



# NAND Flash memory 소자 기술 동향

## I. 서론

USB 메모리, Digital Camera, MP3, Smart Phone, Tablet, SSD 등 대부분의 Digital 기기에 적용하고 있는 Flash 메모리는 과거 20년 넘게 Floating Gate Cell 의 scale-down 이 NAND Flash market 성장의 주 견인차 역할을 하였습니다. 그러나 1Xnm 미만 size 을 향한 Floating Gate technology 의 개발은 몇몇 물리적인 문제와 전기적인 특성 문제로 scale down 한계선에 접근하고 있는 상황입니다. 본 논문에서는 NAND Flash Memory 의 scale down 시

발생하는 Critical issue 를 짚어보면서, 기존 해결 방안과 향후 Flash 메모리의 연속성을 확보하는 방안들을 살펴보겠습니다. 제 2장

에서는 Flash Cell 의 일반적인 특성들을 설명하고, 제 3장에서는 Floating Gate 형 Flash memory 의 Scale-down 문제와 대응기술, 제 4장에서는 Flash 의 미래인 Charge Trap 형 3D NAND Cell 에 관한 내용으로 전개하겠습니다.

**NAND Flash 메모리는 밀폐된 용기와 같은 역할을 하는 Floating gate를 사용하는 구조이며, 최근 scale-down의 한계에 직면하고 있다.**



이희열  
SK하이닉스

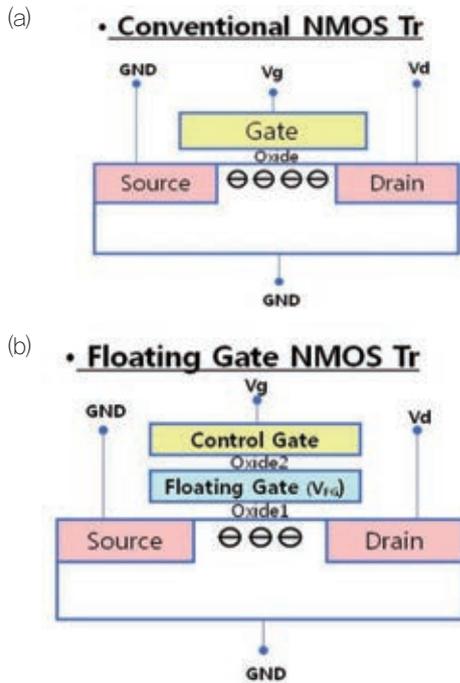


박성계  
SK하이닉스

## II. Flash Memory Cell

### 1. Flash Cell 기본 특성

Flash Cell 도 Normal NMOS 의 한 종류이며, 기본적인 Transistor(Tr) 동작 원리는 동일합니다. 다만 가장 큰 차이는 Bottom Gate 가 존재한다는 점입니다. Top Gate 는 Cell 선택 기능을 하는 Word Line (WL) 으로 형성되어 있어 Control Gate (C.G) 로 호칭합니다. Poly-Si으로 구성된 Bottom



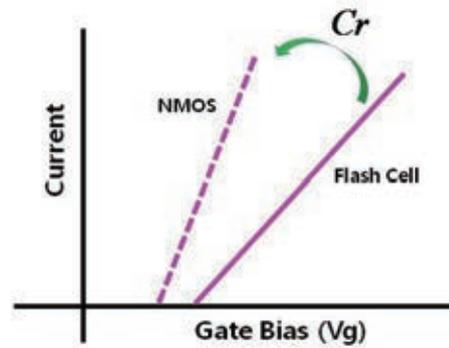
〈그림 1〉 NMOS Tr 과 Flash Cell 구조 비교

Gate 는 유전 물질로 둘러 싸여 있어 Floating Gate (F.G) 로 호칭하며 이 곳에 저장된 전하량에 따라 Digital 신호인 “0”, “1” 을 구분하는 방식으로 data 를 저장합니다. Floating Gate 가 마치 밀폐된 보관용 용기와 같은 역할을 하고 있어 Flash 는 전원이 꺼지더라도 저장된 Data 를 유지하고 있는 비 휘발성 메모리의 특성을 갖고 있습니다. (〈그림 1〉)

