High-Speed Ultra-Energy-Efficient Memristor-Based Massive MIMO SIC Detector Circuit With Hybrid Analog-Digital Computing Architecture

Jia-Hui Bi , Shaoshi Yang , *Senior Member, IEEE*,
Sheng Chen , *Life Fellow, IEEE*, and Ping Zhang , *Fellow, IEEE*

Abstract—The emerging memristor crossbar array based computing circuits exhibit computing speeds and energy efficiency far surpassing those of traditional digital processors. This type of circuits can complete high-dimensional matrix operations in an extremely short time through analog computing, making it naturally applicable to linear detection and maximum likelihood detection in massive multiple-input–multiple-output (MIMO) systems. However, the challenge of employing memristor crossbar arrays to efficiently implement other nonlinear detection algorithms, such as the successive interference cancellation (SIC) algorithm, remains unresolved. In this paper we propose a memristor-based circuit design for massive MIMO SIC detector. The proposed circuit comprises several judiciously designed analog matrix computing modules and hybrid analog-digital slicers, which enables the proposed circuit to perform the SIC algorithm with a hybrid analog-digital computing architecture. We show that the computing speed and the computational energy-efficiency of the proposed detector circuit are 43 times faster and 110 times higher, respectively, than those of a traditional 8-core digital signal processor (DSP), and also advantageous over the benchmark high-performance field programmable gate array (FPGA) and graphics processing unit (GPU).

Index Terms—Analog matrix computing, in-memory computing, memristor crossbar array, massive MIMO, successive interference cancellation.

I. Introduction

In modern wireless communication systems, the massive multiple-input multiple-output (MIMO) technology is a key enabler, which employs a large number of antennas to simultaneously serve multiple users, thus significantly enhancing the transmission rates and spectral efficiency. However, the extensive use of radio frequency (RF) chains in massive MIMO incurs substantial power consumption. Additionally, the large number of antennas significantly increases the complexity of baseband signal processing algorithms, such as signal detection algorithms [1], resulting in high processing latency and energy consumption in the baseband. Therefore, designing low-latency, energy-efficient massive MIMO communication systems has long been

a research hotspot in the field of wireless communications. To address this challenge, extensive research efforts have been undertaken. For example, the hybrid analog/digital architecture has been proposed as an effective means of reducing the number of RF chains, yielding a valuable solution for enhancing the energy efficiency of massive MIMO systems [2], [3]. On the other avenue, the emerging memristor devices have recently gained significant attention due to their great potential in realizing low-latency, high-energy-efficiency massive MIMO baseband signal processors.

Memristors are typically integrated into crossbar arrays, which enables high-dimensional matrix operations, such as matrix multiplication and inversion [4], to be completed within an extremely short time, generally in the range of tens of nanoseconds. The underlying principle of memristor-based matrix computing involves mapping the matrix operand onto the conductance matrix of the memristor crossbar array, mapping the vector operand onto the input voltages or currents, and obtaining computational results by measuring the output voltages of the circuit [5]. This form of analog matrix computing is an in-memory computing approach, which bypasses the von Neumann bottleneck, thereby achieving significantly higher processing speed and computational energy efficiency than the traditional digital computing approach.

Memristor-based analog matrix computing technology was initially used to accelerate matrix multiplications in neural network training and has been applied to massive MIMO baseband signal processing, especially in MIMO detection, in recent years. The work [6] first applied memristor-based analog matrix computing technology to MIMO signal detection, by utilizing memristor crossbar arrays to perform the matrix multiplication operations in the minimum mean square error (MMSE) detection. The study [7] proposed a memristor-based ridge regression computing circuit and applied it to perform linear detection algorithms, and the study [8] introduced a similarly structured memristor-based linear detector circuit. The work [9] proposed a memristor-based MMSE detection scheme by converting the MMSE algorithm into a linearized iterative algorithm and using memristor crossbar arrays to accelerate it. The work [10] employed memristor crossbar arrays to perform the matrix multiplication operations in the maximum likelihood (ML) detection with ultra-high energy efficiency. The superior performances of memristor-based massive MIMO detectors have been demonstrated in the aforementioned works.

Existing works utilizing memristor crossbar arrays for analog matrix computing can naturally achieve linear detection and ML detection, because the processes of linear and ML detection algorithms primarily involve matrix computations. However, the processes of other nonlinear detection algorithms are often more complex and do not rely solely on matrix computations. For instance, the successive interference cancellation (SIC) detection, a widely used detection algorithm with better detection performance than linear detection, involves iterative matrix computations and slicing operations. Existing memristor-based computing circuits are limited to performing matrix computations but cannot perform slicing operations. Therefore, efficiently performing the SIC algorithm using memristor crossbar arrays remains an unresolved challenge.

In this paper, we present a memristor-based massive MIMO SIC detector circuit. The proposed detector circuit comprises several memristor-based matrix computing modules and associated slicers. The proposed matrix computing modules perform the matrix computations in the SIC algorithm through analog computing. The proposed hybrid analog-digital slicers first convert each of the analog voltages to be

quantized into a digital (binary) vector and then output the corresponding quantization voltage levels based on these binary vectors. Therefore, the proposed detector circuit implements the SIC algorithm through a hybrid analog-digital processing approach. We investigate the impact of the precision of memristors on detection performance of the proposed detector. We also evaluate the computing speed and computational energy efficiency of the proposed circuit. Our results show that they are 43 times faster and 110 times higher, respectively, than those of a traditional 8-core digital signal processor (DSP) in our considered scenario, and also advantageous over the benchmark high-performance field programmable gate array (FPGA) and graphics processing unit (GPU).

