


Low-power edge detection based on ferroelectric field-effect transistor

Received: 17 April 2024

Accepted: 3 December 2024

Published online: 10 January 2025

 Check for updates

Jiajia Chen^{1,2}, Jiacheng Xu³, Jiani Gu³ , Bowen Chen³, Hongrui Zhang¹, Haoji Qian^{1,2}, Huan Liu¹, Rongzong Shen³, Gaobo Lin³, Xiao Yu^{1,2}, Miaomiao Zhang^{1,2}, Yi'an Ding¹, Yan Liu^{1,2} , Jianshi Tang⁴, Huaqiang Wu⁴, Chengji Jin^{1,3}  & Genquan Han^{1,2}

Edge detection is one of the most essential research hotspots in computer vision and has a wide variety of applications, such as image segmentation, target detection, and other high-level image processing technologies. However, efficient edge detection is difficult in a resource-constrained environment, especially edge-computing hardware. Here, we report a low-power edge detection hardware system based on HfO₂-based ferroelectric field-effect transistor, which is one of the most potential non-volatile memories for energy-efficient computing. Different from the conventional edge detectors requiring sophisticated hardware for the complex operation such as convolution and gradient, the proposed edge detector is analogue-to-digital converter free and loaded into a multi-bit content addressable memory, which only needs one 4 × 4 ferroelectric field-effect transistor NAND array. The experimental results show that the proposed hardware system is able to achieve efficient image edge detection at low power consumption (~10 fJ/per operation), realizing no-accuracy-loss, low-power and analogue-to-digital-converter-free hardware system, providing a feasible solution for edge computing.

Edge intelligence refers to the processing and decision-making of data at the device or sensor level, without the need for data transmission to cloud or other remote data centers, thereby enabling more efficient, real-time, secure, and privacy-preserving applications^{1–5}. Image edge detection, as a key technology in edge intelligence, finds extensive applications in fields such as intelligent security, intelligent transportation, and intelligent manufacturing that require real-time and accurate identification of edge information in images to make intelligent decisions^{6,7}. Although the development of technologies in recent years such as event-based cameras has significantly reduced hardware power consumption⁸. Low-power and high-speed edge detection is still a challenge due to the complex computational steps involved in conventional edge detectors, such as gradient and convolution

operations. In order to overcome this challenge, in-memory computing (IMC) with emerging non-volatile memories (eNVMs), such as resistive random-access memory (ReRAM)^{9–11}, ferroelectric field-effect transistor (FeFET)^{12–16}, etc. have been widely explored in edge detection. These eNVMs not only achieve on-chip storage with high density, but also show great potential in low-power and high-speed multiply-accumulate computations, greatly accelerating the computing process and significantly reducing computational power consumption in edge detection.

The main computing core of most eNVMs-based edge detection lies in convolution operations, such as Prewitt, Sobel and Laplacian detector as shown in Fig. 1a. By adjusting the conductance of eNVMs in a cross-bar array to represent the elements of the convolutional kernel

¹Hangzhou Institute of Technology, Xidian University, Hangzhou 311231, China. ²Faculty of Integrated Circuits, Xidian University, Xi'an 710071, China.

³Research Center for New Materials Computing, Zhejiang Lab, Hangzhou 311121, China. ⁴School of Integrated Circuits, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China. ✉ e-mail: jiani_gu@zhejianglab.edu.cn; cjjin@ieee.org

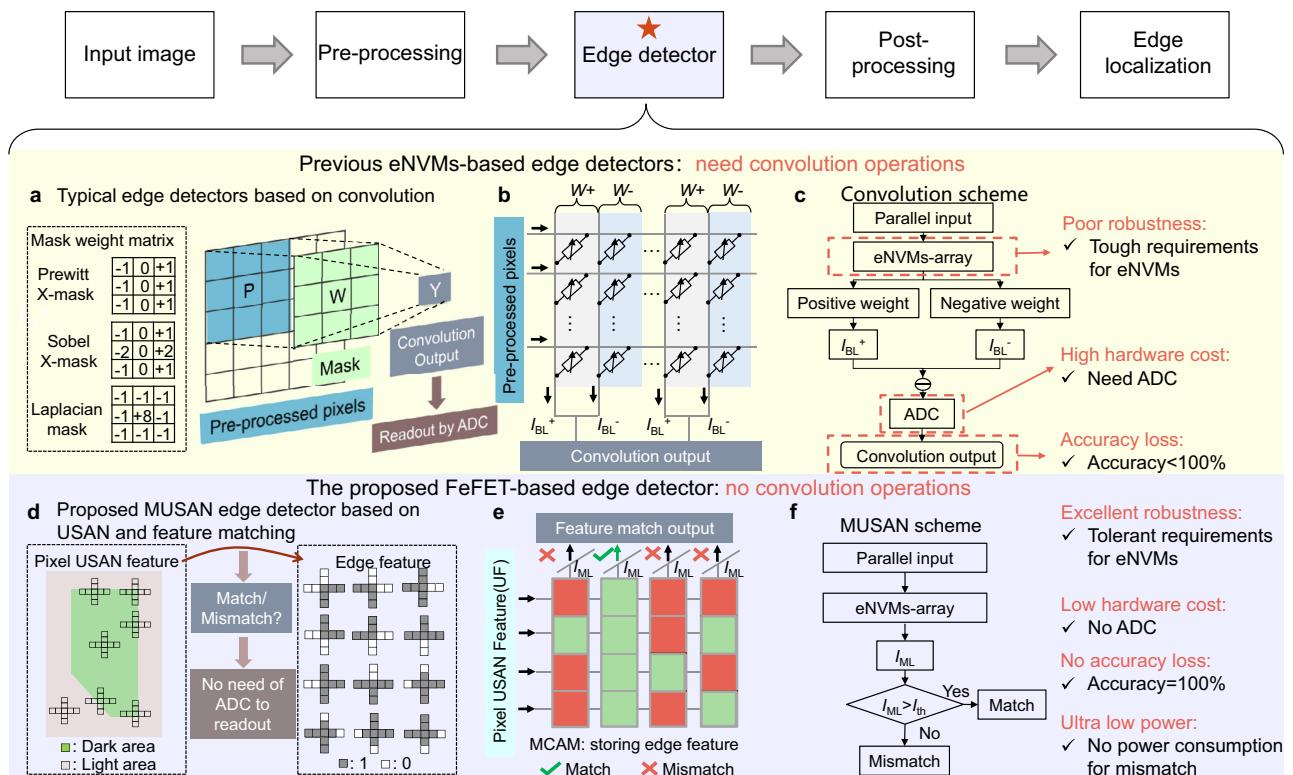


Fig. 1 | The comparison of the typical edge detectors and the proposed MUSAN edge detector. a Typical edge detectors: Prewitt, Sobel, Laplacian detectors, need convolution operations and ADC to readout. **b** The eNVMS-based hardware implementation of the typical edge detectors. **c** The corresponding scheme of

eNVMS-based convolution operation. **d** The proposed MUSAN edge detector, combining USAN and feature matching, does not need convolution operations and ADC to readout. **e** The MCAM-based hardware implementation of the proposed edge detector. **f** The corresponding scheme of MUSAN operation.

