



공학석사 학위논문

Charge-storage synapse using In-Ga-Zn-Oxide Thin Film Transistor for training Deep Neural Network

인듐-갈륨-징크-옥사이드 박막 트랜지스터를 사용한 딥 뉴럴 네트워크 학습용 전하저장형 시냅스

2021년 8월

서울대학교 대학원

재료공학부

홍 상 준

Charge-storage synapse using In-Ga-Zn-Oxide Thin Film Transistor for training Deep Neural Network

지도 교수 김 상 범

이 논문을 공학석사 학위논문으로 제출함 2021년 8월

> 서울대학교 대학원 재료공학부 홍 상 준

홍상준의 공학석사 학위논문을 인준함 2021년 8월

위 원	· 장	황 철 성
부위	원장	김 상 범
위	원	이 형 민

초 록

실리콘 트랜지스터와 커패시터를 활용한 CMOS 기반의 전하 저장형 시냅스 구조는 비휘발성 소자 기반 저항 변화형 시냅스 구조 대비 선형적이고 대칭적인 가중치 업데이트가 가능해 딥 뉴럴 네트워크(DNN) 온 칩 학습용 가속기 시냅스 요구사항을 만족하여 소프트웨어 수준의 학습 정확도를 달성하였음이 알려져 있다. 하지만 CMOS 기반의 시냅스는 트랜지스터 누설전류로 인해 학습된 가중치를 잃어버리게 되면 학습 정확도가 하락하는 문제를 가지고 있어 커패시터 크기 축소가 어렵다. 이번 연구에서는 누설전류가 낮은 IGZO TFT를 활용하여 선형적이고 대칭적 가중치 업데이트 특성을 가지도록 새로운 구조의 전하저장형 시냅스를 제안하고, 제작하였다. 또한 합리적인 커패시터 면적 축소가 가능한 IGZO TFT 누설전류 수준을 제안하고 제작한 소자의 실력치와 비교하였다. 그 결과 IGZO TFT를 활용한 2T-1C 구조에서 3mV/step Depression 동작을 검증했고, 제작한 IGZO TFT의 누설전류 수준이 ~2.5x10⁻¹⁶A/um 읶을 확인하였다. 그리고 IGZO TFT를 사용한 5T-1C 시냅스 구조에서 100 레벨의 Potentiation 및 Depression 동작을 검증하였다.

주요어 : 전하저장형 시냅스, 선형적이고 대칭적 가중치 갱신, 낮은 누설전류, 긴 가중치 전하 보유시간, 인듐-갈륨-징크-산화물 박막 트랜지스터

학 번 : 2019-23695

목 차

초록	i
목차	ii
List of Figures	iv
List of Tables	vi

1		1. 서
·항·1	상적인 인공 신경망 시냅스 소자 요구사형	1.1
3	MOS 기반 시냅스 소자의 장점과 한계	1.2

••••••		2. 문헌연구	2.
	IGZO TFT 소자	2.1	
	커패시터 소자	2.2	

3. 실험 및 분석 방법6

3.1

3.2	커패시터 소자	제작	9

- 3.3 시냅스 소자 제작......11
 - 3.3.1 CMOS 기반 시냅스 소자와의 비교11

IGZO TFT 소자 제작......6

- 3.3.2 IGZO TFT 활용 시냅스 소자 회로 구조......15
- 3.4 소자 레이아웃 및 공정 프로세스 디자인......18

4.	결과 및 논의	
	4.1 IGZO TFT	기반 전하저장형 시냅스 소자 동작특성25
	4.1.1	기초 소자 특성25
	4.1.2	Off Leakage 특성27
	4.1.3	Potentiation update37
	4.1.4	Depression update
	4.1.5	시냅스 소자 가중치 읽기 동작41
	4.2 IGZO TFT	기반 전하저장형 시냅스 가중치 거동42
	4.2.1	Update 특성42
	4.2.2	Retention 특성50

제	5	장	결	론	. 5	6
---	---	---	---	---	-----	---

감사의말	58
참고문헌	59

Abstract	62
----------	----

List of Figures

Figure 1.1 MAC 연산 가속용 Crossbar Array 구조 예1	
Figure 3.1.1 IGZO TFT 소자 단면도7	,
Figure 3.1.2 IGZO TFT 소자 특성8)
Figure 3.2.1 커패시터 단면도10)
Figure 3.3.1.1 3T-1C 시냅스 회로 구조12)
Figure 3.3.1.2 IGZO TFT Transfer 특성13)
Figure 3.3.1.3 IGZO TFT를 활용한 시냅스의 Potentiation 동	작
	÷
Figure 3.3.2.1 IGZO TFT를 활용한 5T-1C 시냅스 회로 구조16	;
Figure 3.3.2.2 Source 노드 전압이 고정된 IGZO TFT 특성17	,
Figure 3.4.1 Retention > 0.2s 확보 가능 커패시터 면적19)
Figure 3.4.2 Step수 > 1000 확보 가능 커패시터 면적)
Figure 3.4.3 트랜지스터 및 커패시터 누설전류 검증 회로21	
Figure 3.4.4 누설전류 측정용 Jumbo FET)
Figure 3.4.5 4Layer 공정 프로세스23)
Figure 3.4.6 8Layer 공정 프로세스24	:
Figure 4.1.2.1 Leakage current 측정용 2T 회로)
Figure 4.1.2.2 Leakage current 측정용 2T 회로 동작 방법30)
Figure 4.1.2.3 Read 트랜지스터 Drain 전류 선형 구간31	
Figure 4.1.2.4 Leakage current 측정 Timing Diagram32)
Figure 4.1.2.5 Program 트랜지스터 Gate Off 전압에 따뜻	른
Storage 전압 변화33)
Figure 4.1.2.6 Program 트랜지스터 Gate Off 전압에 따뜻	른
Leakage 환산	

Figure 4.1.2.7 Jumbo FET Transfer curve
Figure 4.1.3.1 5T-1C 시냅스 소자 Potentiation update 38
Figure 4.1.4.1 5T-1C 시냅스 소자 Depression update 40
Figure 4.1.5.1 5T-1C 시냅스 소자 가중치 읽기 동작42
Figure 4.2.1.1 Read ADC 회로 개념도45
Figure 4.2.1.2 5T-1C 시냅스 50 Step 업데이트 거동46
Figure 4.2.1.3 5T-1C 시냅스 100 Step 업데이트 거동47
Figure 4.2.1.4 5T-1C 시냅스 100 Step 업데이트 거동(10회)48
Figure 4.2.1.5 시냅스 가중치에 따른 업데이트 전류량 비 49
Figure 4.2.2.1 Leakage current 측정용 2T-1C 회로52
Figure 4.2.2.2 HfO2 절연체 기반 커패시터 누설전류53
Figure 4.2.2.3 IGZO TFT 기반 시냅스 Retention54
Figure 4.2.2.4 IGZO TFT 누설전류 수준별 예상 Retention55

List of Tables

Table 4.1.1.1 | IGZO TFT wafer 특성1

1. 서 론

1.1 이상적 인공 신경망 시냅스 소자 요구사항

인공 신경망에 기반한 인공지능을 활용한 산업의 발전에 따라 인공 신경망 연산에 최적화된 컴퓨팅 구조가 널리 제안되고 있다. 특히, 대규모로 병렬화된 곱셈 및 누산 (Multiply-and-accumulation, MAC) 연산 효율을 극대화할 수 있는 하드웨어 구조가 심층 신경망 컴퓨팅을 저전력, 고효율화 할 수 있다고 알려져 있다. 이러한 구조로 Crossbar Array 구조의 Analog Resistive Processing Unit (RPU) 아키텍처가 제안되어 알려져 있다.^[1] RPU 아키텍처의 Unit Cell은 입력 펄스 신호에 따라 전기전도도나 저장된 전하량을 변화시켜 인공 신경망 시냅스 가중치 연결 강도를 조정한다. 이것은 인공 신경망 시냅스가 곱셈 및 누산 연산과 가중치 저장을 동시에 수행할 수 있게 해준다.