II. MIMO SYSTEM AND SIC DETECTION

A. System Model

We consider a massive MIMO system, in which the base station (BS) is equipped with R antennas to support K single-antenna user terminals (UTs) with $R > K$. Let $\lambda_1, \lambda_2, \ldots, \lambda_K$ be the transmitting power values of UTs, respectively. The uplink received signals are given by:

$$\mathbf{y} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{y} \in \mathbb{C}^{R \times 1}$ is the received signal vector, $\mathbf{s} \in \mathbb{C}^{K \times 1}$ is the transmitted signal vector sent by the UTs, $\mathbf{H} \in \mathbb{C}^{R \times K}$ is the channel matrix, the diagonal matrix $\boldsymbol{\Lambda} = \mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_K})$, and $\mathbf{n} \in \mathbb{C}^{R \times 1}$ is a zero-mean complex additive white Gaussian noise (AWGN) vector having the covariance matrix $\mathbb{E}[\mathbf{n}\mathbf{n}^{\mathrm{H}}] = \sigma_n^2 \mathbf{I}$, with $(\cdot)^{\mathrm{H}}$ denoting the Hermitian transpose operator and $\mathbf{I}$ denoting the identity matrix of appropriate dimension. The transmitted signals are normalized as $\mathbb{E}[|s_k|^2] = 1, 1 \leq k \leq K$, where s_k is the kth element of $\mathbf{s}$. By defining $\mathbf{F} = \mathbf{H}\boldsymbol{\Lambda} \in \mathbb{C}^{R \times K}$, we obtain:

$$\mathbf{y} = \mathbf{F}\mathbf{s} + \mathbf{n}. \tag{2}$$

B. Operations of SIC Detection

The core idea of the SIC detection algorithm is to detect the transmitted symbols one by one using a linear detection algorithm, with the aid of cancelling the corresponding detected symbols from the received signals after each detection [11], [12]. We consider the MMSE-SIC algorithm in this paper.

Let $\{m_1, m_2, \ldots, m_K\}$ be the ordered set representing the detection sequence, and $\mathbf{G}$ be the reordered transfer matrix, i.e., $\mathbf{G} = [\mathbf{f}_{m_1}, \mathbf{f}_{m_2}, \ldots, \mathbf{f}_{m_K}]$ with $\mathbf{f}_k$ being the kth column of $\mathbf{F}$. Denote by $\mathbf{G}_{(k)}$ the first to kth columns of $\mathbf{G}$, and by $\mathbf{G}_{\langle k \rangle}$ the kth to Kth columns of $\mathbf{G}$, i.e., $\mathbf{G}_{(k)} = [\mathbf{f}_{m_1}, \ldots, \mathbf{f}_{m_k}]$, $\mathbf{G}_{\langle k \rangle} = [\mathbf{f}_{m_k}, \ldots, \mathbf{f}_{m_K}]$. Let e_{m_k} be the kth estimated symbol, i.e., the estimated value of s_{m_k}. Denote by $\mathbf{e}_{(k)}$ the vector consisting of the first to kth estimated symbols, i.e., $\mathbf{e}_{(k)} = [e_{m_1}, \ldots, e_{m_k}]^{\mathrm{T}}$, where $(\cdot)^{\mathrm{T}}$ denotes the transpose operator.

The SIC algorithm performs the matrix computation

$$\mathbf{b}_1 = \left(\mathbf{G}^{\mathrm{H}}\mathbf{G} + \sigma_n^2 \mathbf{I}\right)^{-1} \mathbf{G}^{\mathrm{H}}\mathbf{y} \tag{3}$$

to estimate e_{m_1}, where $\mathbf{b}_1$ is the result vector, and $(\cdot)^{-1}$ denotes the inverse operator. Specifically, e_{m_1} is obtained by performing a slicing (i.e., quantization) operation on the first element of $\mathbf{b}_1$.

The SIC algorithm performs the matrix computation

$$\mathbf{b}_k = \left(\mathbf{G}_{\langle k \rangle}^{\mathrm{H}}\mathbf{G}_{\langle k \rangle} + \sigma_n^2 \mathbf{I}\right)^{-1} \mathbf{G}_{\langle k \rangle}^{\mathrm{H}} \left(\mathbf{y} - \mathbf{G}_{(k-1)}\mathbf{e}_{(k-1)}\right) \tag{4}$$

to estimate $e_{m_k} (1 < k \leq K)$, where $\mathbf{b}_k$ is the result vector. Similarly, e_{m_k} is obtained by performing a slicing operation on the first element of $\mathbf{b}_k$.
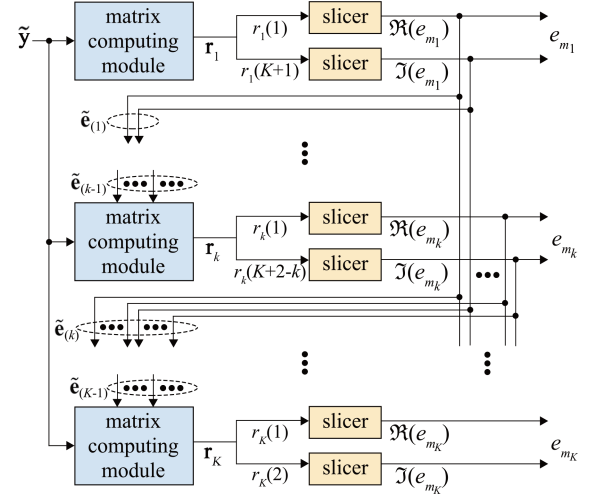

Fig. 1.   Structure of the proposed memristor-based SIC detector circuit.

