Input-aware Flow-based Computing on Memristor Crossbars with Applications to Edge Detection

Dwaipayan Chakraborty, Student Member, IEEE, Sunny Raj, Student Member, IEEE, Steven L. Fernandes, Member, IEEE, and Sumit K. Jha, Member, IEEE

Abstract—Sneak paths in nanoscale memristor crossbars have traditionally been viewed as a problem in the use of memristor crossbars as non-volatile replacements of traditional volatile RAM memories. We show that the sneak paths in a memristor crossbar can be employed to perform computation that exploits device-level parallelism. Our computation can be performed in the memory and does not require data to be moved between a processor and a memory unit – thereby, avoiding the von Neumann bottleneck. We demonstrate the potential of our approach by applying it to a basic problem in computer vision: edge detection in an image. Our results show that the flow-based computing approach on nanoscale memristor crossbars can be used to obtain high-quality approximations of edge detection. We have synthesized multiple 8x8 crossbar circuits for this purpose – a single crossbar circuit for detecting edges between all possible pixel pairs with ~85% accuracy, and another family of input-aware crossbars with higher performance over real-world images. The family of input-aware crossbars together performs approximate edge detection for a subset of pixel pairs obtained from analyzing the BSD500 database, and the resultant images are of a quality comparable to exact edge detection.


I. INTRODUCTION

The rise of digital data acquisition such as digital cameras has led to a deluge of high-volume high-velocity big data whose subsequent processing requires both the ability to process this high-velocity data quickly and the capability to rapidly transfer this high-volume data from the memory to the processor. In many intelligent applications such as smart cameras, the computation being performed is fixed a priori such as edge detection or face identification. Our proposed approach enables the design of nanoscale memristor crossbars for implementing such a priori known computations using naturally occurring sneak paths in two-dimensional arrays of nanoscale memristors.

Artificial Intelligence (AI) applications often rely on high-dimensional image data and there is an urgent need to eliminate or reduce the von Neumann bottleneck that requires this data to be moved from memory to the processor and back. Significant progress has been made towards reducing the von Neumann bottleneck by creating high-performance memories. The crossbar circuit topology is making its way into mainstream technologies [1]–[4], because of its design as a two-dimensional uniform array of devices and the subsequent ease of fabrication. Significant efforts have been made towards efficient fabrication of the crossbar circuits [5]–[10]. A combination of the above developments has led to the production of high-performance memory devices on the crossbar fabric [11]–[14]. While these advancements have definitely yielded performance and scalability improvements, the separation between the processor and the memory remains a bottleneck in both computing speed and energy efficiency.

Our flow-based computing approach stores data on a nanoscale memristor crossbar in such a pattern that the flow of current through the naturally-occurring sneak paths in the crossbar performs the desired Boolean computation. The nanoscale memristor crossbar is used both for the storage of data and for performing the desired computation – thus, eliminating the von Neumann bottleneck.

Flow-based computing designs involving only a few memristors can be obtained by implicitly or explicitly enumerating the space of all possible designs [15]. However, flow-based computing memristor crossbar designs for even simple AI applications such as edge detection cannot be
synthesized using such direct enumeration approaches. An overview of our proposed approach is illustrated in Figure 1. In this article, we make the following contributions:

- We demonstrate how flow-based computing on nanoscale memristor crossbars [16] coupled with a simulated annealing search based on model counting [17] can be used to design crossbars that approximately implement edge detection in images. Our approach [17] is the first to use model counting to search the space of possible designs.
- We show that the results of a number of approximately correct input-aware memristor crossbars can be combined using elementary logical operations implemented on memristor crossbars to enable high-quality edge detection in real-world images.

Our results indicate that the flow-based computing approach may be applicable to more interesting AI applications, such as deep learning. The rest of the paper is organized as follows: Section II discusses related work and presents it in the context of the current manuscript. Section III presents the idea of flow-based computing on nanoscale memristor crossbars. Our approach for searching the space of possible nanoscale memristor crossbar designs using model counting is presented in Section IV. Section V presents the results of applying flow-based computing to the problem of edge detection, and we conclude with ideas for future work in Section VI.

II. RELATED WORK

The expected demise of Moore’s law has fueled active research in the area of emerging computer architectures over the last decade. This problem has been further exaggerated by the rise of big data and the consequent narrowing of the von Neumann bottleneck between the processor and the memory. We briefly survey both these topics in Sections II-A and II-B.