Figure 1.1 MAC 연산 가속용 Crossbar Array 구조 예

또한, 인공 신경망 시냅스는 학습 동작과 추론 동작에서 각각 그 요구사항이 다르다. 학습을 위한 소자는 비 휘발성 보다 연산 정확도 구현이 보다 중요하다. 반면 추론을 위한 소자는 학습된 가중치를 잘 보존하는 것이 보다 중요하다. 이런 연산 기능 수행과의 관계에 따른 소자 요구사항의 변화는 현재 널리 활용되는 폰노이만 기반 컴퓨팅 체계의 메모리 소자에서도 유사하게 나타난다. 연산 장치와 가장 가까운 소자로 사용되는 SRAM과 더불어 프로그램을 보조 기억장치에서 가져와 주 기억장치로 사용되는 DRAM 등은 휘발성이지만, 빠른 속도와 넓은 대역폭을 특징으로 하고 있다. 반면 보조 기억장치는 연산이 완료된 데이터를 잘 보존하는 것이 중요하고, 대용량의 데이터를 저장하는 것이 중요하다. 이와 같이 현재의 메모리 소자도 연산장치와 상호작용하기 위해 그 요구사항을 달리하며 계층구조를 가진다.^[2]

마찬가지로 인공신경망 학습에 필요한 시냅스 소자는 연산 기능을 주요하게 수행해야 하기 때문에 다음의 추가적인 요구사항들을 가진다.^[3]

Synapse Device Requirement for DNN Training

- 1. Analog multi-level states
- 2. On-off Ratio
- 3. Linearity
- 4. Symmetry
- 5. Endurance

1.2 CMOS 기반 시냅스 소자의 장점과 한계

앞서 소개한 대규모로 병렬화 된 Crossbar Array에 사용될 시냅스 소자를 구현하기 위해 다양한 저항 변화형 소자를 활용한 연구들이 이루어져왔다.^{[4][5]} 그러나 대부분의 저항변화형 소자들은 비 휘발성이고, 작은 소자 크기를 구현하는데 적합한 특성을 가지지만, 선형적이고 대칭적 업데이트 특성을 구현하기 어렵거나, 다단계 구현이 어렵고 소자간 산포가 지수적으로 분포한다. 이것은 결국 곱셈과 누산 연산의 정밀도를 떨어뜨리게 되어 학습 정확도의 손실을 가져오는 원인이 된다.

한편 인공 신경망 학습에 필요한 요구사항들을 만족하는 시냅스 소자 구조로 CMOS 기반의 전하저장형 시냅스 소자인 3T-1C, 2T-2R 구조가 알려져 있다.^[6] 기존에 널리 연구되던 저항변화 기반의 Phase-Change Memory^[7], RRAM^[8], CBRAM^[9] 등의 소자들과는 달리 CMOS 기반의 3T-1C 구조는 인공신경망 학습에 필요한 선형적이고, 대칭적이며, 점진적 다단계 업데이트 특성을 탁월한 수준으로 달성하여 Software에 필적하는 학습 정확도를 보인다.

그러나 전하 저장형 시냅스 소자는 CMOS 트랜지스터를 포함한 다양한 누설전류로 학습한 가중치를 잃어버리게 되면 학습 정확도 저하가 발생하게 된다. 충분한 정전용량을 가지는 커패시터를 제작하여 이 문제를 해결할 수 있지만, CMOS 트랜지스터의 누설전류 수준에서는 필요한 커패시터의 면적이 지나치게 커지게 되어 시냅스 소자 크기를 합리적으로 줄이기 어렵게 만든다.

한편, 학습 정확도 저하가 없는 CMOS 기반 3T-1C 구조 시냅스 가중치 Retention 요구사항은 Fully-connected network기준으로 학습 주기의 10⁶배 이상 수준으로 연구되어 있고, 가중치 공유를 사용하는 컨볼루션 인공 신경망 (CNN)의 경우 추가로 600배 더 큰 수준으로 알려져 있다. 이 것은 학습 주기를 200ns로 가정했을 때 0.2s~120s 수준이다.^[10]

요약하면, 학습 정확도 손실이 없는 Retention 요구사항을 만족하는 전하저장형 소자를 합리적인 소자 크기로 구현하기 위해서는 충분히 작은 트랜지스터 누설전류를 바탕으로 커패시터의 면적 축소가 필요하다.

2. 문헌연구

2.1 IGZO TFT 소자

In-Ga-Zn-Oxide Thin Film Transistor (IGZO TFT)는 충분한 Mobility와 매우 낮은 누설전류를 바탕으로 디스플레이 구동회로의 Switch 트랜지스터나 Driving 트랜지스터로서 활용되고 상용화되었다.^{[11][12]} 밴드갭이 실리콘 (Eg ~1.1eV) 대비 큰 IGZO (Eg ~3.2eV)는 Channel Accumulation-mode로 작동하며, Hole tunneling current가 거의 없고 Hole 유효질량이 커 intrinsic NMOS로 활용이 적합하다.^{[13][14]}

그러나 IGZO TFT를 CMOS와 같이 dopant를 변경하여 PMOS로 활용하는 것은 inversion channel이 아닌 intrinsic accumulation channel로 작동하는 재료적 성질 때문에 어렵다. 따라서 IGZO TFT를 디스플레이 등에서 활용하는 경우 NMOS만으로 구성된 구동회로를 설계하거나 IGZO TFT의 장점인 매우 낮은 누설전류 특성이 필요한 Switch 트랜지스터로 활용하고 있다.

반면 Logic^{[15][16][17]} 이나 메모리 소자로 IGZO TFT를 활용하는 다양한 연구들이 있다. 특히 DRAM과 연결된 Switch 트랜지스터를 IGZO TFT로 제작하는 경우 수일 이상 Refresh 동작을 요구하지 않는 수준의 높은 Retention을 보인다.^[18] 이번 연구의 목표 중 하나인 트랜지스터 누설전류를 효과적으로 제어하여 커패시터 소자 크기를 축소하면서도 학습 정확도의 손실없는 충분한 Retention 수준의 시냅스 소자를 개발하는데 IGZO TFT가 적합한 후보 물질이라고 할 수 있다.

또한 IGZO는 산화물 반도체의 한 종류로 수소와 산소결함 농도에 따라 그 특성이 매우 민감하다.^{[19][20][21][22]} 이 것은 IGZO 증착 이후 후공정에서 발생하는 수소에도 민감한 상호작용을 하고 있어 IGZO TFT 제작 공정이 주로 BEOL단에서 이루어지게 되는 한 원인이다.^[23] IGZO 증착 후 Passivation 막질 개발을 통해 이를 방지하고자 하는 다양한 연구들이 이루어지고 있다.^[24]

4

2.2 커패시터 소자

Atomic Layer Deposition (ALD) 공정을 이용한 High-κ dielectric을 Insulator로 활용하는 Metal-Insulator-Metal (MIM) 커패시터 소자는 DRAM Cell 등에 널리 활용되고 있다.^{[25][26]} 커패시터의 Electrode 역할을 하는 Metal 물질과 Insulator 역할을 하는 High-κ dielectric의 재료적 특성 및 제작한 커패시터의 구조, 커패시터 양단에 가해진 전압 조건 등에 따라 커패시터가 저장가능한 전하량과 누설전류가 결정된다.

$$C = \varepsilon_0 \varepsilon_r \frac{A}{d}$$
(1)
$$Q = CV$$
(2)

이때 커패시터에 저장한 전하를 잘 보존하는 Retention 특성을 개선하기 위해서는 insulator의 dielectric constant를 증가시키면서도 다양한 누설전류 원인 매커니즘을 제어할 수 있는 높은 Energy Gap을 가지는 물질을 제작하는 것이 필요하다. 하지만 이것은 일반적으로 반비례 관계에 놓여있어 낮은 누설전류와 높은 정전용량을 동시에 달성하는 커패시터 insulator를 제작하기 위해 다양한 재료적 연구들이 이루어지고 있다.^{[27][28]}

이번 연구에서는 트랜지스터 누설전류를 IGZO TFT로 제어한 시냅스 제작이 주된 연구목표로 커패시터 insulator 막질은 ALD 공정을 활용하여 10nm thickness의 hafnium oxide를 활용했다. 하지만, 학습 정확도 손실없는 전하저장형 시냅스의 Retention 요구 수준인 200ns 학습 주기 기준 0.2s~120s를 달성하기 위해서는 커패시터 누설전류의 효과적 제어와 높은 정전용량 달성을 위한 재료적 연구와 개선이 반드시 필요하다.

