



# Photonic computing to accelerate data processing in wireless communications

MAHSA SALMANI,<sup>1,2</sup>  ARMAGHAN ESHAGHI,<sup>1,3</sup> ENXIAO LUAN,<sup>1</sup>  
AND SREENIL SAHA<sup>1</sup>

<sup>1</sup>Huawei Technologies Canada, ON, L3R 5A4, Canada

<sup>2</sup>mahsa.salmani@huawei.com

<sup>3</sup>armaghan.eshaghi@huawei.com

**Abstract:** Massive multiple-input multiple-output (MIMO) systems are considered as one of the leading technologies employed in the next generations of wireless communication networks (5G), which promise to provide higher spectral efficiency, lower latency, and more reliability. Due to the massive number of devices served by the base stations (BS) equipped with large antenna arrays, massive-MIMO systems need to perform high-dimensional signal processing in a considerably short amount of time. The computational complexity of such data processing, while satisfying the energy and latency requirements, is beyond the capabilities of the conventional widely-used digital electronics-based computing, i.e., Field-Programmable Gate Arrays (FPGAs) and Application-Specific Integrated Circuits (ASICs). In this paper, the speed and lossless propagation of light is exploited to introduce a photonic computing approach that addresses the high computational complexity required by massive-MIMO systems. The proposed computing approach is based on photonic implementation of multiply and accumulate (MAC) operation achieved by broadcast-and-weight (B&W) architecture. The B&W protocol is limited to real and positive values to perform MAC operations. In this work, preprocessing steps are developed to enable the proposed photonic computing architecture to accept any arbitrary values as the input. This is a requirement for wireless communication systems that typically deal with complex values. Numerical analysis shows that the performance of the wireless communication system is not degraded by the proposed photonic computing architecture, while it provides significant improvements in time and energy efficiency for massive-MIMO systems as compared to the most powerful Graphics Processing Units (GPUs).

© 2021 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

## 1. Introduction

The next generations of wireless communication networks, i.e., 5G and Beyond (5GB), are designed to accommodate a large number of smart devices, connected to each other, that are being served by base stations equipped with a massive number of antennas while the requirements of individual devices on the reliability, latency, and energy consumption are met [1,2]. Massive-MIMO systems are one of the key enablers of 5GB to provide high-data-rate and low-latency connectivity [3]. Such dense connectivity together with the demands for higher data rate and lower latency will increase the complexity, and accordingly, the cost of data processing in 5GB systems. In massive-MIMO systems, due to the large number of antennas at the BS and massive number of mobile devices in the network, parallel signal processing for missions such as channel estimation, precoding, and signal detection will become increasingly complex and time-consuming. This complexity has been considered as one of the main bottlenecks of realizing massive-MIMO systems. Accordingly, efficient methods for reducing this complexity and improving the efficiency of the radio transceiver architectures are required. Different optimization techniques have been proposed in the literature to address the 5GB challenges from both algorithmic and hardware implementation perspectives [4–8]. Moreover, machine learning techniques have been recently

exploited to reduce the ever-increasing computational complexity and to satisfy ultra-reliability, high data rate, and low latency requirements of 5G [9,10]. It has also been shown that a careful co-design of algorithms and hardware parameters can result in an even more energy-efficient signal processing in MIMO systems [11,12].

While the proposed methods can result in a less complex signal processing for massive-MIMO systems, the limitations imposed by the digital electronics-based hardware preclude full exploitation of the improvements offered by those algorithms. In particular, in order to provide the desired connectivity in 5G massive-MIMO systems, the base stations are potentially equipped with more than thousands of antennas to serve more than hundreds of mobile users in the network. In these scenarios, precoding the signals that are to be transmitted to the users or detecting the signals received from the users at the BS requires billions of MAC operations per second. For millimeter-wave (mmWave) 5G, in which the latency requirements limit the slot length to be as short as tens of millisecond, the required rate for MAC operations to perform tasks such as precoding or detection is in the order of hundreds of Tera MACs per second. This number can easily reach to tens of Peta MACs per second for beyond-5G wireless communications. In addition, due to the mobility of the users and the dynamics of the environment, the channel state information matrix, and accordingly, the precoding and/or detection matrix need to be updated frequently [13]. Such computational requirements can be hardly met with current power- and bandwidth-limited base stations with digital electronics-based processing units. Computing based on digital electronics hardware which is accompanied by components such as analog-to-digital (ADC) and digital-to-analog (DAC) converters faces fundamental challenges in terms of processing rate and energy consumption [14,15]. On the other hand, the fact that analog electronic components are frequency-dependent with poor reconfigurability in the radio frequencies (RF) limits their application in 5G systems [16].

Photonics-based computing has been proposed as a promising approach to provide high-performance and low-latency systems for large-scale signal processing [17–19]. In particular, an integrated optical platform comprising of both active elements such as modulators, lasers and photodetectors, and passive elements such as waveguides and couplers has shown orders of magnitude improvement in computation time and throughput as compared to the electronics counterparts [20]. Leveraging unique features of light, optical computing has been regarded as one of the emerging technologies to address the “von-Neumann bottleneck” [21].

One of the recently-developed photonics-based computing protocols is Broadcast-and-Weight (B&W) [22], in which wavelength-division multiplexing (WDM) scheme, a bank of microring modulators (MRM), and balanced photodetectors (PD) are utilized to implement weighted addition in a photonic platform. This photonic MAC unit can provide significant potential improvements over digital electronics in energy, processing speed, and compute density [20]. However, the proposed architecture [22] has limitations that makes it unqualified for communication systems. One of the most important limitations is that the B&W architecture can only realize real-valued vectors or matrices. Moreover, it is not capable of operating MAC over two negative-valued inputs. Finally, there exist constraints on the number of MRMs and parallel wavelength channels that can be realized in the system. Therefore, the typical large matrices in communication systems need to be partitioned in an efficient way before processing.

In this paper, we exploit the B&W architecture to develop a photonic computing platform that meets the stringent requirements of next-generation wireless communication systems, including massive-MIMO-enabled networks. The proposed photonic computing platform tackles the aforementioned limitations of the B&W architecture, and hence, it is capable of supporting wireless communication networks. In particular, we devise simple preprocessing steps by which inputs (vectors or matrices) with arbitrary values (real or imaginary, positive or negative) can fit into the proposed computing platform. Furthermore, by utilizing different algorithmic approaches such as matrix inversion approximation and parallelization techniques, the efficiency of the

proposed architecture for matrix inversion and large-size matrix multiplication is improved. Several numerical analyses show that while the performance of the proposed photonic architecture is comparable to the performance of the digital-electronics processing units such as GPUs, the time efficiency of the proposed architecture is significantly improved.

## 2. Massive-MIMO systems

### 2.1. System model

The computational complexity required for signal detection in a massive-MIMO system is formulated in this section and the capability of digital electronics and optics to address such computational requirement is explored accordingly. Consider a massive-MIMO uplink system with  $K$  single-antenna users and a BS equipped with  $M$  antennas, where the channels between the users and the BS are modelled as block Rayleigh fading channels. Moreover, over each channel coherence time, the channels are assumed to be perfectly known at the BS based on the channel estimation approach employed prior to data transmission phase [23–25]. If  $\mathbf{x} \in \mathbb{C}^{K \times 1}$  denotes the vector of the transmitted symbols, that are selected from finite modulation set  $\mathcal{S}$ , i.e.,  $\mathbf{x} = \{x_i | x_i \in \mathcal{S}\}$ , and  $\mathbf{n}$  denotes the white symmetric Gaussian noise, i.e.,  $n \sim \mathcal{CN}(0, \sigma^2)$ , then the received signal,  $\mathbf{y} \in \mathbb{C}^{M \times 1}$ , can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (1)$$

where matrix  $\mathbf{H} \in \mathbb{C}^{M \times K}$  denotes the channels between the users and the BS over each channel coherence time.

