wavelengths of the microrings in the first row are 1550.450, 1552.850, 1555.182, and 1557.494nm, respectively. When a forward bias voltage is applied, a blue shift occurs to these wavelengths; see the red trace in Figure 3(b). The carrier injection induces absorption while changing the refractive indices of the waveguides. The extinction ratios at the working wavelengths are about 24dB (troughs in the black trace), guaranteeing the computational accuracy of the system.

To perform MVM, the four continuous waves are fed from tunable lasers into four commercial MZ modulators to generate the elements of $B$. The four modulated signals are then multiplexed and coupled into the bus waveguide of the first row of microring modulators by a lensed fiber. Since the wave-guides are polarization-dependent, each wavelength channel is controlled independently by a polarization controller before they are multiplexed to a fiber. To achieve this, a total of eight electrical driving signals are generated by four synchronized two-channel arbitrary function generators. Four of those signals are applied to the MZ modulators to generate the four elements of $B$ ($b_1, b_2, b_3, b_4$). The other four signals, representing the first row of $A$ ($a_{11}, a_{12}, a_{13}, a_{14}$), are applied to the first row of microring modulators. Figure 4 shows the waveforms of the eight electrical driving signals and, in the bottom trace, the final result of a vector-vector multiplication ($A_1 \cdot B$). These waveforms show that the multiplication is performed at a speed of $10^7$ MAC/s (multiplications and accumulations per second), corresponding to a data-processing speed of 20Mb/s.

For the larger optical MVM we are proposing, the scale of the matrix is determined by the fan-out number $M_{\text{fan-out}}$ and the multiplexed wavelength channel number $N_{\text{mux}}$. The fan-out number is determined by both the number of the output ports on the optical splitter and the ‘power budget’, which limits the number of times the optical signal can be split due to the the limits of sensitivities of the photodetectors. If the total optical power is split into too many parts (i.e., $M$ is a larger number), the photo-detectors will not be able to distinguish the increment. $N_{\text{mux}}$ is determined by the FSR of the microring resonator and the channel spacing of the WDM signals. Smaller
Figure 4. Waveforms of the eight driving voltages, and optical output (bottom). The first four voltages are applied to the microring modulator row representing the elements in the first row of matrix $A$. The second four voltages are applied to the four MZ modulators to generate the optical vector representing the elements of vector $B = (b_1, b_2, b_3, b_4)$.  

Microring resonators can be used to increase the FSR: for example, it has been reported that the microring resonator with a radius of $1.5\mu m$ has a FSR of $62.5\text{nm}^5$. Reducing the channel spacing for a given FSR can also increase $N_{\text{mux}}$, but the channel spacing is itself limited by the spectral broadening of the optical signals after the MZ modulators. Suppose that the microring modulator with a radius of $1.5\mu m$ has the same modulation speed with the reported $25\text{Gb/s}$ microring modulator; the $N_{\text{mux}}$ will be $\sim 78$ and the computation speed of the optical MVM will be $78 \times 78 \times 2.5 \times 10^{10} = 1.52 \times 10^{14}$ MAC/s. In addition, multi-level modulation technology can be adopted to further improve the performance of our optical MVM system.

In summary, we have designed and demonstrated a prototype of an optical signal processor that can perform matrix-vector multiplication. This consists of a laser-modulator array, multiplexer, splitter, microring modulator matrix, and photo-detector array. An operation rate of $8 \times 10^7$ MAC/s has been achieved, corresponding to a data-writing speed of $20\text{Mb/s}$. Ultimately, these units can be monolithically integrated on a chip, contributing to the development of silicon photonics and leading to high-performance computing systems. In future we will adopt a faster modulating scheme, and construct a larger modulator matrix to improve the operating speed. Further, we plan to integrate the peripheral components onto the same platform as the device described here in order to realize a working integrated optical-computing system.

Author Information

Lin Yang, Lei Zhang, and Ruiqiang Ji  
State Key Laboratory for Integrated Optoelectronics  
Institute of Semiconductors, Chinese Academy of Sciences  
Beijing, China

Lin Yang is currently a professor. He is the author or coauthor of more than 50 refereed journal papers. His current research interests include silicon-based photonic devices for optical interconnects, optical computing, and optical communication.

Lei Zhang is currently an assistant professor. His research interests include silicon photonics, microwave photonics, on-chip optical interconnects, and information processing.

Ruiqiang Ji is a PhD candidate. His research interests include the modeling, design, and characterization of nanophotonic components, especially in the area of sub-micrometer silicon waveguides on the silicon-on-insulator platform.

References