(weights), and summing up the currents of each column through Kirchhoff's current law with the voltage input related to the pixel value, the total current obtained is read out as the result of the convolution (Fig. 1b). Therefore, theoretically, the eNVMS-based edge detection hardware with convolution operation can achieve low-power and high-speed IMC. However, three kinds of basic challenges exist in hardware implementations of eNVMS-based convolution operations, making it difficult to widely promote eNVMS-based edge detection in practical applications. The first one is the poor robustness, which is due to the convolution operations with eNVMS-based array are of low tolerance for variation and read-out disturbance. The second one is high hardware cost for read-out. Typically, the results of analog convolution need be converted into digital read-out signals with additional analog-to-digital converter (ADC), as shown in Fig. 1c, which increases power consuming and introduces additional latency. The last one is the accuracy loss due to the error accumulation of eNVMS-based convolution operations. The situation with low efficiency becomes even worse when floating-point-based approaches are used. Therefore, it is urgent to explore alternative approaches to realize the eNVMS-based edge detection which does not lie in convolution operations.

In this regard, a matching univalue segment assimilating nucleus (MUSAN) method without convolution operations is proposed in this work. The proposed MUSAN combines univalue segment assimilating nucleus (USAN) operator and feature matching method, as shown in Fig. 1d. The hardware system corresponding to MUSAN method is very simple and can be loaded into a multi-bit content addressable memory (MCAM) array as shown in Fig. 1e. Figure 1f shows the proposed MUSAN scheme. Different from the typical eNVMS-based edge detection hardware system in which each convolution operation exhibits power consuming and the analog outputs need to be converted into digital read-out by ADC, the proposed one only exhibits obvious power consuming when feature matching occurs and the match/mismatch

results can be readout by distinguishing high/low current with general purpose input/output (GPIO) instead of ADC. Moreover, the hardware system based on the proposed MUSAN method exhibits high tolerance of performance degradation caused by variation of eNVMS, which results in excellent robustness. The architecture and details of the MUSAN method are presented in this paper, and its high performance and low-power consumption are verified by the hardware implementation with MCAM array based on FeFET. The measurement results show that the proposed FeFET-based MUSAN edge detection is able to achieve image edge detection with low-power consumption (~10 fJ/per operation), ADC-free, and no-accuracy-loss at low hardware cost, providing a feasible solution for edge computing.

Results

The proposed MUSAN method

The basic idea of MUSAN is matching USAN feature. Here, USAN feature of pixels is defined as 0/1 for the neighboring pixel with dissimilar/similar brightness. Figure 2a shows the proposed USAN feature extraction method. We use a cross mask consisting of 9 pixels including 4 vertical ones ($P_1 - P_4$), 4 horizontal ones ($P_5 - P_8$), and a central one (P_0) in finding USAN area. The vertical/horizontal USAN feature ($I_1 - I_4/I_5 - I_8$) can be extracted by comparing the gray scale values of $P_1 - P_4/P_5 - P_8$ with the one of P_0 . $I_i = 1$ represents the similar brightness with central pixel, while $I_i = 0$ is considered as dissimilar brightness. Figure 2b summarizes the vertical/horizontal USAN feature (UF) of 5 typical image edge locations ($L_2 - L_6$) and 2 non-edge locations (L_1, L_7) in Fig. 2a. Fig. S1 shows several MUSAN schemes with different matched edge feature (EF), and the corresponding simulation results of edge detection. Among them, 0011/1100 and 1110/0111 are the exact matching EF, while 00XX/XX00 is the fuzzy matching EF ($X = 0/1$). Compared to 0011/1100, 00XX/XX00 can describe more conditions, i.e. 0011/1100, 0000, 0001/1000, showing better robustness. Here, we