### 2.2. Signal detection in MIMO systems

After the signals transmitted by the users are received, the BS is to obtain the best estimation of the signal transmitted by each user,  $\hat{\mathbf{x}}$ , by solving the following optimization problem,

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|. \quad (2)$$

The optimal solution of the optimization problem in 2 can be obtained by the Maximum Likelihood (ML) detection. However, since the computational complexity of the ML detection increases exponentially by increasing the number of antennas, alternative detection schemes, such as Zero Forcing (ZF) or Minimum Mean-Square Error (MMSE) have been proposed in the literature [26]. Depending on the detection scheme that is employed by the massive-MIMO system, a detection matrix, namely  $\mathbf{A}$ , is constructed and used to detect the transmitted signal through linear processing as follows

$$\hat{\mathbf{x}} = \mathbf{A}\mathbf{y} = \mathbf{A}\mathbf{H}\mathbf{x} + \mathbf{A}\mathbf{n}, \quad (3)$$

where

$$\begin{cases} \mathbf{A} = (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H, & \text{in ZF,} \\ \mathbf{A} = (\mathbf{H}^H \mathbf{H} + \sigma^2 \mathbf{I})^{-1} \mathbf{H}^H, & \text{in MMSE.} \end{cases} \quad (4)$$

### 2.3. Computational complexity of signal detection in massive-MIMO systems

The complexity of signal processing in 5GB can be mostly attributed to computationally-complex matrix operations such as multiplications and inversions of matrices with large sizes (see 4). In order to measure the complexity of signal processing required in massive-MIMO systems, the number of Floating Point operations (FLOPs) [27] or the number of MAC operations can be calculated (note that each MAC operation includes two FLOPs). As can be seen in 4, the linear detection process for a single channel  $\mathbf{H}$  involves matrix inversions and matrix multiplications

which both have complexity of the order  $O(K^3)$  MACs (FLOPs). Accordingly, the computational complexity in a wireless communication system scales with the number of antennas at the BS, or with the number of users in the cellular network, or both. This is specifically a fundamental challenge for 5G, where massive-MIMO is an essential part of the development.

Furthermore, one of the key features of the next generations of wireless communications is the spectrum expansion by operating at mmWave frequencies which provides large transmission bandwidths and hence very high data rates. In order to support a wide range of spectrum from large cells with sub-1 GHz carrier frequency up to mmWave, 5G systems will adopt flexible subcarrier spacing from 15 kHz up to 480 kHz in an Orthogonal Frequency Division Multiplex (OFDM) scheme [28], which can decrease the slot duration down to 30  $\mu$ s [28,29]. Consider an OFDM-based mmWave massive-MIMO system with 3300 subcarriers in which a BS equipped with more than a thousand antennas, i.e.,  $M = 1024$ , is serving more than a hundred mobile users, i.e.,  $K = 128$ . Such system is supposed to perform the signal detection (see 4), i.e., one million MAC operations, for each subchannel in tens of microsecond time duration. That implies that the required MAC operations rate to complete the detection task would be in the order of 400 TMACs/s. Accordingly, the BS must be capable of processing the data at such a rate while keeping the power consumption and the size of the data processing unit equitable.

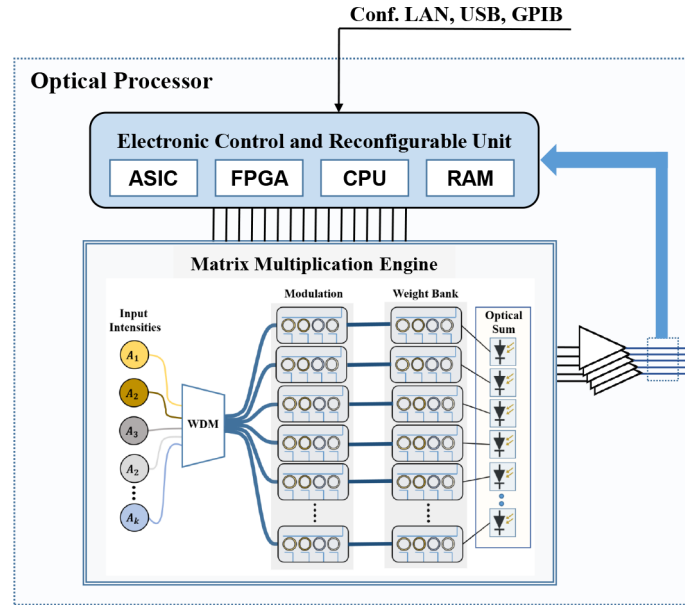
### 2.3.1. Digital and analog electronics

In recent years, GPUs and FPGAs have been developed to encompass general purpose tasks in the high-performance computing arena. The most powerful GPU architecture to date is *NVIDIA VOLTA* with 640 tensor cores and 21 billion transistors which can deliver more than 50 TMACs/S. However, tens of these units are needed to meet the required computational power. In these systems, I/O latency and sequential processing capabilities cannot exceed the time resolution of the processor which is ultimately bounded by its clock rate.

To tackle the speed limitation of digital-electronics-based processors while maintaining a reasonable area and power consumption, an optical computing approach is proposed in this paper as a revolutionary computing paradigm. It allows very complex operations to be performed in real time, which can significantly offload electronic post-processing and provide a technology to make RF decisions on-the-fly.

## 3. Photonic computation for massive-MIMO systems

The architecture of the proposed photonic computing platform for ultra-fast signal processing in the next generations of wireless communication networks is depicted in Fig. 1. The processor core which is based on the B&W architecture [22] is a photonic-integrated circuit (PIC) fabricated on a silicon photonics (SiPh) platform. It contains a matrix-multiplication engine where the input vectors are loaded using microring resonators (MRRs) in the modulation and weight bank sections. The photodetector array performs the optical summation. Wavelength-division multiplexing scheme is adopted where all optical inputs, spatially separated by wavelength, lies in a single waveguide. The Electronic Control and Reconfigurable Unit (ECRU) is composed of interconnected ASIC, FPGA, central processing unit (CPU) and random-access memory (RAM) modules. Its main function is to generate analog control signals for setting the weights of the microring photonic modulators. Another important feature of the ECRU unit is to make sure that the processor core is well calibrated by correcting for the fabrication variations and regulating the controls signals against any thermal fluctuations. By means of the General-Purpose Input/Output (GPIO) or Universal Serial Bus (USB), the ECRU unit maintains a high-bandwidth communication link with a computer motherboard.



