Cognitive Sensing for Energy-Efficient Edge Intelligence

Minah Lee, Sudarshan Sharma, Wei Chun Wang, Hemant Kumawat, Nael Mizanur Rahman, and Saibal Mukhopadhyay
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA
{minah.lee, ssharma497, johnson.wang, hkumawat6, naelmr, smukhopadhyay6}@gatech.edu

***Abstract*—Edge platforms in autonomous systems integrate multiple sensors to interpret their environment. The high-resolution and high-bandwidth pixel arrays of these sensors improve sensing quality but also generate a vast, and arguably unnecessary, volume of real-time data. This challenge, often referred to as the analog data deluge, hinders the deployment of high-quality sensors in resource-constrained environments. This paper discusses the concept of cognitive sensing, which learns to extract low-dimensional features directly from high-dimensional analog signals, thereby reducing both digitization power and generated data volume. First, we discuss design methods for analog-to-feature extraction (AFE) using mixed-signal compute-in-memory. We then present examples of cognitive sensing, incorporating signal processing or machine learning, for various sensing modalities including vision, Radar, and Infrared. Subsequently, we discuss the reliability challenges in cognitive sensing, taking into account hardware and algorithmic properties of AFE. The paper concludes with discussions on future research directions in this emerging field of cognitive sensors.**

***Index Terms*—Autonomous System, Edge intelligence, Smart Sensor, Compute-in-Memory**

I. Introduction

Artificial Intelligence (AI) platforms, particularly deep neural networks (DNNs), are increasingly being employed in real-time, safety-critical, and autonomous systems at the edge, including self-driving cars, unmanned aerial vehicles, drones, and robots [1], [2]. These systems typically use high-resolution and high-bandwidth pixel arrays to enhance sensing quality, which leads to a substantial volume of real-time data processing and transmission. However, unlike AI in cloud environments, AI in real-time edge devices must operate within strict energy and time constraints. Therefore, a key challenge in implementing edge intelligence is to efficiently conserve resources.

To utilize raw sensor data at a processing node for a specific perception, traditional sensors require analog-to-digital (A/D) conversion of the raw sensor data, which, along with data transmission, consumes significant energy (Fig. 1(a)). Although techniques like feature encoding [3], [4] and network pruning [5] have reduced communication loads, they still involve energy-intensive full data digitization. In contrast, analog compute-in-memory (ACIM) utilizes vector matrix multiplications (VMM) in crossbar arrays and enables in-sensor data processing [6], [7]. Analog feature extraction (AFE) using ACIM eliminates the need for extensive digitization and transmission, leading to considerable energy savings compared to conventional methods that digitize and transmit the entire dataset (Fig. 1(b)).

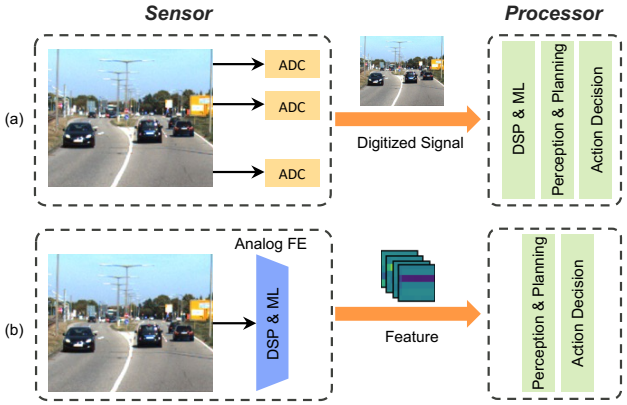

Fig. 1. (a) Conventional sensor platform and (b) AFE-based cognitive sensor platform.

This paper discusses the concept of cognitive sensing, which focuses on task-relevant features, leading to reductions in both digitization power and data volume. This paper first introduces an Analog-to-Feature Extraction (AFE) engine that leverages mixed-signal Computing-In-Memory (AFE-CIM) with 8T-SRAM, specifically designed for wide-band and multi-element sensor arrays, such as MIMO radars. This paper's design strategy involves an analog-vector digital-matrix multiplication (A/D-VMM) engine that processes lower-dimensional digital features from high-dimensional analog inputs, significantly reducing power consumption. The AFE-CIM test chip is fabricated in 28nm CMOS.

Subsequently, this paper introduces an AFE-based cognitive sensor framework that integrates AFE hardware with processing algorithms. We present the capability of cognitive senor for efficient feature extraction, achieving a reduced output data rate while ensuring accuracy on par with direct digitization. This efficiency is demonstrated on a variety of sensing modalities, including vision, Radar, and Infrared (IR), and encompasses various tasks such as beamforming and image classification. Furthermore, this paper examines how hardware factors like non-linearity error, voltage and temperature variations, as well as algorithmic factors such as signal compression rate, affect classification accuracy and energy consumption.