Our proposed solution is based on the use of nanoscale memristors as non-volatile switches storing the data values and guiding the flows of current through the nanoscale memristor crossbar to implement a desired computation. Section II-C introduces memristors and their dynamics, and Section II-D briefly describes other efforts that have used nanoscale memristor crossbars for implementing in-memory computing. Our survey of related work is certainly neither exhaustive nor representative; it is only an attempt at providing a general perspective to the potential readers of this manuscript.

A. Moore’s law and its limitations

Over the last five decades, Moore’s Law [18], [19] has been a driving factor towards immense technological and societal progress. Akin to how Moore’s Law paints a picture for the future of chip design, Dennard scaling [20] examines possibilities for the scaling of devices. Dennard scaling hit a roadblock as CMOS transistors started replacing MOSFETs as this has been marked by a shift from micrometer (\(\mu m\)) scale to nanometer (nm) scale devices.

Interesting breakthroughs have been made in the area by using self-aligned double gate MOSFET structures (FinFET) [21], which are projected to scale down to about 10nm or smaller. The chip density and the energy efficiency of CMOS-based architectures have been increased by three dimensional layering of active devices [22] and by intuitively utilizing voltage scaling techniques [23], respectively. As the device dimensions are reduced, the traditional challenges posed by capacitance and resistance become more daunting, dielectric properties of materials cause more interference, and physical properties like orientation of silicon lattice structures, wafer thickness and mechanical strain start having more effects on the device behavior.

Our flow-based computing approach does not require the repeated switching of devices to perform computation and exploits the natural flow of current through devices storing data to produce results. By not relying on repeated device switching during a computation, our approach is less affected by Dennard scaling and dark silicon. We believe that this property makes our flow-based computing approach different from many other memristor-based approaches that require repeated switching of memristor devices [4], [24]–[26] for logical computations.

B. John von Neumann bottleneck and beyond

John von Neumann’s “First Draft of a Report on the EDVAC” [27] first circulated in 1945 defined a new computer architecture that has thrived for the last 74 years. The stored-program computer became the de facto computer and was widely replicated for commercial use. The architecture can be explained very simply by breaking it down into its modular components – primarily the central processing unit and the memory. The von Neumann computer can be interfaced with input and output devices. The mathematical and engineering foundations for this architecture were subsequently laid in [28].

A weakness in the von Neumann architecture is the processor-memory bottleneck, as discussed in [29], [30]. The bus connecting the processor and the memory will always have a limited bandwidth, thereby throttling performance. Despite the dynamic random access memory (DRAM) performance improving exponentially, it has been shown [31] that the processor-memory gap will also increase in an exponential manner – ultimately hitting a memory wall.

An inelegant but useful solution to the von Neumann bottleneck was the use of on-chip cache memory as a component of the existing memory hierarchy [32], which has since become an industry standard for designing microprocessors. Alternative approaches towards the nullification of the von Neumann bottleneck include the data-flow based programming paradigm [33], the processor-in-memory (PIM) architecture [34], in-memory computation of logic operations [35], and the end-to-end computation-in-memory (CIM) model [36].

Our approach completely nullifies the von Neumann bottleneck by storing data in a specific pattern on the memristor crossbar and using only the flow of current through naturally occurring sneak paths in the crossbar to perform the desired computation. Our solution still uses von Neumann’s stored program concept and relies on a classical von Neumann computer to compile a program into the desired pattern once
such that the data should be loaded onto the crossbar using such a pattern to implement the desired program [37].

C. Memristor – the fourth fundamental electrical element

In 1971, Leon Chua used symmetry arguments to hypothesize that there exists a fourth fundamental electrical element in addition to the known elements: resistors, capacitors and inductors. This new element, which demonstrated a relation between flux linkage and electrical charge, was termed as a “memristor” [38] – an amalgamation of memory and resistor. This name arose from the fact that a memristor’s resistance varies with the net current flowing through it and the resistance state can be maintained even if the device is disconnected from a power source, thereby giving it the property of non-volatile memory. Given the voltage \(v\), the current \(i\), the charge \(q\) and the flux linkage \(\phi\), one can define \(d\phi(q)\) as the change in flux linkage. The memristance \(M(q)\) of a memristor is defined as follows:

\[
M(q) = \frac{d\phi(q)}{dq}
\]  

(1)

The voltage across a memristor at time \(t\) is characterized as:

\[
v(t) = M(q(t))i(t)
\]  

(2)

The voltage drop across the memristor decreases and the current across it increases as it transitions from a high resistance state (HRS) to a low resistance state (LRS), when an external positive voltage is applied across the memristor.

