

저 전력 8+T SRAM을 이용한 인 메모리 컴퓨팅 가산기 설계

홍창기* · 김정범**

Design of In-Memory Computing Adder Using Low-Power 8+T SRAM

Chang-Ki Hong* · Jeong-Beom Kim**

요약

SRAM 기반 인 메모리 컴퓨팅은 폰 노이만 구조의 병목 현상을 해결하는 기술 중 하나이다. SRAM 기반의 인 메모리 컴퓨팅을 구현하기 위해서는 효율적인 SRAM 비트 셀 설계가 필수적이다. 본 논문에서는 전력 소모를 감소시키고 회로 성능을 개선시키는 저 전력 차동 감지 8+T SRAM 비트 셀을 제안한다. 제안하는 8+T SRAM 비트 셀은 SRAM 읽기와 비트 연산을 동시에 수행하고 각 논리 연산을 병렬로 수행하는 리플 캐리 가산기에 적용한다. 제안하는 8+T SRAM 기반 리플 캐리 가산기는 기존 구조와 비교 하여 전력 소모는 11.53% 감소하였지만, 전파 지연 시간은 6.36% 증가하였다. 또한 이 가산기는 PDP(Power Delay Product)가 5.90% 감소, EDP(Energy Delay Product)가 0.08% 증가하였다. 제안한 회로는 TSMC 65nm CMOS 공정을 이용하여 설계하였으며, SPECTRE 시뮬레이션을 통해 타당성을 검증하였다.

ABSTRACT

SRAM-based in-memory computing is one of the technologies to solve the bottleneck of von Neumann architecture. In order to achieve SRAM-based in-memory computing, it is essential to design efficient SRAM bit-cell. In this paper, we propose a low-power differential sensing 8+T SRAM bit-cell which reduces power consumption and improves circuit performance. The proposed 8+T SRAM bit-cell is applied to ripple carry adder which performs SRAM read and bitwise operations simultaneously and executes each logic operation in parallel. Compared to the previous work, the designed 8+T SRAM-based ripple carry adder is reduced power consumption by 11.53%, but increased propagation delay time by 6.36%. Also, this adder is reduced power-delay-product (PDP) by 5.90% and increased energy-delay-product (EDP) by 0.08%. The proposed circuit was designed using TSMC 65nm CMOS process, and its feasibility was verified through SPECTRE simulation.

키워드

In-memory Computing, Low-Power Circuit, Process-In-Memory (PIM), SRAM, von Neumann Architecture
인 메모리 컴퓨팅, 저 전력 회로, Process-In-Memory (PIM), SRAM, 폰 노이만 구조

* 강원대학교 전자공학과(hong6010@kangwon.ac.kr)