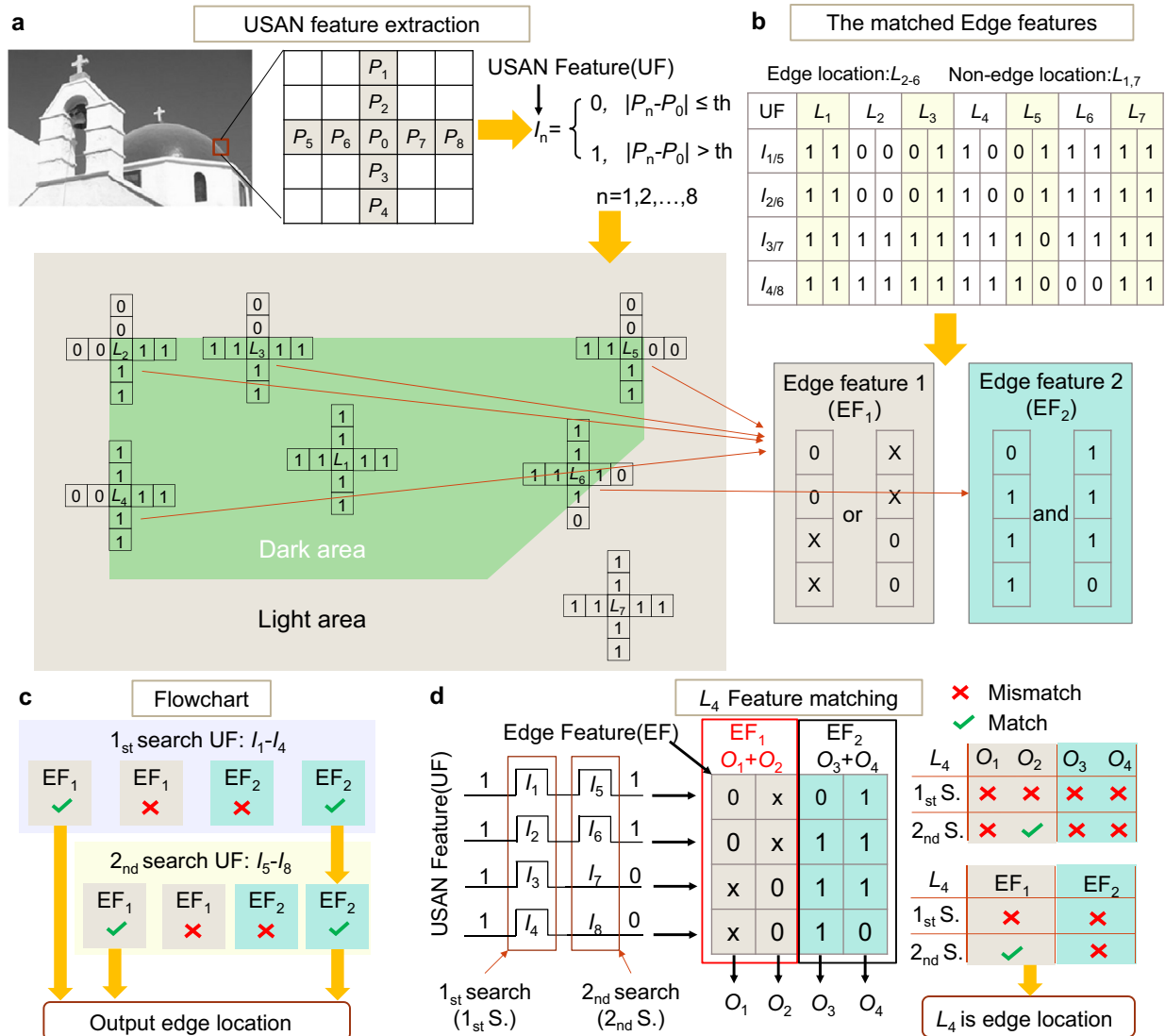


Fig. 2 | Scheme of the proposed MUSAN method. a USAN feature extraction. **b** The matched edge features. **c** The flowchart of the proposed edge detector. **d** Feature matching process using L_4 as an example.

combine the fuzzy matching EF with 00XX/XX00 as the primary EF (EF_1) and the exact matching EF with 0111/1110 as secondary EF (EF_2). The primary EF means that one of vertical and horizontal UF satisfies 00XX/XX00, while secondary EF means that both vertical and horizontal UF satisfy 1110/0111. For example, the vertical or horizontal UF of $L_2 - L_5$ satisfies 00XX or XX00, while both vertical and horizontal UF of L_6 satisfy 1110/0111. According to the analysis above, the flowchart of the proposed feature matching process can be described in Fig. 2c. The feature matching consists of two search-matching processes. The first (1st) search-matching one is for vertical UF ($I_1 - I_4$). If the input $I_1 - I_4$ matches with EF_1 , then output the pixel location as the edge location. Otherwise, the second (2nd) search-matching one for horizontal UF ($I_5 - I_8$) is performed. If the input $I_5 - I_8$ matches with EF_1 in the 2nd search-matching process, then output the pixel location as the edge location. If the input $I_1 - I_4$ matches with EF_2 in the 1st search-matching process, and the input $I_5 - I_8$ matches with EF_2 in the 2nd search-matching one, then also output the pixel location as the edge location. We take L_4 as an example shown in Fig. 2d. For the 1st search-matching process, the input $I_1 - I_4$ of L_4 (1111) does not match with EF_1 (00XX/XX00). Then, the 2nd search-matching process is needed. And the 2nd input $I_5 - I_8$ of L_4 (1100) matches with EF_1 (00XX/XX00). Therefore, L_4 is considered as the edge location. In summary, the proposed MUSAN

edge detection method includes UF extraction, feature matching, and processing of matching results.

To evaluate the performance of the proposed MUSAN method, we compare it with various classical ones, e.g. Prewitt, Sobel, Roberts, Laplacian of Gaussian and SUSAN (smallest USAN) detectors. The masks and schemes of these methods are described in Fig. S2. As mentioned earlier, these detectors are based on convolution operations. In this work, the BSDS500 dataset (200 images) is used to evaluate the ability of the proposed MUSAN detector. Note that the pixel resolution of all the images in BSDS500 dataset is 481×321 . Here, the precision¹⁷, recall¹⁷, feature similarity (FSIM)¹⁸, and Pratt's figure of merit (FOM)¹⁹ are used as indicators, and their detailed definitions are described in Supplementary. Table. S1 shows the average quality assessment of different edge detection methods for 200 images (BSDS500 dataset). Through these quality assessments, it can be observed that the ability of the proposed MUSAN detector is not inferior to convolution-based detectors. Moreover, in terms of subjective visual perception, the proposed MUSAN detector outperforms SUSAN detector in distinguishing some details (taking the comparison of simulation results in Fig. S3 as an example). Nonetheless, there are still some limitations of the proposed MUSAN detector, i.e. it does not perform as well on higher pixel images. For example, if we adjust the