3. 실험 및 분석 방법

3.1 IGZO TFT 소자 제작

이번 연구에서는 IGZO TFT의 탁월하게 낮은 누설전류 특성을 활용하여 가중치 업데이트 동작을 수행하는 전하 저장형 시냅스를 제작하였다. IGZO TFT는 Top-Gate Staggered 구조로 제작하였다. (Figure 3.1.1) 그리고 Amorphous IGZO 증착을 위해 In:Ga:Zn = 1:1:1로 구성된 Target을 장착한 Sputtering 설비를 활용해 1Pa Ar100sccm 분위기에서 2.44W/cm RF Bias로 Sputtering 하였다.

IGZO TFT 소자 제작 공정

- 1. Silicon Substrate Oxidation: Wet Oxidation 5000 Å
- 2. Tungsten 200Å deposition
- 3. Source, drain patterning & dry etching
- 4. Channel deposition: Amorphous IGZO 10nm Sputtering
- 5. Channel IGZO Patterning & wet etching
- 6. Gate insulator deposition: hafnium oxide 10nm ALD
- 7. Tungsten 300Å Deposition
- 8. Gate electrode patterning & dry etching

이때, IGZO의 수소 농도를 효과적으로 제어하기 위해 IGZO TFT는 후술할 커패시터 소자를 먼저 제작한 후에 제작하였다. 이는 TMA(trimethylaluminum) 기반 ALD(Atomic Layer Deposition) 공정 과정에서 발생하는 수소에 의해 IGZO 채널 특성이 영향을 받아 IGZO TFT의 특성이 열화 되는 것을 방지하기 위함이다. IGZO Channel에 과다한 수소가 주입되는 경우 electron carrier가 증가하여 Conducting한 특성을 나타내게 되므로, 이를 효과적으로 조절하는 공정 조건 확보가 중요하다. (Figure 3.1.2)



Figure 3.1.1 | IGZO TFT 소자 단면도



Figure 3.1.2 | IGZO TFT 소자 특성

3.2 커패시터 소자 구조

전하 저장형 시냅스 소자의 가중치는 IGZO TFT를 통해 커패시터에 저장된다. 이번 연구에서는 IGZO TFT를 활용한 새로운 구조의 시냅스의 동작 검증에 문제가 없는 수준의 High Capacitance 확보를 위해 High- κ dielectric을 활용한 Metal-Insulator-Metal (MIM) 커패시터를 제작하였다. (Figure 3.2.1) High- κ dielectric은 TMA기반 Source와 O3 Reactant를 활용해 250℃에서 증착하였다.

그리고 다양한 크기의 커패시터 소자를 제작하여 소자 크기별 누설전류 수준과 정전용량을 측정할 수 있도록 하였다. 제작한 커패시터 소자는 um단위에서 1x1, 2x2, 5x5, 10x10, 25x25, 50x50, 100x100, 200x200 등이다.

커패시터 소자 제작 공정

- 1. Silicon 기판 상부 Oxidation: Wet Oxidation 5000Å
- 2. Tungsten 200Å deposition
- 3. Bottom electrode Patterning & dry etching
- 4. Gate insulator deposition: hafnium oxide 10nm ALD
- 5. Tungsten 300 Å Deposition
- 6. Top electrode patterning & dry etching



Figure 3.2.1 | 커패시터 소자 단면도

3.3 시냅스 소자 제작

3.3.1 CMOS 기반 시냅스 소자와의 비교

전하 저장형 시냅스 소자의 누설전류를 제어하기 위해 CMOS 트랜지스터 대신 IGZO TFT를 사용하기 위해서는 시냅스 소자 회로 구조의 변경이 필요하다. 먼저 CMOS 3T-1C 구조의 PMOS를 단순히 IGZO TFT로 변경 (Figure 3.3.1.1)하는 경우를 생각해보자.

IGZO TFT는 Intrinsic NMOS로 동작하므로, Silicon 기반 FET처럼 Dopant를 활용하여 Hole carrier를 운반하는 PMOS로 구성하여 활용하기 어렵다. 그러므로 3T-1C 시냅스에서 PMOS가 자리했던 경우 시냅스 가중치에 해당하는 노드 전압은 Drain전압으로 활용되고 PMOS의 Source 전압은 고정되었던 것과는 달리 IGZO TFT는 NMOS와 같은 특성을 가지고 동작하기 때문에 가중치 노드를 트랜지스터의 Drain이 아니라 Source로 사용하게 된다.

다음으로 고정된 1.5V Drain 전압에서 Source 노드 전압을 증가시킴에 따라 Gate 전압에 따른 Drain 전류의 IV 특성을 확인해보자. (Figure 3.3.1.2) 가중치 업데이트 동작 상황인 동일한 Gate 전압을 가했을 때 Drain 전류의 거동을 보면 Source 전압의 변화에 따라 Drain 전류의 변화폭이 심하고 심지어 Drain 전압과의 관계에 따라 전류의 방향도 바뀌는 것을 확인할 수 있다.

결국, 동일한 Gate 전압으로 업데이트를 수행하여야 하는 시냅스의 동작조건에서는 업데이트된 가중치 수준에 따라 Source 전압이 변화하게 되면 업데이트 전류인 IGZO TFT의 On-Current의 변화폭이 매우 커져 균일한 가중치 업데이트에 실패하고 수회의 단방향 업데이트 만으로도 가중치 업데이트 량이 줄어들어 수렴하게 된다. (Figure 3.3.1.3) 이것은 가중치 Potentiation 업데이트 동작과 Depression 업데이트 동작 간의 차이점을 가져와 선형적이고 대칭적인 업데이트 구현을 어렵게 만든다.

11



Figure 3.3.1.1 | 3T-1C 시냅스 회로 구조



Figure 3.3.1.2 | IGZO TFT Transfer 특성 : Drain 전압을 1.5V로 고정하고 Source 전압을 변화



Figure 3.3.1.3 | IGZO TFT를 활용한 시냅스의 Potentiation 동작

3.3.2 IGZO TFT 활용 시냅스 소자 회로 구조

3절에서 상술한 이유로 IGZO TFT를 활용하여 시냅스 가중치를 선형적이고 대칭적으로 업데이트 하기 위해서는 Potentiation 동작과 Depression 동작 모두에서 트랜지스터 On-Current 량이 균일하게 제어가 가능한 구조가 필요하다. 앞서 3T-1C 구조에서 Potentiation 동작에서 Source 노드 전압이 고정되지 않는 문제를 해결하기 위해 Potentiation과 Depression 동작 모두 트랜지스터의 Source 노드가 고정되는 아래와 같은 5T-1C 시냅스 구조를 살펴보자. (Figure 3.3.2.1)

Potentiation 동작시 먼저 N1 트랜지스터의 Gate에 V_{DD} 전압을 인가하고 N2 트랜지스터의 Gate에 적절한 V_{on} pulse 전압을 가해 커패시터 하단 노드 전압과 Ground 전압 간의 차이를 기준으로 N2 트랜지스터에서 Pulse 전류를 흘려준다. 이로써 커패시터 양단 전위차를 양의 방향으로 증가시킬 수 있다.

반대로 Depression 동작시 먼저 N3 트랜지스터의 Gate에 V_{DD} 전압을 인가하고 N4 트랜지스터의 Gate에 적절한 V_{on} pulse 전압을 가해 커패시터 상단 노드 전압과 Ground 전압 간의 차이를 기준으로 N4 트랜지스터에서 Pulse 전류를 흘려준다. 이것은 커패시터 양단 전위차를 음의 방향으로 증가시킬 수 있다.