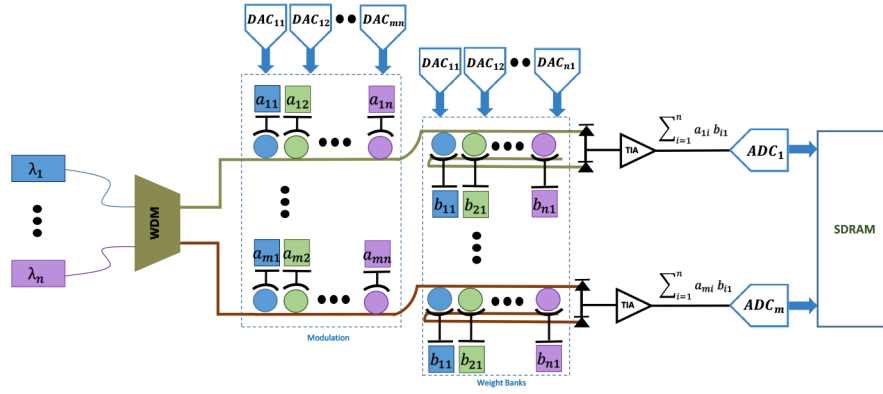
**Fig. 1.** Schematic of the proposed photonic computing platform.

### 3.1. Matrix-multiplication engine

As depicted in Fig. 2, in order to implement matrix multiplication, i.e.  $\mathbf{A} \times \mathbf{B}$ , where  $\mathbf{A} \in \mathbb{R}_+^{m \times n}$  and  $\mathbf{B} \in \mathbb{R}^{n \times k}$ , in the proposed optical architecture, the elements of the first vector-to-be-multiplied are loaded using all-pass MRMs and are encoded in the intensities of the wavelength-multiplexed signals (Modulation section). The elements of the second vector-to-be-multiplied are encoded as weights using add-drop MRMs (Weight Bank section). The interfacing of optical components with electronics are facilitated by the use of mixed-signal integrated circuit blocks such as DACs and ADCs, integrated inside the ASIC in the ECRU unit. In Fig. 2, it is assumed that there is sufficient number of parallel waveguide channels and MRRs in the optical computation unit such that the matrix-to-vector multiplication  $\mathbf{A}_{m \times n} \times \mathbf{b}_{n \times 1}$  can be completed in a single usage of the computation unit. The more general case, where multiple usages of the computation unit are required, is discussed in Section 3.3.3. We would like to note that incorporating a large number of MRMs (and equivalently a large number of wavelengths) in the architecture can lead to crosstalk which is caused by small Free Spectral Range (FSR) of the ring resonators. One of the solutions to address this issue is to utilize the wavelength-fixed intensity-tuning modulation scheme [30,31].

The multiplication is performed by linking the elements of the first and the second vectors via an optical waveguide, and the accumulation is performed by the photodetector followed by a transimpedance amplifier (TIA) to provide electronic gain, which is also integrated in the same ASIC. For heterogeneous integration, the different analog and digital electronic control circuitry such as ADCs, DACs, TIAs are fabricated in a standard complementary metal–oxide–semiconductor (CMOS) process and interfaced with the corresponding SiPh chip by means of wire-bonding or flip-chip bonding. The MRMs are controlled by the DACs, while the interfacing with ADC is required to compute the digital representation of the analog output, which can then be stored in the SDRAM and processed by the CPU or FPGA.

In our numerical simulations, the number of wavelength channels is considered to be equal to the number of rows of the left-hand-side (LHS) matrix,  $m$ , and we set the number of MRRs in the



**Fig. 2.** Implementing matrix-to-vector multiplication in a B&W architecture with  $D = m$  parallel waveguide channels and  $2R = 2n$  MRRs at each channel.

Modulation and Weight Bank sections, equal to the number of columns of the LHS matrix or number of rows of the right-hand-side (RHS) matrix,  $n$ .

### 3.2. Photonic matrix inversion

In addition to matrix multiplication, matrix inversion is a widely-used operation in wireless communication networks. In 5GB, the inverse of an arbitrary, potentially large, and complex matrix needs to be calculated on-the-fly. Although the computational complexity of matrix multiplication and matrix inversion are in the same order, matrix multiplication is preferred from the hardware implementation perspective [13]. Accordingly, several algorithms have been proposed in the literature to approximate the inverse of a matrix with a set of matrix multiplications, e.g., conjugate gradients [32] and iterative algorithms such as Newton [33] and Neumann series method [34,35]. Cholesky factorization can also be considered as another technique that can be used to implement matrix inversion with a number of matrix multiplications [11]. Iterative algorithms can outperform other approaches by taking advantage of reusing available resources in the processing unit, and hence are considered to be more hardware-friendly [36].

In massive-MIMO systems calculating the inverse of the Gram matrix,  $\mathbf{H}^H \mathbf{H}$ , is an essential operation. In a massive-MIMO system with  $K$  users and  $M$  antennas at the BS, where  $M \gg K$ , the Gram matrix becomes diagonally dominant, and that leads to an accurate approximation of inverse matrix in both Newton and Neumann series approximations.

In the following, calculating the inverse of matrix  $\mathbf{A} \in \mathbb{C}^{K \times K}$  using Newton iterative method and Neumann-series approximation techniques is explained.

#### 3.2.1. Neumann-series approximation

The Neumann-series expansion of the inverse of an arbitrary matrix  $\mathbf{Z}$  is given as [34,35]

$$\hat{\mathbf{Z}}^{-1} = \sum_{n=0}^{\infty} (\mathbf{X}^{-1}(\mathbf{X} - \mathbf{Z}))^n \mathbf{X}^{-1}, \quad (5)$$

where  $\mathbf{X}$  is a pre-conditioning matrix, and  $\hat{\mathbf{Z}}^{-1}$  is guaranteed to converge to the exact inverse of matrix  $\mathbf{Z}$  when

$$\lim_{n \rightarrow \infty} (\mathbf{I} - \mathbf{X}^{-1} \mathbf{Z})^n = \mathbf{0}. \quad (6)$$

The asymptotically orthogonal column vectors of the channel matrix  $\mathbf{H}$  in massive-MIMO systems result in a Gram matrix, i.e.,  $\mathbf{Z} = \mathbf{H}^H \mathbf{H}$ , which is diagonally dominant [11]. It has been



shown in [37] that, considering the diagonally-dominant structure of the Gram matrix, matrix  $\mathbf{X}$  in 6 can be chosen as a diagonal matrix with diagonal elements equal to the diagonal elements of the Gram matrix. In that case, if matrix  $\mathbf{Z} = \mathbf{H}^H \mathbf{H}$  is rewritten as

$$\mathbf{Z} = \mathbf{Z}_{\text{diag}} + \mathbf{Z}_{\text{off-diag}}, \quad (7)$$

where matrix  $\mathbf{Z}_{\text{diag}}$  is a diagonal matrix with diagonal elements of  $\mathbf{Z}$  and matrix  $\mathbf{Z}_{\text{off-diag}}$  holds all elements of the matrix  $\mathbf{Z}$  and has zeros on the main diagonal, the  $K$ -term Neumann series approximation is

$$\hat{\mathbf{Z}}_K^{-1} = \sum_{n=0}^L (-\mathbf{Z}_{\text{diag}}^{-1} \mathbf{Z}_{\text{off-diag}})^n \mathbf{Z}_{\text{diag}}^{-1}. \quad (8)$$

It has been shown that in massive-MIMO systems, the Gram matrix inversion approximation in 8 converges to the exact matrix inversion value after  $L = 5$  iterations [36].