** 교신저자 : 강원대학교 전자공학과

• 접수일 : 2023. 02. 21

• 수정완료일 : 2023. 03. 18

• 게재확정일 : 2023. 04. 17

• Received : Feb. 21, 2022, Revised : Mar. 18, 2023, Accepted : Apr. 17, 2023

• Corresponding Author : Jeong-Beom Kim

Dept. of Electronics Engineering, Kangwon National University

Email : kimjb@kangwon.ac.kr

I. 서론

현재 머신러닝 및 인공지능과 같은 데이터 집약적 응용분야의 증가로 인해 높은 처리량과 에너지 효율적인 컴퓨팅 구조에 대한 많은 수요가 발생하고 있다. 폰 노이만 구조의 경우 프로세스 장치와 메모리 간의 데이터 전송 중 많은 에너지 소비가 발생하고 병목 현상 또는 메모리 월 (memory wall) 현상이 발생하면서 컴퓨팅의 성능을 저하시킨다. 이러한 문제점을 해결하는 방안 중 하나로 메모리 내에서 연산 동작을 수행하여 프로세스 장치와 메모리 사이의 데이터 전송을 감소시키는 방법으로 컴퓨팅의 성능을 향상하는 인 메모리 컴퓨팅 (IMC, In-Memory Computing)이 연구 되어왔다. SRAM(Static Random Access Memory) 기반 인 메모리 컴퓨팅 회로는 두 개의 메모리 워드라인 (WL, Word Line)을 동시에 활성화한 후 메모리에 저장된 값을 이용하여 컴퓨팅 동작을 진행한다. 표준 6T SRAM의 경우 서로 다른 값이 저장된 상황에서 컴퓨팅 동작 시 읽기 및 쓰기 비트라인 (BL, Bit Line)을 공유하기 때문에 메모리에 저장된 값이 뒤바뀌는 읽기 장애 문제가 존재한다. 이러한 문제를 해결하기 위해 별도의 읽기/쓰기 단자가 존재하는 8+T SRAM이 등장하였다. 8+T SRAM은 별도의 읽기/쓰기 단자를 이용하여 안정적인 컴퓨팅 동작을 할 수 있다. 하지만 늘어난 트랜지스터의 개수로 인해 컴퓨팅 동작 시 표준 6T SRAM과 비교하여 전력 소모가 증가한다는 단점을 가지고 있다[1-9].

이러한 문제를 해결하기 위해 본 논문에서는 기존 8+T SRAM 기반 인 메모리 컴퓨팅 회로의 컴퓨팅 동작 시 발생하는 전력 소모를 감소시키는 저 전력 차동 감지 8+T SRAM 비트 셀을 제안하며 제안한 회로를 이용하여 인 메모리 컴퓨팅 리플 캐리 가산기를 설계한다.

본 논문은 다음과 같이 구성한다. 먼저 II장에서 기존 인 메모리 컴퓨팅 회로의 8+T SRAM 셀에 대해 기술하고, III장에서는 제안하는 저 전력 8+T SRAM 셀과 제안하는 구조를 이용한 인 메모리 컴퓨팅 전가산기 및 리플 캐리 가산기 설계에 대해 기술한 후, IV장에서 시뮬레이션 결과를 비교 및 분석하며, V장에서 결론을 맺는다.

II. 8+T SRAM 인 메모리 컴퓨팅 회로

2.1 기존 구조

그림 1 (a)는 기존 8+T SRAM 셀의 구조이다. 기존 구조의 경우 읽기 액세스 (access) 트랜지스터가 NMOS(N-type metal-oxide-semiconductor)로 구성되어 있다. 기존 구조의 읽기 동작은 읽기 단자 (RBL, Read Bit Line)와 $\overline{\text{RBL}}$ 를 VDD로 사전 충전 후 N7 트랜지스터에 '1'의 값이 인가되면 진행된다. N5와 N6 트랜지스터는 SRAM 셀에 저장된 값에 따라 RBL과 $\overline{\text{RBL}}$ 를 GND로 방전 또는 VDD 상태를 유지한다.

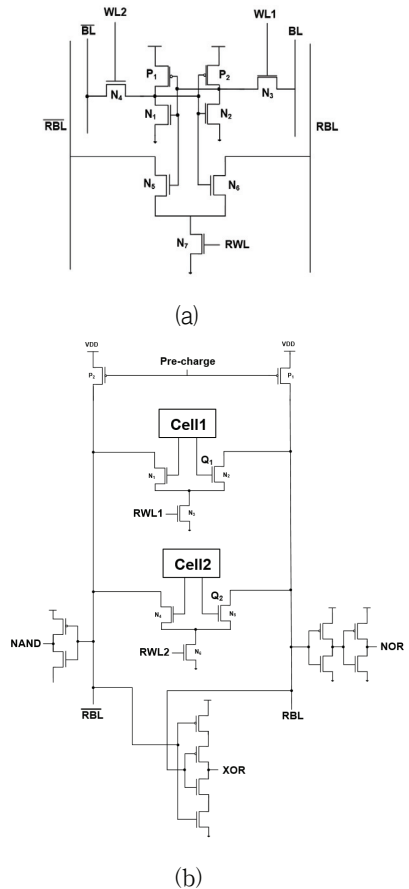


그림 1. (a) 기존 8+T SRAM 셀
(b) 기존 인 메모리 컴퓨팅 회로
Fig. 1 (a) Previous 8+T SRAM cell
(b) Previous IMC circuit