공통적으로 N1, N3 트랜지스터의 경우 각각 Potentiation과 Depression 동작이 가능하도록 커패시터 한쪽 노드 전압의 기준점을 형성하는 업데이트 Enable의 역할을 담당한다. N2, N4 트랜지스터는 업데이트 Enable 상태에서 Gate Pulse 전압을 받아 Pulse 전류를 커패시터 노드에서 Ground 노드로 흘려 커패시터 양단에 전위차가 변경되도록 하는 업데이트 역할을 담당한다.

이 구조는 3T-1C 구조와 비교했을 때 결과적으로 Potentiation과 Depression 동작에서 업데이트 역할을 담당하는 N2, N4 트랜지스터의 Source를 Ground 전압으로 고정시켜주므로 보다 균일한 업데이트 전류를 형성할 수 있게 해준다.

15

고정된 Source 노드 전압에서 Drain 노드 전압을 증가시킴에 따라 Gate 전압에 따른 Drain 전류의 전형적인 IV 특성을 확인해보자. (Figure 3.3.2.2) 고정된 Gate Bias를 가했을 때 포화 영역에서 Drain 전압의 변화에도 Drain 전류를 일정한 수준으로 안정적으로 가할 수 있음을 알 수 있다.



Figure 3.3.2.1 | IGZO TFT를 활용한 5T-1C 시냅스 회로 구조



Figure 3.3.2.2 | Source 노드 전압이 고정된 IGZO TFT 특성 a. Transfer 특성 b. Output 특성

3.4 소자 레이아웃 및 공정 프로세스 디자인

이번 연구에서는 삼성종합기술원이 보유 중인 8inch Stepper 노광 장비를 활용하여 Photo Mask를 제작했다. 연구 목표에 맞도록 소자 레이아웃 단계 부터 반영하였다. 그리고 소자 제작에 필요한 공정 순서를 고려하여 Layer별로 역할을 부여하였다.

주요 연구 목표

1. IGZO TFT를 활용한 전하 저장형 시냅스 소자 동작 검증 2. IGZO TFT 누설 전류 검증

Photo Layer 및 Process별 구분

[4Layer] : 26 Step

: IGZO Gate oxide와 커패시터 insulator를 동시 형성

- 1. Source / Drain Metal Line
- 2. Active IGZO
- 3. VIA for Top gate oxide remove
- 4. Top Gate Metal Line

[8Layer]: 52 Step

: IGZO Gate oxide와 커패시터 insulator를 분리 형성

- 1. Bottom electrode Metal Line
- 2. VIA1 for Capacitor insulator remove
- 3. Top electrode Metal Line
- 4. VIA2 for Bottom gate oxide remove
- 5. Source / Drain Metal Line
- 6. Active IGZO
- 7. VIA3 for Top gate oxide remove
- 8. Top Gate Metal Line





Figure 3.4.1 | Retention > 0.2s 확보 가능 커패시터 면적



Figure 3.4.2 | Step수 > 1000 확보 가능 커패시터 면적



Figure 3.4.3 | 트랜지스터 및 커패시터 누설전류 검증 회로





Figure 3.4.4 | 누설전류 측정용 Jumbo FET



Figure 3.4.5 | 4Layer 공정 프로세스



Figure 3.4.6 | 8Layer 공정 프로세스

4. 결과 및 논의

4.1 IGZO TFT 기반 전하 저장형 시냅스 소자 동작특성

4.1.1 기초 소자 특성

앞서 기술한 공정을 진행하여 IGZO TFT와 커패시터를 제작하였고, 측정을 통해 확보된 IGZO TFT 특성은 다음과 같다. 산포는 웨이퍼내 측정 결과를 바탕으로 도출된 결과로 Chip내 산포는 더 줄어들 수 있다.

확보된 소자 특성에서 Threshold voltage 값이 음의 값을 가지는 것에 주목해보면, IGZO TFT를 Turn-off 시켜 누설전류를 충분히 제어하기 위해서는 더 많은 음의 전압을 가해야 함을 알 수 있다. 그 외 다른 특성은 일반적인 CMOS 소자의 논리를 대부분 준용할 수 있고, 이번 연구의 시냅스 소자 동작 관점에서 필요한 논의는 4.2 절에서 다루기로 한다.

IGZO 특성 파라미터 계산 방법

Subthreshold Swing : $SS = \min(\frac{\partial V_G}{\partial \log I_d})$ Threshold voltage : $V_{th} = V_{gs} @I_d = 1 \times 10^{-11}A$ Gm max : $G_m \max = \max(\frac{\partial I_d}{\partial V_G})$ Mobility : $\mu_{FE}^{sat} = \frac{2L}{WCox} \left(\frac{\partial \sqrt{I_d}}{\partial V_g}\right)^2$, $\mu_{FE}^{lin} = \frac{L}{WCox} \frac{\partial I_d}{\partial V_g} \frac{1}{V_d}$ DIBL : DIBL = $\frac{\partial V_{th}}{\partial V_d} = \frac{V_{th} (V_d = 0.1V) - V_{th} (V_d = 1.5V)}{0.1 - 1.5V}$

IGZO TFT 2/5	Vds=0.1[V]	Vds=1.5[V]	
Subthreshold	Median	117.7	112.5
Swing [mV/dec]	Variation [3 ₀]	11.9	14.0
Threshold	Median	-0.93	-0.81
voltage [V]	Variation [3 ₀]	0.54	0.54
gm max	Median	2.8.E-07	3.9.E-06
[S]	Variation [3 ₀]	4.7.E-08	5.3.E-07
mobility	Median	4.76	4.38
[cm^2/Vs]	Variation [3 ₀]	0.79	0.54
DIRI	Median	-	0.09
	Variation [3σ]	-	0.06

Table 4.1.1.1 | IGZO TFT wafer 특성

4.1.2 Off leakage current 특성

A. 2T 회로를 활용한 Leakage 측정

IGZO TFT의 탁월하게 낮은 Off leakage current는 측정장비의 한계로 인해 수십~수백 fA 수준 이하를 측정하기 어렵다.^[29] 따라서 측정을 위한 회로를 별도로 구성하여 측정하였다. (Figure 4.1.2.1)

Leakage current 측정용 회로는 Program 트랜지스터의 Source 노드와 Read 트랜지스터의 Gate 노드를 연결한 것을 특징으로 한다. Leakage current 측정 회로의 동작 방법은 다음과 같다. (Figure 4.1.2.2)

i. Read 트랜지스터 선형 구간 확인

Program 트랜지스터의 Gate 전압으로 3V를 인가한 상태에서 Drain 전압을 -3~3V로 Sweep하여 Read 트랜지스터의 Gate 전압에 따른 Drain 전류를 먼저 확인한다. Read 트랜지스터의 포화 영역 구간인 0~2V 구간에서 매우 선형적으로 거동함을 확인할 수 있다. (Figure 4.1.2.3)

후술할 Leakage 확인 과정에서 Storage 노드 전압을 1.5V로 Program후 0V를 향해 서서히 내려가는 것을 Sensing 하는 것을 감안하면, 이는 Storage 노드 전압을 Read 트랜지스터의 Drain 전류를 통해 추정하기 충분한 수준으로 매우 정확한 선형 구간이 확보되었음을 의미한다.

