SCIENTIFIC REPORTS
natureresearch

OPEN

Exploiting Dual-Gate Ambipolar CNFETs for Scalable Machine Learning Classification

Farid Kenarangi[1*], Xuan Hu[2], Yihan Liu[3], Jean Anne C. Incorvia[3], Joseph S. Friedman[2] & Inna Partin-Vaisband[1]

Ambipolar carbon nanotube based field-effect transistors (AP-CNFETs) exhibit unique electrical characteristics, such as tri-state operation and bi-directionality, enabling systems with complex and reconfigurable computing. In this paper, AP-CNFETs are used to design a mixed-signal machine learning logistic regression classifier. The classifier is designed in SPICE with feature size of 15 nm and operates at 250 MHz. The system is demonstrated in SPICE based on MNIST digit dataset, yielding 90% accuracy and no accuracy degradation as compared with the classification of this dataset in Python. The system also exhibits lower power consumption and smaller physical size as compared with the state-of-the-art CMOS and memristor based mixed-signal classifiers.

Power consumption and physical size of integrated circuits (ICs) is an increasing concern in many emerging machine learning (ML) applications, such as, autonomous vehicles, security systems, and Internet of Things (IoT). Existing state-of-the-art architectures for digital classification, are highly accurate and can provide high throughput[1]. These classifiers, however, exhibit high power consumption and occupy a relatively large area to accommodate complex ML models. Alternatively, mixed-signal classifiers have been demonstrated to exhibit orders of magnitude reduction in power and area as compared to digital classifiers with prediction accuracy approaching the accuracy of digital classifiers[2–9]. While significant advances have been made at the ML circuit and architecture levels (*e.g.*, in-SRAM processing[2], comparator based computing[3], and switched-capacitor neurons[4]), the lack of robust, ML-specific transistors is a primary concern in ML training and inference with all conventional CMOS technologies. To efficiently increase the density and power efficiency of modern ML ICs while enabling complex computing, emerging technologies should be considered.

While the non-volatility of memristors has proven quite attractive for storing the weights required for feature-weight multiplication[10–13], field-effect transistors provide several advantages over memristors for ML. First, transistors provide a broader range of linear tuning of resistance, thereby better matching ML models. Second, transistors are not subject to the deleterious aging that deteriorates memristor behavior over time. Finally, the connectivity between feature-weight multiplication layers requires electrical signal gain, which cannot be provided by memristors; transistors can be used for such interlayer connections, thereby enabling a monolithic integrated circuit that can be fabricated efficiently. Of particular interest for on-chip classification are ambipolar devices. Owing to unique electrical characteristics, as described in the next section, ambipolar devices are expected to provide efficient on-chip training and inference solutions and reduce design and routing complexity of ML circuits.

A carbon nanotube (CNT) ambipolar device has been reported as a potential candidate for controllable ambipolar devices because of its satisfying carrier mobility and its symmetric and good subthreshold ambipolar electrical performance[15]. Based on the dual-gate CNT device's electrical performance, a library of static ambipolar CNT dual-gate devices based on generalized NOR-NAND-AOI-OAI primitives, which efficiently implements XOR-based functions, has been reported, indicating a performance improvement of ×7, a 57% reduction in power consumption, and a ×20 improvement in energy-delay product over the CMOS library[16].

Existing results exploit the switching characteristics of the AP-CNFETs for enhancing digital circuits[16,18]. In this work, we repurpose the AP-CNFET device for neuromorphic computing. Owing to the dual gate structure, AP-CNFET significantly increases the overall density of analog ML ICs, simplifies routing, and reduces power consumption. To the best of the authors knowledge, the AP-CNFET based ML framework is the first to

[1]University of Illinois at Chicago, Chicago, 60607, Illinois, USA. [2]University of Texas at Dallas, Richardson, 75080, Texas, USA. [3]University of Texas at Austin, Austin, 78712, Texas, USA. *email: fkenar2@uic.edu

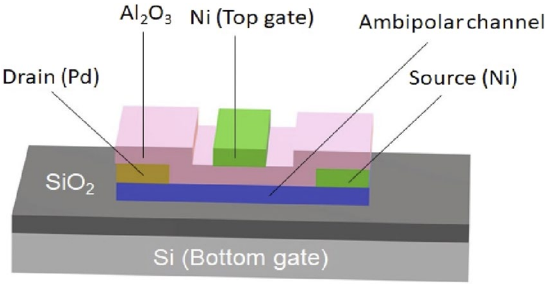

Figure 1. Schematic of a top-bottom dual-gate ambipolar device.

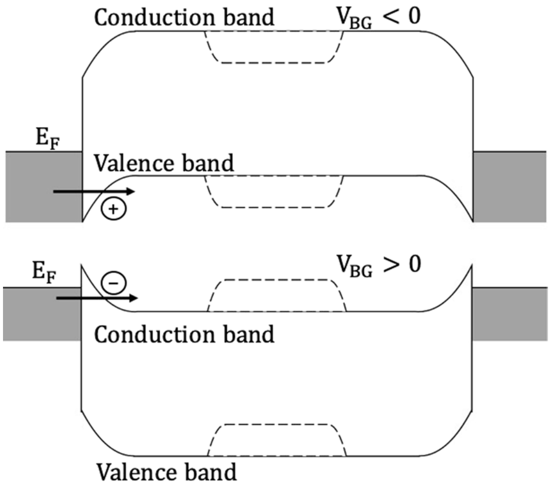

Figure 2. Schematic band diagrams of a dual-gate for bottom gate voltage <0 and >0. The solid and dashed lines at the middle region show how the top gate voltage switches the operating state by allowing or stopping the current flow[15].

demonstrate a multiplication-accumulation (MAC) operation with single-device-single-wire configuration. Note that in CMOS classifiers at least two sensing lines are required to separately accumulate results of multiplication with positive and negative weights. Furthermore, additional circuitry is required to process the signals from the individual sensing lines into a final prediction. Alternatively, in memristor based classifiers, a crossbar architecture is typically used with a single sensing line per class. With this configuration, only positive (or negative) feature weights are, however, utilized. Thus, additional non-linear thresholding circuits are required for extracting the final decision.