### 3.2.2. Newton approximation method

For an arbitrary invertible matrix  $\mathbf{Z}$ , with an initial rough estimation of its inverse  $\mathbf{X}_0^{-1}$ , the estimated inverse matrix at the  $n^{\text{th}}$  iteration of Newton approximation technique is [33]

$$\mathbf{X}_n^{-1} = \mathbf{X}_{n-1}^{-1} (2\mathbf{I} - \mathbf{Z}\mathbf{X}_{n-1}^{-1}), \quad (9)$$

where  $\mathbf{Z}_{\text{diag}}^{-1}$ , which is a diagonal matrix with diagonal elements of  $\mathbf{Z}$ , can be used as the first rough estimation,  $\mathbf{X}_0^{-1}$ . The main advantage of the Newton method is that it converges quadratically to the inverse matrix if  $\|\mathbf{I} - \mathbf{Z}\mathbf{X}_0^{-1}\| < 1$ . This condition is satisfied for the diagonally-dominant Gram matrix in the uplink data detection in massive-MIMO systems, and it has been shown [36] that the Newton approximation method converges to the exact value of the inverse matrix through 5 or 6 iterations.

### 3.3. Algorithm-hardware co-design for photonic computing

In B&W architecture, the elements of the LHS matrix are encoded into the light intensities. This implies that only real positive values can be realized in this architecture and it limits the application of the B&W-based photonic MAC in a variety of cases including wireless communication networks. In order to tackle this issue, in the following sections, preprocessing steps are proposed such that any arbitrary matrix can be represented by the optical architecture.

#### 3.3.1. Preprocessing step 1: Addressing complex-valued matrices

In order to represent complex-valued matrices with the proposed photonic computing platform, the real representation of complex-valued matrices is explored. Any arbitrary complex-valued matrix,  $\mathbf{A} \in \mathbb{C}^{m \times n}$ , can be written as the summation of the real and imaginary parts,

$$\mathbf{A} = \mathbf{A}_r + j\mathbf{A}_i, \quad (10)$$

where  $\mathbf{A}_r \in \mathbb{R}^{m \times n}$  and  $\mathbf{A}_i \in \mathbb{R}^{m \times n}$  are real-valued matrices denoting the real and imaginary parts of  $\mathbf{A}$ , respectively. According to 10, the multiplication of two complex-valued matrices, namely  $\mathbf{A} \in \mathbb{C}^{m \times n}$  and  $\mathbf{B} \in \mathbb{C}^{n \times k}$ , can be obtained as

$$\mathbf{A} \times \mathbf{B} = (\mathbf{A}_r + j\mathbf{A}_i) \times (\mathbf{B}_r + j\mathbf{B}_i) \quad (11a)$$

$$= \underbrace{(\mathbf{A}_r \times \mathbf{B}_r - \mathbf{A}_i \times \mathbf{B}_i)}_{\text{real part}} + j \underbrace{(\mathbf{A}_r \times \mathbf{B}_i + \mathbf{A}_i \times \mathbf{B}_r)}_{\text{imaginary part}}. \quad (11b)$$

Therefore, in order to multiply two complex-valued matrices, four parallel real-valued matrix multiplications of the same size as that of the original matrices can be considered.

### 3.3.2. Preprocessing step 2: Addressing negative-valued matrices

The proposed solution to represent negative-valued LHS matrices in this architecture is to project the negative sign of the elements of the LHS matrix,  $\mathbf{A}$ , to the sign of the RHS matrix,  $\mathbf{B}$ . In doing so, the negative-valued matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is rewritten as a subtraction between two positive-valued matrices,  $\bar{\mathbf{A}} \in \mathbb{R}_+^{m \times n}$  and  $|a_{\min}|\mathbf{1}$ , where  $a_{\min}$  is the element of matrix  $\mathbf{A}$  with the smallest value, and  $\mathbf{1}$  is an all-one matrix of size  $m \times n$ . The negative sign of the subtraction can then be projected into the sign of the RHS matrix elements as follows

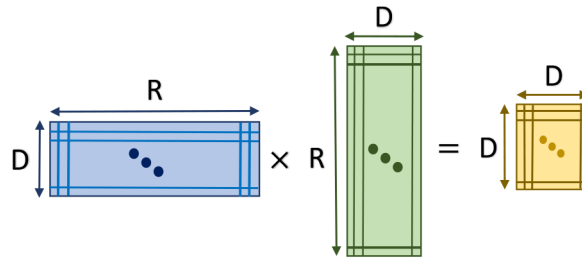
$$\mathbf{A} \times \mathbf{B} = (\bar{\mathbf{A}} - |a_{\min}|\mathbf{1}) \times \mathbf{B} = \bar{\mathbf{A}} \times \mathbf{B} + |a_{\min}|\mathbf{1} \times (-\mathbf{B}). \quad (12)$$

Therefore, negative-valued matrix multiplication can be performed by summation of two positive-valued matrix multiplications which can be processed in parallel in the proposed photonics-based computing architecture.

### 3.3.3. Preprocessing step 3: Parallelization and matrix tiling based on the photonic computing architecture

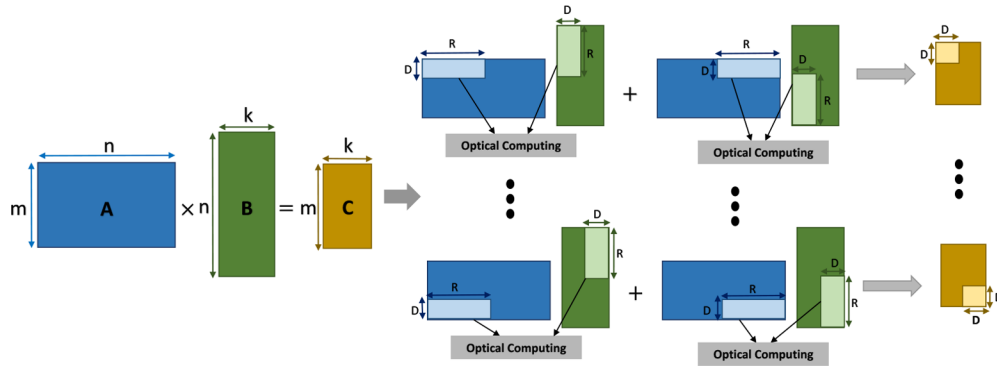
The last preprocessing step is proposed to implement parallelization and matrix tiling methods. This step is required so that matrices with any arbitrary size can be processed efficiently using the proposed photonic computing system with limited number of cascaded modulators and parallel channels.

Consider a computing unit with  $D$  parallel wavelength channels,  $R$  all-pass MRRs in the Modulation section (representing LHS matrix), and accordingly,  $R$  add-drop MRRs in the Weight Bank section (representing RHS matrix) (Fig. 2). Each single usage of such architecture can compute multiplication of matrices with sizes that lie within the range of parameters  $D$  and  $R$ ; see Fig. 3. In order to optimally utilize this architecture to perform the multiplication between arbitrary matrices,  $\mathbf{A} \in \mathbb{R}_+^{m \times n}$  and  $\mathbf{B} \in \mathbb{R}_+^{n \times k}$ , the matrices need to be partitioned based on  $D$  and  $R$ , following the mapping discussed in Section 3.1. The results of the partial multiplications are recorded in the memory and in the last step, corresponding parts are added together to generate the final result. Figure 4 illustrates the implementation of the multiplication of matrices  $\mathbf{A}$  and  $\mathbf{B}$ , in which, without loss of generality, it is assumed that  $m$  and  $n$  are dividable into  $D$  and  $R$ , respectively.