Obviously, the SIC detection processes consist of K matrix computations, and each of them is followed by a slicing operation. Since the memristor-based matrix computing circuit can only perform real-valued matrix computations, we define two conversion operators $\mathcal{V}(\cdot)$ and $\mathcal{M}(\cdot)$ for converting a complex-valued vector to a real-valued vector and a complex-valued matrix to a real-valued matrix, respectively:

$$\mathcal{V}(\cdot) = \begin{bmatrix} \Re(\cdot) \\ \Im(\cdot) \end{bmatrix}, \ \mathcal{M}(\cdot) = \begin{bmatrix} \Re(\cdot) & -\Im(\cdot) \\ \Im(\cdot) & \Re(\cdot) \end{bmatrix},$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts of the corresponding vector or matrix, respectively.

Let $\mathbf{r}_k$, satisfying $\mathbf{r}_k = \mathcal{V}(\mathbf{b}_k)$, be the real-valued result vector of the kth matrix computation. Then the first matrix computation in SIC detection, i.e., (3), can be expressed as:

$$\mathbf{r}_1 = \left(\tilde{\mathbf{G}}^{\mathrm{T}}\tilde{\mathbf{G}} + \sigma_n^2 \mathbf{I}\right)^{-1} \tilde{\mathbf{G}}^{\mathrm{T}}\tilde{\mathbf{y}}, \tag{5}$$

where $\tilde{\mathbf{G}} = \mathcal{M}(\mathbf{G})$, $\tilde{\mathbf{y}} = \mathcal{V}(\mathbf{y})$. The kth $(1 < k \leq K)$ matrix computation, i.e., (4), can be equivalently expressed as:

$$\mathbf{r}_k = \left(\tilde{\mathbf{G}}_{\langle k \rangle}^{\mathrm{T}}\tilde{\mathbf{G}}_{\langle k \rangle} + \sigma_n^2 \mathbf{I}\right)^{-1} \tilde{\mathbf{G}}_{\langle k \rangle}^{\mathrm{T}} \left(\tilde{\mathbf{y}} - \tilde{\mathbf{G}}_{(k-1)}\tilde{\mathbf{e}}_{(k-1)}\right), \tag{6}$$

where $\tilde{\mathbf{G}}_{\langle k \rangle} = \mathcal{M}(\mathbf{G}_{\langle k \rangle})$, $\tilde{\mathbf{G}}_{(k-1)} = \mathcal{M}(\mathbf{G}_{(k-1)})$, $\tilde{\mathbf{e}}_{(k-1)} = \mathcal{V}(\mathbf{e}_{(k-1)})$.

Let $\mathcal{Q}(a_1, a_2)$ be the slicing operator, with the real values a_1 and a_2 corresponding to the real and imaginary parts of the sliced complex value, respectively. Denote by $r_k(k')$ the k'th element of $\mathbf{r}_k$. The kth $(1 \leq k \leq K)$ slicing operation can be expressed as:

$$e_{m_k} = \mathcal{Q}\left(r_k(1), r_k(K + 2 - k)\right). \tag{7}$$

III. PROPOSED SIC DETECTOR CIRCUIT DESIGN

A. Circuit Structure

Structure of the proposed SIC detector circuit is illustrated in Fig. 1, which consists of K stages, each comprising a matrix computing module and two slicers. The matrix computing modules are employed to compute (5) and (6), while the slicers are employed to perform (7). In the kth stage, the matrix computing module outputs $\mathbf{r}_k$, where $r_k(1)$ and $r_k(K + 2 - k)$ serve as inputs to two slicers. The two slicers output the real and imaginary parts of e_{m_k}, respectively. The outputs of the
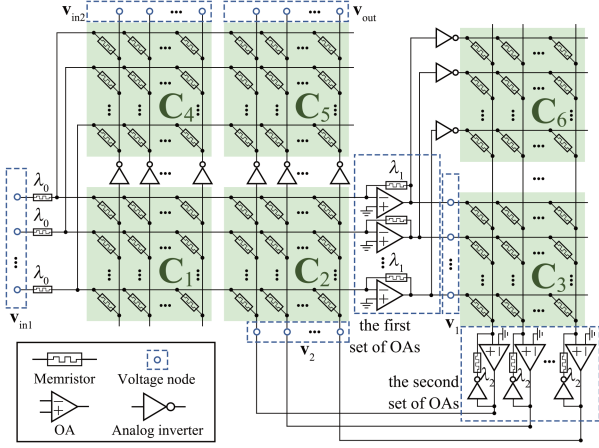
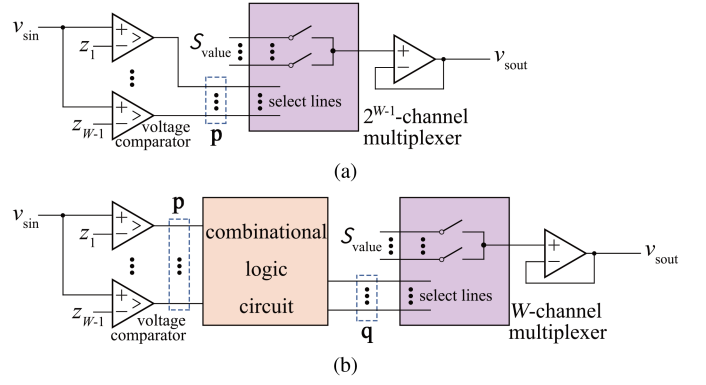Fig. 2. Circuit design of the proposed matrix computing module.