The rest of the paper is organized as follows. The device background and electrical characteristics of AP-CNFET device are presented. The proposed scheme for utilizing AP-CNFETs for on-chip classification is described. The classifier is evaluated based on classification of commonly used Modified National Institute of Standards and Technology (MNIST) dataset and finally, the paper is concluded.

Background

AP-CNFET device.

Depending on the gate voltage, an ambipolar device allows both electrons and holes to flow from source to drain because of its narrow Schottky barrier width, as small as a couple of nanometers, between metal contacts and the channel[15]. Figure 1 shows the schematic of a top-bottom dual-gate ambipolar device. In contrast to a normal single-gate-control transistor, the bottom gate plays an important role in determining the type of majority carrier and on current. This dual-gate device's electrical characteristic can be understood by the schematic band diagrams shown in Fig. 2. For a sufficiently negative (positive) bottom gate voltage, the Schottky barrier is thinned enough to allow for holes (electrons) tunneling from the source contact into the channel to the drain. By tuning the top gate voltage more positive (negative) to alter the barrier height for carrier transport across the channel, the top gate can switch between the ON and OFF operating states.

Besides CNTs, some 2D semiconductors such as MoS_2, WS_2, WSe_2, and black phosphorus (BP) are also reported as ambipolar semiconductors at room temperature. Surface transfer doping and using different source and drain contact metal are reported as two effective ways to modulate its ambipolar characteristics to move the subthreshold curve's symmetric point to $V_{BG} = 0$ and reduce the difference between the n-branch and p-branch saturation current. Fig. 3 indicates that undoped multilayer BP FETs' saturation current and mobility in the n-branch are much lower than that of the p-branch. However, Cs_2CO_3 layers deposited over BP serve as an efficient n-type surface dopant to improve the electron transport in the BP devices, thereby inducing either a more balanced ambipolar or even-transport-dominated FET behavior[14]. Also, by using Ni as the source metal and Pd

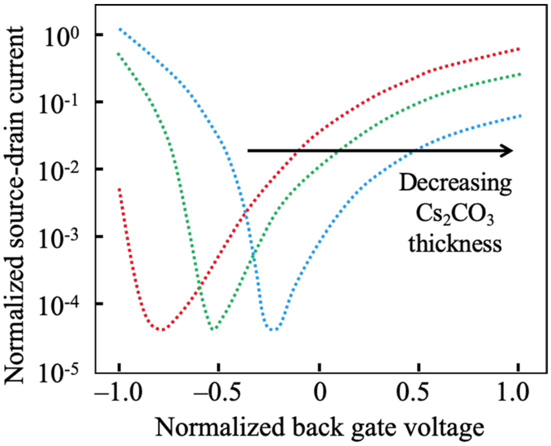

Figure 3. Forward-transfer characteristics (bottom gate voltage (V_{BG}) varies from −80 V to 80 V) evolution of a BP FET measured at V_{sd} = 100 mV in logarithmic scale with increasing Cs_2CO_3 thickness from 0 to 1.5 nm[14].

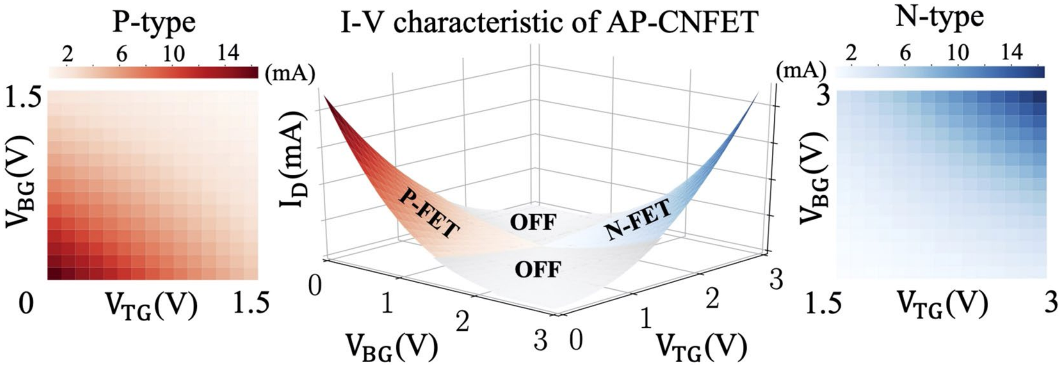

Figure 4. IV-characteristics of the AP-CNFETs.

as the drain contact, the experimental transfer characteristics curve of a multilayer WSe_2 FET indicates this configuration allows for ambipolar characteristics with both the electron and hole conduction current levels being similar[17].

Electrical characteristics of the AP-CNFET.

As compared with a conventional MOSFET, an AP-CNFET exhibits two unique characteristics as explained below.