**Fig. 3.** Size of the LHS and RHS matrices supported for multiplication using an architecture with  $D$  parallel waveguide channels and  $2R$  MRRs in each channel.





**Fig. 4.** Matrix tiling facilitates multiplication of matrices with any arbitrary size.

#### 4. Numerical analysis

The time and power efficiency of the proposed photonic computing approach mainly depend on the number of parallel multiplexed waveguide channels,  $D$ , and the number of different wavelengths that can be realized in each of those channels,  $R$ . An optimally designed MRR is capable of supporting up to 108 WDM channels, taking both the finesse of the resonator and the channel spacing in linewidth-normalized units into account [38]. Here, we consider a finesse of  $\mathcal{F} = 368$  for the MRRs, and a minimum channel spacing of  $3.41 \times$  linewidth, based on the assumption that a 3dB cross-weight penalty is allowed [30]. Accordingly, the maximum number of the MRRs, in each of the modulation and weight-bank parts, is considered as  $R = 100$  to provide an insight into the ultimate capabilities of the system. However, we acknowledge that realizing 100 WDM channels, which is equal to the number of MRRs in each part of the system, is a challenging task. Particularly, if the MRRs have a high quality factor (Q-factor), high-speed EO-modulation is difficult to achieve due to the long oscillation time in the cavity, and the MRRs become very sensitive to the peripheral temperature fluctuations.

An architecture with  $D$  multiplexed waveguide channels contains  $2RD$  MRRs in total. However, assuming that a maximum of 1024 MRRs can be manufactured in the optical architecture [39], there is a finite set of feasible values for each of  $D$  and  $R$ . The optimal arrangement of the number of parallel waveguide channels and the number of MRRs fundamentally depends on the computational and energy requirements.

In this section, the performance and the efficiency of the proposed photonic computing platform in different practical scenarios are evaluated and compared to those of the conventional digital electronics-based processing units.

##### 4.1. Power consumption

The total power consumption of the proposed photonic computing architecture can be obtained by adding up the power usage of different photonic and electronic components. The power consumption of the lasers is mainly determined by i) the number of lasers utilized in the architecture, and ii) the minimum power that can be detected by the photodetector. In particular, the number of lasers follows the number of wavelengths in the WDM scheme,  $R$ . Furthermore, the laser power must be detectable by the photodetector after splitting among all  $D$  parallel waveguide channels. In this work, lasers are considered to have 100 mW power consumption, and their optical power is sufficiently large to be detectable by the photodetectors after splitting into  $D$  parallel channels.

The architecture contains  $2DR$  MRRs and  $2DR$  DACs with 19.5 mW [40] and 26 mW [41] power consumption, respectively. The utilized DACs have 5 bits resolution operating at the

conversion rate of 10 GS/s, and are implemented in 0.18- $\mu\text{m}$  SiGe BiCMOS process [41]. Finally, the TIAs with 17 mW [39] and the ADCs each with 76 mW [42] power usage are integrated at the output of each waveguide. The ADCs utilized in the proposed architecture are 1V-supplied design fabricated in 65-nm CMOS technology by aggregating four 8-bit two-stage time-domain ADCs operating at a conversion rate of 10 GS/s [42]. Hence, an upper bound for the total power consumption can be obtained as

$$P_{\text{Total}}(mW) = 100R + 91DR + 93D. \quad (13)$$

Using 13, the power usage of the proposed architecture is calculated and reported in Table 1, along with those of other computation hardware baselines [39]. The results show that the power consumption of the photonic system is close to that of digital processing hardware such as GPUs. Furthermore, based on the computational requirements of the task to be executed, the power consumption of the photonic computing system can be 1/3 of that of the best digital electronic processors in the literature.

**Table 1. Power consumption of the proposed photonic computing system with different parameters, and the benchmarked GPUs**

GPU	Power Usage (W)
AMD Vega FE	375
AMD M125	300
NVIDIA Tesla V100	250
NVIDIA GTX 1080 Ti	250
Photonic System ( $D = 32, R = 32$ )	100
Photonic System ( $D = 64, R = 32$ )	195
Photonic System ( $D = 64, R = 64$ )	385

#### 4.2. Time efficiency

In order to evaluate the time efficiency of the proposed computing platform, the latency and the throughput of the whole system have been investigated. Propagation time (delay) is the time that it takes for the input data to propagate through the system to produce the result, while the throughput of the computing system is the number of MAC operations executed per unit of time, which mainly depends on the bandwidth of the components.

The propagation time of the proposed photonic core can be calculated as the total time that it takes for the light to propagate through different components in the architecture. The propagation time after multiplexing, when  $2R$  MRRs are considered in each waveguide channel, is estimated by

$$t_p = \frac{2r_{\text{MRR}} \times 2R + 2 \times 2\pi r_{\text{MRR}} \times (\mathcal{F}/2\pi)}{(c/n_{\text{eff}})}, \quad (14)$$

where  $r_{\text{MRR}}$  is the radius and  $\mathcal{F}$  is the finesse of the MRRs. Moreover, the parameters  $c$  and  $n_{\text{eff}}$  denote the speed of light and the effective refractive index of the waveguide, respectively. For an architecture with  $2R$  cascaded MRRs, light propagates through the shared bus waveguide with the minimum length of  $2r_{\text{MRR}} \times 2R$  and will be trapped by the in-resonance MRRs (only two in total, one in the modulation section and one in the weigh-bank section)  $\mathcal{F}/2\pi$  times [43]. Accordingly, considering  $2R = 200$ ,  $r_{\text{MRR}} = 10 \mu\text{m}$ ,  $\mathcal{F} = 368$ , and  $n_{\text{eff}} = 2.4$ , the total propagation time through MRRs in the modulation and weight bank parts is 110 ps. While this number shows the data propagation time, the throughput demonstrates how much data can be processed by the computing architecture within a given time frame. The throughput of the proposed computing

platform depends on the bandwidths of different components which are listed in Table 2 along with their latencies [39]. The throughput of the proposed computing architecture is limited by the component with the lowest bandwidth. As shown in Table 2, ADCs, DACs, and TIAs have the minimum bandwidth equal to 10 GHz. Hence, the average processing time of the optical computing architecture, limited by these components, can be obtained as  $T_{\text{single use}} = \frac{1}{10 \text{ GHz}} = 100 \text{ ps}$ .