Flash Cell 의 동작은 2가지 중요 요소가 있습니다. 외부 적으로 볼 때 Bias 는 Control gate 에 인가되지만, Cell 내 부적으로는 Substrate Silicon (Sub-Si) Surface 에 인접 한 Floating gate 의 Potential (VFG) 에 의해 Tr 로써 동작합니다. VFG 는 ① Control Gate 와 Floating Gate 간 Coupling Capacitance 에 의해 F.G 에 유기된 Potential 과 ② Floating Gate 에 저장된 전하량으로 정해집니다.

상기 항목을 자세히 설명하면, ① 항목은 Control Gate 와 Floating gate 간 dielectric layer (Oxide2) 가 존재하여 상호간 Capacitance 로 작용하여 발생합니다. 간단하게 생각하면 2<sup>nd</sup> Gate Oxide 가 존재하므로 Control Gate Bias 가 특정 비율만큼만 VFG로 전달되는 것입니다.

VFG 는 수식적으로는  $V_{FG} = Cr * V_g$  표현하며, Cell 동작의 핵심 factor 인 Cr 은 Gate Coupling Ratio



〈그림 2〉 NMOS 와 Flash Cell 간 Vg-Id 비교

(또는 Coupling Ratio) 로 수식적으로는

$$Cr = \frac{C_{Oxide2}}{C_{Oxide1} + C_{Oxide2}}$$

으로 표현합니다.

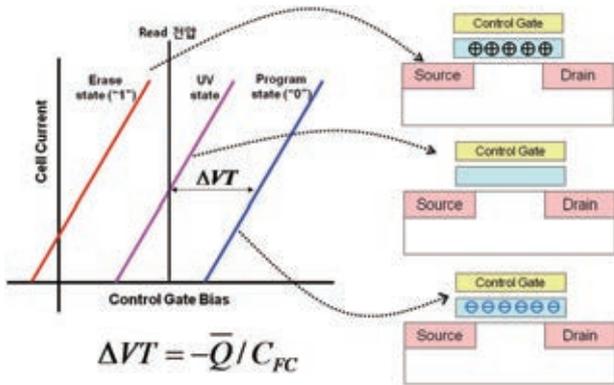
[Oxide2 는 C.G-F.G 간 유전물질이며 IPD (Inter-Poly Dielectric) 로 호칭합니다.

Oxide1 은 F.G - Sub. Si 간 유전물질로써 Tunnel Oxide 혹은 약칭으로 Tox 로 호칭합니다.]

즉, Cr 은 C.G-F.G 간 capacitance 를 F.G와 연계된 모든 Capacitance들의 총합으로 나눈 값입니다. 종래에는 통상 0.6 정도의 level 이지만 Cell size 가 감소하면서 Control Gate 와 Floating Gate 간 Capacitance 의 감소 비율이 증가하여 0.5 또는 그 이하 level 로 감소하는 추세입니다.

동일한 dimension 의 일반 NMOS 와 Flash Cell 간 Vg-Id 을 비교하면 (〈그림 2〉) Coupling Ratio (Cr) 비율만큼 Flash Cell 의 문턱 전압 증가와 완화된 Vg-Id slope 으로 Current 감소 현상이 발생합니다.

② 항목은 Floating Gate 에 저장된 전하량에 의한 Cell 문턱 전압의 변화 ( $\Delta VT$ ) 입니다. 〈그림 3〉에서 Q는 Floating Gate 에 저장된 전하량이고,  $C_{FC}$  는 Control Gate 와 Floating Gate 간의 Capacitance 입니다. UV state 는 Floating Gate 가 중성 상태로써 Flash Cell 동작의 기준점 역할을 하고 있습니다. [물리적 의미로는 Oxide 와 Silicon 간 Conduction Band 는 3.1eV, Valence Band는 4.7eV 의 potential Barrier 가 존재합니다. 4.9eV 에너지를 갖는 자외선 (UV)이 Flash Cell 에 조사 (expose)되면 Floating Gate 에 저장된 Electron 또는 Hole 이 외부로 배출 되어 중성 상태가 되므로, 이 상태를 UV state로 호칭합니다.]