II. Analog-to-Feature Extraction

A. AFE using mixed-signal Compute-in-Memory

Analog-to-Feature Extraction (AFE) engines generate low-dimensional digital features from high dimensional analog

979-8-3503-4859-0/DATE24/© 2024 EDAA

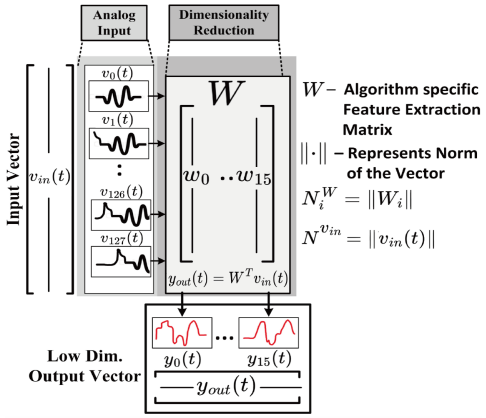

Fig. 2. Overview of Analog-to-Feature Extraction. Reprinted with permission from [8].

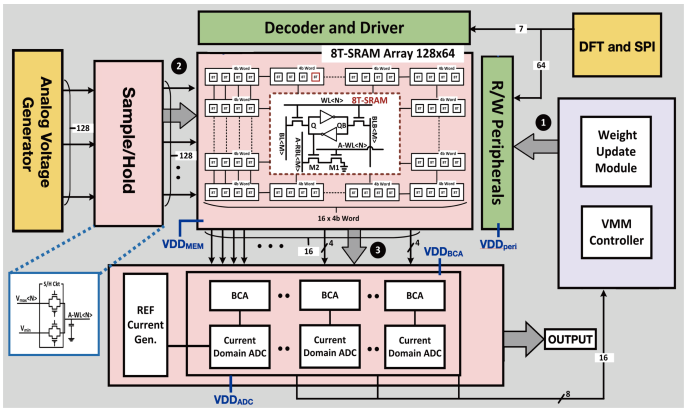

Fig. 3. Top-level microarchitecture and dataflow, consisting of SRAM bit cell structure, S/H circuit, binary-weighted bit-current accumulator (BCA), current domain ADC, and peripherals. Reprinted with permission from [8].

inputs [6], [7]. This is achieved using a single data converter, which implements a signal processing or Machine Learning (ML) algorithm, instead of multiple ADC units. The value of this feature extraction lies in its ability to maintain the quality of subsequent tasks.

AFE using mixed-signal Compute-in-Memory (AFE-CIM) [8] is a spatial AFE as illustrated in Fig. 2 which performs analog-vector digital matrix multiplication using CIM techniques. It computes a weighted linear combination of high dimensional input analog signals ($\boldsymbol{v}_{\text{in}}(\boldsymbol{t})$) to generate lower dimension digital features ($\boldsymbol{y}_{\text{out}}(\boldsymbol{t}) = \boldsymbol{W}^T \cdot \boldsymbol{v}_{\text{in}}(\boldsymbol{t})$). The design comprises of a bit-level (digital weights) and word-level (analog inputs) multiply-and-accumulate (MAC) operations, followed by analog-to-digital conversion in current domain.

As illustrated in Fig. 3, AFE-CIM consists of an SRAM array with 8-T bitcells storing the digital weight matrix with 4-bit weights. The 8-T bitcell's 2T path is used for analog word and digital weight bit multiplication via the Analog-Read BitLine (A-RBL) bit line. The read current from each bitcells in a column is accumulated in the A-RBL. The column currents of the 4 A-RBLs are then binary weighted and combined using the Bit-Current Accumulator (BCA). This combined current is converted into an 8-bit digital format by a current domain ADC [9].

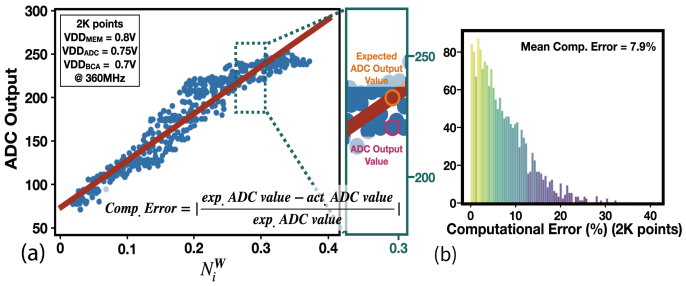

Fig. 4. Linearity measurement results (a) Linearity and (b) Computational error histogram of VMM with random $\boldsymbol{W}$ and varying $\boldsymbol{N_i^W}$ (red line shows the expected ADC outputs). Reprinted with permission from [8].