Tri-State operation. Due to ambipolarity of an AP-CNFET, the majority carriers in the device can be either electrons or holes, depending upon the gate biases. The same device can, therefore, operate as n-type (majority carriers are electrons) or p-type (majority carriers are holes) FET. The I-V characteristics (*i.e*., drain current I_D versus top gate voltage V_{TG} and bottom gate voltage V_{BG}) of a single AP-CNFET are shown in Fig. 4, exhibiting p-type (red), n-type (blue), and OFF (gray) operational regions. In typical mixed-signal ML classifiers, positive and negative classification decisions are separately accumulated on the individual sensing lines[2,5,6,19]. Thus, at least two wires are required for a single MAC operation. Alternatively, with the proposed structure, the unique tri-state operation of AP-CNFETs is leveraged for merging the sensing lines, significantly reducing the routing complexity and area overhead, as described in the next section.

Bi-directionality. Another unique characteristic, is the bi-directionality of the device[20,21]. The drain and source terminals are determined based on the potential difference across the device. The terminals with higher and lower potentials act as, respectively, the drain and source nodes. Exploiting the bi-directionality of the AP-CNFETs, highly reconfigurable systems can be targeted. One approach is using bi-directionality for on-chip training, where the current direction within the individual AP-CNFETs is adjusted during each training iteration. By interchanging the drain and source terminals, similar current values yet in opposite directions can be generated, giving rise to fundamentally new, efficient, and reconfigurable implementation of ML algorithms.

Leveraging AP-CNFET for machine learning classification

In this section, AP-CNFET based ML classification is demonstrated. In the following subsections, the ML background is provided and the circuit level operation principles are explained.

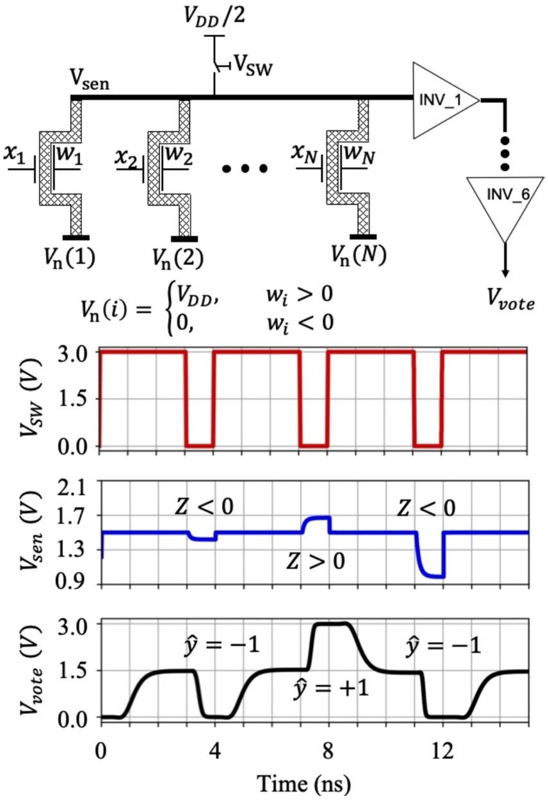

Figure 5. The overall schematics of the proposed N-dimensional, binary classifier. The top and bottom gates of the AP-CNFETs are connected to, respectively, the corresponding features, x_i, and feature weights, w_i. To encode the weight sign, the individual AP-CNFETs are biased as either p-type ($w_i > 0$) or n-type ($w_i < 0$) states. The circuit operates in two stages: precharge and classification (see the red V_{SW} waveform). During the precharge state, the sensing line is charged to $V_{DD}/2$. The classification decision is made during the classification phase based on the line voltage level as compared to $V_{DD}/2$ (see the blue V_{sen} waveform). The final decision between two classes is made based on the output of the buffers chain (see the black V_{vote} waveform).

ML background. Owing to its dual gate structure and electrical properties, AP-CNFET is a natural choice for embedded AI. The superiority of the AP-CNFET based learning over the traditional approaches is demonstrated in this paper based on the linear ML classification problem. Linear predictors are commonly preferred for on-chip classification due to their simplicity, satisfactory performance in categorizing linearly separable data, and low design complexity and hardware costs. Note that the proposed scheme is robust and can be utilized with more complex systems, such as non-linear ML classifiers and deep neural networks. With a multivariate linear classifier, the system response Z is an accumulated dot product of N input features $x = (x_1, x_2, \ldots, x_N)$ and model weights $w = (w_1, w_2, \ldots, w_N)$,

$$Z = \sum_{i=1}^{N} w_i \cdot x_i, \quad Z \in \mathbb{R}. \tag{1}$$

Logistic regression (LR) algorithm is utilized in this paper to train the classifier based on gradient descent algorithm[22]. The model weights, w, are determined during supervised training by minimizing the prediction error between the system response, Z, and labeled training dataset. In inference, the probability threshold is used for predicting system response to unseen input data, exhibiting a simple on-chip implementation,

$$\hat{y} = \text{sign}(Z) = \text{sign}\left(\sum_{i=1}^{N} w_i \cdot x_i\right) = \begin{cases} 1, & Z \geq Z_{th} \\ -1, & Z < Z_{th}. \end{cases} \tag{2}$$

The described logistic regressor with the probability threshold of $Z_{th} = 0$ is referred to as logistic classifier. The accuracy of the proposed logistic classifier is evaluated as the percentage of all the correct predictions out of the total number of the unseen test data points.