**Table 2. The bandwidth and latency of different components in the proposed architecture. \*SDRAM is connected to a computer and is considered as digital memory before DACs and after ADCs.**

Component	Bandwidth (GHz)	Latency (ps)
MRR	60	17
ADC	10	100
DAC	10	100
Balanced PD	25	40
TIA	10	100
GDDR6 SDRAM*	16	60

In order to obtain the total processing time for multiplication of matrices  $\mathbf{A} \in \mathbb{C}^{m \times n}$  and  $\mathbf{B} \in \mathbb{C}^{n \times k}$ , the number of times that the architecture is used to calculate the final result needs to be calculated which depends on the dimensions of the matrices. Moreover, the upper-bound of the processing time is obtained by assuming that both the first and the second preprocessing steps in Section 3.3 are required. Considering an optical chip with  $D$  parallel waveguide channels, and  $R$  MRRs to represent the corresponding elements of each of the matrices, the number of times that chip should be used to compute the multiplication of  $\mathbf{A}$  and  $\mathbf{B}$ , namely  $N_{\text{use}}$ , can be obtained (see Fig. 4) as

$$N_{\text{use}} = 8 \times k \times \left\lceil \frac{m}{D} \right\rceil \times \left\lceil \frac{n}{R} \right\rceil, \quad (15)$$

and accordingly, the total processing is calculated as

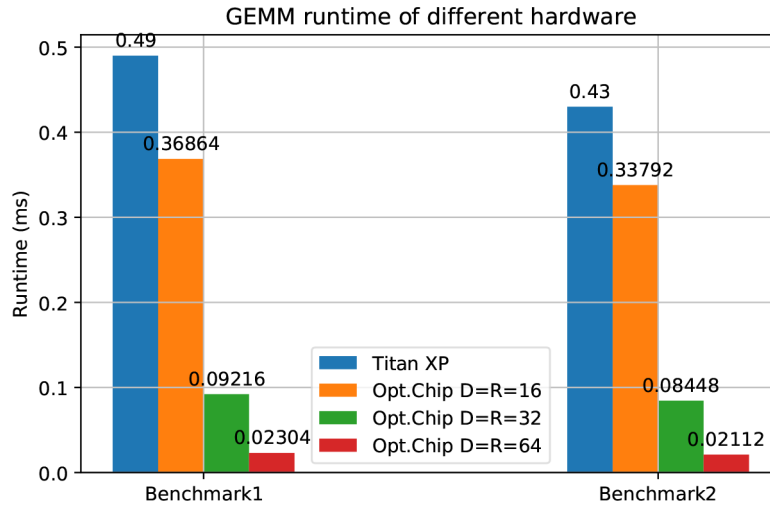
$$T_{\text{total}}(\text{ps}) = N_{\text{use}} \times T_{\text{single use}} = 8 \times k \times \left\lceil \frac{m}{D} \right\rceil \times \left\lceil \frac{n}{R} \right\rceil \times 100. \quad (16)$$

The total processing time in 16 and the power usage calculated in 13 show that designing the optimal architecture requires optimization over  $D$  and  $R$  parameters considering the requirements of the target application.

In this work, General Matrix Multiplication (GEMM) is considered as a benchmark to evaluate the computation speed of the proposed architecture. Figure 5 illustrates the computation time of the proposed photonic system as compared to the runtime of *Titan XP* GPU [44]. The parameters of the benchmarked scenarios are the dimensions of the matrices as listed in Table 3. As shown in Fig. 5, by increasing the number of parallel waveguide channels and/or the number of MRMs in the proposed architecture, the processing time can be reduced notably. Moreover, Fig. 5 highlights the fact that the processing time of the photonic computing platform is lower than that of *Titan XP* GPU, and the gap between the processing times exponentially increases when the photonic architecture scales up.

#### 4.3. Massive-MIMO wireless communication systems

In this section, the performance and the efficiency of the proposed photonic computing system when employed in massive-MIMO wireless communication systems is studied. For this purpose, MMSE signal detection in an uplink scenario, where  $K$  single-antenna users transmit their signals to a BS equipped with  $M$  antennas, is considered. The Channel State Information (CSI) is



**Fig. 5.** Runtime of the proposed photonic computing architecture as compared to Titan XP GPU [44].

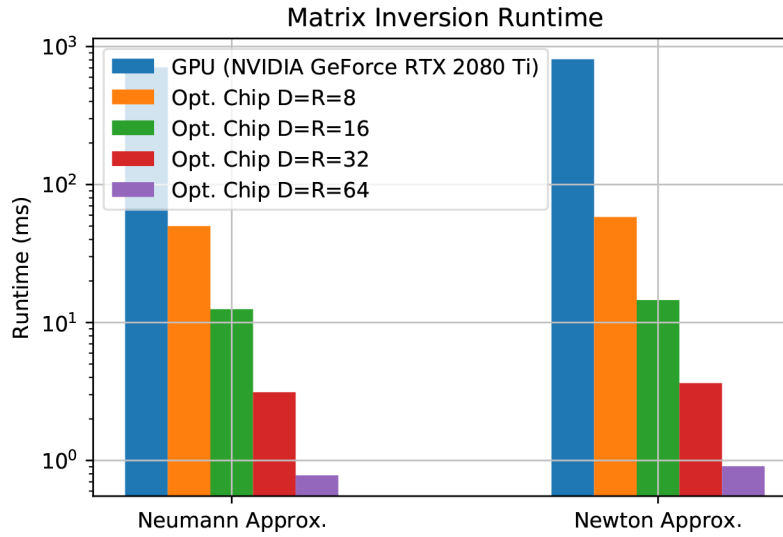
**Table 3. Benchmarked parameters for  $A_{m \times n} \times B_{n \times k}$ .**

Parameters	$m$	$n$	$k$
Benchmark 1	7680	1500	2560
Benchmark 2	10752	1	3584

assumed to be perfectly known at the transmitter and the receiver. The performance, in terms of the Symbol Error Rate (SER), and the processing time are compared with one of the most powerful GPUs, namely *NVIDIA GeForce RTX 2080 Ti*.

#### 4.3.1. Time efficiency analysis

In this numerical study, we consider a massive-MIMO system that consists of a 1024-antenna BS which is serving  $K = 64$  users. Accordingly, the size of the Gram matrix that we seek to compute the inverse of is  $64 \times 64$ . The processing time of this system for different parameters of the photonic system, i.e., different numbers of waveguide channels and different numbers of MRRs in each channel, is evaluated using 16. Additionally, the computation time for two matrix inversion approximation methods, namely Neumann-series and Newton approximations, explained in Sections 3.2.1 and 3.2.2, is reported. In both approximation approaches, we consider 5 iterations over which the methods have been shown to converge to the exact value of the inverse matrix; see Section 3.2.1. As shown in Fig. 6, the processing time associated with the proposed photonic computing system is significantly less than that of GPU in both approximation approaches. Furthermore, as expected, increasing the number of parallel waveguide channels and modulators in each of those channels can further reduce the runtime. However, other limitations must be taken into account when increasing the number of components in the photonic system. It can be seen in Fig. 6 that the computation time of Neumann-series approximation is marginally lower than that of the Newton approximation method. Moreover, the study in [36] shows that both methods perform equally in terms of SER. Hence, Neumann approximation method is adopted in the next numerical experiments.



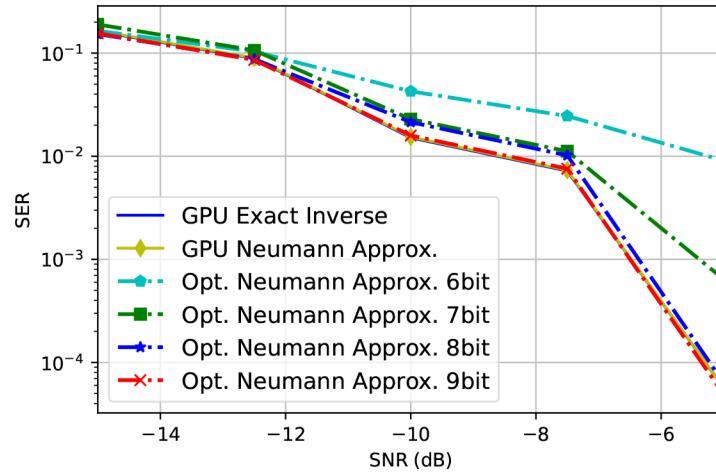
**Fig. 6.** Matrix inversion processing time using Neumann series and Newton approximation approaches.