In contrast, the voltage drop across the memristor increases and the current across it decreases while transitioning from a low resistance state (LRS) to a high resistance state (HRS), when an external negative voltage is applied across the memristor. Hence, the memristor device can function as a switch, and can transition from LRS to HRS (and vice versa). A memristor in the LRS allows current to flow across it, akin to a closed circuit or a turned-on switch. A memristor in the HRS allows no current to flow, similar to a turned-off switch.

The dynamics of memristors were further investigated and characterized using Lissajous figures by Chua [39]. A memristor’s transitional behavior is identified as a pinched hysteresis loop, which underlies the special dynamics associated with a memristor. This makes it possible to explore the possibilities of using a memristor as an interesting non-linear element for computation, even though our approach uses memristor only as a digital switch and a non-volatile memory element.

D. Applications of memristors

Due to their non-volatile property, memristors and other ReRAMs are primarily employed as memory elements in novel memory architectures. A compelling design for caches has been proposed [40] that leverages 3D stacked ReRAMs. A framework for performing stateful logic using memristor-based nanoscale crossbar circuits has been investigated [41]. The crossbar fabric has also been shown to support parallel computation [42]. It has been demonstrated in [43] that Boolean logic operations can be reliably implemented using a single memristor – the trick lies in applying the input bits periodically in a sequential manner instead of applying them all at once. One of the impressive displays of using memristors at the core of parallel computation can be found in [25]. In [44], a promising solution has been suggested in order to unleash the potential of in-memory computation with memristors for big data applications. A search-based synthesis procedure has been used to design edge detection crossbars [45]. However, these designs are about 3.5 times larger than our designs, and no PSNR-based comparison of the resultant images has been performed in the aforementioned publication. Their approach directly uses our earlier work [37] but does not leverage model counting.

There exists exciting contemporary research in neuromorphic computing, both in the design of neuromorphic systems with non-volatile resistive components, as well as the development of new devices targeted towards neuromorphic architectures [46], [47]. A well-known technique for neuromorphic computation is the usage of spiking signals. A spiking based neuromorphic design was proposed in [48] that lead to more than 50% in energy savings with insignificant decrease in recognition probability. A system based on similar operating principles, in which both training and classification is performed on the crossbar array was proposed in [49]. It is well known that neurons are dynamical systems; an alternative computational paradigm which takes advantage of such behavioral characteristics using Mott memristors was introduced in [50].

III. Flow-based computing with crossbars

The key challenge in our flow-based computing approach [51] is to design the pattern in which data should be loaded onto a two-dimensional array of nanoscale memristors such that the flow of current through the crossbar can perform a desired computation. Figure 2 illustrates the flow-based computing approach using the simple Boolean formula “a AND b”. All the memristors in the crossbar are first turned off; then, the data corresponding to the values of Boolean variables are loaded onto the memristors labeled “a” and “b” respectively. Finally, a flow of current introduced in one (say, rightmost) nanowire reaches another (say, leftmost) nanowire if and only if the formula “a AND b” is true.

In general, a Boolean formula is mapped onto a crossbar design with the following properties [51], [52]:

- Each memristor in the crossbar is mapped to a literal in the Boolean formula or the logical constants True and False.
- The memristors are switched ON or OFF depending on the patterns in which data is to be loaded for a specific computation and the specific input instance.
- A current flow is introduced in an \(a \text{ AND } b\) specified input nanowire.
- If the Boolean formula evaluates to TRUE for a given input instance, then a current flow is observable at a predefined output nanowire. On the other hand, if the Boolean formula evaluates to FALSE for a given input instance, then a current flow is not observable at a predefined nanowire.
An example of a crossbar implementing the Boolean formula \( \neg A \land \neg B \land \neg C \) is shown in Figure 3. The blue circles represent memristors which assume the value of literals, the green circles represent memristors in the ON state and the grey circles represent memristors in the OFF state. In Figure 3(b) and 3(c), the red arrows represent current flow and the crosses represent a lack of current flow.