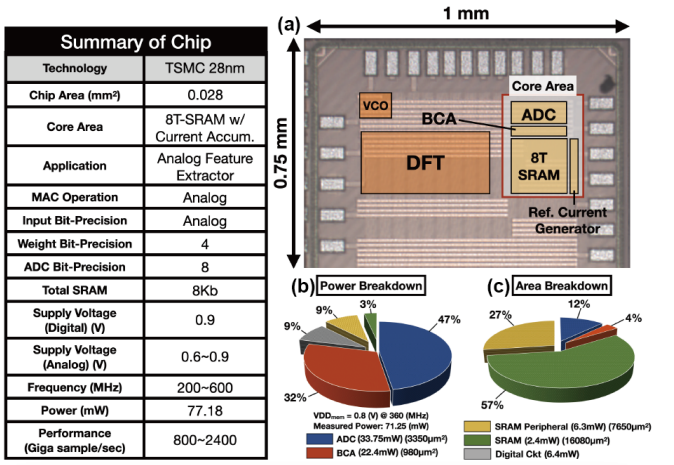

Fig. 5. AFE-CIM chip summary in 28nm CMOS: (a) Die shot and breakdown of (b) power and (c) area. Reprinted with permission from [8].

The overall computation error in AFE-CIM arises from the following three factors:

(a) The linearity error of the N-MOS in the 2T path of the 8T bitcell, which is used for CIM operation.
(b) The current mismatch and accumulation errors occurring in the BCA.
(c) The inherent error introduced by the current domain ADC also adds to the overall computation error.

Fig. 4(a) depicts the impact of these three errors in the measured linearity plot, showing deviations of the ADC output from the ideal value on the linearity plot of AFE-CIM. An error model linked to the ADC output is constructed to incorporate these errors shown in Fig. 4(b). This process includes identifying the ideal output for each level of ADC output, followed by calculating the mean and the standard deviation of the discrepancies between the measured and ideal ADC outputs. This information is used to create an error model, where error distribution in each ADC output level is characterized by Gaussian distribution.

As described in Fig. 5, the majority of the chip's core area is taken up by the SRAM macro, making up almost 85% of the total core area, but only 12% of the overall power consumption. On the other hand, the analog peripherals, namely the ADC and BCA, exhibit a more modest area footprint but contribute significantly to the chip's power consumption by approximately 80%.

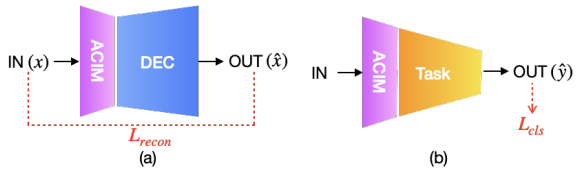

Fig. 6. Training framework of AFE-based cognitive sensor for (a) data dimension reduction and (b) finetuning with task. Reprinted with permission from [10].

TABLE I
DESIGN OF AFE-BASED FEATURE EXTRACTOR FOR DIFFERENT C_{out}. REPRINTED WITH PERMISSION FROM [10].

C_{out}	8	16	32
# Rows	147	147	147
# Columns	64	128	256
# AFE Tiles	3	6	12

III. AFE-BASED COGNITIVE SENSING

In this section, we introduce a cognitive sensor leveraging our AFE-CIM design for multiple sensing modalities. Utilizing an ADC out-referred AFE error model (Fig. 4), we simulate AFE error by adding it to the digitized outputs from AFE. For computational efficiency, the simulation considers only the AFE error distribution corresponding to the average feature value. Error sampled from this distribution is added to the digitized feature, which is subsequently used for task processing. This framework allows to examine the effect of AFE linearity errors on cognitive sensor accuracy and the potential for retraining ML algorithms.

A. Case Study 1: AFE for Vision Image based perception

We present a method that combines DNNs with AFE to extract high-level information from high-dimensional data for complex vision-based perception [10]. The training of cognitive sensor involves a two-step process. Firstly, it is trained in an auto-encoder framework, where the reconstruction loss $\left(\mathcal{L}_{\text{recon}} = -\sum ||\boldsymbol{x} - \hat{\boldsymbol{x}}||^2\right)$ is minimized, as shown in Fig. 6(a). This step aims to reconstruct the input data accurately. Subsequently, the feature extractor is fine-tuned with a task-specific network, optimizing the classification loss $(\mathcal{L}_{\text{cls}} = -\sum y_c \ln(\hat{y_c}))$ as shown in Fig. 6(b). This ensures that the feature extractor learns task-relevant representations.

We compare our AFE-based sensor with traditional sensor as described in Fig. 1. Traditional sensor (baseline) consists of an ML backend running a floating point classifier operating on the digitized input from the pixel array. On the other hand, cognitive sensor is trained using the reduced features from the AFE-CIM from Section II. The first layer of the classifier is implemented using the AFE-CIM, and it is trained using the 8-bit weight precision. Both sensors are trained and evaluated based on a high-resolution version of the CIFAR10 dataset [11], generated using [12].