Design of AP-CNFET classifier. The overall schematics of the classifier is shown in Fig. 5. The circuit is designed to perform a N-feature binary classification. A single sensing line is used for storing the classification results produced by N AP-CNFETs attached to the line. The top and bottom gates of the transistors are connected to, respectively, the corresponding features and feature weights. Consequently the current through each transistor is proportional to the feature-weight product. The tri-state attribute of the AP-CNFETs is exploited to encode

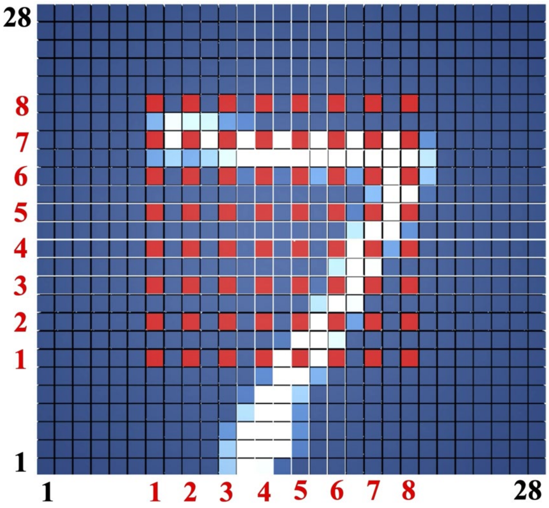

Figure 6. An example of MNIST digit image. Out of the total 784 features, 720 features are dumped during the downsampling stage. The selected 64 (8 × 8) features are shown by red squares.

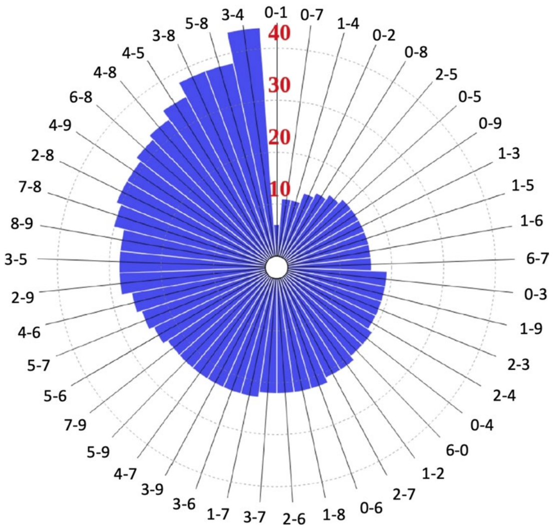

Figure 7. The selected number of features for each binary classifier (*i.e.*, i-j classifier).

the sign of the individual products, facilitating the accumulation of the multiplication results on a single sensing line. To encode the sign, the individual AP-CNFETs are biased as either p-type (for $w_i > 0$) or n-type (for $w_i < 0$) states. As a result, a certain amount of charge, as determined by the feature-weight product, is injected into (by p-type) or removed from (by n-type) the sensing line. To enable the tri-state operation, each device is connected between the common sensing line and either the V_{DD} (p-type) or ground (n-type) supply. Note that while biasing individual transistors in two operation regions requires additional wire connections, the advantages of single-line sensing prevail over the additional interconnect cost, as explained below.

The circuit operates in two stages: precharge and classification. During the precharge state, the sensing line is charged to $V_{DD}/2$ and transistors are gated with appropriate top gate biases (*i.e.*, V_{DD} for p-type and ground for n-type). During classification, the transistors are biased with their corresponding feature and feature weight values. As a result, the line is further charged (with p-type FETs) or discharged (with n-type FETs). Depending on the signs of the individual feature-weight products, the sensing line accumulates or dissipates a certain amount of charge, saturating at, respectively, higher ($>V_{DD}/2$) or lower ($<V_{DD}/2$) voltage V_{sen}. Consequently, the classification decision is determined as,

$$\hat{y} = \begin{cases} +1, & V_{sen} > V_{DD}/2 \\ -1, & V_{sen} < V_{DD}/2. \end{cases} \tag{3}$$

To extract the final decision, a non-inverting buffer is connected to the sensing line, as shown in Fig. 5. For $V_{sen} > V_{DD}/2$, the output of the sensing line is forced to $V_{vote} = V_{DD}$. The line is forced to $V_{vote} = 0$ for $V_{sen} < V_{DD}/2$.