ii. Program 트랜지스터 Gate Off 전압에 따른 Leakage 확인

먼저 Program 트랜지스터의 Drain 전압을 1.5V로 Gate 전압을 3V로 수 초간 인가하여 Storage 노드에 양전하를 주입해 전압을 1.5V 인근까지 충분히 올려준다. (Figure 4.1.2.2) 이후 Leakage 수준을 측정할 Program 트랜지스터의 Gate 전압을 적절한 Off 전압으로 변경한다. 잠시 후 Program 트랜지스터의 Drain 노드 전압을 0V로
변경한 후 Storage 노드 전압에 의해 변화하는 Read 트랜지스터의 Drain 전류를 Sensing하여 Storage 전압의 시간에 따른 변화를 계산한다. (Figure 4.1.2.2) 여기서 계산되는 Leakage current는 시간에 따른 Storage 노드 전압의 변화와 Capacitance와의 곱으로 생각할 수 있다. (수식 3, 4)

$$I_{leakage} = I_{off.Program} + I_{leak.Read} = C_{Storage} \times \frac{dV_{Storage}}{dt}$$
(3)
$$C_{Storage} = C_{Read.Tr} + C_{Program.Tr}$$
(4)

충분한 off 전압을 Program 트랜지스터에 가하는 경우 Storage 노드 전압이 시간에 따라 떨어지는 속도가 느려져 기울기가 완만해진다. (Figure 4.1.2.5) 이 것은 Program 트랜지스터의 누설전류가 충분한 off 전압에 의해 효과적으로 제어되어 Storage 노드의 Charge를 보존하는 것으로 생각된다. 이 것을 바탕으로 제작한 트랜지스터의 Capacitance를 측정하여 누설전류로 환산해보면 수백 aA(1aA = 10⁻¹⁸ A) 수준의 낮은 누설전류를 가짐을 확인할 수 있다. (Figure 4.1.2.6, 수식 3)



Figure 4.1.2.1 | Leakage current 측정용 2T 회로



Figure 4.1.2.2 | Leakage current 측정용 2T 회로 동작 방법 a. Storage node Program 동작 b. Storage node Retention 동작



Figure 4.1.2.3 | Read 트랜지스터 Drain 전류 선형 구간



Figure 4.1.2.4 | Leakage current 측정 Timing Diagram



Figure 4.1.2.5 | Program 트랜지스터 Gate Off 전압에 따른 Storage 전압 변화



Figure 4.1.2.6 | Program 트랜지스터 Gate Off 전압에 따른 Leakage 환산

B. Jumbo FET (Width 10000 µm 트랜지스터) 측정

Width 10µm, Length 0.5µm의 트랜지스터 1000개를 병렬 연결하여 누설 전류를 증폭해 측정 장비 한계 수준 이상에서 직접 측정해보았다.

Width 2µm, Length 0.5µm 트랜지스터 대비 높은 Off current값이 측정된 것으로 보아 Width 10µm, Length 0.5µm 트랜지스터 1000개 병렬연결 유닛에서 측정된 누설전류를 실제 누설전류로 생각할 수 있다.

이 것을 바탕으로 Width 2µm로 환산해보면, 수십~수백 aA(1aA = 10⁻¹⁸ A) 수준의 누설전류를 보일 것으로 예상되고, 이는 앞서 2T 회로에서 예측한 수준과 수십~수백 aA 정도의 근소한 차이를 보이며 다소 낮다. 2T 회로의 경우 Read 트랜지스터의 Gate 누설전류와 Program 트랜지스터의 Drain 누설전류의 합으로 누설전류가 계산되는 것에 반해 Width 10µm, Length 0.5µm 1000개를 병렬 연결해 직접 측정하는 경우 Drain 누설전류만 측정된다. 따라서 이 차이는 Read 트랜지스터의 Gate 누설전류에 의한 것으로 판단된다. (Figure 4.1.2.7)

한편 Width 10µm, Length 0.5µm 트랜지스터 1000개를 연결한 Oncurrent 수준이 Width 2µm, Length 0.5µm 트랜지스터에서 측정된 수준의 134배로 Layout상의 Width 차이인 5000배 대비 낮은 원인은 500Ω 수준의 측정회로 배선 저항에 의한 것으로 판단된다.



Figure 4.1.2.7 | Jumbo FET Transfer curve (Drain 0.1V) : Width 10, Length 0.5, m 1000개 병렬 연결

4.1.3 Potentiation update

인공 신경망의 시냅스 가중치 연결 강도를 강하게 하는 방향으로 시냅스 소자에 양전하를 저장하여 커패시터 전압을 Positive update 하는 것을 Potentiation update라 한다.

5T-1C 시냅스 소자 구조에서는 Potentiation 동작이 가능하게 하기 위해 N1 트랜지스터의 Gate에 전압을 가해 Turn-on시켜 커패시터 상단 노드 전압을 $V_{DD}/_2$ 로 Boost up시키고, (Potentiation Enable) 이와 Coupling 된 커패시터 하단 노드 전압을 Drain으로 하여 Source가 Ground 전압과 연결된 N2 트랜지스터의 Gate에 적절한 Pulse 전압을 가해 Turn-on 시켜 커패시터 하단 노드에 전자를 주입해 커패시터 양단에 전위차를 양의 방향으로 증가시키게 된다. (Potentiation, 수식 5, Figure 4.1.3.1)

$$\Delta V_{Capacitor.Bottom} = -\frac{I_{on,N2} \times t_{pulse,N2}}{C_{Capacitor}}$$

$$\Delta V_{Capacitor} = V_{Capacitor.Top} - V_{Capacitor.Bottom}$$
(5)

2장에서 상술한대로 기존의 CMOS 트랜지스터 기반 3T-1C 대비 IGZO TFT 기반 5T-1C 구조는 PMOS 역할이 가능한 트랜지스터 없이 NMOS 역할만 가능한 트랜지스터로 시냅스 소자 구성이 필요하다. 선형적이고 점진적인 업데이트 특성 확보를 위해 NMOS 트랜지스터의 source가 고정 전압일 필요가 있고, 5T-1C 구조에서 N1 트랜지스터는 커패시터 하단 노드 전압을 Ground 보다 높거나 같아지도록 하는 역할을 수행하여, N2 트랜지스터가 가중치 update를 Gate Pulse 전압에 따라 일정하게 수행할 수 있도록 돕는다.



Figure 4.1.3.1 | 5T-1C 시냅스 소자 Potentiation update

4.1.4 Depression update

인공 신경망의 시냅스 가중치 연결 강도를 약하게 하는 방향으로 시냅스 소자에 음전하를 저장하여 커패시터 전압을 Negative update 하는 것을 Depression update라 한다.

5T-1C 시냅스 소자 구조에서는 커패시터의 상단 노드와 연결된 N4 트랜지스터 Gate에 적절한 Pulse 전압을 가해 Turn-on 시켜 커패시터 상단 노드에 전자를 주입해 커패시터 양단에 전위차를 음의 방향으로 증가시키게 된다. (Depression, 수식 7, Figure 4.1.3.1)

$$\Delta V_{Capacitor.Top} = -\frac{I_{on,N4} \times t_{pulse,N4}}{C_{Capacitor}}$$
(7)

이 때, N3 트랜지스터를 켜서 커패시터의 하단 노드 전압을 $V_{DD}/_2$ 로 Boost up시켜 커패시터의 상단 노드 전압이 Ground 보다 낮아지지 않도록 한다. (Depression Enable)

3T-1C와 비교하면 커패시터 상단 노드와 Ground 노드 사이에 연결된 N4 트랜지스터를 활용하여 Pulse 전류를 통해 업데이트를 진행하는 점은 동일하다. 하지만, N3 트랜지스터를 활용한 Depression Enable 작동이 커패시터 하단 노드 전압을 $V_{DD}/_2$ 로 Boost up하기 때문에 $\Delta V_{capacitor}$ 가 $-V_{DD}/_2$ 까지 이론적으로 Depression이 가능하다는 점에서 3T-1C 구조와는 차이를 보인다.