〈그림 3〉 Flash Cell 의 Program / Erase 시 문턱 전압 변화

Floating Gate 에 electron 이 저장되어 있으면 문턱 전압이 증가하여 Read 전압 인가 시 전류 흐름이 없는 Program 상태 (“0”) 가 되고, Hole 이 저장되어 있으면 문턱 전압이 감소하여 Read 전압 인가 시 전류 흐름이 있는 Erase 상태 (“1”) 가 됩니다. Flash Cell 의 문턱전압 shift 는 공정적인 요소가 아닌 Floating gate 의 전위차에 기인한 것으로 문턱전압간에 slope 차이 없이 동일한  $V_g$ - $I_d$  curve 특성을 보여주고 있습니다. Flash Cell Tr 의 문턱 전압 shift 량은 F.G의 전하 변화 량을 C.G-F.G 간 Capacitance 로 나눈 값입니다. C.G-F.G 간 Cap. 이 감소하면 동일한 문턱전압 shift 를 위해 필요한 전하량도 같이 감소하기 때문에 Tech. scale-down 으로 인한 C.G-F.G 간 Capacitance 감소가 큰 문제로 나타납니다.

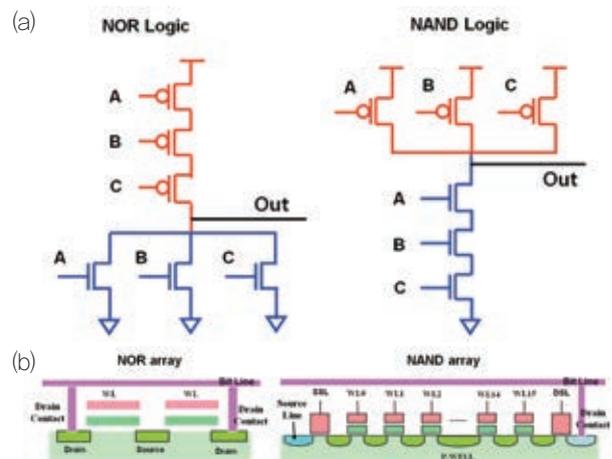
비 휘발성 메모리의 개념상 메모리 특성인 ① Data 의 Erase / Program 반복 가능 횟수 (Cycling or EW Cycling or Cycling endurance) 와 비 휘발성 특성인 ② Data의 보존 능력 (Retention)은 Flash 메모리의 매우 중요한 성능 지표입니다. 여기서 ① Cycling 은 Program 과 Erase 동작시  $T_{ox}$  를 통한 electron tunneling 중 일부 electron 들이  $T_{ox}$  내 trap site 또는 Channel -  $T_{ox}$  interface 의 dangling bond 에 trap 되어 문턱전압의 Positive shift, Current 감소, Program  $V_t$  분포 증가와 같은 특성 열화 현상이 발생합니다. 이 현상은 P/E 동작이 반복할수록 trap 이 증가되어 열화 현상이 심화되는 특징으로, 정상동작이 가능한 반복 횟수가 지표이기 때문에 Cycling 이라 표현합니다. 그리고 ②

NAND Flash 메모리는 Logic 회로인 NOR logic와 NAND logic을 참조하여 NOR array와 NAND array로 구분하며 NAND 플래시가 고용량 데이터에 적합한 구조로서 널리 사용되고 있다.