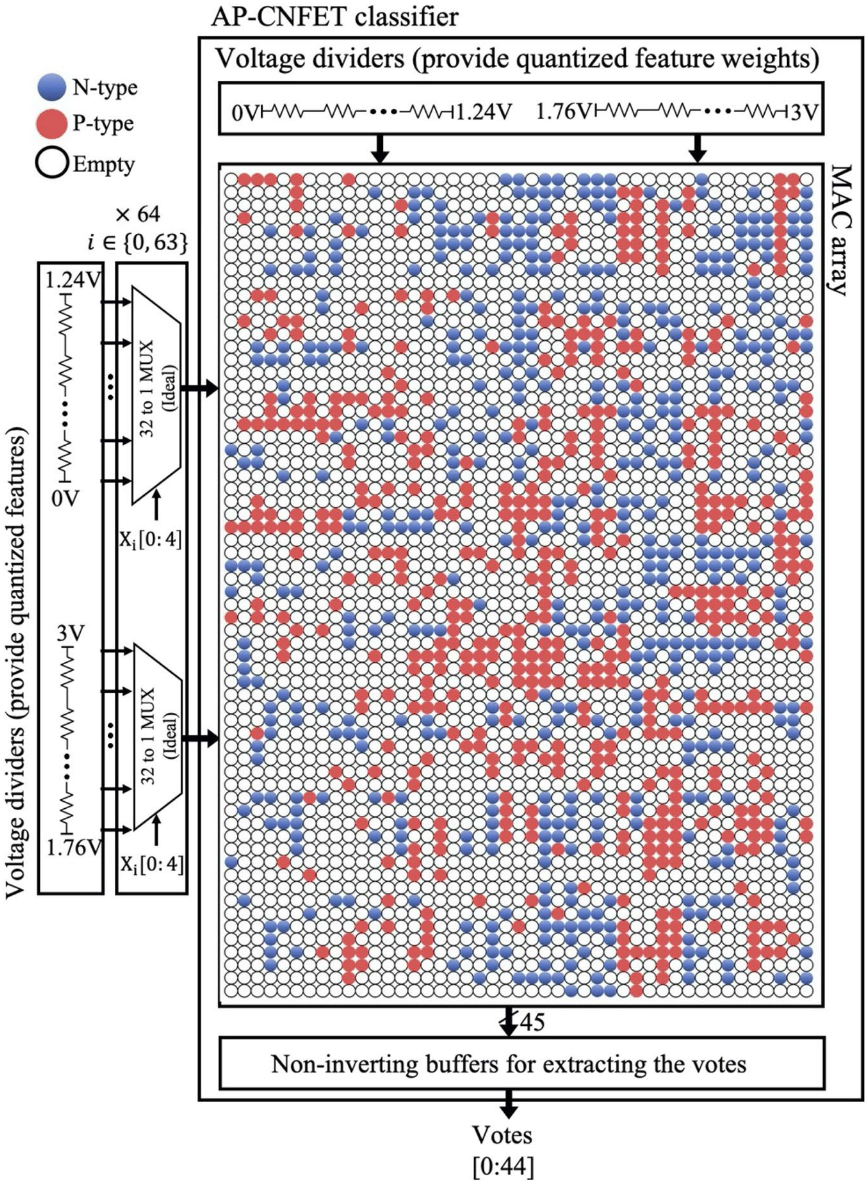

Figure 8. A schematic diagram of the proposed AP-CNFET based 10-class classifier. An array of 64 × 45 AP-CNFETs is used to perform the feature-weight products. The transistors biased in n-type and p-type regions are shown by, respectively, the blue and red circles. The transistors that have been removed as a result of SBS feature reduction are shown by white circles.

Note that using a single sensing line eliminates the critical conventional classification stage where the voltage levels at the positive and negative sensing lines are compared for determining the final binary classification result. The proposed circuit is also more resilient to process, voltage, and temperature (PVT) variations as compared with state-of-the-art mixed-signal classifiers, since it only uses inverters for extracting the final decision. Alternatively, in state-of-the-art classifiers[2,3], comparators, are utilized to extract the final decision. The correct operation of such comparators highly relies on symmetry of the circuit, exhibiting significant offset under PVT variations. To mitigate the sensitivity to PVT variations, additional complex circuitry is utilized in the existing ML classifiers (*e.g.*, compensating rows are used for this purpose[2]), increasing the overall power consumption, area, and routing complexity of these systems.

Linearity of the individual AP-CNFETs is critical for correct classification of the input features. Albeit the semi-linear dependence of output current on the gate biases across the full voltage range of operation (see Fig. 4), AP-CNFET exhibits no degradation in classification accuracy as compared to classification accuracy in Python, as presented in the next section. Alternatively, utilization of the wide bias region allows for quantization of features and feature weights with larger quantization step, increasing the resilience of the circuit to PVT variations. In this paper, five-bit resolution is considered for quantizing features and feature weights with a 40 mV step size. To quantize the features and feature weights, resistive voltage dividers are used. While feature weight connections are set to fixed values, multiplexer (MUX) units are used to update the features within each classification period. Note

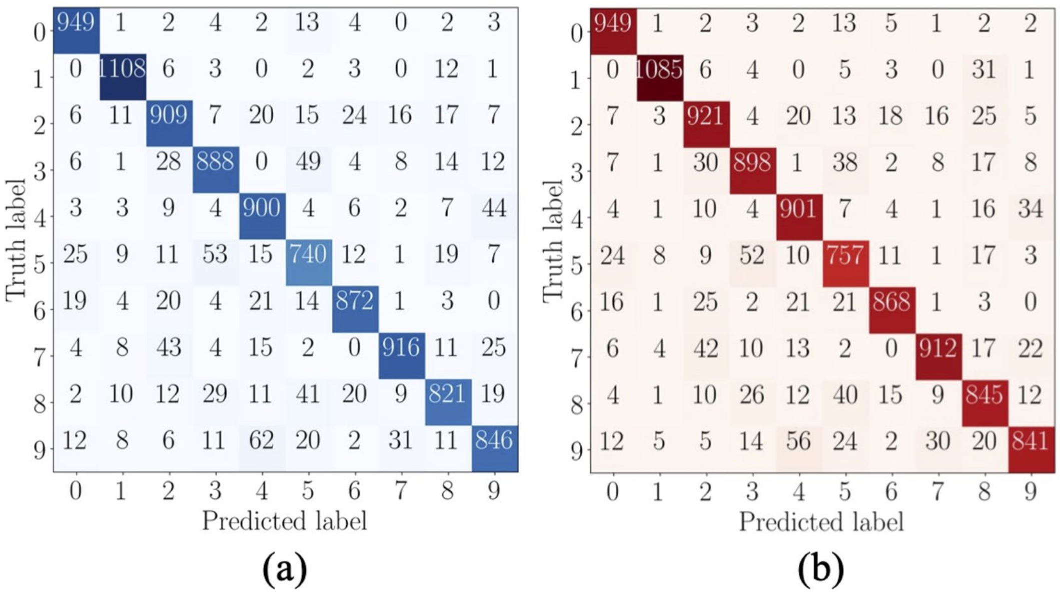

Figure 9. Confusion matrices obtained by classifying MNIST in, (**a**) Python, and (**b**) SPICE.