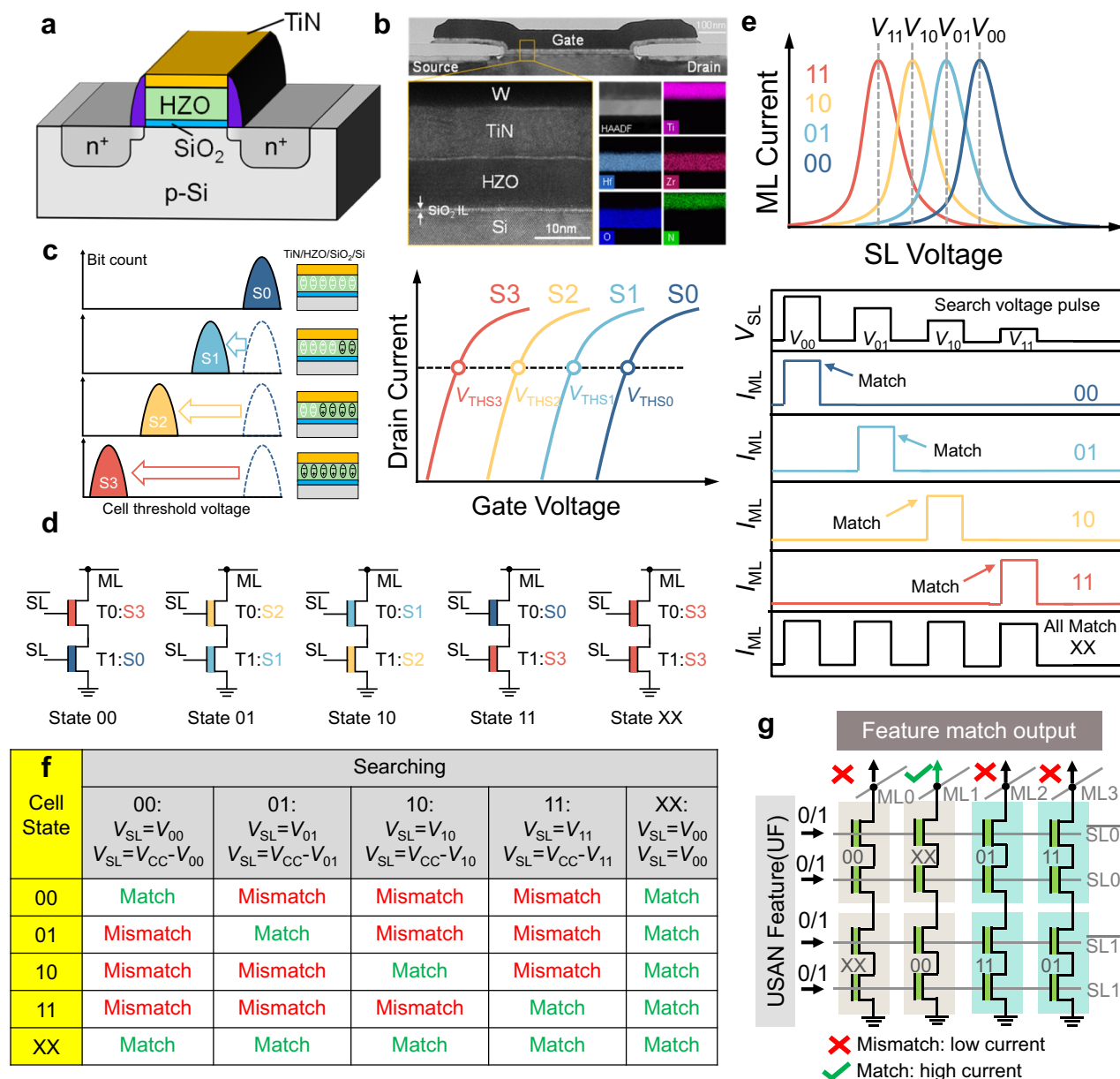


Fig. 3 | Hardware implementation of edge detector: 4×4 FeFET NAND array.

a Schematic illustration of a fabricated FeFET. **b** Cross-sectional TEM image of W/TiN/HZO/SiO₂/Si gate-stack. **c** Operation mechanisms of MLC FeFET. **d** MCAM cell

based on two FeFETs connected in series. **e** Operations Mechanisms of the multi-bit CAM cell. **f** The operation table of a MCAM cell. **g** The edge detector loaded in 4 × 4 FeFET NAND array.

input image pixels (481 × 321) in Fig. S3a into higher pixels (1924 × 1284), the results of the proposed MUSAN detector will lose some details compared to the ones of SUSAN detector. The simulation results of 6 kinds of edge detectors are shown in Fig. S4. This is due to that the cross mask of MUSAN detector consisting only 8 feature pixels, while the SUSAN detector's mask has 37 pixels. Therefore, if we want to perform edge detection for high pixel images with higher accuracy, we need to expand the MUSAN detector to more feature pixels.

Hardware implementation of MUSAN edge detector by FeFET NAND array

Next, we introduce hardware implementation of the proposed edge detector. As mentioned above, the proposed edge detector focuses on feature matching including fuzzy matching and exact matching, which can be realized by MCAM. In our design, the proposed edge detector is loaded into a 4 × 4 FeFET NAND array as MCAM. The MCAM functions

originate from the multiple states of ferroelectric by partial polarization switching, which results in multilevel cell (MLC) operations of FeFETs. Figure 3a schematically illustrates the structure of fabricated FeFETs. The details of the device fabrication are shown in Methods. Figure 3b shows the cross-sectional transmission electron microscope (TEM) image of the W/TiN/HZO/SiO₂/Si gate stack of a fabricated FeFET. The poly-crystalline nature of ferroelectric HZO is confirmed. Figure 3c illustrates operation mechanisms of MLC FeFET. Four polarization states (i.e., 'S0', 'S1', 'S2', and 'S3'), and thus threshold voltages (V_{TH}) can be tuned by different program pulse amplitudes or widths. The measured read drain current (I_d)- gate voltage (V_g) after different write pulse and the corresponding write pulse schemes are shown in Methods and Supplementary. MLC operations are confirmed by the well separated 4 V_{TH} states of a transistor.

Here, we designed a CAM cell consisting of two FeFETs connected in series (Fig. 3d), where T0 and T1 are upper and lower FeFETs with

threshold voltages V_{TH0} and V_{TH1} , respectively. Comparing the complementary input with the stored state in a CAM cell through search lines (SLs), the match/mismatch results are sensed according to the current of each match line (ML). By utilizing MLC operations of FeFET, multi-bit CAM functions, including fuzzy matching (XX) and exact matching (00, 01, 10, and 11), can be realized. For fuzzy matching, both T0 and T1 are programmed to the lowest V_{TH} state (S3) shown in Fig. 3d. Whatever the input search data is, matching always occurs. For exact matching, T0 and T1 are programmed to the different V_{TH} states shown in Fig. 3d. The input search voltages of SL and \overline{SL} (V_{SL} and $\overline{V_{SL}}$) in a cell are complementary, which satisfied $V_{SL} + \overline{V_{SL}} = V_{CC}$. According to the above-mentioned algorithm and the definition of different states of a CAM cell (i.e., 00, 01, 10, and 11), ML current (I_{ML}) - SL voltage (V_{SL}) exhibits a current peak (Fig. 3e). The input V_{SL} in search operation for each state is set as the corresponding V_{SL} where current peak happens (i.e., V_{00} , V_{01} , V_{10} , and V_{11}). When the input search data exactly matches the stored state of a CAM cell, there will be a high current response. Otherwise, the current response can be negligible. The operation table of the multi-bit CAM cell is summarized in Fig. 3f. The corresponding detailed write and search schemes are shown in Methods and Supplementary.

The proposed edge detector requires parallel fuzzy and exact matching of UF with 4-bits data. Therefore, we choose the 4×4 FeFET NAND array (Fig. 3g) for hardware implementation of the proposed algorithm. The stored EFs in 4 columns of FeFET array are 00XX, XX00, 0111, 1110, respectively. Each EF is stored in two FeFET-based MCAM cells in a NAND string. Only the input 4-bit UF fully match the two cells of one column, high current flows in the corresponding ML. Otherwise, these 4 columns will retain negligible current. The schematic diagram is shown in Fig. 3g. This is a key reason why the proposed edge detector hardware consumes low energy. Note that, the proposed MUSAN detector is applicable to various devices, including SRAM (static random-access memory)²⁰, ReRAM^{21–28}, FeFET^{29,30} and so on, as long as these devices can successfully implement CAM. The detailed comparison of MUSAN detectors based on different devices is given in Supplementary.

