SCIENCE CHINA
Information Sciences

• Supplementary File •

An Energy-efficient FeFET-based Computing-in-memory Macro using BEOL-Integrated HZO Ferroelectric Capacitors

Weizeng LI[1,2], Zhidao ZHOU[1,2], Linfang WANG[1,2], Junyu ZHU[1,2], Junzhe SHEN[1,2], Hongyang HU[1], Baihan WANG[1,2], Zhi LI[1,2], Wang YE[1,2], Zhongze HAN[1,2], Hanghang GAO[1,2] & Chunmeng DOU[1*]

[1]*Institute of Microelectronics of the Chinese Academy of Sciences, Beijing 100029, China;*
[2]*University of Chinese Academy of Sciences, Beijing 100049, China*

Appendix A Write operation table

As shown in table A1 and figure A1, to program a memory cell (MC), the N-well and source line (SL) are biased at 4.5 V, while the word line (WL) voltage is selectively set to from 0.5 V to 0 V, allowing precise programming. For half-selected MCs, appropriate biasing ensures inhibition against unintended programming; unselected SLs and WLs are maintained at 2.25 V for programming inhibition. Row-wise erase is accomplished by applying 4 V to the WL while grounding all other terminals.

Table A1 Program(PRG) and Erase(ERS) Operation Table

	BL	SL	WL	Nwell
PRG(V)	Floating	4.5	0.5-0	4.5
PRG Row Inhibition (V)	Floating	2.25	0.5 0	4.5
PRG Column Inhibition (V)	Floating	4.5	2.25	4.5
Row-wise ERS (V)	0	0	4	0

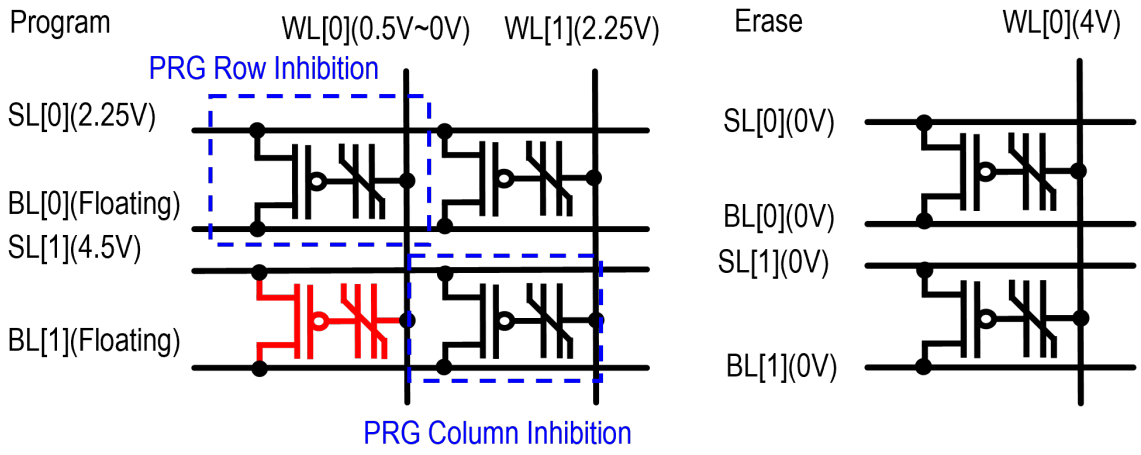

Figure A1 (Color online) Schematic of Program scheme and Erase scheme

* Corresponding author (email: douchunmeng@ime.ac.cn)

Appendix B MAC operation table

Table B1 presents the MAC operation table. In this table, positive weights are encoded on even-numbered bit-lines (BLs), while negative weights are encoded on odd-numbered BLs. For a weight of 0, both cells on the even and odd BLs are set to HVT. When the input voltage V_{IN} is applied, both $I_{BL}[2N]$ and $I_{BL}[2N+1]$ remain at 0. For a weight of -1, the cell on the even BL is set to HVT, and the cell on the odd BL to LVT, causing $I_{BL}[2N+1]$ to output I_{OUT}. Conversely, for a weight of +1, the cell on the even BL is set to LVT, and the cell on the odd BL to HVT, resulting in $I_{BL}[2N]$ equaling I_{OUT}

Table B1 MAC Operation Table

V_{IN}	Weight	Mem. Cell[2N]/[2N+1]	I_{BL}[2N]/[2N+1]
0	X	X	0
V_{IN}	0	HVT/HVT	0
V_{IN}	-1	HVT/LVT	$0/I_{OUT}$
V_{IN}	+1	LVT/HVT	I_{OUT} /0

Appendix C Macro Measurement and chip summary

Figure C1(a) shows the energy efficiency (EF) of the system at various frequencies. At 32 MHz and 50% sparsity, the peak EF of 171.34 TOPS/W is achieved. Figure C1(b) provides a detailed breakdown of the chip's area and power consumption.