As an illustration, Figure 4(a) shows the design of a 1-bit full adder that computes both the sum and the carry-out bits. A grey circle indicates a memristor that is always turned off; a blue circle is labeled with the literal whose value decides whether the corresponding memristor is turned on or off. A memristor that is always turned on is indicated by a green circle.

Figure 4(b) shows the operation of the flow-based computing approach for the inputs \( A = 0, B = 0 \), and \( C_{in} = 0 \). The black circles indicate memristors that are turned off while the green memristors indicate memristors that are turned on. The red arrows and the black crosses indicate that there is no path that allows the current to flow from the input to the output – thereby, yielding a sum of 0 and a carry-out of 0.

Figure 4(c) shows the memristor crossbar when the data corresponding the inputs \( A = 1, B = 1 \), and \( C_{in} = 1 \) have been loaded into the nanoscale memristor crossbar. Again, the turned-on memristors are indicated by green circles and the turned-off memristors are shown by black circles. The red arrows show how the current can flow from the input to the outputs – thereby, yielding both a sum and a carry-out of 1.

In order to synthesize crossbar designs successfully, it is inefficient and cumbersome to stick to electrical models of crossbar circuits. Hence, we encapsulate crossbar circuits in a digital abstraction that can be easily manipulated using Boolean algebra to aid in the synthesis procedure:

Definition 1 (CROSSBAR). A memristor-based crossbar is a 3-tuple \( C = (M, W_r, W_c) \) where
\[ M = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{pmatrix} \]

is a two-dimensional array of memristors with \( l \) rows and \( n \) columns, where \( m_{ij} \in \{0,1\} \) denotes the state of the memristor (ON or OFF) connecting row \( i \) with column \( j \).

- \( \mathbb{W}_r = \{r_1, \ldots, r_l\} \) is the set of horizontal nanowires such that the wire \( r_i \) provides the same input voltage to every memristor in row \( i \).
- \( \mathbb{W}_c = \{c_1, \ldots, c_n\} \) is the set of vertical nanowires such that the wire \( c_j \) provides the same input voltage to every memristor in column \( j \).

The memristor \( m_{ij} = 0 \) is said to be in the high-resistance state or OFF state and \( m_{ij} = 1 \) denotes a memristor in the low-resistance state or ON state.

**Definition 2 (CROSSBAR DESIGN).** A crossbar design \( D(M) \) maps each memristor \( m_{ij} \) in the crossbar \( M \) to one of the following: an input Boolean variable \( v_1, \ldots, v_n \), its negations \( \neg v_1, \ldots, \neg v_n \) or the logical constants True or False.

For a crossbar with \( l \) rows, \( n \) columns and \( v \) different input variables, each memristor can be mapped to \( 2^v + 2 \) different values. Hence the number of possible crossbar designs is as large as \( (2v + 2)^{nxl} \). As an illustration, the edge detection example in this manuscript has 16 Boolean variables on a 8x8 crossbar; this corresponds to a search space involving 1.158e+77 designs. The number of designs is comparable to the number of atoms in the known universe.

**IV. SYNTHESIS USING MODEL COUNTING**

In order to search the exponentially growing space of possible crossbar designs, our synthesis procedure exploits a symbolic representation of Boolean functions and employs model counting [53] to guide a simulated annealing based search procedure. Our approach for synthesizing flow-based computing crossbar circuits relies on approximating the distance between a given crossbar design and the target Boolean function using model counting.

The Boolean satisfiability problem is regarded as one of the foundational problems in computer science. A propositional formula is deemed to be satisfiable if there exists an assignment of values to its variables for which the formula evaluates to true. By extension, model counting is the problem of computing the number of models for a given propositional formula, or finding the number of distinct input variable assignments for which the propositional formula evaluates to true. For a given crossbar design, we would want the output flow states to match the corresponding evaluations of the Boolean formula \( \phi \), for all possible input assignments. Hence, our objective is to count the number of models which satisfy the equivalence of the Boolean formula with the candidate crossbar design. This problem is readily solved using existing model counting algorithms. Extended discussions about approaches to model counting have been presented in a handbook [53].

Algorithm 1 summarizes our methodology. In line 1, we first pick a random crossbar design of size \( l \) rows and \( n \) columns. Each memristor in the design is mapped to either True, False, one of the variables \( v_1, \ldots, v_k \) or one of the negated variables \( \neg v_1, \ldots, \neg v_k \), as sampled from a uniform random distribution.