Performance evaluation of MUSAN edge detection hardware system

Figure 4a, b demonstrate the low-power hardware system with 4×4 FeFET NAND array as the computing core. This system controlled by a field-programmable gate array (FPGA) implements the pre-processing operations based on the UF extraction scheme and realizes post-processing operations and edge output. The 4×4 FeFET NAND array (Fig. 4b) is connected with PCB (Printed Circuit Board) through probe card, and stores the EF of 00XX/XX00/0111/1110. Due to the parasitic capacitance and resistance of devices such as probe card, there is a certain delay time in the hardware system. According to the current responses under different search voltage pulse width (1–100 μ s) in Fig. S7, we choose pulse width of 10 μ s for the search operation. Note that, search operation with pulse width of 1 μ s can be also realized if stable current response is not strictly required. Fig. S8 gives the 2D mapping of measured I_{ML} with all 16 kinds of search UF corresponding to 4 kinds of stored EF. In order to distinguish the match and mismatch I_{ML} more intuitively, we plot the 3D bar chart corresponding to Fig. S8 in Fig. 4c. The measured minimum match I_{ML} is 44.57 nA, and the maximum mismatch I_{ML} is 6.62 nA. In this work, the match I_{ML} and the mismatch I_{ML} can be considered as effective signal and noise signal, respectively. Therefore, the maximum amplitude of the noise signal in search operations does not exceed 15% ($6.62 \text{ nA}/44.57 \text{ nA} = 14.85\%$) of the effective signal amplitude, and the noise signal can be easily avoided at the hardware level. The detailed hardware design (Fig. S9) and noise analysis are given in Supplementary. Note that the hardware implementation in this work uses only a 4×4 FeFET array (16 devices), which is exactly suitable for storing the 8 feature pixels of the

proposed MUSAN detector. However, compared with SUSAN detector of 37 feature pixels, the fewer feature pixels lead to the proposed MUSAN detector not performing well on higher pixel images, as mentioned above. Expanding the MUSAN detector to images with higher pixel resolution requires FeFET array with larger size. The corresponding research will be carried out in our future work.

To further estimate the feasibility of the 4×4 FeFET NAND array as the computing core of proposed edge detector, the retention, the cycle-to-cycle variation and the endurance have been studied systematically. Figure 4d plots the retention of 4 kinds of stored EF (00XX/XX00/0111/1110). According to the retention characteristic of the match and mismatch currents, the functions of computing core can retain even after >10 years. Fig. S10 shows 10 times current responses under search pulse width of 10 μ s with 4 kinds of stored EF and all the matched search UF, illustrating negligible variation of multiple search operations. Fig. S11 depicts the endurance of search-match operation with stored EF (00XX) and search UF (0011). After 10^7 cycles, the match current still can be distinguished well, indicating the excellent endurance of the proposed edge detection hardware system.

Here, we take one image with 481×321 pixels (from BSDS500 dataset) shown in Fig. 4e as an example. The pixel information of input image is supplied by the PC (personal computer) terminal. Here, we do not artificially introduce noise into the input images. Figure 4f, g depict the ideal results by software simulation and the experimental results using the proposed edge detection, respectively. Compared with software simulation, the proposed hardware system does not lose any accuracy. Moreover, 200 images (BSDS500 dataset) are input into the proposed hardware system with the FeFET-based MUSAN detector. By point to point pixel comparison, all the hardware outputs show 100% accuracy compared with the simulation results, indicating that the proposed FeFET-based hardware system is the no-accuracy-loss edge detection hardware system. However, the hardware outputs based on convolution operators cannot achieve such high accuracy^{16,31–33}. One of the main reasons is that, for the proposed MUSAN detector, the hardware output of the edge position is 0 and the one of the non-edge position is 1, which is determined by feature matching or not, and is digital in a sense. This leads to higher accuracy, compared to the analog output of convolution operation with the need of ADC to read out.

The power analysis of the proposed FeFET-based MUSAN edge detector is provided in Methods. The power of the proposed FeFET-based MUSAN is relatively low due to two reasons. The first one is that in-memory computing with eNVMs exhibits lower power consumption compared to the computing core with storage-compute separation architecture. The other one is that different from the conventional eNVMs-based edge detectors, in which convolution operations are the computing core and each convolution operation outputs quantitative current, the 4×4 FeFET NAND array based on MUSAN only exhibit power consuming (10 fJ/per operation) when feature matching occurs. This means that the proposed hardware system consumes the most power at the edge pixels, resulting in lower power consumption than the conventional eNVMs-based edge detection hardware system.

The comparison between this work and previous eNVMs-based edge detection works is shown in Fig. 4h. To summarize, the hardware power consumption of computing core in this work is as low as 10 fJ/per operation which is the lowest one reported so far. The power estimation method is described in Methods. Moreover, the hardware system of MUSAN is ADC-free, while the ones of other NVMs-based methods are not. The performance requirements (such as device-to-device variation) for NVMs of MUSAN is tolerant, while the ones of other NVMs-based methods are tough due to the exact values of readout current considered as the convolution results. This is also the reason that the hardware accuracy compared to simulation of NVMs-