그림 1 (b)는 기존 인 메모리 컴퓨팅 회로이다. 기존 구조의 경우 8+T SRAM 셀 두 개를 기반으로 하고, 인버터와 Muller C-element로 구성되어 있다 [10]. 기존 구조의 컴퓨팅 동작은 먼저 PMOS로 구성된 사전 충전 회로를 이용하여 VDD로 사전 충전 후 두 개의 읽기 워드라인 (RWL, Read Word Line)을 모두 활성화한다. 다음으로 Q1과 Q2에 저장된 값에 따라 RBL과 RBL를 GND로 방전 또는 VDD 상태 유지를 진행한다. 마지막으로 출력값은 RBL과 RBL의 상태에 따라 인버터와 Muller C-요소를 통해 값을 얻을 수 있다. 기존 구조의 경우 RBL을 통해 NOR 연산, RBL를 통해 NAND 연산, Muller C-element를 이용하여 XOR 연산이 가능하다. 기존 구조는 표준 6T SRAM 기반 인 메모리 컴퓨팅 회로에 비해 SRAM 셀에 트랜지스터가 증가하여 컴퓨팅 동작 시 전력 소모가 증가한다는 문제점이 존재한다. 이러한 문제를 해결하기 위해 전력 소모를 감소시키는 구조를 제안한다.

III. 저 전력 인 메모리 컴퓨팅 회로

3.1 저 전력 8+T SRAM 셀

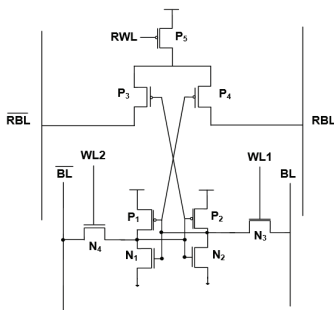


그림 2. 제안하는 저 전력 8+T SRAM 셀
Fig. 2 Proposed Low-power 8+T SRAM cell

그림 2는 제안하는 저 전력 8+T SRAM 셀의 구조이다. 기존 셀 구조와 달리 읽기 액세스 트랜지스터가 PMOS(: P-type metal-oxide-semiconductor)로 구성되어 있다. 제안하는 구조의 읽기 동작은 RBL과

RBL를 GND로 사전 방전 후 P5 트랜지스터에 '0'의 값이 인가되면 진행된다. P3와 P4 트랜지스터는 SRAM 셀에 저장된 값에 따라 RBL과 RBL를 VDD로 충전 또는 GND 상태 유지를 진행한다. 그림 3은 컴퓨팅 동작 시 기존 구조와 제안하는 구조의 읽기 액세스 트랜지스터를 통한 최대 전류이다. 제안하는 구조가 기존 구조에 비해 컴퓨팅 동작 시 최대 전류를 17.8% 감소하여 전력 소모를 감소시킨다.