**Algorithm 1: Crossbar Synthesis Algorithm**

**Input:** Target Boolean Formula \( \phi \) over variables \( \{v_1, v_2, \ldots, v_k\} \)
- Size of crossbar \( C \): \( l \) rows and \( n \) columns
- Initial temperature for simulated annealing \( T \)
- Cooling rate \( c \)

**Output:** Crossbar Design \( D(M) \) mapping each memristor \( m_{ij} \in M \) to the set \( \{True, False, v_1, \ldots, v_k, \neg v_1, \ldots, \neg v_k\} \)

1. \( D_1 \leftarrow \text{PickRandomCrossbarDesign}(l, n, v_1, \ldots, v_k) \)
2. \( D(B(D_1)) \leftarrow \text{BooleanFlow}(D_1) \)
3. \( \Delta_i \leftarrow \text{ModelCount}(B(D_1) \oplus \phi) \)
4. While \( \Delta_i > 0 \) do
5. \( D_{t+1} \leftarrow \text{PerturbCrossbarDesign}(D_t, \phi) \)
6. \( B(D_{t+1}) \leftarrow \text{BooleanFlow}(D_{t+1}) \)
7. \( \Delta_{t+1} \leftarrow \text{ModelCount}(B(D_{t+1}) \oplus \phi) \)
8. If \( \text{rand}(0, 1) < e^{-(\Delta_{t+1} - \Delta_t)/T} \) then
9. \( i \leftarrow i + 1 \)
10. \( T \leftarrow c \times T \)
11. \end
12. \end
13. \text{Return} crossbar design \( D_i \)

For each crossbar design \( D \), we assume that a flow of current is injected into the top horizontal nanowire. Then, we compute the number of assignments to the Boolean variables which cause the flow of current to reach the lowermost horizontal nanowire of the crossbar design \( D \).

In order to evaluate the model, we must first compute the Boolean formula realized by the crossbar. Let \( i^{(0)}_i \) denote the flow value of row \( i \) at time \( t \), and \( c^{(t)}_j \) denote the flow value of column \( j \) at time \( t \). At \( t = 0 \), only the first row has flow; i.e., \( i^{(0)}_0 = True \) and all other rows and columns are set to False. For all \( t > 0 \), we define the following transitions for each nanowire based on the ability of turned-on memristors to create a path between their horizontal and vertical nanowires. In this way, we symbolically compute the Boolean formula representing the values of the memristors for which a flow reaches the topmost nanowire of the crossbar and denote it by \( \text{BooleanFlow}(D_i) \) in line 2 of the algorithm.

\[
\forall i \in \{2, \ldots, n\}, \quad i^{(t+1)}_i \iff (i^{(t)}_i \lor \bigvee_{1 \leq j \leq n} (m_{ij} \land c^{(t)}_j))
\]

\[
\forall j \in \{1, \ldots, m\}, \quad c^{(t+1)}_j \iff (c^{(t)}_j \lor \bigvee_{1 \leq i \leq l} (m_{ij} \land i^{(t)}_i))
\]

The above transitions of the rows and columns in the crossbar have been represented using Boolean functions and can be described succinctly using Boolean Decision Diagrams (BDDs), And-Inverter-Graphs (AIGs) or other representations.
Line 3 of the algorithm computes the number of satisfiable instances $\Delta_1$, which corresponds to the symmetric difference between the target Boolean formula $\phi$ and the formula corresponding to the computation performed by the crossbar design $D_1$. Several competitive implementations of approximate model counting algorithms are available and our approach is agnostic to the choice of the model counting strategy as long as the algorithm produces the count of 0 feasible models only for unsatisfiable formula [53].

The loop in line 4 through line 12 continues to modify the current crossbar design, evaluate the function computed by this perturbed design, and count the number of satisfiable instances to the Boolean formula corresponding to the symmetric difference of the target Boolean formula and the formula computed by the current crossbar design. Lines 8 through 10 show the probabilistic acceptance step of the simulated annealing algorithm. New crossbar designs are always accepted if they are better than the existing crossbar design. New crossbar designs that are worse than the existing crossbar design are accepted with a probability that is a function of both the quality of the designs and the current temperature of the simulated annealing algorithm.

At every iteration of the loop, the temperature of the simulated annealing algorithm is slightly lowered. When the number of satisfiable instances for the symmetric difference becomes zero (or falls below an assigned threshold), our algorithm stops and reports the synthesized crossbar design.