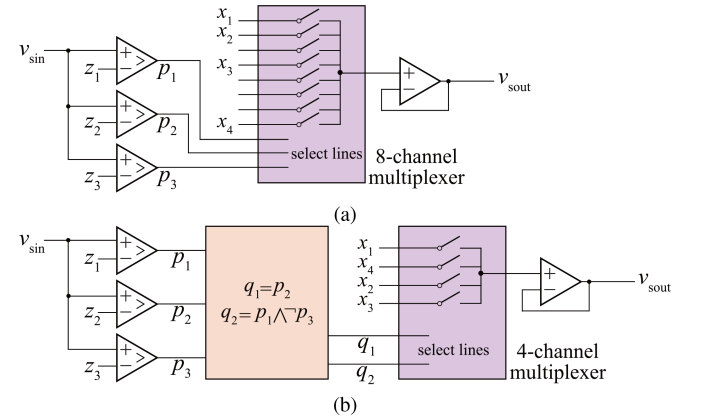Fig. 3. Structures of the proposed slicer: (a) The directly select structure, and (b) the indirectly select structure.

Fig. 4. The proposed slicer with: (a) The directly select structure, and (b) the indirectly select structure, using 16 QAM as an example.

kth stage serve as inputs to the $(k+1)$th to Kth matrix computing modules, with $\tilde{\mathbf{y}}$ simultaneously serving as an input to each matrix computing module.

B. Proposed Matrix Computing Module

The circuit design of the proposed matrix computing module is illustrated in Fig. 2, which comprises six memristor crossbar arrays, three sets of analog inverters and two sets of operational amplifiers (OAs).

Let $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{C}_3$, $\mathbf{C}_4$, $\mathbf{C}_5$ and $\mathbf{C}_6$ be the conductance matrices of the six arrays, respectively, $\mathbf{v}_{\text{in1}}$ and $\mathbf{v}_{\text{in2}}$ be the two sets of input voltages, and $\mathbf{v}_{\text{out}}$ be the output voltages of the circuit, as shown in Fig. 2. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be the voltages at the output nodes of the two sets of OAs, respectively. Clearly, $\mathbf{v}_{\text{out}} = -\mathbf{v}_2$. The conductance values of the memristors that are connected to the voltage nodes of $\mathbf{v}_{\text{in1}}$, the feedback memristors of the first set of OAs, and the feedback memristors of the second set of OAs are λ_0, λ_1 and λ_2, respectively.

The high gain of an OA results in a minimal voltage difference between its inverting and noninverting input nodes, making the voltages at the inverting input nodes of the first set of OAs and the noninverting input nodes of the second set of OAs approximate zeros. Additionally, the inherent properties of OAs result in approximate zero currents flowing into their inverting and noninverting input nodes. Based on Ohm's law and Kirchhoff's current law, the voltages in Fig. 2 satisfy:

$$\lambda_0 \mathbf{v}_{\text{in1}} + (\mathbf{C}_4 - \mathbf{C}_1)\mathbf{v}_{\text{in2}} + (\mathbf{C}_2 - \mathbf{C}_5)\mathbf{v}_2 + \lambda_1 \mathbf{v}_1 = \mathbf{0}, \quad (8)$$

and

$$(\mathbf{C}_3 - \mathbf{C}_6)^{\mathrm{T}}\mathbf{v}_1 - \lambda_2 \mathbf{v}_2 = \mathbf{0}. \quad (9)$$

Upon substituting (9) into (8) we obtain:

$$\mathbf{v}_{\text{out}} = (\mathbf{D}_3^{\mathrm{T}}\mathbf{D}_2 + \lambda_1\lambda_2\mathbf{I})^{-1}\mathbf{D}_3^{\mathrm{T}}(\lambda_0\mathbf{v}_{\text{in1}} - \mathbf{D}_1\mathbf{v}_{\text{in2}}), \quad (10)$$

where $\mathbf{D}_1 = \mathbf{C}_1 - \mathbf{C}_4$, $\mathbf{D}_2 = \mathbf{C}_2 - \mathbf{C}_5$, $\mathbf{D}_3 = \mathbf{C}_3 - \mathbf{C}_6$.

The conductance value of a memristor can be adjusted by a specific program [4] to any value within a certain range. For the first stage, by mapping $\tilde{\mathbf{y}}$ onto $\lambda_0\mathbf{v}_{\text{in1}}$, mapping $\tilde{\mathbf{G}}$ onto $\mathbf{D}_2$ and $\mathbf{D}_3$, mapping σ_n^2 onto $\lambda_1\lambda_2$, removing $\mathbf{C}_1$ and $\mathbf{C}_4$, the result of (5), i.e., $\mathbf{r}_1$, can be obtained by measuring $\mathbf{v}_{\text{out}}$. For the kth ($1 < k \leq K$) stage, by mapping $\tilde{\mathbf{y}}$ onto $\lambda_0\mathbf{v}_{\text{in1}}$, mapping $\tilde{\mathbf{G}}_{(k-1)}$ onto $\mathbf{D}_1$, mapping $\tilde{\mathbf{e}}_{(k-1)}$ onto $\mathbf{v}_{\text{in2}}$, mapping $\tilde{\mathbf{G}}_{\langle k \rangle}$ onto $\mathbf{D}_2$ and $\mathbf{D}_3$, and mapping σ_n^2 onto $\lambda_1\lambda_2$, the result of (6), i.e., $\mathbf{r}_k$, can be obtained by measuring $\mathbf{v}_{\text{out}}$. Since the mapped matrix may contain both positive and negative elements, we map the computed matrix onto the difference between two positive conductance matrices.