#### 4.3.2. Precision analysis

In this section, we seek to evaluate the impact of the modulator control precision on the performance of the proposed optical platform. In order to consider the practical imperfections caused by the limited precision of the MRRs in the numerical analysis, we first normalize the elements of the matrices and then quantize them to the realizable values. Those realizable values are obtained based on the precision bits of the MRRs, and are uniformly distributed within  $[0,1]$  and  $[-1,1]$  for the elements in the LHS and RHS matrices, respectively. The normalization factors are applied to the final results through the utilized TIAs at the output of each waveguide channel. We consider a MIMO system with  $K = 8$  users being served with a BS equipped with  $M = 64$  antennas. The signal detection process in the BS is performed utilizing a photonic system with  $D = 8$  parallel waveguide channels, each of which with  $2R = 16$  MRRs. Figure 7 illustrates the SER for different precisions over  $10^5$  channel realizations. (In that figure GPU Exact indicates the performance of the GPU when the matrix inversion is computed using the built-in functions in the considered software, rather than approximating that using Neumann approximation method.) It can be seen that in the low-SNR regime, where the effect of transmission noise is dominant, the performance of the proposed photonic computing platform is similar to that of the GPU, regardless of the precision bits. However, as the SNR increases, the low-precision error becomes more dominant. As shown in Fig. 7, for relatively higher SNR values, there is a notable gap between the performance of the photonic computing platform with only 6-bit modulator precision and that of the GPU. However, photonic computing can achieve the same performance as that of the GPU when the precision reaches to 8 bits (plus the sign bit). Therefore, in the next numerical experiments, an 8-bit precision is considered for the performance analysis.

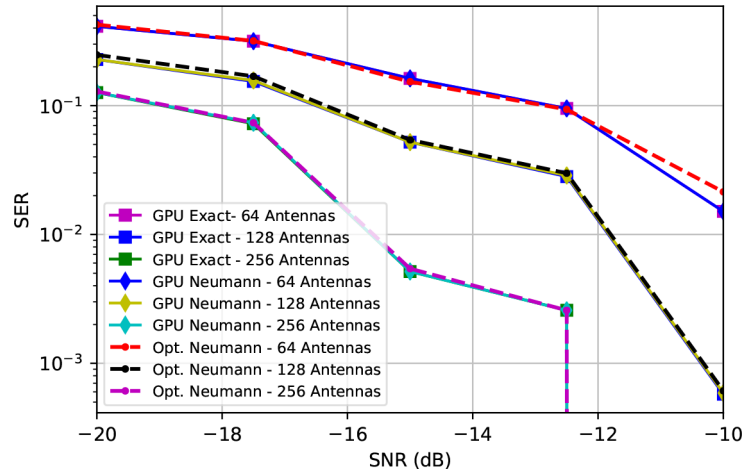
#### 4.3.3. Performance analysis

To evaluate the performance of the proposed photonic computing system when the number of BS antennas increases in a massive-MIMO wireless communication network, an uplink system with  $K = 8$  users and BS with different antenna array length is modelled. The signal detection in the BS is performed using MMSE detection method. Figure 8 summarizes the simulation results for a photonic computing system with  $D = 8$  parallel waveguide channels,  $2R = 16$  MRRs



**Fig. 7.** SER for MMSE detection in GPU and the proposed photonic computing platform for different precision bits. Neumann-series approximation is used for matrix inversion.

in each channel, and 8-bit precision (plus the sign bit) over  $10^5$  channel realizations. It can be seen in Fig. 8 that the proposed photonic computing platform has comparable performance to the GPU for all SNR values and different number of antennas. moreover, as expected, increasing the number of antenna elements (antenna array gain) can significantly improve the performance of the system in both cases. (Note that by increasing SNR, SER improves to the extent that for SNR greater than  $-14$  dB, no symbol error has been found.)



**Fig. 8.** SER for MMSE detection in GPU and the proposed photonic computing platform for different number of BS antennas. Neumann-series approximation is used for matrix inversion.

Considering the above evaluation of the performance, power consumption, and the processing time of one of the most powerful GPUs, i.e., *NVIDIA GeForce RTX 2080 Ti*, and those of the proposed optical computing platform, it can be seen that optical computing can provide the same performance in a relatively lower power consumption as conventional GPUs, while it can significantly reduce the processing time. That makes the proposed optical architecture as a promising candidate which is capable of supporting high computational complexity required in



the next generations of wireless communication networks, while it can meet low latency and high reliability requirements of those systems.

## 5. Conclusion

In this paper, a photonic computing architecture is proposed to be employed in next-generation massive-MIMO wireless communication systems to address their stringent computational requirements. The proposed computing approach is based on the B&W architecture to implement MAC operations in the optical domain exploiting the light speed and lossless propagation. Preprocessing steps are developed so that the proposed computing system can represent matrices with any arbitrary values. Numerical experiments confirm that the proposed photonic computing architecture can offer the same performance as those of the most powerful digital electronics-based data processing units such as GPUs, while its time efficiency is shown to be several orders of magnitude better than that of the modern state-of-the-art GPUs. Based on the simulation results, the proposed photonic platform can be integrated into the 5GB base stations as a power- and cost-efficient solution to enable ultra-fast data processing for the next-generation wireless communication networks.

**Acknowledgments.** The authors would like to thank Dr. Bhavin J. Shastri, assistant professor at Queen's University, Ontario, Canada, for the fruitful discussions that the authors had with him.

**Disclosures.** The authors declare no conflicts of interest.

**Data availability.** Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

## References

1. A. Gupta and R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access* **3**, 1206–1232 (2015).
2. W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network* **34**(3), 134–142 (2020).
3. J. Navarro-Ortiz, P. Romero-Diaz, S. Sendra, P. Ameigeiras, J. J. Ramos-Munoz, and J. M. Lopez-Soler, "A survey on 5G usage scenarios and traffic models," *IEEE Commun. Surveys Tuts.* **22**(2), 905–929 (2020).
4. C. H. Doan, S. Emami, D. A. Sobel, A. M. Niknejad, and R. W. Brodersen, "Design considerations for 60 GHz CMOS radios," *IEEE Commun. Mag.* **42**(12), 132–140 (2004).
5. H. Prabhu, J. Rodrigues, L. Liu, and O. Edfors, "Algorithm and hardware aspects of pre-coding in massive MIMO systems," *49th Asilomar Conf. Signals, Syst. Computers*, (IEEE, 2015), pp. 1144–1148.
6. F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.* **10**(3), 501–513 (2016).
7. R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.* **10**(3), 436–453 (2016).
8. K. Li, R. R. Sharan, Y. Chen, T. Goldstein, J. R. Cavallaro, and C. Studer, "Decentralized baseband processing for massive MU-MIMO systems," *IEEE Trans. Emerg. Sel. Topics Circuits Syst.* **7**(4), 491–507 (2017).
9. Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao, "Application of machine learning in wireless networks: Key techniques and open issues," *IEEE Commun. Surveys Tuts.* **21**(4), 3072–3108 (2019).
10. M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Commun. Surveys Tuts.* **21**(4), 3039–3071 (2019).
11. L. Van der Perre, L. Liu, and E. G. Larsson, "Efficient DSP and circuit architectures for massive MIMO: State of the art and future directions," *IEEE Trans. Signal Process.* **66**(18), 4717–4736 (2018).
12. P. Zhang, L. Liu, G. Peng, and S. Wei, "Large-scale MIMO detection design and FPGA implementations using SOR method," in *8th IEEE Int. Conf. Commun. Softw. Netw.* (IEEE, 2016), pp. 206–210.
13. H. Prabhu, O. Edfors, J. Rodrigues, L. Liu, and F. Rusek, "Hardware efficient approximative matrix inversion for linear pre-coding in massive MIMO," in *IEEE Int. Symp. Circuits Syst.* (IEEE, 2014), pp. 1700–1703.
14. T. Sundstrom, B. Murmann, and C. Svensson, "Power dissipation bounds for high-speed Nyquist analog-to-digital converters," *IEEE Trans. Circuits Syst., I, Reg. Papers* **56**(3), 509–518 (2009).
15. P. Jebashini, R. Uma, P. Dhavachelv, and H. K. Wye, "A survey and comparative analysis of multiply-accumulate (MAC) block for digital signal processing application on ASIC and FPGA," *J. Appl. Sci.* **15**(7), 934–946 (2015).
16. R. K. Mongia, J. Hong, P. Bhartiya, and I. J. Bahl, *RF and microwave coupled-line circuits*, Artech house (2007).
17. A. Mekis, S. Gloeckner, G. Masini, A. Narasimha, T. Pinguet, S. Sahni, and P. De Dobbelaere, "A grating-coupler-enabled CMOS photonics platform," *IEEE J. Sel. Topics Quantum Electron.* **17**(3), 597–608 (2011).