1) Impact of Signal Compression Rate: AFE-CIM employs a single convolutional neural network (CNN) to encode input data $(W_{in} \times H_{in} \times C_{in})$ into smaller dimensions $(W_{out} \times H_{out} \times C_{out})$, while preserving task-relevant information. The signal compression rate, r, defined as the ratio

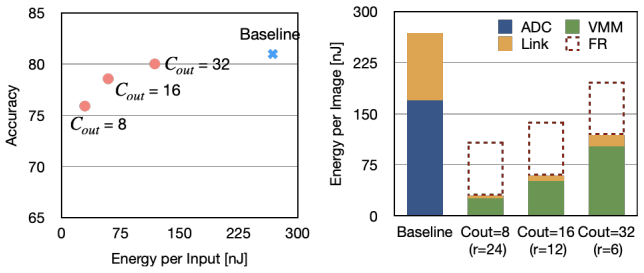

Fig. 7. (left) Trade-off between task accuracy and sensor energy under various number of filters in AFE and (right) sensor energy breakdown. FR is included in the energy analysis only if it is implemented in the sensor node. The ADC design for traditional sensor is based on [13]. Reprinted with permission from [10].

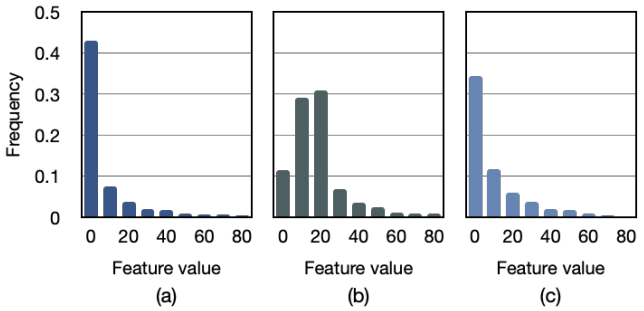

Fig. 8. Distribution of output feature map from AFE (a) without VMM error, (b) with VMM error, (c) and VMM error with FR. Reprinted with permission from [10].

of input to output product dimensions. Since the output dimensions W_{out} and H_{out} are task-specific, C_{out} is a key design factor affecting compression. To meet the computational needs of this increased number of filters, multiple AFE array tiles of 64×64 are used (Table I).

Signal compression rate of AFE plays a significant role in the AFE-based sensing. As shown in Fig. 7(a), the AFE-based sensor with a compression setting of $C_{out} = 32$ (6x compression) achieves an energy saving of 55.9% while experiencing a minimal accuracy drop of less than 1% compared to the baseline. This notable reduction in energy consumption mainly results from the lesser volume of data needing digitization and transmission. Additionally, cognitive sensor, utilizing 8-bit precision for feature extraction directly from raw data, effectively preserves important features, thus maintaining accuracy levels comparable to the baseline.

Fig. 7(b) shows that by further reducing the data dimensions via a decrease in C_{out}, we can achieve even more significant energy savings, at the expense of task reliability. This energy efficiency is reflected in the reduced requirements for processing and digitizing (VMM - green) as well as for transmitting (Link - yellow) the resultant features. As a result, energy savings are in the range of 77.9%-89% compared to the baseline. However, a lower C_{out} also indicates more information is lost, leading to a decrease in task accuracy ranging from 2.4% to 5%. Therefore, selecting C_{out} provides a flexible balance between maintaining task reliability and achieving energy efficiency.

2) Impact of Non-ideal Characteristics: AFE offers an energy-efficient feature extraction method, but its non-ideal

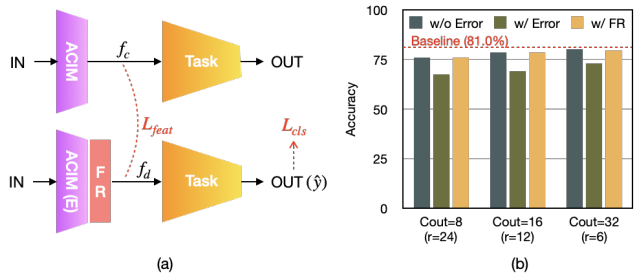

Fig. 9. (a) Training framework of FR and (b) classification accuracy changes under AFE VMM error. ACIM(E) implies VMM error model added from [8]. Reprinted with permission from [10].

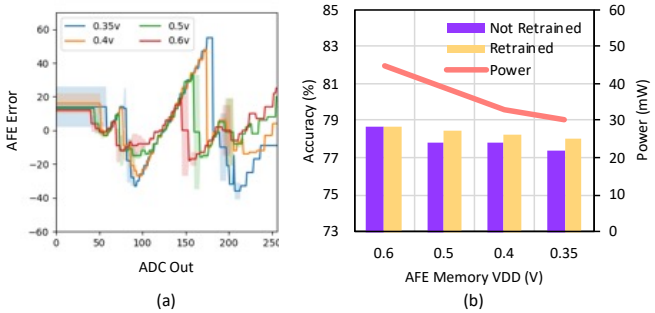

Fig. 10. (a) AFE error models and (b) Accuracy and power of cognitive sensor under different AFE voltage. (Purple: sensor trained at 0.6V only, Yellow: sensor trained at each voltage.)