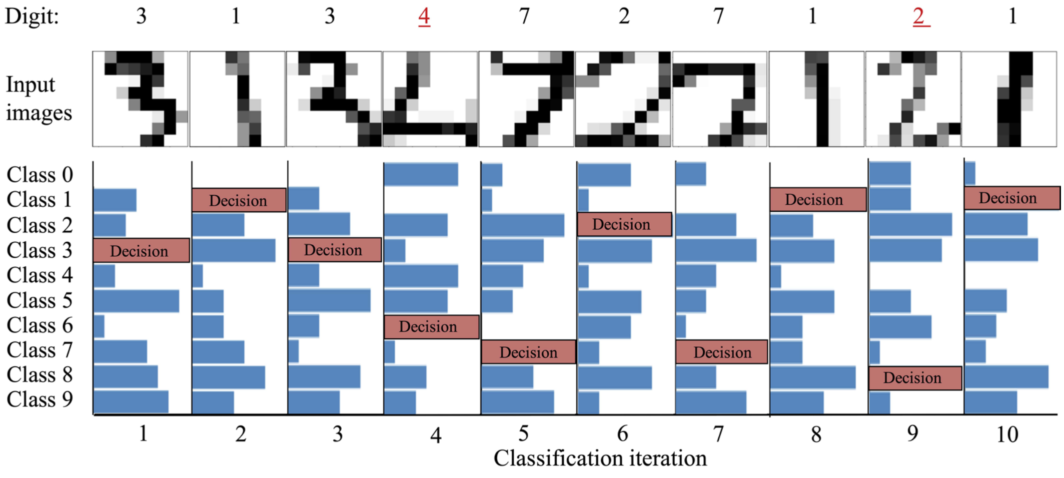

Figure 10. Ten consecutive classifications of digits in MNIST dataset, as extracted based on SPICE simulations. For each classification, the height of the bars corresponds to the number of votes collected to each class. Note that total number of votes is 45 and equals to the total number of binary classifiers. Within a single classification period, the i^{th} class, can get up to nine votes.

that to support reconfigurable feature weights, memory units (for storing the weights) along with multiplexer units (for selecting the desired weights) can be utilized. While power overhead is negligible with this approach, the overall area is expected to be increased by a factor of four.

System demonstration and results

The classifier is designed in SPICE and evaluated based on a commonly used MNIST dataset. Details of the simulation environment are presented in the Supplementary Information. In the following subsections, the dataset and preprocessing steps are described and the overall system and SPICE simulation results are presented.

Dataset and preprocessing steps. MNIST is a 70,000-image dataset of handwritten decimal digits. Each digit image is represented at 28 × 28 resolution, yielding a total of 784 pixels. Out of the 70,000 digits, 60,000 images are utilized for training the proposed ML circuits. The trained system is evaluated based on the remaining, unseen 10,000 images.

One versus one classification scheme[23], is used to discriminate the 10-class MNIST dataset. A K-class, one versus one classifier is designed with $K(K-1)/2$ binary classifiers for pairwise discrimination of the digits. Each binary classifier votes for a single class and the class with highest number of votes is selected as the final

		Krestinskaya *et al.*[13]	Wang *et al.*[3]	Current work
Device		Memristor	MOSFET	AP-CNFET
Technology		180 nm	130 nm	15 nm
Dataset		MNIST	MNIST	MNIST
Algorithm		ANN	Ada-boost	LR
Accuracy		92%	90%	90%
Offset from the ideal accuracy		−1%	0%	0%
Number of features		784	48	64
Supply voltage		1 V	1.2 V	3 V
Speed		3.33 KHz	1.3 MHz	250 MHz
Costs	Energy**	1.25 μJ*	481 pJ	295 pJ
	Energy per MAC	37.5 pJ	0.05 pJ	0.29 pJ
	Area**	1,133 μm^{2*}	246,792μm^2	3.8 μm^2
	Area per MAC	0.034 μm^2	26 μm^2	0.004 μm^2
	Energy × Area per MAC	1.27 pJμm^2	1.31 pJμm^{2*}	0.001 pJμm^2

Table 1. System characteristics of the proposed and a state-of-the-art CMOS and memristor based ML classifier. *Extrapolated based on the numbers reported for the memristive classifier[13] for a network with four input neurons and ten output neurons. **Overhead reported for MAC array.

classification decision. Utilizing the full set of features (*i.e.*, 784 features), accuracy of 94% can be achieved on MNIST test set with $10(10-1)/2=45$ binary logistic classifiers. Alternatively, a subset of the 784 features (*i.e.*, 23 features on average) is used in this paper, trading off the performance (less than 4.6% accuracy degradation) for power and area efficiency (×(784/23) = ×34 less transistors).