### TABLE I: Computation delay (picoseconds) to compute the MSB of $n$-bit addition for area-optimized crossbars.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>425</td>
<td>3400</td>
<td>1020</td>
<td>340</td>
<td>125%</td>
</tr>
<tr>
<td>3</td>
<td>765</td>
<td>5100</td>
<td>1530</td>
<td>340</td>
<td>225%</td>
</tr>
<tr>
<td>4</td>
<td>1020</td>
<td>6800</td>
<td>2040</td>
<td>425</td>
<td>240%</td>
</tr>
</tbody>
</table>

### TABLE II: Power (in $\mu W$) required to compute the MSB of binary addition for area-optimized crossbars.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>240</td>
<td>1200</td>
<td>360</td>
<td>300</td>
<td>80%</td>
</tr>
<tr>
<td>3</td>
<td>420</td>
<td>1800</td>
<td>540</td>
<td>390</td>
<td>107.7%</td>
</tr>
<tr>
<td>4</td>
<td>600</td>
<td>2400</td>
<td>720</td>
<td>720</td>
<td>81.3%</td>
</tr>
</tbody>
</table>

The compact crossbar designs for a 4-bit addition and a 4-bit comparator obtained from model counting are shown in Figure 5. Both the crossbars compute the most significant bit of $n$-bit binary computations. The correctness of the designs has also been verified using ngSPICE-26 simulations [56]. Table I compares the crossbar designed using model counting with those designed using other competing approaches. The last column in Table I represents the speedup achieved through model counting with respect to the prior best method.

Similarly, Table II presents a comparison between the power required by crossbars designed using our model counting approach and compares it to memristor-based designs obtained using other approaches. The last column in Table II represents the power consumption ratio of crossbars produced through model counting with respect to the prior best method.

### V. EDGE DETECTION WITH FLOW-BASED INPUT-AWARE CROSSBAR COMPUTING

Applied problems in domains such as image processing and vision require reasonably accurate but not necessarily exact computations. Such soft applications naturally provide an opportunity for the exploration of approximate input-aware computing with a focus on correctness over the likely inputs as opposed to all possible inputs.

One of the fundamental procedures in image processing is edge detection, that is, the demarcation of boundaries between the different objects in a given image. While there
exist numerous algorithms to perform efficient and highly sophisticated edge detection in images [57], we can say that an “edge” exists if the difference between the gray-scale values of two pixels exceeds a predefined threshold value. Since a pixel can have any value between 0 and 255, the simplest edge detection operation involves computing the difference between two 8-bit binary numbers followed by a comparison of the difference with a constant 8-bit binary number representing the threshold at which an edge is said to exist.

We employ the model counting approach explained in the previous section to synthesize crossbars for approximate computation of Boolean formulas. In this setting of approximate computation, we minimize the symmetric difference between the threshold-based exact edge detection and the Boolean function computed by the synthesized crossbar circuit. The approximate detection of edges is sufficient for practical purposes, and we do not require the number of satisfiable instances for the symmetric difference to reach zero – a fairly low value close to zero is acceptable.

Since computation is performed on two 8-bit pixels for threshold-based edge detection, the truth table consists of 65536 entries. Employing our model counting approach, we are able to obtain a crossbar design which has an accuracy of ~85% over all possible input values.

Figure 6 illustrates the approximate crossbar design. In the crossbar image, the dark circles represent permanently OFF memristors, the green unlabeled circles represent permanently ON memristors, and the light blue circles represent memristors whose values correspond to the literals. The variables \( A_0, A_1, \ldots, A_7 \) and \( B_0, B_1, \ldots, B_7 \) represent the bits of the two pixels, and the variables \( \neg A_0, \neg A_1, \ldots, \neg A_7 \) and \( \neg B_0, \neg B_1, \ldots, \neg B_7 \) represent their respective negations.

We employ the BSD500 database [58] for testing the performance of the obtained crossbar. The peak signal to noise ratio (PSNR) metric is commonly used for quality assessment between two perceptually equivalent images [59]. The PSNR between the edge images produced by exact computation and those produced by the approximate crossbar provides an estimate of the quality of edge detection performed by the crossbar circuit. The synthesized crossbar design is provided in Figure 6 and the histogram of the PSNR for all images in BSD500 are presented in Figure 8.