Retention 은 “1”/ “0” 으로 판별 가능한 누출 허용시간이 지표로서 유전물질로 둘러 싸인 storage 에 저장된 electron 의 보존 능력입니다. 국부적으로 얇아진 유전물질의 두께, defect, trap site 등 취약지점을 통해 electron 이 누출되어 문턱전압의 하향 shift 를 유발합니다. 특히 Cycling 후에는 유전물질의 damage 증가 + trap electron 의 de-trap 까지 더해져 특성 열화가 더욱 심해집니다. 따라서 Cycling 후 Retention 특성 (Post Cycling Retention) 지표인 Cycling 횟수 + Retention 이 Flash 메모리에서 매우 중요한 성능조건입니다.

## 2. Array 구조 (NOR vs. NAND)

Flash 메모리는 NOR array 와 NAND array로 크게 구분하는데, 그 호칭은 설계 기본 Logic 회로인 NOR 와 NAND 로 부터 왔습니다. 일반적으로 Memory Cell Tr 은 current drive-in 능력이 좋은 NMOS 를 적용합니다. 〈그림 4-a〉의 NOR 와 NAND logic 구성에서 data output 단자와 NMOS의 배열을 보면 NOR는 병렬 연결 구조이고 NAND 는 직렬 연결 구조입니다. 이와 같은 NMOS 배열을 참조로 하여 Data 입/출력 통로인 Bit Line 에 memory Cell 들이 병렬 연결되어 있으면 NOR array 구조이고 직렬 연결되어 있으면 NAND array 구



〈그림 4〉 NOR / NAND type 비교  
(a). NOR / NAND logic 구성도, (b). NOR/NAND array 구조

조로 호칭 하였습니다.((그림 4-b)) 이 중에서 그 구조가 단순하여 High density 구현에 알맞은 NAND Flash 는 고 용량 Data 저장 매체로 널리 사용하고 있어 현재는 Flash = NAND 로 통용되는 상황입니다.

### 3. SLC / MLC / TLC

Flash Memory 를 사용한 저장매체들에서 동일한 저장용량 이지만 Write 속도, Read 속도 및 가격에서도 차이가 발생하는 것을 본적이 있을 것입니다. 그와 관련하여 SLC, MLC, TLC 와 같은 용어를 접할 수 있습니다. 이는 데이터를 저장하는 최소 단위인 Cell 에 데이터 저장 bit 수에 따라 그 종류를 나눈 것입니다. <그림 5>에서 보듯이 Cell 상태가 Erase "1" 과 Program "0" 로 2단계로 구분하여 1bit data를 저장하는 SLC (Single Level Cell), Program 상태를 3개로 분할하여서 4단계로 구분하여 2bit data 를 저장하는 MLC (Multi Level Cell), Program 상태를 7개로 분할하여서 8단계로 구분하여 3bit data 를 저장하는 TLC (Triple Level Cell) 가 상용화 되었습니다. 동일한 물리적 Cell 밀집도 이지만 동작 방법에 따라 정보 저장 bit 수가 증가함으로 고 용량 구현이 가능해집니다.

예를 들면 동일 외형상 12Gbit 용량이면 실제 물리적 Cell 수는 12G bits (@SLC) / 6G bits (@ MLC) / 4G bits (@TLC) 로 구성되어 가격과 성능 면에서 각각 차이가 존재합니다. 가격은 SLC > MLC > TLC 순으로 저렴해집니다. SLC 는 1단계 프로그램 및 Read 동작 1회 이지만, TLC 는 7단계 프로그램 과 Read 동작 7회를 진행함으로 SLC, MLC, TLC 순으로 Data

Write time 과 Data Read time 이 증가합니다. 그리고 각각의 프로그램 단계는 max. Program 문턱전압 범위 내에서만 분할할 수 있습니다. 그에 따라 각 program 단계별 Cell 들의 문턱전압 분포를 좁게 형성 해야 하고, 판별 해야 할 Cell 들의 문턱전압과 Read 전압간 간격이 좁아져서 신뢰성 특성도 열화 됩니다. 그래서 사용자는 첫 번째 고려 사항이 가격이면 TLC, 고 성능이 필요하면 SLC, 가격과 성능의 적절한 조합이 필요하면 MLC Flash 를 적용한 제품을 선택할 수 있습니다. 최근에는 Program 상태를 15개로 분할하여서 16단계로 구분하여 4bit data 를 저장하는 QLC (Quadruple Level Cell) 도 개발 중입니다. TLC 보다 낮은 가격이지만 성능도 낮아짐으로, 고성능이 필요하지 않은 USB memory와 사용빈도가 높지 않지만 다른 저장장치 (예, HDD) 에 비해 성능 대비 구매 비용 또는 유지비용이 작은 분야(예, Server) 등에 적용될 것으로 보입니다.