18. J. Capmany, J. Mora, I. Gasulla, J. Sancho, J. Lloret, and S. Sales, "Microwave photonic signal processing," *J. Lightwave Technol.* **31**(4), 571–586 (2013).
19. Y. A. Vlasov, "Silicon CMOS-integrated nano-photonics for computer and data communications beyond 100G," *IEEE Commun. Mag.* **50**(2), s67–s72 (2012).
20. T. F. De Lima, B. J. Shastri, A. N. Tait, M. A. Nahmias, and P. R. Prucnal, "Progress in neuromorphic photonics," *Nanophotonics* **6**(3), 577–599 (2017).
21. M. Miscuglio and V. J. Sorger, "Photonic tensor cores for machine learning," arXiv preprint arXiv:2002.03780 (2020).
22. A. N. Tait, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Broadcast and weight: An integrated network for scalable photonic spike processing," *J. Lightwave Technol.* **32**(21), 4029–4041 (2014).
23. E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.* **54**(2), 114–123 (2016).
24. A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE J. Sel. Topics Signal Process.* **8**(5), 831–846 (2014).
25. F. Bellili, F. Sotiraki, and W. Yu, "Generalized Approximate Message Passing for Massive MIMO mmWave Channel Estimation With Laplacian Prior," *IEEE Trans. Commun.* **67**(5), 3205–3219 (2019).
26. F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.* **30**(1), 40–60 (2013).
27. R. Hunger, *Floating point operations in matrix-vector calculus*, Munich Univ. of Technol., Munich, Germany (2005).
28. A. A. Zaidi, R. Baldemair, V. Moles-Cases, N. He, K. Werner, and A. Cedergren, "OFDM Numerology Design for 5G New Radio to Support IoT, eMBB, and MBSFN," *IEEE Commun. Standards Mag.* **2**(2), 78–83 (2018).
29. S. Parkvall, E. Dahlman, A. Furuskar, and M. Frenne, "NR: The New 5G Radio Access Technology," *IEEE Commun. Standards Mag.* **1**(4), 24–30 (2017).
30. A. N. Tait, "Silicon photonic neural networks," Ph.D. dissertation, Princeton University (2018).
31. B. Marquez, H. Morison, Z. Guo, M. Filipovich, P. Prucnal, and B. Shastri, "Graphene-based photonic synapse for multi wavelength neural networks," *MRS Adv.* **5**(37–38), 1909–1917 (2020).
32. M. Wu, C. Dick, J. R. Cavallaro, and C. Studer, "High-throughput data detection for massive MU-MIMO-OFDM using coordinate descent," *IEEE Trans. Circuits Syst. I, Reg. Papers* **63**(12), 2357–2367 (2016).
33. C. Tang, C. Liu, L. Yuan, and Z. Xing, "High precision low complexity matrix inversion based on newton iteration for data detection in the massive MIMO," *IEEE Commun. Lett.* **20**(3), 490–493 (2016).
34. H. Prabhu, J. Rodrigues, O. Edfors, and F. Rusek, "Approximative matrix inverse computations for very-large MIMO and applications to linear pre-coding systems," in *IEEE Wireless Commun. Netw. Conf.* (IEEE 2013), pp. 2710–2715.
35. L. Fang, L. Xu, and D. D. Huang, "Low complexity iterative MMSE-PIC detection for medium-size massive MIMO," *IEEE Wireless Commun. Lett.* **5**(1), 108–111 (2016).
36. A. Thanos, "Algorithms and hardware architectures for matrix inversion in massive MIMO uplink data detection," M. Sc. Thesis (2017).
37. M. Wu, B. Yin, G. Wang, C. Dick, J. R. Cavallaro, and C. Studer, "Large-Scale MIMO Detection for 3GPP LTE: Algorithms and FPGA Implementations," *IEEE J. Sel. Topics Signal Process.* **8**(5), 916–929 (2014).
38. A. N. Tait, T. F. De Lima, M. A. Nahmias, H. B. Miller, H.-T. Peng, B. J. Shastri, and P. R. Prucnal, "Silicon photonic modulator neuron," *Phys. Rev. Appl.* **11**(6), 064043 (2019).
39. V. Bangari, B. A. Marquez, H. Miller, A. N. Tait, M. A. Nahmias, T. F. de Lima, H. Peng, P. R. Prucnal, and B. J. Shastri, "Digital electronics and analog photonics for convolutional neural networks (DEAP-CNNs)," *IEEE J. Sel. Topics Quantum Electron.* **26**(1), 1–13 (2020).
40. J. Sun, R. Kumar, M. Sakib, J. B. Driscoll, H. Jayatilaka, and H. Rong, "A 128 Gb/s PAM4 Silicon Microring Modulator With Integrated Thermo-Optic Resonance Tuning," *J. Lightwave Technol.* **37**(1), 110–115 (2019).
41. J. I. Jamp, J. Deng, and L. E. Larson, "A 10 GS/s 5-bit ultra-low power DAC for spectral encoded ultra-wideband transmitters," in *Proc. IEEE Radio Frequency Integr. Circuits Symp.*, (IEEE, 2007), pp. 31–34.
42. M. Zhang, Y. Zhu, C. H. Chan, and R. P. Martins, "A 4× interleaved 10GS/s 8b time-domain ADC with 16× interpolation-based inter-stage gain achieving >37.5dB SNDR at 18GHz input," *IEEE Int. Solid State Circuits Conf. Dig. Tech. Papers* (IEEE, 2020), pp. 252–254.
43. W. Bogaerts, P. De Heyn, T. Van Vaerenbergh, K. De Vos, S. Kumar Selvaraja, T. Claes, P. Dumon, P. Bienstman, D. Van Thourhout, and R. Baets, "Silicon microring resonators," *Laser & Photonics Rev.* **6**(1), 47–73 (2012).
44. Baidu Research, "DeepBench," [Online]. <https://github.com/baidu-research/DeepBench>