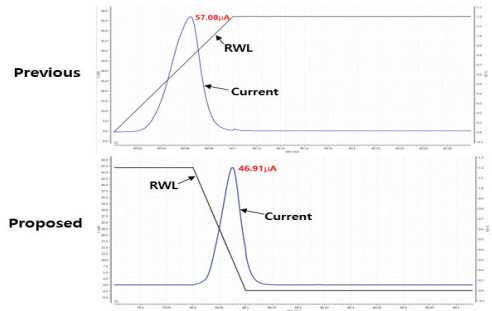


그림 3. 읽기 액세스 트랜지스터를 통한 최대 전류 비교

Fig. 3 Comparison of the maximum current through the read access transistor

3.2 제안하는 인 메모리 컴퓨팅 회로

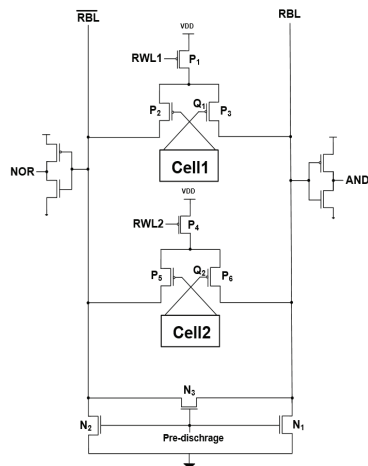


그림 4. 제안하는 인 메모리 컴퓨팅 회로
Fig. 4 Proposed IMC circuit

그림 4와 같이 제안하는 인 메모리 컴퓨팅 회로는 두 개의 저 전력 8+T SRAM 셀을 기반으로 하고, NMOS로 구성된 사전 방전 회로와 인버터로 구성되었다. 제안하는 인 메모리 컴퓨팅 회로의 연산 동작을 진행하기 위해서는 먼저 NMOS로 구성된 사전 방전 회로를 이용하여 RBL과 \overline{RBL} 를 GND로 사전 방전을 진행한다. 그런 다음 P1과 P4 액세스 트랜지스터에 모두 '0'의 값을 인가하여 활성화한다. Q1과 Q2에 인가되는 값에 따라 RBL과 \overline{RBL} 는 VDD 충전 또는 GND 상태 유지를 결정한다. 표 1 진리표를 확인해보면 Q1과 Q2에 인가되는 값에 따라 출력값이 다른 것을 확인할 수 있다. Q1과 Q2에 '0'의 값이 인가된 경우 RBL은 VDD로 충전되고 \overline{RBL} 는 GND 상태를 유지하며 인버터를 통해 RBL은 '0', \overline{RBL} 는 '1'의 값을 출력한다. Q1과 Q2에 서로 다른 값이 인가되는 경우 RBL과 \overline{RBL} 모두 VDD로 충전되기 때문에 인버터를 통해 RBL과 \overline{RBL} 모두 '0'의 값을 출력한다. 마지막으로 Q1과 Q2에 '1'의 값이 저장된 경우 RBL은 GND 상태를 유지하고 \overline{RBL} 는 VDD로 충전되면서 인버터를 통해 RBL은 '1'의 값, \overline{RBL} 는 '0'의 값을 출력한다. 출력된 값을 통해 RBL은 AND, \overline{RBL} 는 NOR 연산 동작이 가능하다는 것을 알 수 있다.

표 1. 제안하는 인 메모리 컴퓨팅 회로 진리표
Table 1. Truth table of proposed IMC circuit

Q1	Q2	RBL (AND)	\overline{RBL} (NOR)
0	0	0	1
0	1	0	0
1	0	0	0
1	1	1	0

3.3 인 메모리 컴퓨팅 8 비트 리플 캐리 가산기

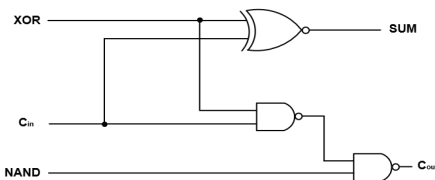


그림 5. 제안한 인 메모리 컴퓨팅 회로를 이용한 전가산기 회로

Fig. 5 Full Adder using proposed IMC circuit

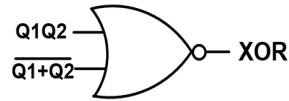


그림 6. NOR 게이트를 이용한 XOR 동작

Fig. 6 XOR operation using NOR gate

그림 5는 제안하는 인 메모리 컴퓨팅 회로를 이용한 전가산기 회로를 보여준다. 인 메모리 컴퓨팅 전가산기의 입력은 총 세 개다. XOR 연산 입력은 그림 6과 같이 RBL과 \overline{RBL} 의 출력에 NOR 게이트를 추가하여 구현한다. NAND 연산 입력은 RBL 출력에 인버터를 추가하여 구현한다. Cin 입력은 전단에 Cout 출력이 입력으로 구현된다. 식 (1)과 식 (2)는 각각 인 메모리 컴퓨팅 전가산기 회로의 Sum과 Carry 연산식이다.