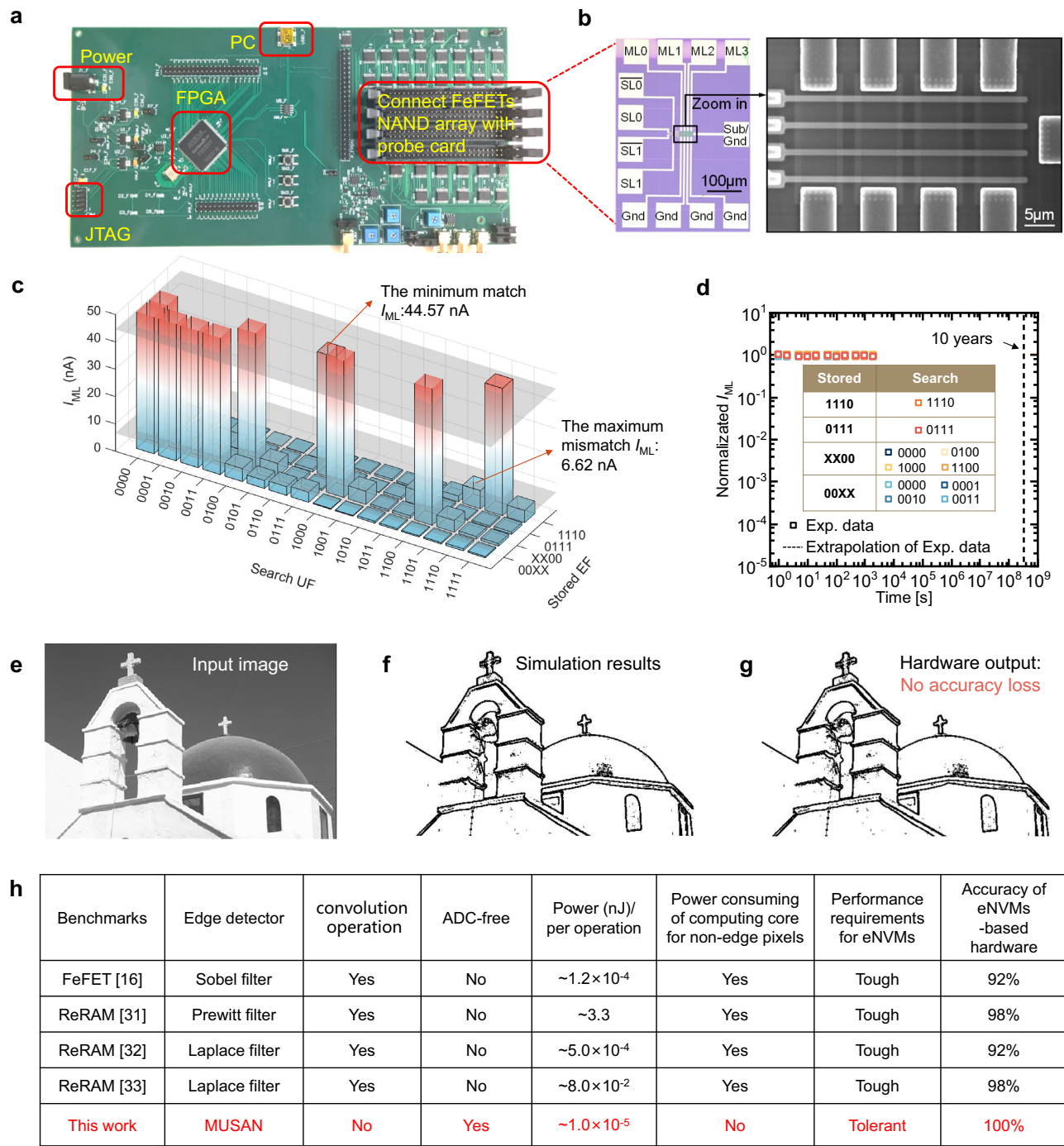


Fig. 4 | Hardware implementation of the proposed MUSUN. a Image of the PCB connected with 4×4 FeFET NAND array through probe card. **b** The microscope image of a 4×4 FeFET NAND array. **c** The measured I_{ML} with all 16 kinds of search UF corresponding to 4 kinds of stored EF. **d** The retention of 4 kinds of stored EF. Symbols are the experimental data, and the dotted lines are the extrapolation of these experimental data. **e** The input image with 481×321 pixels (from BSDS500 dataset). **f** The simulation results of the proposed MASUN detector corresponding to the input image in Fig. 4e. **g** The experimental output of the proposed hardware system corresponding to the input image in Fig. 4e. **h** Comparison of this work with other edge detection hardware.

based convolution methods cannot achieve 100%, while the one of MUSAN can.

Discussion

We have reported a FeFET-based MUSAN method for edge detection, which combines USAN operator and feature matching method. Our IMC hardware system achieves low-power (10 fJ/per operation), which is much lower than the previous eNVMs-based edge detector with convolution operations. Due to the proposed feature matching

scheme, ADC-free and no-accuracy-loss edge detector hardware can also be realized. In summary, the proposed hardware system has higher robustness, stability and accuracy compared with conventional convolution-based schemes. Our proposed edge detector scheme is also applicable to hardware systems based on other emerging NVMs technologies, which is potentially useful for other AI edge computation that may suffer from stability issues of analog convolution operations. Future work will further study the more universal approach to transforming unstable analog

convolution operations into feature matching operations based on NVMs.

Methods

Fabrication of MCAM arrays based on FeFET in NAND configuration

Fabrication started from a lightly doped p-type silicon (Si) substrate. After active area (AA) patterning, shallow trench isolation (STI) was performed for electrical isolation. Then, a dummy gate was formed by thermal oxidation of Si and chemical vapor deposition (CVD) poly-Si. After the Si₃N₄ spacer and lightly doped drain (LDD) formation, arsenic ions with 40-keV energy and $4 \times 10^{15} \text{ cm}^{-2}$ dose were implanted for source/drain (S/D). Dopant activation was performed by a spike annealing at 1050 °C. After dummy gate removal, SiO₂ interfacial layer (IL) was grown by ozone oxidation for better interface property. Next, 10 nm Hf_{0.5}Zr_{0.5}O₂ (HZO) was deposited by atomic layer deposition (ALD) using tetrakis (ethylmethylamido) hafnium (TEMAHf) and tetrakis (ethylmethylamino) zirconium (TEMAZr) as precursor, and water vapor (H₂O) as oxidant. Subsequently, 20 nm TiN and 75 nm W were deposited by sputter and CVD, respectively. To crystallize HZO and induce ferroelectricity, rapid thermal annealing (RTA) was carried out in N₂ ambient at 550 °C for 1 min. Finally, interconnects and contact pads are formed for electrical characterization. The gate width/length (W/L) of FeFET devices as well as MCAM arrays characterized in this study is 8 μm/0.5 μm.

Operation mechanism of MCAM cells

In our design, MCAM functions originate from MLC operations of FeFET cells. Figure 3c illustrates operation mechanisms of MLC FeFET. Four polarization states (i.e., 'S0', 'S1', 'S2', and 'S3'), and thus threshold voltages (V_{TH}) can be tuned by different program pulse amplitudes or widths. Fig. S5a illustrated the optimized write schemes for four different states of fabricated FeFET cells. The corresponding read $I_{\text{d}}-V_{\text{g}}$ curves for ten different FeFET cells are plotted in Fig. S5b, which shows tightly controlled four V_{TH} states. Fig. S5c plots retention and endurance of a FeFET cell.

The designed MCAM cells with two FeFET connected in series and the defined operation rules are illustrated in Fig. S6a. The input voltages of SL and $\bar{\text{SL}}$ (V_{SL} and $V_{\bar{\text{SL}}}$) in a cell are complementary. In other words, the sum of V_{SL} and $V_{\bar{\text{SL}}}$ is a fixed value (V_{CC}) determined by V_{TH} of two FeFETs and an overdrive voltage (V_{OV}). ML current versus SL voltage ($I_{\text{ML}}-V_{\text{SL}}$) exhibits a current peak at different V_{SL} according to the stored states as defined in Fig. 3d. The input V_{SL} in search operation for each state is set as the corresponding V_{SL} where current peak happens (i.e., V_{O0} , V_{O1} , V_{I0} , and V_{I1}). Only if the input search data fully match the stored state of a MCAM cell, both T0 and T1 are on, and high current flows through ML (match). Otherwise, at least one of FeFETs in a cell is off and negligible current flows through ML (mismatch). Figure S6b plots measured $I_{\text{ML}}-V_{\text{SL}}$ for 5 different states of a MCAM cell. The search voltage of SL determined from Fig. S5b are 1.75, 1.3, 0.8 and 0.3 V for data 00, 01, 10 and 11, respectively. The search operations are further verified by transient measurement of a MCAM cell as shown in Fig. S6c. I_{ML} shows behaviors exactly same as predicted by the operation table illustrated in Fig. 3f. Note that, the operating conditions are crucial for the NAND arrays³⁴. Generally, to prevent program disturbance of the SL or word-line (WL) half selected cells, a program inhibition voltage is applied to the MLs or bit-lines (BLs) of unselected NAND strings. In this work, the program inhibition voltage is set as 3 V, which can ensure the program operations are disturb-free³⁵.