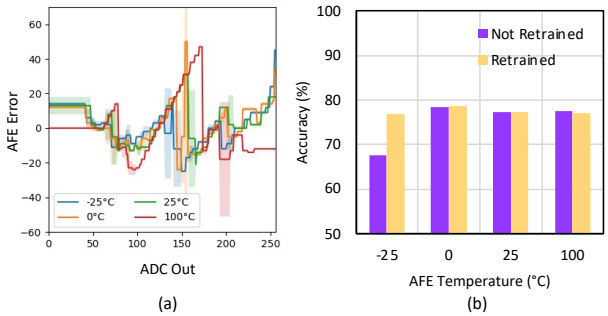

Fig. 11. (a) AFE error models and (b) Accuracy of cognitive sensor at different AFE temperature. (Purple: sensor trained at 25°C only, Yellow: sensor trained at each temperature.)

characteristics can introduce errors in VMM computations (Section II-A). As Fig. 8 shows, AFE's output feature map, when using 8-bit precision, is usually centered around zero. VMM errors, however, tend to shift this distribution (Fig. 8(b)), leading to low task accuracy. To tackle this, we implement feature restoration (FR) using an auto-encoder network. FR is trained with feature reconstruction and cross-entropy loss functions to correct distorted AFE features and improve task performance (Fig. 9(a)). FR restores the feature distribution to its centered state (Fig. 8(c)).

Fig. 9(b) illustrates that AFE VMM errors reduce classification accuracy by 7.0%-8.6%, especially worse at smaller C_{out} values. FR improves accuracy back to 76.0%-79.52%. FR can be implemented in the sensor or processing node, considering additional energy for FR processing. Including FR, AFE-based feature extraction saves 26.4%-59.5% in total sensor energy (Fig. 7(b)).

3) Impact of AFE Memory Voltage: We investigate how reducing the supply voltage for SRAM (VDD_{memory}) and the reference current control for the current domain ADC (VDD_{ref}) affects the accuracy of our cognitive sensor. Reducing VDD_{memory} and VDD_{ref} can save energy, but also changes the error behavior as shown in Fig. 10(a). Specifically, we observe a shift in the maximum error at mid ADC values under low power, due to the ADC's reduced sensitivity in these regions. Low VDD_{memory} and VDD_{ref} lead to a modest accuracy decrease of 0.8-1.3% in the cognitive sensor (Fig. 10(b)), despite the sensor does not previously observed the error behavior at low voltage and the change in error behavior is not trivial. This is primarily due to a change in the AFE error behavior at different voltages, which is effectively mitigated by batch normalization in Resnet-based image classifier. By retraining the classification algorithm with the error model at low voltage, we further reduce the accuracy drop to less than 0.65%. Therefore, voltage variations in cognitive sensors can be effectively managed by retraining the classification algorithm.

4) Impact of AFE Temperature: Understanding temperature variations is crucial in ACIM-based AFE systems, as ACIM are significantly impacted by temperature [14]. We develop AFE error models at various temperatures (Fig. 11(a)) to assess their impact on the accuracy of cognitive sensors. Fig. 11(b) shows that sensors trained solely at 25°C maintain similar accuracy across different temperatures, except at -25°C, where accuracy drops by 9.6%. This drop is due to the notably different error variance behavior at -25°C (shown in blue) compared to 25°C (in red). However, when sensors are trained with an error model specific to each temperature, they exhibit consistent accuracy, with deviations of less than 2%. Thus, temperature variations in cognitive sensors can be effectively managed by retraining the classification algorithm.

B. Case Study 2: AFE for RF Beamforming

As elucidated in [15], Beamforming via Linear Embedding (Beam-LE) simplifies an array of input signals into a compact, lower-dimensional representation, formalized as $W(t) = \Phi^T X(t)$. In this expression, $X(t) = [X_0(t), \ldots, X_M(t)]$ captures a momentary array of M complex (I/Q) baseband analog signals at a specific time instant t. The transformation is facilitated by the complex matrix $\Phi_{M\times N}$, which adeptly minimizes information loss while deriving $W(t) = [W_0(t), \ldots, W_N(t)]$, a succinct set of N-dimensional complex features where $N \ll M$. Such condensation from M input signals, denoted by $\{X_m(t)\}_{m=1}^{M}$, to N output signals, $\{W_n(t)\}_{n=1}^{N}$, permits the execution of beamforming operations on the more compact set $W(t)$ rather than the expansive $X(t)$.