Unlike traditional processors based on digital computing approach, the proposed matrix computing module employs the aforementioned analog computing approach to perform matrix computations in the SIC algorithm, thereby avoiding excessive time and computational resource consumption typically incurred by digital processor based matrix inversion operations.

C. Proposed Hybrid Analog-Digital Slicer

Let $\mathcal{S}_{\text{value}} = \{x_1, \ldots, x_W\}$ be the set of voltage values corresponding to all the possible values of $\Re(e_{m_k})$ or $\Im(e_{m_k})$, where W represents the number of elements in the set and it depends on the modulation scheme. Let $\mathcal{S}_{\text{threshold}} = \{z_1, \ldots, z_{W-1}\}$ be the threshold set, i.e., $z_w = \frac{x_w + x_{w+1}}{2}$ ($1 \leq w < W$). The function of the proposed slicer is to slice a voltage into the element of $\mathcal{S}_{\text{value}}$ that is the closest to it. We propose two slicer structures, one termed the directly select structure, and the other termed the indirectly select structure.

The directly select structure is illustrated in Fig. 3(a), which consists of $W-1$ voltage comparators, a 2^{W-1}-channel analog multiplexer and an OA. When the voltage at the noninverting input node of a comparator is greater than that at its inverting input node, it outputs a high level, otherwise it outputs a low level. Let v_{sin} and v_{sout} be the input and output voltage of the proposed slicer, respectively. The comparators compare v_{sin} with all the threshold voltages. Let $\mathbf{p}$ be the binary vector of length $W-1$ formed by the outputs of the comparators, which are also employed as the select lines of multiplexer. Voltages with values of $\mathcal{S}_{\text{value}}$

TABLE I
RELATIONSHIP BETWEEN v_{sin}, $\mathbf{p}$, $\mathbf{q}$, AND v_{sout} FOR $W = 4$

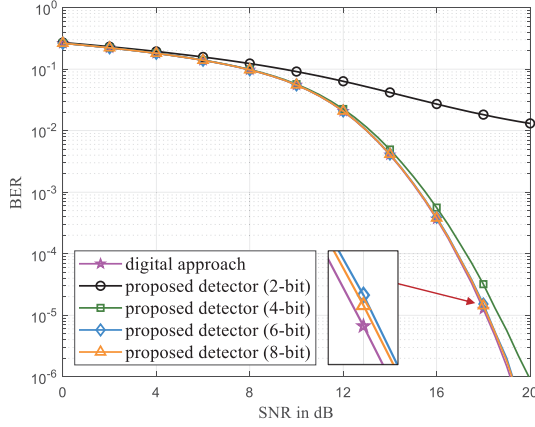| v_{sin}	$\mathbf{p}$	$\mathbf{q}$	v_{sout}
$< z_1$	$[0, 0, 0]$	$[0, 0]$	x_1
$z_1 \sim z_2$	$[1, 0, 0]$	$[0, 1]$	x_2
$z_2 \sim z_3$	$[1, 1, 0]$	$[1, 1]$	x_3
$> z_3$	$[1, 1, 1]$	$[1, 0]$	x_4

Fig. 5. BERs of the proposed detector circuit under different values of memristor precision, using the digital approach as the benchmark.

serve as inputs to the appropriate input channels of multiplexer. The different magnitude relationships between v_{sin} and threshold voltages result in different $\mathbf{p}$, causing the multiplexer to select different channels and output the corresponding values of $\mathcal{S}_{value}$. The OA is employed to construct a voltage follower to ensure output voltage stability.

The indirectly select structure is illustrated in Fig. 3(b), which consists of $W-1$ voltage comparators, a combinational logic circuit, a W-channel analog multiplexer and an OA. The combinational logic circuit transforms $\mathbf{p}$ into a shorter binary vector $\mathbf{q}$ of length $\log_2 W$, so as to indirectly select a channel of the multiplexer. The indirectly select structure can enhance the utilization of multiplexer channels, especially when the modulation order is high. For instance, when utilizing 64 quadrature amplitude modulation (QAM), i.e., $W = 8$, the directly select structure necessitates a 128-channel analog multiplexer, whereas the indirectly select structure only requires an 8-channel multiplexer.