Hardware system set-up

The hardware system consists of three main parts:

- A power management system that supplies power to other modules.
- A PCB with FeFETs NAND array and its peripheral circuits. FeFETs NAND array is connected to PCB with probe card. The peripheral

circuits including several sensitive amplifier circuits, row and column decoder circuits, write driver circuits, selective switching circuits and partial digital circuits are bonded and packaged in a stacked way, and they communicate with the outside through the PCB. Due to that the FeFETs NAND array is connected to peripheral circuits by probe cards, a relatively large RC delay is introduced, leading to a large read latency (~5 μs). We estimate that the maximum throughput of the search task is $\sim 4/(5 \mu\text{s}) = 3.2 \text{ Mbps}$. It is worth noting that the proposed hardware system is just a prototype system. Previous literatures have claimed that the read latency of on-chip FeFET arrays can be within 1 ns³⁶. This means that the corresponding throughput will be greatly improved if RC delay is reduced by avoiding probe card. Therefore, the throughput will be greatly increased if we encapsulate the FeFETs NAND array and peripheral circuits into a chip. The corresponding research will be carried out in our future work.

- An FPGA board with an embedded processor is responsible for data management between the FeFETs NAND array and the host computer as well as the control of the peripheral circuits for FeFETs NAND array.

Power estimation

The hardware system of proposed MUSUN detector based on FeFETs NAND array is advantageous in energy consumption.

- Power analysis of FeFETs array.

Firstly, the power consumption per operation of the proposed FeFET-based MUSUN detector is very low. Fig. S12 gives the methods of power estimation per operation for this work and previous eNVMS-based edge detection works^{16,31–33}. In this work, per operation means high current is captured on a certain ML in 4×4 FeFETs NAND array, i.e. feature matching condition. Therefore, the power consumption per operation can be represented as $E = V_{\text{ML}} \times I_{\text{ML}} \times t = -0.2 \text{ V} \times 50 \text{ nA} \times 1 \mu\text{s} = -10 \text{ fJ}$, where I_{ML} is the current of ML for feature matching, V_{ML} is the voltage of ML, and t is pulse width, respectively. For NVM-based convolutional detector, per operation means one single convolution operation. Taking FeFET-based convolution in Ref. 16 as an example, there are two convolution kernels (X-mask and Y-mask). The power consumption per operation is defined as the energy of performing one convolution with these two convolution kernels in 9×4 FeFET array, and can be evaluated as $E = V_{\text{BL}}^2 \times G \times t \times n_{\text{NVM}} = (0.1 \text{ V})^2 \times -1 \mu\text{S} \times 1 \mu\text{s} \times 12 = -120 \text{ fJ}$ (Fig. S12). Here, V_{BL} is the voltage of BL, G and n_{NVM} is conductance and number of NVMS under high conductance state, respectively. And n_{NVM} depends on the number of non-zero elements in the mask. Note that, we uniformly adopt a pulse width of 1 μs to estimate the power consumption of this work and previous eNVMS-based edge detection works for fair. Secondly, the 4×1 FeFETs connected in series are conductive only when feature matching occurs. For example, for the image with 481×321 pixels in Fig. 4e, the feature matching occurs 27960 times, resulting in the total power consuming of $10 \text{ fJ} \times 27960 = 0.28 \text{ nJ}$. While the classical method based on convolution in Ref. 16 needs $481 \times 321 = 154401$ times convolution operations, resulting $120 \text{ fJ} \times 154401 = 18.5 \text{ nJ}$. Therefore, for one whole image, the proposed FeFET-based MUSAN detector exhibits more significant advantage in power consumption.

- Power analysis of peripheral circuits.

Different from the typical eNVMS-based edge detection hardware system where the analog outputs need to be converted into digital read-out by ADC, in the proposed one the outputs can be readout by distinguishing high/low current with general purpose input/output (GPIO) instead of ADC. This leads to both direct and indirect power savings can be achieved in the following circuits in actual application scenarios. Firstly,

the direct power savings. The power consumption of driving an ADC pin and peripheral circuits is about 150 mW which can be referred to the device handbook of FPGA (CYCLONE IV). However, the power consumption of driving a GPIO pin and its peripheral circuits can usually be controlled within 5 mW, much <150 mW. Therefore, at this point, more than 145 mW of power consumption can be saved. Secondly, the indirect power savings. For conventional DC-DC chips, the working efficiency is usually around 90%. The power consumption savings on the chip side can be further reflected indirectly in the power supplies at all levels. The 145 mW chip-side power consumption savings means about 15 mW or more of the power chip thermal power consumption savings.

Data availability

The data supporting this study and the source data of all the plots of main manuscript and supplementary information are available via Zenodo at <https://doi.org/10.5281/zenodo.13808606>. The BSDS500 dataset³⁷ used to evaluate the edge detection methods discussed in this study is publicly available and can be accessed at <https://github.com/BIDS/BSDS500>.

Code availability

The codes for all the simulations supporting this study are available via Zenodo at <https://doi.org/10.5281/zenodo.13808606>.