Mixed signal architecture of AFE-CIM with digital FIR engine (BeamCIM), enables the design of energy-efficient and scalable beamforming accelerators with Linear Embedding. It overcomes TTD-DBF scalability issues by reducing ADC and FIR filter requirements. By employing ACIM-VMM for linear embedding ($\Phi^T X(t)$) of analog inputs ($X(t)$) to derive digital features ($W(t)$) as in Fig. 12, BeamCIM DBF cuts the ADC count from two per input (M) to two per intermediate

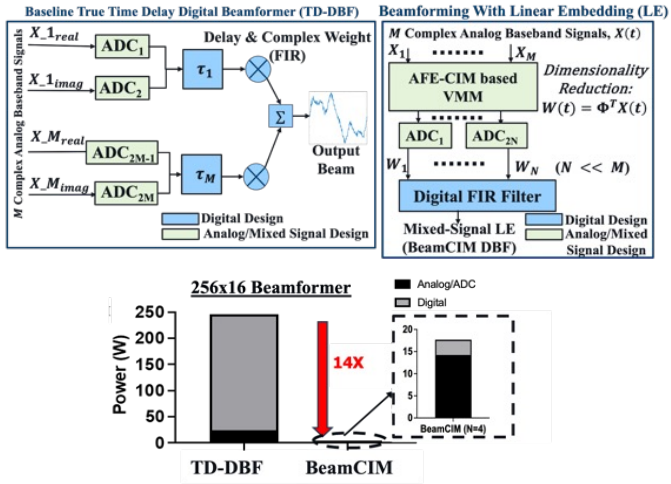

Fig. 12. (top right) Mixed Signal Beamforming with linear embedding reduces number of required ADC's and FIR filters compared to (top left) baseline True Time Delay Digital Beamformers. (bottom) Power of 256 × 16 BeamCIM is 14× lower than TD-DBF, in 28nm CMOS at 1GHz bandwidth. Reprinted with permission from [16].

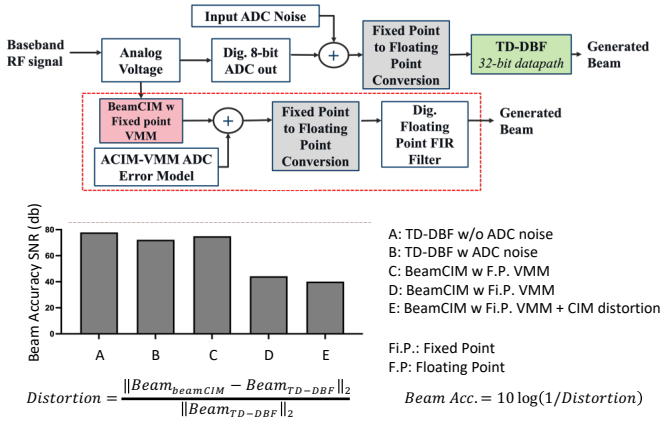

Fig. 13. (top) Simulation model for BeamCIM and True Time Delay Digital Beamformer (TD-DBF). (bottom) Distortion of 256 × 16 BeamCIM and TD-DBF for various non-idealities.

dimension (N), with $N \ll M$, and minimizes FIR filter computations within the embedded domain. The digital features are then transformed into beams via a digital computing engine. This approach also diminishes the computational load for FIR filters compared to baseline TTD-DBF, given the operations are now confined to the reduced-dimensional embedded domain.

Fig. 13(a), shows the simulation framework contrasting BeamCIM with TD-DBF. An analog voltage input, derived from a test-case software emulation and scaled to 0-1V, undergoes linear sampling through an 8-bit ADC, yielding digitized baseband signals. Subsequently, we introduce ADC distortion noise (SNR of 50db) to these signals and convert them from a Fixed Point (Fi.P) to Floating Point (F.P) format, before processing it with a 32-bit TD-DBF digital engine. BeamCIM utilizes the analog input (converts it into an 8-bit floating point representation of the input signal) and executes a Vector Matrix Multiplication (VMM) operation with an 8-bit quantized fixed point representation of the matrix Φ, thereby producing a fixed point VMM outcome. This result, after the addition of

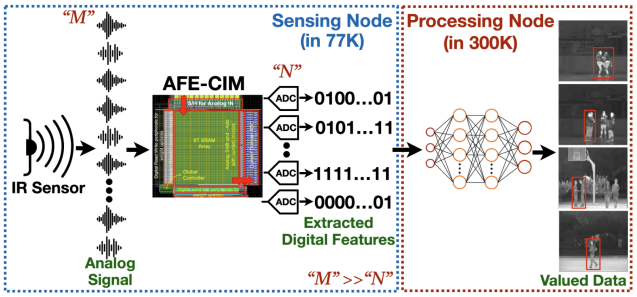

Fig. 14. AFE-CIM-based infrared sensing saves energy by operating at cryogenic temperature. Reprinted with permission from [17].