## III. Floating Gate Flash 와 Scale-down

NAND Flash 의 scale-down 은 공격적으로 진행되었습니다. 과거 10년 동안 120nm 에서 1Xnm 로 급격하게 shrink 되어 chip 생산 비용이 매년 40% 정도 감소하였고, 저렴한 가격으로 인해 NAND Flash 를 탑재하는 전자기기의 수요도 급증하였습니다. 그러나 conventional Floating Gate Cell 의

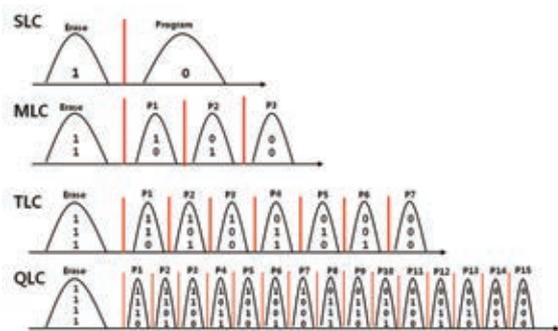
**NAND Flash 메모리는 최소 단위인 Cell에 저장되는 데이터 bit 수에 따라 SLC, MLC, TLC 및 QLC로 나누어진다.**

scale-down 은 3종류의 큰 벽에 직면하고 있으며 이는 ① 물리적인 dimension 요소, ② Device 적 제한 요소, ③ Read 동작 시 판별 Cell 간의 좁은 간격입니다<sup>[1]</sup>. 대안으로 3차원 구조의 NAND Flash Array 구조가 등장하여 scale-down 대신 적층 수를 증가하는 방법으로 집적도 증가를 이어갈 수 있게 되었습니다<sup>[2-4]</sup>.

### 1. 물리적 dimension 한계

#### (1) 노광 장비 문제

Scale-down 은 노광장비 의존성이 매우 큽니다. 지난 20

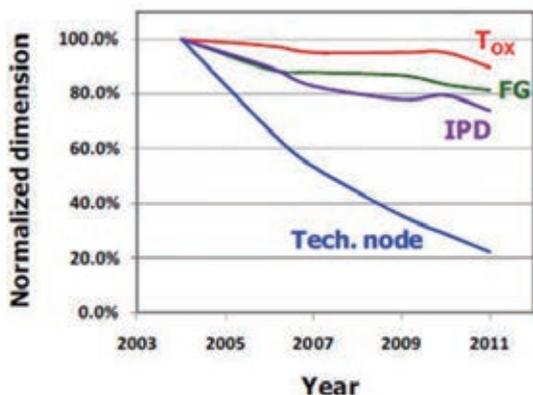


<그림 5> SLC, MLC, TLC, QLC 구성

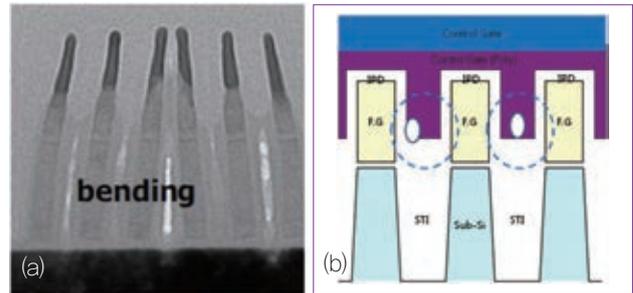
여 년간 노광 장비는 광원 개발로 빛 파장을 줄이는 방향으로 436nm (G-line) → 365nm (I-line) → 248nm (KrF) → 193nm (ArF) 로 진화하였습니다. 광원은 동일하지만 물의 굴절율을 이용하여 해상력을 높이는 방법으로 Immersion ArF 는 38nm 선폭 형성이 가능하여 4Xnm tech. 까지 적용하였지만, 그 미만은 물리적 한계로 인하여 동일 노광 장비에서 Double Patterning 공정으로 3Xnm, 2Xnm tech, Triple Patterning 공정으로 1Xnm tech. 를 개발 & 생산하고 있습니다. 10nm 이하를 위한 차세대 노광 장비는 빛 파장이 13.5nm 인 EUV (Extreme Ultra Violet)입니다만 웨이퍼 처리량이 작고 미 해결 과제들이 있어 아직은 실험용만으로 사용하고 있는 상황입니다.