$$Sum = A \oplus B \oplus C_{in} \quad \dots (1)$$

$$Carry = \overline{\overline{(A \oplus B)} \cdot C_{in} \cdot A \cdot B} \quad \dots (2)$$

인 메모리 컴퓨팅 전가산기 회로는 두 개의 RWL이 동시에 '0'의 값이 인가되어 활성화된 경우 SRAM에 저장된 값을 이용하여 연산 동작을 진행한다.

그림 7은 제안하는 인 메모리 컴퓨팅 전가산기 회로를 이용한 8 비트 리플 캐리 가산기이다[10]. 8 비트 리플 캐리 가산기를 이용하여 기존 8+T SRAM과 동일한 경우에서 성능을 비교하기 위하여 식 (3)의 경우에서 비교를 진행하였다. 식 (3)의 경우는 한 개의 SRAM 셀을 이용하여 RBL을 충전하는 연산 동작이기 때문에 두 개의 SRAM 셀을 이용하여 RBL을 충전하는 연산에 비해 최악의 상황이라 할 수 있다. 컴퓨팅 동작 시작 전 기존 구조의 경우 VDD로 사전 충전을 진행하고, 제안하는 구조의 경우 GND로 사전 방전을 진행하는데 그때의 8 비트 리플 캐리 가산기는 동일한 '1 1111 1110' 값이 연산 결과값으로 출력된다. 사전 충·방전 후 식 (3)의 연산을 진행하여 '0 1111 1111'의 결과값을 출력한다.

$$\begin{array}{r} 1111\ 1111 \\ + 0000\ 0000 \\ \hline 0\ 1111\ 1111 \end{array} \quad (3)$$

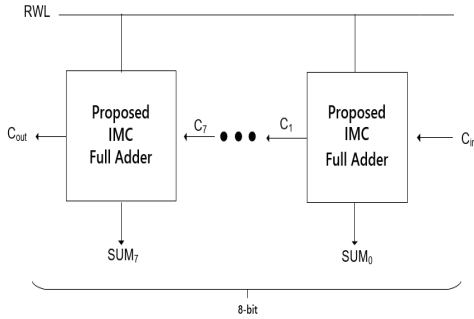


그림 7. 제안하는 인 메모리 컴퓨팅 전가산기를 이용한 8 비트 리플 캐리 가산기
Fig. 7 8 bit ripple carry adder using the proposed IMC full adder

IV. 시뮬레이션 결과 및 비교

본 논문에 구성된 회로는 TSMC 65nm CMOS 공정을 이용하여 SPECTRE로 시뮬레이션을 진행하였다. 그림 8은 제안하는 8 비트 인 메모리 컴퓨팅 리플 캐리 가산기 타이밍 다이어그램을 나타낸다. RWL1, 2의 상승 시간 (rising time), 하강 시간 (falling time)은 100ps이며 컴퓨팅 동작을 시작했을 때 C1~C7은 '0'의 값을 출력하고, Cout도 '0'의 값을 출력하는 것을 알 수 있다. Sum0~7은 '1' 값을 정상적으로 출력하는 것을 알 수 있다. 표 2는 기존 8+T SRAM 인 메모리 컴퓨팅 리플 캐리 가산기와 제안하는 8+T SRAM 인 메모리 컴퓨팅 리플 캐리 가산기 비교표이다. 본 논문에서 구성된 회로는 TT(Typical Typical) 코너에서 1.2V 공급전압, 출력단 0.2pF 부하, 동작 온도 25℃ 조건에서, 전파 지연 시간은 0.685ns, 전력 소모 7.748μW, PDP(Power Delay Product) 5.307fJ과 EDP(Energy Delay Product) 3.635E-24Js 결과값을 얻을 수 있었다. 기존 회로에 비해 전파 지연 시간이 6.36% 증가하였고, 전력 소모는 11.53%, PDP는 5.90% 감소하였고, EDP는 0.08% 증가하였다.