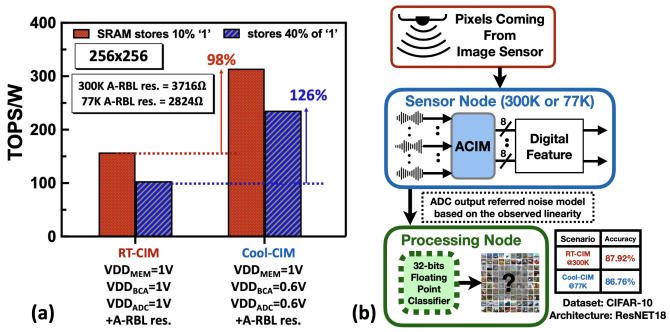

Fig. 15. (a) TOPS/W for AFE-CIM under room and cryogenic temperature. Cool-CIM shows 126% higher TOPS/W over RT-CIM. (b) Simulation of AFE-CIM based feature extraction followed by image classification. Reprinted with permission from [17].

circuit distortion and conversion to a floating point format, is send into a floating point digital FIR filter engine for beam generation. Fig. 13(b) presents the resulting beam accuracy, benchmarked as SNR against an ideal 32-bit datapath TD-DBF devoid of ADC distortion or quantization error. The TD-DBF exhibits beam accuracy degradation (SNR reduction) post ADC distortion addition. BeamCIM, assuming an ideal floating point VMM implementation (as assumed in [16]) with an 8-bit Φ, incurs a modest SNR reduction compared to a TD-DBF without ADC noise. However, approximating the VMM operation using fixed point arithmetic within the SRAM array substantially diminishes SNR, with further reduction upon integrating errors into the VMM output modelled through the AFE-CIM linearity error model.

C. Case Study 3: AFE for Cryogenic Infrared Sensor

The cryogenic operation of infrared (IR) sensors yields enhanced detector performance and imaging quality by mitigating thermal excitation and minimizing noise interference at lower temperatures [18]. Despite the impediment posed by the significant data volume generated by high-resolution sensors, the efficient real-time processing of IR sensor data is hindered. The in-sensor analog processing of high-dimensional sensor signals for extracting low-dimensional features presents a potential solution to mitigate the data load produced by a sensor. While numerous previous studies have delved into cryogenic logic [19], SRAM [20], and digital compute-in-memory [21], the feasibility of in-sensor processing for cryo-cooled IR sensors remains unexplored in the current literature (Fig. 14).

We present a Design-Technology-Co-Optimization (DTCO) methodology to design the cryogenic operation of AFE-CIM platform for IR sensing [17]. Utilizing measurement-calibrated 14nm FinFET models, we conducted optimization on an AFE-CIM to capitalize on the exponentially reduced I_{off}, higher I_{on}, and enhanced transconductance of cryo-cooled transistors. We demonstrate that the cryo-optimized AFE-CIM (Cool-CIM) achieves an 126% enhancement in power efficiency (TOPS/W) when incorporating the A-RBL interconnect resistance compared to the room temperature operation of AFE-CIM (RT-CIM) at the same technology node (Fig. 15(a)). Furthermore, we present simulation results assessing the performance of Cool-CIM and RT-CIM in image classification tasks. The training incorporates 4-bit weights and considers noise introduced in the AFE-CIM-generated features. The simulation outcomes reveal that Cool-CIM maintains a high accuracy of approximately 87% for CIFAR10 tasks [11] (Fig. 15(b)), showcasing the feasibility of in-sensor processing of analog signals to diminish data dimensions in cryogenic IR sensors.

IV. Future Directions

Cognitive sensor optimizes energy efficiency by reducing power in the ADC and transmission. Challenges include slight ML accuracy reduction and SNR degradation in signal processing. While ML models can be made robust to linearity errors during retraining, signal processing algorithms face performance decline due to determinism and floating-point reliance. To improve, we plan to integrate adaptive signal processing for resilience to linearity errors and compatibility with fixed-point representations.

Moreover, we plan to employ our cognitive sensor in a closed-loop perception system for enhanced resource utilization in autonomous systems. Recent studies [22] propose a closed-loop perception system where a real-time controller modifies sensor data collection based on task feedback. This system enhances motion planning by using a feedback loop from motion planning to perception and an Attention module to adjust DNN activations in object detector based on identified Regions of Interest (RoI) in images. Integrating our cognitive sensor with this closed-loop perception could lead to more aggressive energy efficiency, while ensuring reliable motion planning and other upstream tasks.