Figure 4.1.4.1 | 5T-1C 시냅스 소자 Depression update

4.1.5 시냅스 소자 가중치 읽기 동작

5T-1C 시냅스 소자의 읽기 동작은 커패시터의 한쪽 노드가 Ground로 고정된 3T-1C와 구별된다. 커패시터 상단 노드와 Read 트랜지스터의 Gate가 연결되어 있고, Read 트랜지스터의 Drain에 0.1V 정도의 낮은 전압을 가하여 커패시터 양단 전위차에 따라 달라지는 Read 트랜지스터의 Drain 전류를 만들어낸다. 이때 N3 트랜지스터의 Gate에 전압을 가해 $V_{DD}/_2$ 전압을 커패시터 하단 노드에 가해주어 Read 트랜지스터가 포화 영역 구간에서 작동하여 Read 전류가 Gate 전압에 따라 선형적으로 거동할 수 있도록 한다. (Figure 4.1.5.1)

단일 시냅스에서의 읽기 동작에서는 Read 전류가 커패시터 가중치 전압 영역 $+ \frac{V_{DD}}{2}$ 에서 선형적으로 거동하는 것이 Read 트랜지스터에 의한 추가적인 왜곡없이 가중치 업데이트 선형성을 보존하는 핵심 요구사항이다.

그러나 실제 Array 구조에서는 Read전류는 Array Row 개수의 합으로 더해지므로 최대한 개별 시냅스의 가중치를 읽어내는데 필요한 전류의 량을 줄이는 것이 필요하다. 제작한 IGZO TFT의 Threshold voltage가 음의 값을 가지므로 읽기 동작에서는 커패시터 하단 노드에 가해주는 전압을 $V_{DD/2}$ 보다 다소 낮게 조정해주거나 Read 트랜지스터의 Threshold voltage를 양의 값으로 조정해주는 Dual gate 공정 트랜지스터 제작 등 추가적인 읽기 동작 개선이 필요할 것으로 예상된다.

41



Figure 4.1.5.1 | 5T-1C 시냅스 소자 가중치 읽기 동작

4.2 IGZO TFT 기반 전하저장형 시냅스 가중치 거동

4.2.1 Update 특성

앞서 소개한 5T-1C 시냅스 업데이트 동작을 검증하기 위해 IGZO TFT와 커패시터를 연결하여 실제 wafer로 제작하여 측정하였다.

이때 Arduino MCU Board를 활용하여 Bit line (BL) 노드와 연결된 별도의 커패시터를 방전시켜 Read 전류를 ADC (Analog to Digital Converter)를 통해 Sensing 하였다. (Figure 4.2.1.1) ADC 회로와 Arduino Board 구성은 공동 연구로 고려대학교 전기전자공학부 이형민 교수님 연구실에서 제작되었다. Read 전류에 따른 MCU의 ADC 값을 대응시키고 Read 전류와 시냅스 커패시터 전압과 연결되는 Read 트랜지스터 Gate 전압을 서로 대응시켜 시냅스 가중치 업데이트에 따른 커패시터 전압을 산출했다.

이를 통해 Potentiation 업데이트 50회 후 방향을 바꾸어 Depression 업데이트 50회를 15세트 반복하여 50Step 업데이트 반복 상황의 시냅스 거동을 측정했다. (Figure 4.2.1.2) 측정 조건은 2µs Pulse width의 OV 전압을 업데이트 동작에서 N2나 N4의 Gate 노드에 인가했다. 그 결과 Potentiation과 Depression 동작이 정상적으로 진행됨을 확인했다.

추가로 Multi-level Step 구현을 위하여 동일 업데이트 전압에서 2µs 미만의 pulse로 업데이트를 진행하지는 못 하였다. 해당 원인은 Arduino MCU의 한계로 pulse width를 줄이기는 어려웠기 때문이다. 대안으로 업데이트 전압을 추가로 낮춰 업데이트 전류를 줄이는 방법으로 100Step 업데이트를 구현했다. (Figure 4.2.1.3, Figure 4.2.1.4)

업데이트 세트를 반복할수록 업데이트 구간에서 평형점을 찾아 동일 구간에서 Potentiation과 Depression을 반복하는 것을 확인할 수 있다. 또한 시냅스 가중치 전압 구간을 좁게 형성할수록 선형성을 극대화할 수 있음을 알 수 있다. 추가로 업데이트 Pulse width를 수십ns 수준까지 줄인다면 동일 가중치 업데이트 전압에서도 더 많은 Step을 구현할 수 있을 것이다.

업데이트시 N1 또는 N3를 통해 Enable 상태에 돌입하면 커패시터 전압은 시냅스 가중치에 따라 0~3V 사이에 위치하게 된다. 이때 커패시터 전압은 N2 또는 N4 IGZO TFT의 Drain전압으로 작용한다. IGZO TFT의 Output 특성을 바탕으로 시냅스 가중치 전압에 따른 업데이트 전류를 최대치와의 비율로 환산할 수 있다. (Figure 4.2.1.5) 제작한 IGZO TFT의 특성에 맞는 적절한 업데이트 전압 선택을 통해 선형성 및 대칭성 확보 구간을 최대화할 수 있고, 그 결과 업데이트 전류 관점에서 2V Range에서도 우수한 업데이트 특성을 예상할 수 있다.

위 수식 8을 살펴보면 시냅스 캐패시터의 용량이 클수록, 업데이트에 필요한 전류량이 작을수록, 업데이트 전류 펄스 폭이 작을수록 시냅스 가중치 레벨 개수를 늘릴 수 있다. 하지만, 시냅스 캐패시터, IGZO TFT On-current, 짧은 펄스 전압인가 모두 각기 다른 제한 요소들을 가지고 있다. 시냅스 캐패시터는 두께가 얇을수록 정전용량을 키울 수 있지만, 너무 얇은 insulator는 tunneling current에 취약해져 누설전류가 높아진다. IGZO TFT의 경우 On-current를 조절하기 위해서 적절한 mobility를 가지는 트랜지스터를 Subthreshold swing 및 Threshold voltage의 열화없이 제작하여야 한다. 짧은 펄스 전압은 현재 수GHz (1GHz=1ns) 수준의 Logic CPU가 상용화 되어있으나 발열 등의 문제로 더 이상 빠르게 스위칭 동작을 수행하기 어렵다.



Figure 4.2.1.1 | Read ADC 회로 개념도



: 2us Pulse width, 0V 50 Step Update



Figure 4.2.1.3 | 5T-1C 시냅스 100 Step 업데이트 거동 : 2us Pulse width, -1V 100 Step Update, 15Set Repeat





4.2.2 Retention 특성

앞서 IGZO TFT의 누설전류 수준을 2T 회로를 통해 측정할 수 있음을 보였다. 이번 연구에서 제작한 시냅스 트랜지스터는 Width 2µm, Length 5µm이고, 해당 크기의 트랜지스터를 동일한 방법으로 측정하여 468 aA(4.68×10⁻¹⁸A) 수준의 누설전류를 확인했다.

한편 시냅스 커패시터의 누설전류 또한 시냅스 가중치 Retention에 중요하다. 따라서 2T-1C 회로를 제작하여 커패시터 크기별로 커패시터의 누설전류를 측정하였다. (Figure 4.2.2.1) 2T 회로의 Storage 노드에 커패시터를 연결하였고, 동작 방식은 2T 회로와 동일하다.