The PSNR histogram in Figure 6 shows that a single approximate computing crossbar over all the truth table entries does not really perform well in the case of edge detection since...
not all pixel pairs occur with the same frequency. Hence, we identify a subset of all possible pixel pairs that occur with high frequency. A given pixel pair belongs in this subset if it occurs more than three times in each image, on an average, over all the images in BSD500. We then design approximate input-aware edge-detection crossbars with a focus on these likely inputs. A few crossbar designs obtained using this approach are illustrated in Figure 7.

The downsized truth table involves only the chosen subset of pixel pairs and contains 16,658 entries. This input-aware truth-table enables the synthesized crossbar designs to perform well on the relevant inputs even though they have relatively lower accuracy over all possible truth table inputs. The subset

![Crossbar Diagrams](image)

**Fig. 7**: Crossbars for edge detection on input-aware pixel pairs.

![Histograms](image)

**Fig. 8**: Histogram of PSNR values corresponding to output images produced by majority-based combination of input-aware crossbar outputs.

A combination of crossbars produce a better (i.e. shifted to the right) PSNR distribution that a single input-aware crossbar.
of pixel pairs is utilized to build a model for the model counting approach, in order to synthesize a family of crossbar circuits for input-aware approximate edge detection.

Since the standalone designs are not accurate, it is intuitive that they may not agree on the same mistakes. Given a family of crossbar designs approximately computing the same function, their outputs should be combined by using a majority operation. In order to test this hypothesis, we have synthesized a number of different crossbar designs with varying degrees of accuracy. The PSNR values between the computed edge images and the output images obtained by majority-based combination are presented in Figure 8. The input-aware crossbar is better i.e. it has a higher PSNR than a crossbar designed using all possible inputs. Similarly, a family of 5 crossbars produces better PSNR values than any crossbar in the family. While the best crossbar has a performance of about 6.275dB, combining it with crossbars of even less performance produces an overall design with a PSNR of 7.02dB. This shows the importance of using a (small) family of approximate designs as opposed to the best approximate design.

The edge images produced by the combination of approximately-correct input-aware crossbars are better than the edge images produced by a single crossbar with higher accuracy over all possible inputs. The input gray-scale images, the computed edge images and the output edge images produced by combining the outputs from a family of five approximate input-aware crossbars are presented in Figure 9.

VI. CONCLUSION AND FUTURE WORK

We have explored the design of input-aware approximately-correct flow-based crossbar circuits for basic computer vision applications, like edge detection. Our proposed approach is driven by model counting to explore the design space of crossbar circuits. The synthesized designs are compact, and approximately compute the necessary functions for the desired applications. We have produced an assortment of 8x8 crossbar designs belonging to two different operational groups. One of the groups can perform threshold-based edge detection with high accuracy on all possible pixel pairs and the other group performs approximate edge detection on an application-specific subset of input values. In case of the second set of crossbar designs, the outputs from individual crossbars are combined using the majority function to yield the final output image. We have tested the performance of our method on the well-known BSD500 database. We utilize the peak-signal-to-noise-ratio (PSNR) metric to evaluate the quality of the output images. The results obtained from the input-aware approximate computation are observably better than those produced by the more accurate general-purpose crossbars.

Our work has focused on the design of crossbar arrays for approximate computations and has not explored the impact of peripheral circuitry. Our approach only requires standard peripheral circuitry required to read and write data on the memristor crossbar but does not need any additional peripheral circuitry to control the flows. Our design automation approach is based on the native use of memristor crossbars as a
Fig. 9: The input grayscale image, the computed edge image and the edge image based on combining approximately correct input-aware crossbar outputs using a majority function.
computational fabric without trying to control the flow of information through the crossbar using external peripheral circuitry. In fact, our design permits all possible flows to occur and the memristor crossbar is designed in a manner where all possible flows put together enable a desired computation. In previous experimental studies [60] we have shown that a 1-bit adder can be implemented in practice on a hardware memristor crossbar and the current theoretical investigation extends the same approach without requiring any additional changes in peripheral circuitry. It should be noted that the efficiency and correctness of the peripheral circuitry will have an impact on the overall performance of our approach, and these important impacts need to be investigated.