To provide a clearer illustration of the two structures, we consider the 16 QAM example. Hence, $x_1 = -\frac{3}{\sqrt{10}} v_0$, $x_2 = -\frac{1}{\sqrt{10}} v_0$, $x_3 = \frac{1}{\sqrt{10}} v_0$ and $x_4 = \frac{3}{\sqrt{10}} v_0$, while $z_1 = -\frac{2}{\sqrt{10}} v_0$, $z_2 = 0$ and $z_3 = \frac{2}{\sqrt{10}} v_0$, where v_0 is a reference voltage. The varying values of v_{sin} lead to the four possible values of $\mathbf{p}$: $[0,0,0]$, $[1,0,0]$, $[1,1,0]$, $[1,1,1]$. The slicer with the directly select structure is illustrated in Fig. 4(a), the four possible $\mathbf{p}$ values select the first, second, fourth and eighth input channels of the 8-channel multiplexer, respectively. So we input x_1, x_2, x_3 and x_4 to the corresponding four channels. The slicer with the indirectly select structure is illustrated in Fig. 4(b). The logic formulas of the combinational logic circuit are $q_1 = p_2$ and $q_2 = p_1 \wedge \neg p_3$, where $\neg$ and $\wedge$ denote the logical NOT and AND, respectively, and it is worth noting that this is not the only option. The four values of $\mathbf{q}$ corresponding to the four possible values of $\mathbf{p}$ are: $[0,0]$, $[0,1]$, $[1,1]$, $[1,0]$, which select the first, third, fourth, second input channels of the 4-channel multiplexer, respectively. So we input x_1, x_2, x_3 and x_4 to the four channels. Table I summarizes the relationship between v_{sin}, $\mathbf{p}$, $\mathbf{q}$ (if exist) and v_{sout} in both the directly select structure and indirectly select structure.
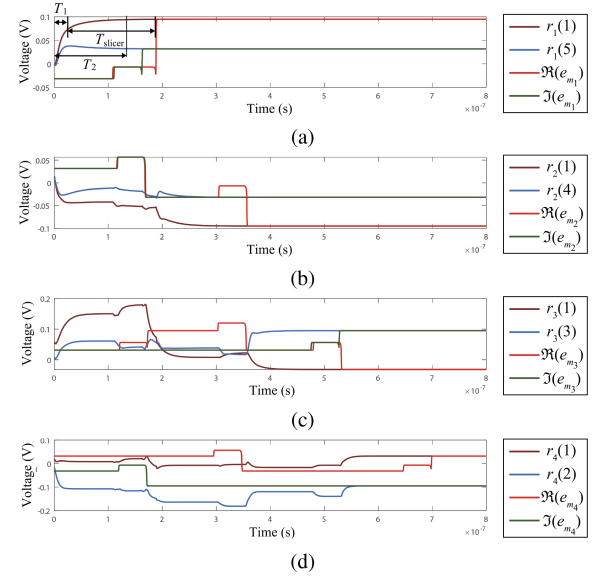

Fig. 6. Voltage waveforms at input and output nodes of slicers in the four stages of the proposed circuit: (a) The first stage, (b) the second stage, (c) the third stage, (d) the fourth stage.
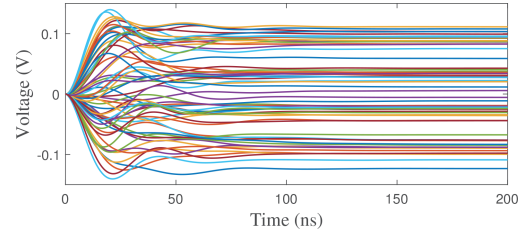

Fig. 7. An example of the voltage waveforms of the 64 output nodes in the first matrix computing module.

IV. SIMULATIONS

Let α_{min} and α_{max} represent the minimum and maximum achievable conductance values of the memristors, respectively. Let $\mathbf{O}$ denote the matrix to be mapped, and let $\mathbf{U}$ and $\mathbf{V}$ denote the two corresponding conductance matrices. When mapping $\mathbf{O}$ onto $\mathbf{U} - \mathbf{V}$, we consider the following mapping scheme:

$$u_{i,j} = \begin{cases} \alpha_{max}, & o_{i,j} > 0 \\ \alpha_{min}, & o_{i,j} \leq 0 \end{cases} \tag{11}$$

and

$$v_{i,j} = u_{i,j} - \beta o_{i,j}, \tag{12}$$

where β is given by $\frac{\alpha_{max} - \alpha_{min}}{\max\{|o_{i,j}|\}}$, ensuring that the conductance range covers all elements of $\mathbf{O}$. In this paper, we consider the memristor conductance range of $0.1\,\mu S \sim 30\,\mu S$.

A. Detection Performance

Unlike digital processors, the computational results of a memristor-based analog matrix computing circuit are not absolutely precise and are particularly affected by the precision of the memristors. In this experiment, we consider a 32×64 massive MIMO system, i.e., $K = 32$ and $R = 64$, using 16 QAM. All UTs are assumed to have the same transmitted power and the column norm-based ordering scheme [12] is adopted. Fig. 5 depicts the bit error rates (BERs) of the proposed

TABLE II
COMPARISON BETWEEN THE PROPOSED CIRCUIT AND DIFFERENT TRADITIONAL DIGITAL COMPUTING APPROACH BASED PROCESSORS

	8-core DSP (TMS320C6678)	FPGA (Virtex-7 690T)	GPU (RTX A1000)	**Proposed circuit**
Computing speed	0.128 TOPS	3.12 TOPS	6.7 TOPS	**5.5 TOPS**
Energy consumption	2.1 mJ	343.1 μJ	199.7 μJ	**18.98 μJ**
Computational energy efficiency	0.0128 TOPS/W	0.078 TOPS/W	0.134 TOPS/W	**1.41 TOPS/W**

The bold values represent the performance metrics of the proposed circuit. These values are highlighted to emphasize the advantages of the proposed design over traditional digital computing approaches.

detector circuit as the functions of the signal-to-noise ratio (SNR) under different memristor bit-precision values, using the digital approach as the benchmark. As expected, the higher the precision of the memristors, the lower the BER of the proposed detector. The simulation results indicate that the memristor bit-precision should be at least 6 bits to ensure detection performance approaching the BER of digital computing.