 $I_{leakage} = I_{leak.Transistor} + I_{leak.Capacitor} = C_{Storage} \times \frac{dV_{Storage}}{dt}$ (9) $I_{leakage.Transistor} = I_{off.Program.Transistor} + I_{leak.Read.Transistor}$ (10) $C_{Storage} = C_{Transistor} + C_{Capacitor}$ (11) $C_{Transistor} = C_{Read.Transistor} + C_{Program.Transistor}$ (12)

커패시터 면적이 충분히 커짐에 따라 전체 누설전류는 트랜지스터의 누설전류 변동에는 둔감해지고 커패시터의 누설전류가 지배하게 된다. 2T-1C 측정결과를 바탕으로 계산한 커패시터의 누설전류는 0.6aA/ um² 수준에 근접함을 알 수 있다. (Figure 4.2.2.2)

가중치 1Step update 전압량에 해당하는 3mV가 RC시정수 *t* 시간만큼 자연 감쇄하는 경우를 가정하면 약 1.9mV가 소실된다. 해당 전압량을 누설전류에 의해 소실하는데 걸리는 시간을 Retention time으로 정의하기로 한다. (수식 13, 14) Retention time은 시냅스 누설전류에 대한 정전용량의 비율과 비례하게 거동한다. (수식 15) $\Delta Q = I_{leakage} \times t_{retention} = C_{synapse} \times \Delta V_{step.retention}$ (13)

 $\Delta V_{step.retention} = V_{step} \times (1 - 1/e) \approx 0.003 \times 0.6321 = 0.0019 \quad (14)$ $t_{retention} = (C_{synapse}/I_{leakage}) \times \Delta V_{step.retention} \quad (15)$

제작한 커패시터의 정전용량은 C-V 측정으로 쉽게 얻을 수 있고, 시냅스 누설전류 또한 앞서 측정회로를 통해 IGZO TFT와 커패시터 각각의 값을 얻을 수 있다. 이 것을 바탕으로 제작한 커패시터의 재료적 특성을 활용하여 구성한 시냅스의 Retention time을 예측할 수 있다. 제작한 시냅스의 예상 Retention time은 >2s 이상으로 5T-1C 동작검증에 문제가 없는 수준임을 확인하였다. (Figure 4.2.2.3)

그리고, 제작한 IGZO TFT의 누설전류 수준에서 1s 이상의 Retention time을 확보하기 위해서는 1pF의 커패시터 정전용량이 필요함을 확인했고, 다른 연구 그룹의 IGZO TFT의 경우 보다 우수한 낮은 누설 전류 수준을 보이며 DRAM 커패시터 정전용량 수준으로 알려진 10fF에서도 1s 이상의 충분한 Retention time 확보가 가능할 것으로 예상된다. (Figure 4.2.2.4)^{[30][31]}



Figure 4.2.2.1 | Leakage current 측정용 2T-1C 회로





Figure 4.2.2.3 | IGZO TFT 기반 시냅스 Retention



Figure 4.2.2.4 | IGZO TFT 누설전류 수준별 예상 Retention

5. 결 론

이번 연구에서는 DNN 학습용 시냅스 소자의 요구사항으로 연산기능인 선형적이고 대칭적인 업데이트 특성이 중요하고, CMOS기반 전하저장형 시냅스 소자는 이를 만족하지만 높은 CMOS 트랜지스터 누설전류로 Retention 요구사항 만족을 위해서는 큰 정전용량이 필요하여 합리적인 수준의 소자 크기 축소가 어려운 문제 상황을 해결하고자 하였다. 그리고 탁월하게 낮은 누설전류 수준을 보이는 IGZO TFT를 활용하여 CMOS 트랜지스터를 대체하기 위한 대안 구조를 제시하고 이를 웨이퍼로 구현하여 동작을 검증하였다. 그 결과 IGZO TFT를 활용한 5T-1C 구조에서 Potentiation과 Depression 업데이트 동작이 정상적으로 수행됨을 확인했고, 업데이트 전류 관점에서 분석해보면 높은 선형성과 대칭성이 달성 가능함을 확인했다. 제작한 IGZO TFT의 누설전류 수준을 측정했고, 이 것을 바탕으로 Retention 요구사항을 달성하기 위한 필요 정전용량을 분석했다. 그리고 다른 연구그룹의 사례를 바탕으로 추가적인 누설전류를 개선하는 경우 Retention 요구사항을 달성하기 위한 필요 정전용량이 DRAM 수준까지 축소가 가능함을 확인하였다.

그러나, 시냅스 커패시터 소자 축소는 커패시터 자체의 개선을 통해 구현될 수도 있고, IGZO TFT의 입력 펄스당 업데이트 전류 량이나, 시냅스 동작 상황에서 트랜지스터 정전용량 등의 기생성분의 영향 등 다양한 소자 및 동작 구성요소들의 조합에 영향을 받는다. 이에 대한 추가 연구를 통해 학습 정확도와의 상호 작용을 분석하여 시냅스 커패시터 소자 축소 가능성을 논할 필요가 있다.

또한, Array에서 동작이 가능한 지에 대하여 본 연구에서 레이아웃으로 구현하였으나, 물리적 시간의 한계로 아직 테스트하지 못 하였다. DNN 연산 가속기에서 MAC 연산에 필요한 파라미터 수는 수 만개 이상이다.^[32] 이 점에서 IGZO TFT는 디스플레이와 같은 대규모로 병렬화 된 회로에서도 제작이 가능하다는 것이 이미 산업에서 증명되어 있지만, 시냅스 요구사항에 부합하는 산포 수준을 보이는 지에 대하여는 추가 연구가 필요하다. 한편, IGZO TFT를 활용하면 누설전류가 높은 CMOS 소자의 단점으로 어려움을 겪는 인공 신경망 가속 연산용 회로들을 개선할 수 있을 것이 기대되며, 다양한 연구들이 진행되고 있다.^{[33][34][35][36]} 디스플레이 반도체에서의 성공을 넘어 Logic 및 Memory 반도체에서 IGZO 트랜지스터를 활용하기 위해서는 소자 안정성 개선, 수소나 온도, 수분 등에 대해 민감한 공정 제약조건의 완화 등이 필요하다.^{[37][38][39]}

다양한 재료에 대한 이해와 연구를 통해 실리콘 기반의 반도체 산업 한계를 극복하고 AI 반도체 등의 확장된 어플리케이션 요구사항에 대응할 수 있을 것이 기대된다.^[40]

감사의 말

이 연구는 삼성전자 주식회사와 정부(과학기술정보통신부)의 재원으로 한국연구재단-지능형반도체선도기술개발사업의 지원을 받아 수행된 연구임. (과제번호: 2020M3F3A2A01081240)

IGZO TFT와 커패시터 소자 제작 및 특성개선 활동을 위해 삼성종합기술원의 연구시설을 활용했고, 무기소재Lab과 협업하였습니다. 연구에 많은 지도와 도움을 주신 김상욱, 이광희, 정문일, 양지은 박사님, 김은태 연구원님과 박성준 Lab장님께 감사드립니다.

시냅스 소자 동작 측정을 위해 고려대학교 전기전자공학부 이형민 교수님 연구실에서 Arduino MCU Board 및 상용 소자를 활용해 측정용 Board를 제작해 주셨습니다. 연구에 도움을 주신 강민일 연구원님께 감사드립니다.

제작한 소자의 측정 환경을 구성하고 측정 데이터 수집 및 결과 분석을 위해 본 연구실의 원종운, 노영채, 박예지 연구원님과 협업하였습니다. 시냅스 소자 측정 환경 구성에 원종운 님, 누설 전류와 기초 소자 특성 데이터 수집 및 분석에 노영채 님, 기초 소자 특성 데이터 수집 및 분석에 박예지 님께서 깊은 토론 및 도움을 주셨고, 감사드립니다.

참고 문헌

[1] Kim, Seyoung, et al. "Analog CMOS-based resistive processing unit for deep neural network training." 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS). IEEE, 2017.

[2] Hennessy, John L., and David A. Patterson. Computer architecture: a quantitative approach. Elsevier, 2011.

[3] Zhang, Wenqiang, et al. "Neuro-inspired computing chips." Nature Electronics 3.7 (2020): 371-382.

[4] Burr, Geoffrey W., et al. "Neuromorphic computing using nonvolatile memory." Advances in Physics: X 2.1 (2017): 89-124.

[5] Xi, Yue, et al. "In-memory learning with analog resistive switching memory: A review and perspective." Proceedings of the IEEE 109.1 (2020): 14-42.

[6] Ambrogio, Stefano, et al. "Equivalent-accuracy accelerated neural-network training using analogue memory." Nature 558.7708 (2018): 60-67.

[7] Wong, H-S. Philip, et al. "Phase change memory." Proceedings of the IEEE 98.12 (2010): 2201-2227.

[8] Wong, H-S. Philip, et al. "Metal-oxide RRAM." Proceedings of the IEEE 100.6 (2012): 1951-1970.

[9] Kund, Michael, et al. "Conductive bridging RAM (CBRAM): An emerging non-volatile memory technology scalable to sub 20nm."
IEEE International Electron Devices Meeting, 2005. IEDM Technical Digest.. IEEE, 2005.