The subset of features is selected in a two-step approach: downsampling and feature selection. First, the features are uniformly downsampled from 28 × 28 pixels to 8 × 8 pixels, as shown in Fig. 6. The downsampling of features significantly reduces the required hardware resources (*e.g.*, 12 times less transistors is required) in exchange for 2.8% accuracy degradation. A greedy feature selection algorithm, sequential backward selection (SBS)[24], is used to select those most informative features (out of the remaining 64 features) for each binary classifier. As a result, the 8 × 8 features are reduced on average to 23 features per digit. Note that the number of selected features varies among the MNIST digits. For example, the digits 3 and 4 tend to look much more alike than the digits 0 and 1. Thus, significantly more features is selected for the 3-vs-4 binary classifier (44 features) than for the 0-vs-1 binary classifier (6 features). Comparing with other well-known feature selection algorithms (*e.g.*, Fisher information[25]), SBS is determined to select the most informative features. Alternatively, SBS is an iterative greedy algorithm and is computationally expensive. For example, completing SBS on the original feature set of 28 × 28 pixels requires 306,936 training iterations and 72 days (as extrapolated on shorter runs) on Intel Core i7-7700 CPU. Alternatively, with downsampled feature space only 2,016 training iterations which are completed within three hours on Intel Core i7-7700 CPU. The preferred set of features for each binary classifier is shown in Fig. 7, exhibiting an average of 23 features per classifier.

System demonstration and simulation results.

The schematic representation of the overall system is shown in Fig. 8, comprising of AP-CNFET array to perform feature-weight products, voltage dividers to provide quantized features and feature weights, and buffers to extract the individual votes of each binary classifier. The proposed system comprises 45 binary classifiers with a total of 1,021 AP-CNFETs utilized for the feature-weight products (shown by red and blue circles in Fig. 8). The product results are accumulated within the 45 sensing lines (one for each binary classifier). The system occupies 3.8 μm^2 as estimated based on transistor count and consumes 295 pJ energy per digit classification. Accuracy of 90% is observed in SPICE on test set of 10,000 unseen digits, as compared with the theoretical classification accuracy of 90% obtained on the low resolution data set in Python. The confusion matrices obtained with Python and SPICE for the 10,000-digit test set are shown in Fig. 9. Note the similarities of the decisions extracted by SPICE and Python. The classifier is designed to operate at 250 MHz, producing a single digit classification per cycle. The extracted votes and resultant decisions are shown in Fig. 10 for ten consecutive classifications, as extracted from SPICE. Each binary classifier votes for a single class. Thus, the total number of votes equals to the number of binary classifiers. Final decision is made based on the class with highest number of votes within each classification period, as shown in the Fig. 10. Note the similarity of the incorrectly classified images ('4' and '2') to the predicted labels ('6'and '8', respectively).

Performance characteristics are listed in Table 1 for the proposed system along with the existing state-of-the-art conventional CMOS and emerging device memristor based classifiers[3,13]. Note that while the proposed ML classifier, CMOS based classifier[3] and memristive classifier[13] are all demonstrated on the MNIST dataset, the memristive system[13] implements an artificial neural network (ANN) as compared with the linear model implemented in the proposed and CMOS based[3] systems. For better performance comparison among the systems in Table 1, the energy and area per MAC are also included in the table. For example, to determine the ANN area per MAC, the overall area of the system is divided by the total number of MAC operations, yielding $1{,}133/(784\times 42+42\times 10)=0.034\ \mu m^2$/MAC. To determine the ANN energy per MAC, the overall energy consumption of the system is divided by the total number of MAC operations. Based on these results, AP-CNFET classifier exhibits orders of magnitude improvement with respect to the combined Energy-Area per MAC figure of merit as compared with both state-of-the-art classifiers.

Conclusions

Increasing the device density and power efficiency has become challenging as the conventional CMOS scaling approaches its physical limits. Alternatively, emerging devices, such as AP-CNFETs are inherently intelligent. The seamless mapping of the AI logic primitives onto AP-CNFET increases by orders of magnitude the embedded AI per transistor.

To the best of the authors knowledge, the proposed system is the first to demonstrate ML classification with a single sensing line. To evaluate the system, a multi-class logistic classifier is designed in SPICE and demonstrated on MNIST dataset. The classifier uses 1,021 AP-CNFETs (×17 reduction in the transistor count as compared with state-of-the-art CMOS based classifiers[2,3]) and generates predictions at 250 MHz. The system exhibits 295 pJ energy consumption and occupies 3.8 μm^2 as estimated based on the transistor count in SPICE. With the proposed configuration, the reduced MNIST dataset is classified with no reduction in the overall prediction accuracy as compared with the theoretical Python results.

Theoretical bounds on classification accuracy are a strong function of the classification algorithm. In this paper, linear classification is demonstrated as a proof of concept of AP-CNFET based AI. The theoretical accuracy with linear classifiers is however limited to 94% using the default dataset of all the 784 MNIST features[26]. Higher accuracy (>98%) can be achieved with more complex algorithms, such as non-linear support vector machines (SVM) and deep neural networks (DNNs). Similar to linear classifiers, the operation of these complex network is dominated by feature-weight multiplication and product accumulation. The proposed framework is expected to significantly increase the AI density, while reducing the power and area overheads in complex ML networks.