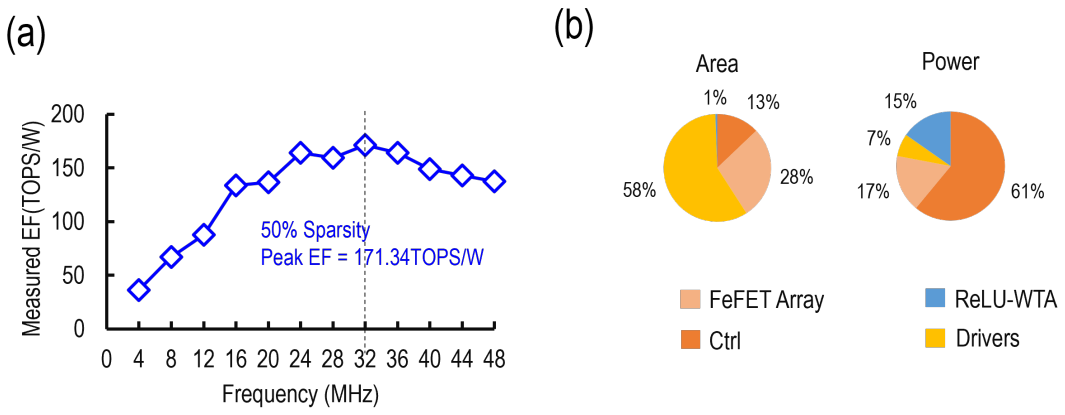

Figure C1 (Color online) (a) Measured system-level EF at different frequency; (b) area breakdown and power breakdown.

Table C1 summarizes the chip's specifications. It is fabricated using 180 nm CMOS process with one poly-silicon layer and six metal layers(1P6M) and incorporates HZO FeFETs as the storage medium. The chip operates within a frequency range of 4 to 48 MHz and a voltage range of 1.3 to 2.2 V. It has a capacity of 8 kb, with a peak throughput of 393.2 GOPS and a peak EF of 171.34 TOPS/W.

Table C1 Chip Summary

Technology	180nm CMOS 1P6M
Memory	HZO FeFET
BEOL Ferroelectric Cap Size	$0.5\mu m \times 0.5\mu m$
Frequency	4-48MHz
Supply Voltage	1.3-2.2V
Chip Area(mm^2)	0.74
IN/W Precision	Analog/Ternary
Capacity	8kb
Input/Output Parallel	256/16
Accuracy	92% @MNIST [1)]
Measured Peak Throughput(GOPS)	393.2(Analog/Ternary) [2)]
Measured Peak Energy Efficiency(TOPS/W)	171.34(Analog/Ternary) [3) 4)]
Measured Peak Area Efficiency(GOPS/mm^2)	531.35(Analog/Ternary)[2)]

[1)]Neural network consists of 2 fully-connected layer (768×256 and 256×10) [2)]Measure at 25°C, 2.2V and 48MHz [3)]Measure at 25°C, 1.7V and 32MHz [4)]With 50% Sparsity

Appendix D Comparison Table

Table D1 presents a comparison between this work and other ferroelectric-based studies. This work employs HZO BEOL FeFETs, which exhibit superior endurance compared to other work. Furthermore, this work achieves the competent energy efficiency among the configurations measured.

Table D1 Comparison Table

	IEDM20 [1]	VLSI22 [2]	VLSI23 [3]	IEDM23 [4]	This work
Technology Node(nm)	28	180	N.A.	28	180
FeFET Technology	FEOL HfO_2 FeFET	BEOL HZO FeFET	BEOL HZO FeFET	FEOL HZO FeFET	BEOL HZO FeFET
Capacity(kb)	4	8	N.A.	0.125	8
Endurance(cycles)	10^5	10^6	N.A.	N.A.	10^7
Energy Efficiency (TOPS/W) @IN/W Precision	13714 (1b/1b) [2)]	N.A.	33 [1)3)]	8.08 (3b/2b)[2)]	171.34 (Analog /Ternary)[1)]
w/ Peripheral Design [4)] and Die-photo	No	No	No	No	Yes

[1)]EF is measured [2)]EF is simulated [3)]IN/W Precision is not provided [4)]Including ctrl, driver, and readout circuit

References

1 Soliman T, Müller F, Kirchner T, et al. Ultra-Low Power Flexible Precision FeFET Based Analog In-Memory Computing. In: Proceedings of IEEE International Electron Devices Meeting, San Francisco, 2020. 29.2.1-29.2.4

2 Seidel K, Lehninger D, Hoffmann R, et al. Memory Array Demonstration of fully integrated 1T-1C FeFET concept with separated ferroelectric MFM device in interconnect layer. In: Proceedings of IEEE Symposium on VLSI Technology and Circuits, Honolulu, 2022. 355-356

3 Du Y W, Tang J S, Li Y J, et al. Monolithic 3D Integration of FeFET, Hybrid CMOS Logic and Analog RRAM Array for Energy-Efficient Reconfigurable Computing-In-Memory Architecture. In: Proceedings of IEEE Symposium on VLSI Technology and Circuits, Kyoto, 2023. 1–2

4 Zhao Z J, Xu Y X, Read J, et al. In-Situ Encrypted NAND FeFET Array for Secure Storage and Compute-in-Memory. In: Proceedings of International Electron Devices Meeting, 2023, 1–4