B. Computing Time, Computing Speed and Computational Energy Efficiency

We consider a 4×4 MIMO system, i.e., $K = R = 4$, in the noise-free environment to investigate the rules of the convergence time of the proposed detector circuit. OAs in the circuit have a gain-bandwidth product (GBP) of 500 MHz, and v_0 is set to 0.1 V. SPICE simulation results presented in this paper are provided by LTspice.

We first consider the directly select structure, employing the voltage comparator LT1016 [13] and the 8-channel multiplexer ADG1608 [14]. Let T_{slicer} be the delay of the proposed slicer, i.e., the delay time between the input of the slicer reaching the threshold and the slicer completing the switching. T_{slicer} is the sum of the propagation delay of the comparator, which is 10 ns typically, and the transition time of the multiplexer, which is 150 ns typically. Fig. 6 shows an example of voltage waveforms at input and output nodes of slicers in the four stages of the proposed detector circuit. For the first stage, the convergence time of $\Re(e_{m_1})$ is $T_1 + T_{\text{slicer}}$, where T_1 is the time required for $r_1(1)$ to reach the corresponding threshold. Let T_2 be the time required for the outputs of a matrix computing module to reach steady states. Obviously, we have $T_1 \leq T_2$, indicating that the maximum convergence time of the first stage in the circuit is $T_2 + T_{\text{slicer}}$. The change in the output of a slicer leads to alterations in the inputs of subsequent slicers, which may result in changes in the outputs of these slicers. Therefore, the maximum convergence time of the proposed circuit is $K(T_2 + T_{\text{slicer}})$, which corresponds to the longest time required for the output voltages of the Kth stage to reach steady state after inputs are applied. As for the indirectly select structure, the delay of the combinational logic circuit is negligible, and therefore the convergence time of the circuit with indirectly select structure also satisfies the above rules.

In the rest of this subsection, we consider a 32×64 massive MIMO system. To further reduce the convergence time, we consider the high-speed voltage comparator MAX903 with 8 ns propagation delay [15], the 4-channel multiplexer ADG709 [16], and the 8-channel multiplexer ADG708 [16]. Both the ADG708 and ADG709 exhibit a latency of 14 ns. Fig. 7 shows an example of voltage waveforms of the 64 output nodes in the first matrix computing module, i.e., $\mathbf{r}_1$. T_2 is typically less than 130 ns and can be further reduced by increasing the GBP of OAs [7]. Since the considered 4-channel and 8-channel multiplexers have the same delay time, there is no difference in the convergence time of the detector circuits with the two slicer structures. Analog to digital converters (ADCs) of [17] with 10 ns delay are used to measure the output voltages of the circuit. OAs, in conjunction with the digital to analog converters (DACs) of [18] whose settling time is 0.4 ns,

are employed to provide stable input voltages for the circuit. In the considered scenario, the total computing time of the proposed detector circuit is approximately 4.87 μs, including the settling time of DACs, the convergence time of the detector circuit, and the delay of ADCs. The computing speed of the proposed circuit is evaluated by the ratio of its equivalent floating-point operation (FLOP) number to its computing time. In the considered scenario, the computing speed of the proposed circuit is about 5.5 tera operations per second (TOPS).

To further evaluate the energy consumption of the proposed circuit, we consider the OA of [19] with a GBP of 500 MHz and power consumption of 12 μW. We consider both the Joule dissipation on each memristor and typical energy consumption of all other components in the circuit during the computing period. In the scenario considered, the energy consumption of the circuits with both the directly select structure and the indirectly select structure is approximately the same, at 18.98 μJ. The computational energy efficiency of a circuit can be expressed as the ratio of its equivalent FLOP number to the energy consumed for completing a given computational task, and it is typically measured in TOPS per watt (TOPS/W), i.e., tera operations per joule. In the scenario considered, the proposed detector circuit achieves the computational energy efficiency of approximately 1.41 TOPS/W.

In Table II we compare the proposed circuit with three different traditional digital computing approach based processors, including a DSP, an FPGA, and a GPU, in terms of equivalent computing speed, energy consumption, and computational energy efficiency. Specifically, we use Texas Instruments' 8-core high-performance DSP TMS320C6678 [20], the high-performance low-power FPGA Xilinx Virtex-7 690T [21], and the powerful GPU NVIDIA RTX A1000 [22] as representative models for these three types of processors, respectively.

The proposed circuit operates at a speed approximately 43 times faster than that of the DSP, with the computational energy efficiency around 110 times greater, i.e., the computing energy consumption is merely less than 1% of that of the DSP. Compared with the Virtex-7 690T FPGA, the proposed circuit exhibits a computing speed and computational energy efficiency approximately 1.76 and 18 times greater, respectively. Although the equivalent computing speed of the proposed circuit is slightly lower than that of the RTX A1000 GPU in the considered scenario, its computational energy efficiency exceeds that of the RTX A1000 GPU by more than an order of magnitude. Furthermore, for massive MIMO systems with a higher number of UTs or BS antennas, the proposed memristor-based detector circuit can achieve even higher equivalent computing speed and computational energy efficiency [7], making it particularly suitable for next-generation wireless communication systems with massive connectivity requirements.