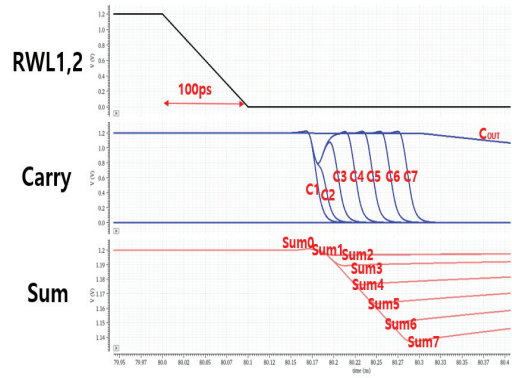


그림 8. 8 비트 리플 캐리 가산기 타이밍 다이어그램

Fig. 8 Timing diagram of 8 bit ripple carry adder

표 2. 8 비트 인 메모리 컴퓨팅 리플 캐리 가산기 비교표

Table 2. Comparison table of 8 bit IMC ripple carry adder

	Previous work	Proposed work
Propagaion delay time [ns]	0.644	0.685
Power consumption [μW]	8.758	7.748
PDP [fJ]	5.640	5.307
EDP [E-24Js]	3.632	3.635

기존 구조와 제안하는 구조의 공급 전압, 동작 온도 변화에 따른 전파 지연 시간, 전력 소모, PDP, EDP의 비교를 진행하였다. 그림 9는 공급 전압이 0.9V로부터 1.5V까지 변화할 때, 그림 10은 동작 온도가 0℃부터 75℃ 범위에서의 변화할 때 비교이다. 그림 9를 통해 공급 전압이 1.4V 이상이 되는 지점에서 기존 구조에 비해 제안하는 구조의 전력 소모, PDP, EDP 측면에서 우수함을 확인하였다. 또한 그림 10을 통해 동작 온도 60℃ 이상이 되는 지점에서 전력 소모, PDP, EDP 측면에서 제안하는 구조가 우수함을 확인하였다.