[10] Li, Y., et al. "Capacitor-based cross-point array for analog neural network with record symmetry and linearity." 2018 IEEE Symposium on VLSI Technology. IEEE, 2018.

[11] Chang, Ting-Kuo, Chin-Wei Lin, and Shihchang Chang. "39-3: invited paper: LTPO TFT technology for AMOLEDs." SID Symposium Digest of technical papers. Vol. 50. No. 1. 2019

[12] Yamazaki, Shunpei, and Tetsuo Tsutsui, eds. Physics and Technology of Crystalline Oxide Semiconductor CAAC-IGZO: Application to Displays. John Wiley & Sons, 2017. [13] Kamiya, Toshio, and Hideo Hosono. "Material characteristics and applications of transparent amorphous oxide semiconductors." NPG Asia Materials 2.1 (2010): 15-22.

[14] Nomura, Kenji, et al. "Room-temperature fabrication of transparent flexible thin-film transistors using amorphous oxide semiconductors." nature 432.7016 (2004): 488-492

[15] Ishizu, Takahiko, et al. "An Energy-Efficient Normally Off Microcontroller With 880-nW Standby Power, 1 Clock System Backup, and 4.69-\$\mu \$ s Wakeup Featuring 60-nm CAAC-IGZO FETs." IEEE Solid-State Circuits Letters 2.12 (2019): 293-296.
[16] Yamazaki, Shunpei, and Masahiro Fujita, eds. Physics and Technology of Crystalline Oxide Semiconductor CAAC-IGZO: Application to LSI. John Wiley & Sons, 2016.

[17] Suzuki, Akio, et al. "Characteristics and Applications of CAAC-IGZO FET with Gate Length of 13nm." ECS Transactions 98.7 (2020): 13.

[18] Atsumi, Tomoaki, et al. "DRAM using crystalline oxide semiconductor for access transistors and not requiring refresh for more than ten days." 2012 4th IEEE International Memory Workshop. IEEE, 2012.

[19] Kamiya, Toshio, and Hideo Hosono. "Roles of hydrogen in amorphous oxide semiconductor." ECS Transactions 54.1 (2013): 103.

[20] Nguyen, Thi Thu Thuy, et al. "Impact of passivation conditions on characteristics of bottom-gate IGZO thin-film transistors."Journal of Display Technology 11.6 (2015): 554-558.

[21] Nam, Yunyong, et al. "Effect of hydrogen diffusion in an In-Ga-Zn-O thin film transistor with an aluminum oxide gate insulator on its electrical properties." RSC advances 8.10 (2018): 5622-5628.
[22] Lin, Yu-Hsien, and Jay-Chi Chou. "Temperature effects on a-IGZO thin film transistors using HfO2 gate dielectric material." Journal of Nanomaterials 2014 (2014).

[23] Cho, D. H., et al. "Passivation of bottom-gate IGZO thin film transistors." Journal of the Korean Physical Society 54.9 (2009): 531-534.

[24] Chu, Yen-Lin, et al. "Fabrication and characterization of a-IGZO thin-film transistors with and without passivation layers."ECS Journal of Solid State Science and Technology 10.2 (2021): 027002.

[25] George, S. M., "Atomic layer deposition: an overview.", Chemical reviews 110, 111-131 (2009)

[26] Knoops, Harm CM, et al. "Atomic layer deposition." Handbook of Crystal Growth. North-Holland, 2015. 1101-1134.

[27] Jeon, Woojin. "Recent advances in the understanding of high-k dielectric materials deposited by atomic layer deposition for dynamic random-access memory capacitor applications." Journal of Materials Research 35.7 (2020): 775-794.

[28] Ding, Shi-Jin, et al. "High-performance MIM capacitor using ALD high-k HfO 2-Al 2 O 3 laminate dielectrics." IEEE Electron Device Letters 24.12 (2003): 730-732.

[29] Sekine, Yusuke, et al. "Success in measurement the lowest off-state current of transistor in the world." ECS Transactions 37.1 (2011): 77.

[30] Kunitake, Hitoshi, et al. "A c-Axis-Aligned Crystalline In-GZn Oxide FET With a Gate Length of 21 nm Suitable for Memory Applications." IEEE Journal of the Electron Devices Society 7 (2019): 495-502.

[31] Belmonte, A., et al. "Capacitor-less, long-retention (> 400s) DRAM cell paving the way towards low-power and high-density monolithic 3D DRAM." 2020 IEEE International Electron Devices Meeting (IEDM). IEEE, 2020.

[32] Joseph, Jan Moritz, et al. "Architecture, Dataflow and Physical Design Implications of 3D-ICs for DNN-Accelerators." 2021 22nd International Symposium on Quality Electronic Design (ISQED). IEEE, 2021.

[33] Saito, D., et al. "IGZO-Based Compute Cell for Analog In-Memory Computing—DTCO Analysis to Enable Ultralow-Power AI at Edge." IEEE Transactions on Electron Devices 67.11 (2020): 4616-4620.

[34] Raman, Siddhartha Raman Sundara, Shanshan Xie, and Jaydeep

P. Kulkarni. "Compute-in-eDRAM with Backend Integrated Indium Gallium Zinc Oxide Transistors." 2021 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2021.
[35] Cosemans, S., et al. "Towards 10000TOPS/W DNN Inference with Analog in-Memory Computing-A Circuit Blueprint, Device Options and Requirements." 2019 IEEE International Electron Devices Meeting (IEDM). IEEE, 2019.

[36] Houshmand, Pouya, et al. "Opportunities and Limitations of Emerging Analog in-Memory Compute DNN Architectures." 2020
IEEE International Electron Devices Meeting (IEDM). IEEE, 2020.
[37] Prasad, Om Kumar, et al. "Role of in-situ hydrogen plasma treatment on gate bias stability and performance of a-IGZO thinfilm transistors." Nanotechnology (2021).

[38] Fuh, Chur-Shyang, et al. "Role of environmental and annealing conditions on the passivation-free in-Ga-Zn-O TFT." Thin Solid Films 520.5 (2011): 1489-1494.

[39] Song, Aeran, et al. "Hydrogen Behavior in Top Gate Amorphous In–Ga–Zn–O Device Fabrication Process During Gate Insulator Deposition and Gate Insulator Etching." IEEE Transactions on Electron Devices 68.6 (2021): 2723–2728.

[40] Tang, Jianshi, et al. "ECRAM as scalable synaptic cell for high-speed, low-power neuromorphic computing." 2018 IEEE International Electron Devices Meeting (IEDM). IEEE, 2018.

Abstract

Charge-storage synapse using In-Ga-Zn-Oxide Thin Film Transistor for training Deep Neural Network

Sangjun Hong Materials Science and Engineering Major The Graduate School Seoul National University

The CMOS-based charge storage synapse structure using silicon transistors and capacitors enables linear and symmetric weight updates, unlike the non-volatile element-based resistance-variable synaptic structure. It is known that a level of learning accuracy has been achieved. However, the CMOS-based synapse has a problem in that the learned weights are lost due to the leakage current of the transistor, so the learning accuracy decreases, thus it is difficult to reduce the size of the capacitor.

In this study, a charge storage type synapse with a new structure was proposed and fabricated using an IGZO TFT with low leakage current to have a linear and symmetric weight update characteristic. In addition, the IGZO TFT leakage current level, which can reduce the area of a reasonable capacitor, was proposed and compared with the capability of the manufactured device. As a result, 3mV/step depression operation was verified in the 2T-1C structure using IGZO TFT, and the leakage current level of the manufactured IGZO TFT
was confirmed to be $\sim 2.5E-16A/um$. And 100-levels of potentiation and depression operations were verified in 5T-1C synaptic structures using IGZO TFT.

Keywords : Charge storage type synapse, linear and symmetric weight update, low leakage current, long retention time, indium-galliumzinc-oxide thin film transistor

Student Number : 2019–23695