(2) Inter-Connection Line issue

Cell Array 에서 필요한 단위 cell 선택과 Data 입력/출력은 각각 Word Line 과 Bit Line 을 통해서 진행합니다. 고성능 동작을 위해서는 낮은 Line 저항 (R) 과 Line간에 낮은 Capacitance (C) 가 필요합니다. 즉 RC delay 에 의한 Loading time 이 작아야 합니다. 그러나 scale-down 진행되어 Line의 선폭 감소로 저항이 증가하고, Line 간 간격 감소로 Capacitance 도 증가하여 Line 에 필요한 동작 전압을 인가하는 Loading time이 길어지고 있습니다. 그에 대응하기 위해 비 저항이 낮은 물질을 Line material 로 적용하고, 유전율이 낮은 물질로 Line 간 space 영역을 매립하는 것이 필요합니다. 그에 따라 점진적으로 ① Word Line 물질은 Poly Si → WSi → CoSi (or NiSi) → W, ② Bit Line 물질은 W →



〈그림 6〉 Floating Gate Cell geometry 의 scaling trend<sup>[5]</sup>



〈그림 7〉 F.G Cell 구조의 주요 물리적 문제요소  
(a) WL-WL 방향 과 (b) BL-BL 방향

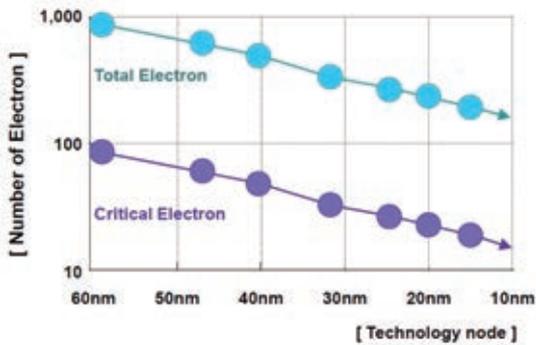
Al → Cu, ③ Space 영역 매립물질은 Nitride → Oxide → Air-Gap 으로 변경 적용하고 있습니다.

(3) F.G Cell 의 수직 - 수평 구조간 scale down 의 문제점

수직 방향의 scale down 은 Cell 기본 동작과 관련되어 수평 방향의 scale down 보다 매우 어렵습니다. Tox (oxide1) 와 IPD (oxide2) 두께는 F.G 에 저장된 전하의 유출을 방지하는 Data 보존 성능의 핵심요소이고, F.G 의 두께는 Gate Coupling Ratio (Cr) 유지를 위해 C.G 와 F.G 간의 중첩면적을 확보하는 요소 임으로 〈그림 6〉에서 보듯이 Technology node 대비 수직 방향 요소(Tox / IPD / FG 두께)는 scaling 이 거의 이루어지지 않고 있습니다.

이러한 수직 방향대비 수평 방향만의 scaling 에 기인한 매우 좁으면서 높은 구조의 WL pattern 은 bending 현상을 유발하고 있습니다(〈그림 7-a〉). 이 현상은 WL pattern시 불균일한 shape 형성으로 WL 좌우측간 화학적 세정액의 모세관력 (Capillary force) 차이로 인해 발생 합니다<sup>[6]</sup>. Pattern shape 의 요소 기술과 모세관력을 최소화하는 화학적 세정액 도입으로 Bending 문제를 해결하여 1Xnm 까지는 진행하고 있는 상황입니다.