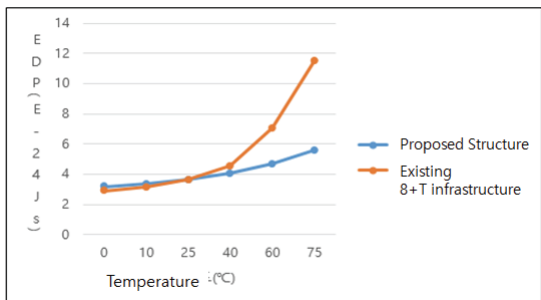
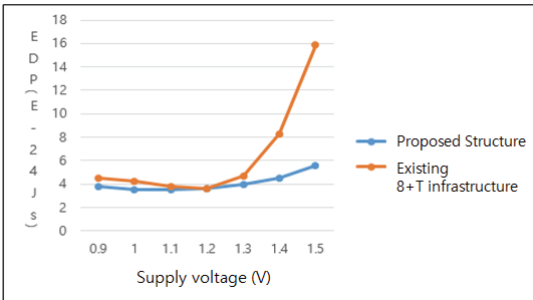
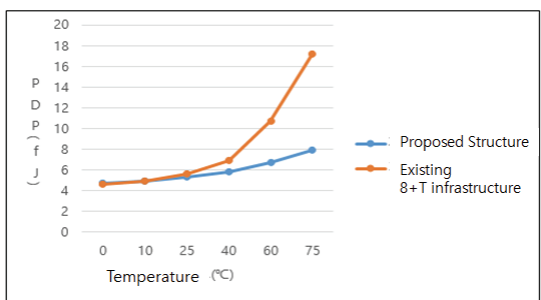
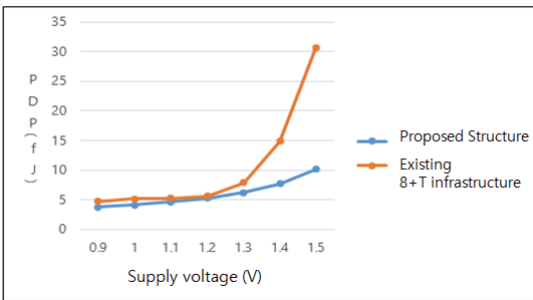
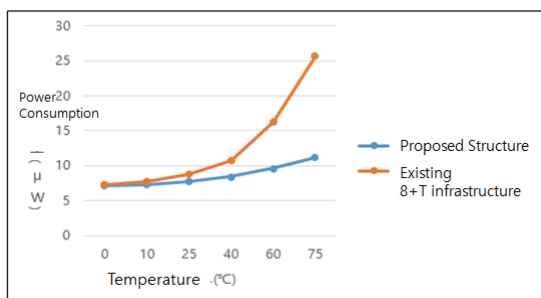
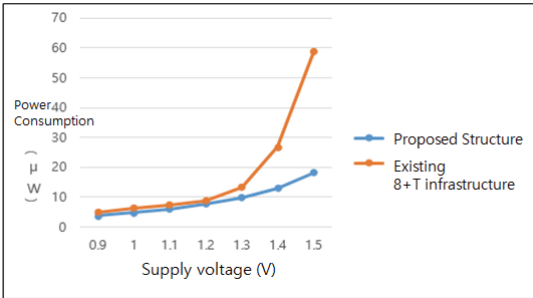
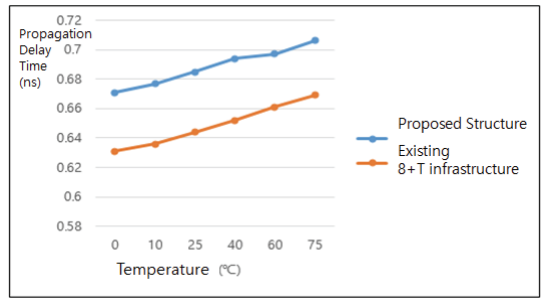
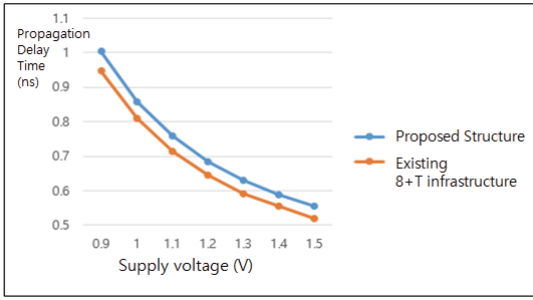


그림 9. 공급 전압 변화에 의한 인 메모리 컴퓨팅 리플 캐리 가산기 특성
 Fig. 9 Characteristics variation of IMC ripple carry adder circuit by supply voltage

그림 10. 동작 온도 변화에 의한 인 메모리 컴퓨팅 리플 캐리 가산기 특성
 Fig. 10 Characteristics variation of IMC ripple carry adder circuit by operating temperature

V. 결 론

본 논문에서는 PMOS 액세스 트랜지스터를 이용한 저 전력 차동 감지 8+T SRAM 셀을 제안하였다. 제안하는 구조를 이용한 리플 캐리 가산기는 기존 구조에 비해 전력 소모가 감소하였지만, 전류 구동 능력이 감소하여 전파 지연 시간이 증가하였다. 향후 연구는 제안하는 인 메모리 컴퓨팅 회로의 전력 소모를 유지하면서 전파 지연 시간을 감소시키기 위해 고속 동작이 가능한 출력단 증폭기에 대한 연구가 필요하다.