V. CONCLUSION

In this article, we have presented a memristor-based massive MIMO SIC detector circuit for K UTs. The proposed detector circuit structure comprises K stages, each containing a matrix computing module and

two slicers. We have explained the circuit structure of the matrix computing modules and proposed two feasible slicer structures. We have also investigated the impact of the precision of memristors on the detection performance of the proposed detector circuit. In a 32×64 massive MIMO scenario, we have evaluated the computing speed and computational energy efficiency of the proposed circuit, demonstrating significant advantages over the representative high performance DSP, FPGA and GPU.

REFERENCES

[1] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surveys Tut.*, vol. 17, no. 4, pp. 1941–1988, Fourthquarter 2015.

[2] R. Zhang, B. Shim, and W. Wu, "Direction-of-arrival estimation for large antenna arrays with hybrid analog and digital architectures," *IEEE Trans. Signal Process.*, vol. 70, pp. 72–88, 2022.

[3] R. Zhang et al., "Channel training-aided target sensing for terahertz integrated sensing and massive MIMO communications," *IEEE Internet Things J.*, vol. 12, no. 4, pp. 3755–3770, Feb. 2025.

[4] Z. Sun, G. Pedretti, E. Ambrosi, A. Bricalli, W. Wang, and D. Ielmini, "Solving matrix equations in one step with cross-point resistive arrays," *Proc. Nat. Acad. Sci.*, vol. 116, no. 10, pp. 4123–4128, Mar. 2019.

[5] X. Yang, B. Taylor, A. Wu, Y. Chen, and L. O. Chua, "Research progress on memristor: From synapses to computing systems," *IEEE Trans. Circuits Syst. I., Reg. Papers*, vol. 69, no. 5, pp. 1845–1857, May 2022.

[6] G. Yuan et al., "Memristor crossbar-based ultra-efficient next-generation baseband processors," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. 2017*, Boston, MA, USA, Aug., 2017, pp. 1121–1124.

[7] P. Mannocci, E. Melacarne, and D. Ielmini, "An analogue in-memory ridge regression circuit with application to massive MIMO acceleration," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 12, no. 4, pp. 952–962, Dec. 2022.

[8] Q. Zeng et al., "Realizing in-memory baseband processing for ultra-fast and energy-efficient 6G," *IEEE Internet Things J.*, vol. 11, no. 3, pp. 5169–5183, Feb. 2024.

[9] Y. Fang, L. Chen, C. You, and H. Yin, "Rethinking massive MIMO detection: A memristor approach," *IEEE Commun. Lett.*, vol. 27, no. 12, pp. 3350–3354, Dec. 2023.

[10] Y.-H. Ren, S. Yang, J.-H. Bi, and Y.-X. Zhang, "Accelerating maximum-likelihood detection in massive MIMO: A new paradigm with memristor crossbar based in-memory computing circuit," *IEEE Trans. Veh. Technol.*, vol. 73, no. 12, pp. 19745–19750, Dec. 2024.

[11] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, Autumn 1996.

[12] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM Wireless Communications With MATLAB*. Hoboken, NJ, USA: Wiley, 2010.

[13] "Data sheet: LT1016," Analog Devices, Inc., 2018. [Online]. Available: https://www.analog.com/media/en/technical-documentation/data-sheets/lt1016.pdf

[14] "Data sheet: ADG1608/ADG1609," Analog Devices, Inc., 2015. [Online]. Available: https://www.analog.com/media/en/technical-documentation/data-sheets/ADG16081609.pdf

[15] "Data sheet: MAX900–MAX903," Analog Devices, Inc., 2005. [Online]. Available: https://www.analog.com/media/en/technical-documentation/data-sheets/MAX900-MAX903.pdf

[16] "Data sheet: ADG708/ADG709," Analog Devices, Inc., 2014. [Online]. Available: https://www.analog.com/media/en/technical-documentation/data-sheets/ADG708709.pdf

[17] Z. Guo, D. Chen, and X. Xue, "Algorithm/hardware co-design configurable SAR ADC with low power for computing-in-memory in 28nm CMOS," in *Proc. IEEE 14th Int. Conf. ASIC 2021*, Kunming, China, Oct., 2021, pp. 1–4.

[18] S. M. I. Huq, S. Islam, N. Saqib, and S. N. Biswas, "Design of low power 8-bit DAC using PTM-LP technology," in *Proc. Int. Conf. Recent Trends Elect., Electron. Comput. Technol. 2017*, Warangal, India, Jul., 2017, pp. 64–69.

[19] B. Feinberg et al., "An analog preconditioner for solving linear systems," in *Proc. IEEE Int. Symp. High-Perform. Comput. Architecture 2021*, Seoul, Korea, Feb./Mar., 2021, pp. 761–774.

[20] B. Ramesh, A. Bhardwaj, J. Richardson, A. D. George, and H. Lam, "Optimization and evaluation of image- and signal-processing kernels on the TI C6678 multi-core DSP," in *Proc. IEEE High Perform. Extreme Comput. Conf. 2014*, Waltham, MA, USA, Sep., 2014, pp. 1–6.

[21] "GPU vs FPGA performance comparison," BERTEN DSP S.L., White Paper, 2016. [Online]. Available: https://www.bertendsp.com/gpu-vs-fpga-performance-comparison/

[22] "Data sheet: NVIDIA RTX A1000," NVIDIA, 2024. [Online]. Available: https://www.nvidia.com/en-us/design-visualization/rtx-a1000/