C.G 와 F.G 간 충분한 Gate Coupling ratio 와 균일한 Cell Transistor 확보를 위해서는 이웃한 F.G 간 space 영역 (〈그림 7-b〉)에 Control Gate Poly 의 안정적인 매립이 필요합니다. 즉 균일한 Poly doping 과 함께 Poly Void 형성이 없어야 합니다. 규칙성 없이 형성되는 Void 와 Poly doping이 부족하여 Cell 동작 중 생성되는 공핍 영역은 F.G - C.G 간 Overlap capacitance 의 level 감소 및 variation 증가를 유



〈그림 8〉 Tech.별 Program 유지를 위한 전자 수<sup>[7]</sup>

발합니다. 그러나 F.G 간 간격이 지속적으로 scale-down 되는 상황임으로 F.G 간 space 확보를 위한 Cell 구조 design, 매립 특성 개선을 위한 Poly 요소 기술 개발 및 Metal liner (shallow Metal 증착 + Poly 증착) 기술 개발 등이 필요합니다. 그리고 F.G 두께를 내리고 High-K 유전물질의 IPD 적용하는 Planar F.G 형성방법은 다른 방식의 해결 방안이지만, High-K 유전물질의 leakage 문제 해결이 매우 중요한 요소입니다.

## 2. Device 적 제한요소

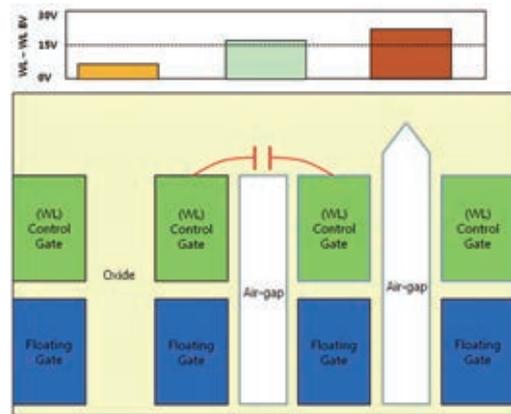
### (1) Storage 의 전자 개수

Scale down 에 따라 F.G-C.G 간 Capacitance 가 감소하면서 문턱전압 shift 를 유발할 수 있는 임계 전자 수가 급격하게 감소하여 오 동작 발생 요인이 되고 있습니다(〈그림 8〉).

수식적으로 표현하면  $\Delta VT = \Delta Q / C_{FG-CG}$  입니다. 동일한 문턱전압 shift 는 Capacitance 의 감소 비율만큼 전하량도 감소하면서 나타납니다. 예를 들면 60nm tech. 에서 100개의 electron loss 로 문턱 전압이 0.2V 감소하였다면 1X nm tech. 에서는 20개의 electron loss 만으로도 0.2V 감소합니다. 결국 이 양상의 변화는 Tech. shrink 되면서 Flash Cell 자체의 Data 보존 능력 저하를 의미하는 것으로 이를 보상하기 위해 다양한 algorithm 적용 및 Controller 의 도움이 더 필요해지고 있습니다.

### (2) 이웃 Cell 간 Leakage path 발생

WL 간 간격이 감소하면서 WL(C.G) - WL(C.G), WL(C.G) - 이웃한 F.G 간 leakage 문제가 나타나고 있습니다.