The design of approximate input-aware application-specific solutions using flow-based computing is an area that is ripe with exciting opportunities. Future work related to the approach presented in this manuscript includes, but is not limited to, the following possibilities:

1. A natural question to investigate is the dependency of the crossbar design on the data set used to train our input-aware approach. Figure 10 illustrates the BSD500 and CIFAR [61] datasets. The axes in the color map represent the intensity of the pixels in the pair and the color bar represents the natural logarithm of the frequency with which the pixel pairs occur in the respective datasets. Because of the remarkable visual similarity of these distributions, one can conjecture that the crossbar design is not sensitive to the choice of a specific data set. Future work is needed to investigate this dependence more thoroughly.

2. The exploration of the design space of crossbars can benefit by leveraging the massively parallel multi-core computing capabilities of general purpose graphics processing units (GPGPUs). In particular, deep learning approaches may be helpful in accelerating the evaluation of a candidate crossbar design.

3. A number of higher level artificial intelligence applications depend on fundamental operations like convolutions, clustering and classification. Embedding such operations on compact crossbar circuits may have a significant impact on designing hardware accelerators for machine learning and artificial intelligence.

4. Our current approach has only employed memristors as binary switches. Memristors and other resistive memory devices exhibit a multitude of discrete memory states in the interval between the on and off states. This interesting property can unlock a novel pathway for doing ternary and other n-ary logic computations on the memristor crossbar fabric.

**ACKNOWLEDGMENT**

We acknowledge support from NSF Awards #182976 and #1422257, an award from the Florida Cybersecurity Center, awards from the Royal Bank of Canada, and the Air Force Young Investigator Award to Sumit Jha. Dr. Steven Fernandes acknowledges support from the University of Central Florida Preeminent Post-doctoral Fellowship Program.

**REFERENCES**


E. J. Merced-Grafals, N. Davila, N. Ge, R. S. Williams, and J. P. Strachan, “Repeatable, accurate, and high speed multi-level programming of memristor 1T1R arrays for power efficient analog computing applications,” Nanotechnology, vol. 27, no. 36, p. 365202, 2016.


Dwaipayan Chakraborty received the Ph.D. degree in Computer Science from the University of Central Florida, Orlando, USA in 2019. His research interests include post Moore’s Law computer architectures, memristor-based computing, stochastic computing, and nanoscale architectures. He has published papers in several reputed international conferences and journal including Design, Automation, and Test in Europe (DATE 2016, 2017), IEEE International Symposium on Circuits and Systems (ISCAS 2017), IEEE International Conference on Nanotechnology (IEEE NANO 2017) and the IEEE Transactions on Circuits and Systems (TCAS) II.

Sunny Raj received the B.Tech degree in Computer Science and Engineering from Indian Institute of Technology (Indian School of Mines), Dhanbad, India in 2011. He is a PhD candidate in Computer Science at the University of Central Florida, Orlando, USA. His research interests include machine learning, quantum computing, and emerging computer architectures. He has published papers at BMC Bioinformatics Journal and reputed international conferences including 16th International Conference on Formal Modelling and Analysis of Timed Systems (FORMATS-2018), 10th International Symposium on Foundations and Practice of Security (FPS 2017), and IEEE International Conference on Nanotechnology (IEEE NANO 2017).

Steven Lawrence Fernandes received the Ph.D. degree in Electronics and Communication Engineering from Karunya Institute of Technology and Sciences, India in 2017. He is a postdoctoral researcher in Computer Science at the University of Central Florida, Orlando, USA. He worked as a postdoctoral researcher in Electrical and Computer Engineering at the University of Alabama at Birmingham, USA. His research interests include machine learning and computer vision. He has served as a Guest Editor of Journal of Computational Science (2016-2017) and Future Generation Computer Systems (2017-2018). He has published papers at Pattern Recognition Letters and Computers in Biology and Medicine.

Sumit Kumar Jha received the Ph.D. degree in Computer Science from Carnegie Mellon University in 2010. He is an Associate Professor with the Computer Science Department at the University of Central Florida, Orlando, USA. Dr. Jha has worked on R&D problems at Microsoft Research India, General Motors, INRIA France and the Air Force Research Lab Information Directorate. His research has been supported by the National Science Foundation, the Air Force Office of Scientific Research, the Oak Ridge National Laboratory, the Royal Bank of Canada, the Florida Center for Cybersecurity, and the Air Force Research Laboratory. Dr. Jha was awarded the prestigious Air Force Young Investigator Award in 2016, and his research has led to three Best Paper awards.