References

- Soliman, T. et al. First demonstration of in-memory computing crossbar using multi-level cell FeFET. *Nat. Commun.* **14**, 6348 (2023).
- Kazemi, A. et al. Achieving software-equivalent accuracy for hyperdimensional computing with ferroelectric-based in-memory computing. *Sci. Rep.* **12**, 19201 (2022).
- Wu, G. et al. Ferroelectric-defined reconfigurable homojunctions for in-memory sensing and computing. *Nat. Mater.* **22**, 1499–1506 (2023).
- Chiu, Y.-C. et al. A CMOS-integrated spintronic compute-in-memory macro for secure AI edge devices. *Nat. Electron* **6**, 534–543 (2023).
- Ning, H. An in-memory computing architecture based on a duplex two-dimensional material structure for in situ machine learning. *Nat. Nanotechnol.* **18**, 493–500 (2023).
- Sun, R. et al. Survey of image edge detection. *Front. Signal Process.* **2**, 826967 (2022).
- Muntarina, K., Shorif, S. B. & Uddin, M. S. Notes on edge detection approaches. *Evol. Syst.* **13**, 169–182 (2022).
- Gallego, G. et al. Event-based vision: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 154–180 (2022).
- Li, C. et al. Analogue signal and image processing with large memristor crossbars. *Nat. Electron.* **1**, 52–59 (2017).
- Huo, Q. et al. A computing-in-memory macro based on three-dimensional resistive random-access memory. *Nat. Electron.* **5**, 469–477 (2022).
- Gong, T. et al. A 128 kb stochastic computing chip based on RRAM flicker noise with high noise density and nearly zero autocorrelation on 28 nm CMOS platform. In *IEEE International Electron Devices Meeting (IEDM)* 12.5.1–12.5.4 (IEEE, 2021).
- Lin, P. et al. Three-dimensional memristor circuits as complex neural networks. *Nat. Electron.* **3**, 225–232 (2020).
- Behbahani, F., Jooq, M. K. Q., Moaiyeri, M. H. & Tamersit, K. Leveraging negative capacitance CNTFETs for image processing: an ultra-efficient ternary image edge detection hardware. *IEEE Trans. Circuits Syst. I* **68**, 5108–5119 (2021).
- Eslahi, H., Hamilton, T. J. & Khandelwal, S. Energy-efficient ferroelectric field-effect transistor-based oscillators for neuromorphic system design. *IEEE J. Explor. Solid State Comput. Devices Circuits* **6**, 122–129 (2020).
- Yu, E. et al. Ferroelectric FET based coupled-oscillatory network for edge detection. *IEEE Electron Device Lett.* **42**, 1670–1673 (2021).
- Kim, M. K., Kim, I. J. & Lee, J. S. CMOS-compatible compute-in-memory accelerators based on integrated ferroelectric synaptic arrays for convolution neural networks. *Sci. Adv.* **8**, eabm8537 (2022).
- Tariq, N., Hamzah, R. A., Ng, T. F., Wang, S. L. & Ibrahim, H. Quality assessment methods to evaluate the performance of edge detection algorithms for digital image: a systematic literature review. *IEEE Access* **9**, 87763–87776 (2021).
- Zhang, L., Zhang, L., Mou, X. & Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* **20**, 2378–2386 (2011).
- Yi, S., Labate, D., Easley, G. R. & Krim, H. A shearlet approach to edge analysis and detection. *IEEE Trans. Image Process.* **18**, 929–941 (2009).
- Hayashi, I. et al. A 250 MHz 18 Mb full ternary CAM with low-voltage matchline sensing scheme in 65 nm CMOS. *IEEE J. Solid-State Circuits* **48**, 2671–2680 (2013).
- Zheng, L. et al. RRAM-based TCAMs for pattern search. In *2016 IEEE International Symposium on Circuits and Systems (ISCAS)* 1382–1385 (IEEE, Montreal, QC, Canada, 2016).
- Pan, K., Tosson, A. M. S., Wang, N., Zhou, N. Y. & Wei, L. A novel cascable TCAM using RRAM and current race scheme for high-speed energy-efficient applications. *IEEE Trans. Nanotechnol.* **22**, 214–221 (2023).
- Li, J., Montoye, R. K., Ishii, M. & Chang, L. 1 Mb 0.41 μm^2 2T-2R cell nonvolatile TCAM with two-bit encoding and clocked self-referenced sensing. *IEEE J. Solid State Circuits* **49**, 896–907 (2014).
- Lin, C.-C. et al. 7.4 A 256b-wordlength ReRAM-based TCAM with 1 ns search-time and 14 \times improvement in wordlength-energyefficiency-density product using 2.5T1R cell. In *2016 IEEE International Solid-State Circuits Conference (ISSCC)* 136–137 (IEEE, San Francisco, CA, USA, 2016).
- Yin, X. et al. FeCAM: A universal compact digital and analog content addressable memory using ferroelectric. *IEEE Trans. Electron Devices* **67**, 2785–2792 (2020).
- Bazzi, J., Sweidan, J., Fouda, M. E., Kanj, R. & Eltawil, A. M. Variability-aware design of RRAM-based analog CAMs. *IEEE Access* **12**, 55859–55873 (2024).
- Li, C. et al. Analog content-addressable memories with memristors. *Nat. Commun.* **11**, 1638 (2020).
- Ni, K. et al. Ferroelectric ternary content-addressable memory for one-shot learning. *Nat. Electron* **2**, 521–529 (2019).
- Das, D. et al. Experimental demonstration and modeling of a ferroelectric gate stack with a tunnel dielectric insert for NAND applications. In *2023 International Electron Devices Meeting (IEDM)* 1–4 (IEEE, San Francisco, CA, USA, 2023).
- Kim, I.-J. & Lee, J.-S. Unlocking large memory windows and 16-level data per cell memory operations in hafnia-based ferroelectric transistors. *Sci. Adv.* **10**, eadn1345 (2024).
- Oh, S. et al. Energy-efficient Mott activation neuron for full-hardware implementation of neural networks. *Nat. Nanotechnol.* **16**, 680–687 (2021).
- Hong, E. et al. Convolutional kernel with PrCaMnOx-based resistive random-access memory for neuromorphic image processing. *AIP Adv.* **13**, 015318 (2023).
- Jeon, S. et al. Hybrid precision in resistive memory-based convolutional kernel for fault-resilient neuromorphic systems. *IEEE Trans. Electron Devices* **70**, 1659–1663 (2023).
- Kim, M.-K., Kim, I.-J. & Lee, J.-S. CMOS-compatible ferroelectric NAND flash memory for high-density, low-power, and high-speed three-dimensional memory. *Sci. Adv.* **7**, eabe1341 (2021).

35. Jin, C. et al. Disturb-free operations of multilevel cell ferroelectric FETs for nand applications. *IEEE Trans. Electron Devices* **70**, 1653–1658 (2023).
36. Genssler, P. R., Van Santen, V. M., Henkel, J. & Amrouch, H. On the reliability of FeFET on-chip memory. *IEEE Trans. Comput.* **71**, 947–958 (2022).
37. Arbeláez, P., Maire, M., Fowlkes, C. & Malik, J. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 898–916 (2011).

Acknowledgements

This work was supported in part by the National Key R&D Program of China (No. 2023YFB4402303) and in part by the National Natural Science Foundation of China (No. 62204229, 62204228, 62204226, and 62025402).

Author contributions

J.C., J.G. and C.J. proposed and implemented the project. J.X., B.C., H.Z., H.Q., H.L. R.S. G.L. X.Y. M.Z. and Y.D. designed the chip and performed the experimental verification of the proposed design. J.C. conducted MATLAB simulations and verification. Y.L., J.T., H.W. and G.H. helped with data analysis. All authors contributed to write up of the manuscript. G.H. organized and supervised the whole project implementation and provided the project financial support.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-55224-8>.

Correspondence and requests for materials should be addressed to Jiani Gu or Chengji Jin.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025