감사의 글

본 논문은 2022년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임(P0017011, 2022년 산업혁신인재성장지원사업)

References

- [1] J. Chen, W. Zhao, Y. Wang, Y. Shu, W. Jiang and Y. Ha, "A Reliable 8T SRAM for High-Speed Searching and Logic-in-Memory Operations," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 30, no. 6, June 2022, pp. 769-780.
- [2] J. Chen, W. Zhao, Y. Wang and Y. Ha, "Analysis and Optimization Strategies Toward Reliable and High-Speed 6T Compute SRAM," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 4, April 2021, pp. 1520-1531.
- [3] M. Nabavi and M. Sachdev, "A 290-mV, 3.34-MHz, 6T SRAM With pMOS Access Transistors and Boosted Wordline in 65-nm CMOS Technology," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 2, Feb 2018, pp.656-667.
- [4] J. P. Kulkarni, A. Goel, P. Ndai and K. Roy, "A Read-Disturb-Free, Differential Sensing 1R/1W Port, 8T Bitcell Array," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 19, no. 9, Sept 2011, pp. 1727-1730.
- [5] K. Lee, J. Jeong, S. Cheon, W. Choi and J. Park, "Bit Parallel 6T SRAM In-memory Computing with Reconfigurable Bit-Precision," 2020 57th ACM/IEEE Design Automation Conference (DAC), San Francisco, CA, USA, 2020, pp. 1-6.
- [6] M. Nabavi and M. Sachdev, "A 290-mV, 3.34-MHz, 6T SRAM With pMOS Access Transistors and Boosted Wordline in 65-nm CMOS Technology," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 2, Feb. 2018, pp. 656-667.
- [7] A. Agrawal, A. Jaiswal, C. Lee and K. Roy, "X-SRAM: Enabling In-Memory Boolean Computations in CMOS Static Random Access Memories," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 12, Dec. 2018, pp. 4219-4232.
- [8] C. Yu, T. Yoo, K. T. C. Chai, T. T. -H. Kim and B. Kim, "A 65-nm 8T SRAM Compute-in-Memory Macro With Column ADCs for Processing Neural Networks." *IEEE Journal of Solid-State Circuits*, vol. 57, no. 11, Nov. 2022, pp.3466-3476.
- [9] H. Kim, T. Yoo, T. T. -H. Kim and B. Kim, "Colonnade: A Reconfigurable SRAM-Based Digital Bit-Serial Compute-In-Memory Macro for Processing Neural Networks," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 7, July 2021, pp. 2221-2233.
- [10] S. Song, & Y. Kim, "Novel In-Memory Computing Adder Using 8+ T SRAM." *Electronics*, vol. 11, no. 6, March 2022.

저자 소개



홍창기(Chang-Ki Hong)

2022년 2월 : 강원대학교 전자공학과
졸업(공학사)

2022년 2월 ~ 현재 : 강원대학교 대
학원 BIT의료융합학과 석사과정

※ 관심분야 : SRAM 설계, 저 전력회로 설계



김정범(Jeong-Beom Kim)

1985년 2월 : 인하대학교 전자공학과
졸업(공학사)

1987년 2월 : 인하대학교 대학원 전
자공학과 졸업(공학석사)

1997년 2월 : 포항공대학교 대학원 전자전기공학과 졸
업(공학박사)

1987년 ~ 1992년 : 금성반도체 중앙연구소 선임연구원

1997년 ~ 1998년 : 현대전자 시스템IC 연구소 책임연구원

1999년 ~ 현재 : 강원대학교 전자공학과 교수

※ 관심분야 : VLSI 설계, 저 전력회로 설계