Received: 13 December 2019; Accepted: 13 March 2020;
Published online: 31 March 2020

References

1. Lee, J. *et al*. UNPU: An energy-efficient deep neural network accelerator with fully variable weight bit precision. *IEEE J. Solid-State Circuits* **54**(1) 173–185 (Oct 2018).
2. Zhang, J., Wang, Z. & Verma, N. In-memory computation of a machine-learning classifier in a standard 6T SRAM array. *IEEE J. Solid-State Circuits* **52**(4) 915–924 (Apr 2017).
3. Wang, Z. & Verma, N. A low-energy machine-learning classifier based on clocked comparators for direct inference on analog sensors. *IEEE Trans. Circuits Syst. I, Reg. Papers* **64**(11) 2954–2965 (Jun 2017).
4. Bankman, D. *et al*. An always-on 3.8 μJ/86% CIFAR-10 mixed-signal binary CNN processor with all memory on chip in 28nm CMOS. in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 222–224 (Feb 2018).
5. Kenarangi, F. & Partin-Vaisband, I. A single-MOSFET MAC for confidence and resolution (CORE) driven machine learning classification. *arXiv preprint arXiv:1910.09597*, (Oct 2019).
6. Kang, K. & Shibata, T. An on-chip-trainable gaussian-kernel analog support vector machine. *IEEE Trans. Circuits Syst. I, Reg. Papers* **57**(7) 1513–1524 (Jul 2009).
7. Gonugondla, S. K., Kang, M. & Shanbhag, N. R. A variation-tolerant in-memory machine learning classifier via on-chip training. *IEEE J. Solid-State Circuits* **53**(11) 3163–3173 (Sep 2018).
8. Kang, M. *et al*. A multi-functional in-memory inference processor using a standard 6T SRAM array. *IEEE J. Solid-State Circuits* **53**(2) 642–655 (Jan 2018).
9. Amaravati, A. *et al*. A 55-nm, 1.0–0.4 v, 1.25-pJ/mac time-domain mixed-signal neuromorphic accelerator with stochastic synapses for reinforcement learning in autonomous mobile robots. *IEEE J. Solid-State Circuits* **54** (1) 75–87 (Dec 2018).
10. Hu, M. *et al*. Memristor-based analog computation and neural network classification with a dot product engine. *Advanced Materials* **30**(9) 1705914 (Jan 2018).
11. Yu, S. *et al*. Scaling-up resistive synaptic arrays for neuro-inspired architecture: challenges and prospect. in *Int. Electron Devices Meeting* 17–3 (Dec 2015).
12. Agarwal, S. *et al*. Resistive memory device requirements for a neural algorithm accelerator. in *Int. Joint Conf. on Neural Networks* 929–938 (Jul 2016).
13. Krestinskaya, O., Salama, K. N. & James, A. P. Learning in memristive neural network architectures using analog backpropagation circuits. *IEEE Trans. Circuits Syst. I, Reg. Papers* **66**(2) 719–732 (Sep 2018).
14. Xiang, D. *et al*. Surface transfer doping induced effective modulation on ambipolar characteristics of few-layer black phosphorus. *Nature communications* **6** 6485 (Mar 2015).
15. Lin, Y.-M. *et al*. High-performance carbon nanotube field-effect transistor with tunable polarities. *IEEE Trans. on Nanotech*. **4**(5) 481–489 (Sep 2005).
16. Ben-Jamaa, M. H., Mohanram, K. & De Micheli, G. An efficient gate library for ambipolar CNTFET logic. *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst*. **30**(2) 242–255 (Feb 2011).
17. Das, S. & Appenzeller, J. WSe2 field effect transistors with enhanced ambipolar characteristics. *Applied physics letters* **103**(10) 103501 (Sep 2013).
18. O'Connor, I. *et al*. CNTFET modeling and reconfigurable logic-circuit design. *IEEE Trans. Circuits Syst. I, Reg. Papers* **54**(11) 2365–2379 (Nov 2007).
19. Kenarangi, F. & Partin-Vaisband, I. Leveraging independent double-gate FinFET devices for machine learning classification. *IEEE Trans. Circuits Syst. I, Reg. Papers* **66**(11) 4356–4367 (Jul 2019).
20. Hu, X. & Friedman, J. S. Transient model with interchangeability for dual-gate ambipolar CNTFET logic design. in *Int. Symp. on Nanoscale Arch. (NANOARCH)* 61–66 (Oct 2017).
21. Closed-form model for dual-gate ambipolar CNTFET circuit design. in *Int. Symp. on Circuits and Syst. (ISCAS)* 1–4 (Sep 2017).
22. Nelder, J. A. & Baker, R. J. Generalized linear models. *Encyclopedia of statistical sciences* **4** (Jul 2004).
23. Aly, M. Survey on multiclass classification methods. *Neural Netw* **19** 1–9 (2005).
24. Raschka, S. & Mirjalili, V. *Python Machine Learning*. Birmingham, U.K.: Packt (2017).
25. Guyon, I. & Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res*. **3** 1157–1182 (Jan 2003).
26. Kaggle Inc. Public Leaderboard: MNIST. [Online]. Available, https://www.kaggle.com/numbersareuseful/public-leaderboard-mnist (2018).

Acknowledgements

J.A.C.I.'s contribution is sponsored in part by the National Science Foundation under CCF award 1910997.

Author contributions

F.K. designed the ML system, performed simulations, and interpreted the data. X.H. characterized the electrical operation of the device in Verilog. J.A.C.I. and Y.L. provided experimental guidance. J.S.F. conceived the idea of exploiting the device in ML ICs. All authors collaborated at the system-circuit and circuit-device design abstraction interfaces. I.P.-V., J.S.F. and J.A.C.I. supervised the research. F.K. and I.P.-V. wrote the manuscript, to which all authors contributed.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-62718-0.

Correspondence and requests for materials should be addressed to F.K.

Reprints and permissions information is available at www.nature.com/reprints.