V. Conclusions

This paper presents an AFE-based cognitive sensor that significantly reduces energy by extracting features and reducing data dimensions directly from analog sensor data. Our AFE-CIM, designed with 8T-SRAM for MIMO radar arrays and fabricated in 28nm CMOS, features an A/D-VMM engine that decreases power consumption. The sensor's effectiveness is demonstrated across multiple sensing modalities, such as vision, Radar, and IR, without compromising accuracy. Additionally, we explore the reliability of cognitive sensor, examining how factors like signal compression and device non-linearity influence both accuracy and energy efficiency.

Acknowledgment: The materials are based on work supported in parts by SRC JUMP2.0 (CogniSense Center, #2023-JU-3133). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of SRC.

References

[1] S. Teng *et al.*, "Motion planning for autonomous driving: The state of the art and future perspectives," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 6, pp. 3692–3711, 2023.

[2] N. Hossein Motlagh *et al.*, "Unmanned aerial vehicles for air pollution monitoring: A survey," *IEEE Internet of Things Journal*, vol. 10, no. 24, pp. 21 687–21 704, 2023.

[3] J. H. Ko *et al.*, "Edge-Host Partitioning of Deep Neural Networks with Feature Space Encoding for Resource-Constrained Internet-of-Things Platforms," in *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2018, pp. 1–6.

[4] J. Shao and J. Zhang, "BottleNet++: An End-to-End Approach for Feature Compression in Device-Edge Co-Inference Systems," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2020, pp. 1–6.

[5] W. Shi *et al.*, "Improving Device-Edge Cooperative Inference of Deep Learning via 2-Step Pruning," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2019, pp. 1–6.

[6] T. Kang *et al.*, "A 650-uw 4-channel 83-dba-sndr speech recognition front-end with adaptive beamforming and feature extraction," *IEEE Solid-State Circuits Letters*, vol. 4, pp. 158–161, 2021.

[7] S. Lee *et al.*, "An eight-element frequency-selective acoustic beamformer and bitstream feature extractor," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 6, pp. 1812–1823, 2022.

[8] S. Sharma *et al.*, "Afe-cim: A current-domain compute-in-memory macro for analog-to-feature extraction," in *ESSCIRC 2023- IEEE 49th European Solid State Circuits Conference (ESSCIRC)*, 2023, pp. 33–36.

[9] D. Nairn and C. Salama, "Current-mode algorithmic analog-to-digital converters," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 4, pp. 997–1004, 1990.

[10] M. Lee *et al.*, "Energy-efficient sensor platform using reliable analog-to-feature extraction," in *2023 IEEE SENSORS*, 2023, pp. 1–4.

[11] A. Krizhevsky *et al.*, "CIFAR-10 (Canadian Institute for Advanced Research)," *Technical report*, 2009.

[12] J. P. S. Schuler. (2021, Dec.) CAI NEURAL API. [Online]. Available: https://doi.org/10.5281/zenodo.5810077

[13] M. Zhan *et al.*, "17.5 A 10mW 10-ENOB 1GS/s Ring-Amp-Based Pipelined TI-SAR ADC with Split MDAC and Switched Reference Decoupling Capacitor," in *2023 IEEE International Solid- State Circuits Conference (ISSCC)*, 2023, pp. 272–274.

[14] P. Kumar *et al.*, "Process, bias, and temperature scalable cmos analog computing circuits for machine learning," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 70, no. 1, pp. 128–141, 2023.

[15] C. DeLude *et al.*, "Broadband beamforming via linear embedding," *arXiv preprint arXiv:2206.07143*, 2022.

[16] N. M. Rahman *et al.*, "Beamcim: A Compute-In-Memory based Broadband Beamforming Accelerator using Linear Embedding," in *IEEE Radio And Wireless Symposium (RWS) 2024*, 2024.

[17] W.-C. Wang *et al.*, "Cool-cim: Cryogenic operation of analog compute-in-memory for improved power-efficiency," in *2023 International Electron Devices Meeting (IEDM)*, 2023.

[18] M. Liu *et al.*, "Cryogenic technology for infrared detection in space," *Scientific Reports*, vol. 12, no. 1, p. 17884, 2022. [Online]. Available: https://www.nature.com/articles/s41598-022-06216-5

[19] H. Bohuslavskyi *et al.*, "Cryogenic characterization of 28-nm fd-soi ring oscillators with energy efficiency optimization," *IEEE Transactions on Electron Devices*, vol. 65, no. 9, pp. 3682–3688, 2018.

[20] V. P.-H. Hu *et al.*, "High-density and high-speed 4t finfet sram for cryogenic computing," in *2021 IEEE International Electron Devices Meeting (IEDM)*, 2021, pp. 8.6.1–8.6.4.

[21] P. Wang *et al.*, "Cryogenic performance for compute-in-memory based deep neural network accelerator," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2021, pp. 1–4.

[22] H. Kumawat and S. Mukhopadhyay, "Radar guided dynamic visual attention for resource-efficient rgb object detection," in *2022 International Joint Conference on Neural Networks (IJCNN)*, 2022, pp. 1–8.