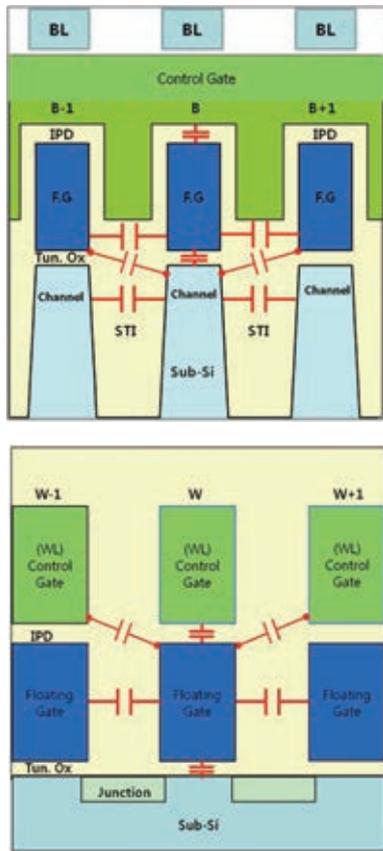


〈그림 9〉 WL 간 space 영역 구성 별 WL-WL BV

〈그림 6〉에서 보았듯이 WL 간 space 감소율 대비 Tunnel Oxide 와 IPD 두께 감소율이 작아서 Program 동작 중 선택된 WL에는 계속해서 높은 Program Bias 가 인가되고 있습니다. 이로 인해 이웃한 비선택 WL에 인가되는 By-Pass bias 와 전압차이 (>15V)도 거의 유사하게 유지되어 WL 간 leakage path 형성 또는 Breakdown 이 발생할 수도 있습니다. 또한 이웃한 F.G 와도 간격도 같이 감소되어 F.G 에 electron 이 저장된 경우에는 electron loss 가 발생할 수도 있습니다. 따라서 〈그림 9〉에서 보듯이 WL 간 isolation 을 강화하기 위해 WL 간 space 영역을 Oxide 매립 대신 유전율이 낮은 빈 공간인 Air-Gap 을 형성하여 E-field 를 완화시키며<sup>[8]</sup>, 더 나가 WL 상단부의 fringe E-field 를 감안하여 Air-Gap shape 을 WL 보다 높게 형성하고 있습니다. 이 방법으로 〈그림 9〉에서 보듯이 WL - WL 간 Breakdown 전압을 개선하여 1X nm tech. 에서도 정상적인 Program 동작을 진행할 수 있습니다. 그러나 가장 낮은 유전율인 Air-Gap 을 이미 적용하고 있어서 1X nm 미만 tech. 에서는 적절한 공정 조건을 확보하기는 어려운 상황입니다.

## 3. Read 동작 시 판별 Cell 간의 좁은 간격

급격한 Scale down 에 따라 Selected Cell 의 Gate 와 이웃한 Cell 의 Gate 간 좁은 간격으로 인해 Coupling Capacitance 에서 IPD Cap. 과 Tunnel oxide Cap. 외에도 이웃한 Gate 간 Coupling Capacitance(WL-F.G & F.G-F.G)가 중요한 요소로 나타나고 있습니다<sup>[9]</sup>. 〈그림 10〉에서 보듯이 이러한 Coupling capacitance 는 Program 동작 중에는

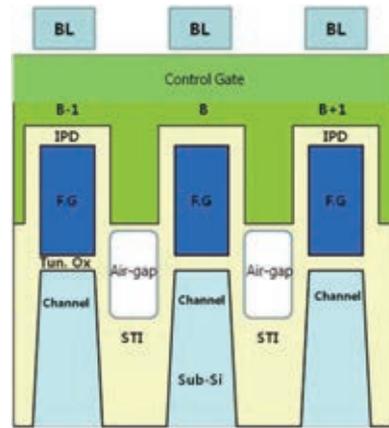


〈그림 10〉 BL과 WL 방향에서 Cell의 구성 요소간 Capacitive Coupling

WL-F.G 간 간섭 효과로 낮은 Program Vt, Read 동작 중에는 F.G - F.G 간 간섭 효과로 Program Cell Vt 분포 증가 현상을 유발하고 있습니다. 예를 들면, B cell의 이웃한 (B+1)th Cell 또는 (B-1)th Cell이 Erase 상태에서는 Positive potential을 B Cell의 F.G에 영향을 주지만, Program 상태로 변경되면 Negative potential로 변경되어 B Cell의 F.G에 영향을 주게 됨으로 B Cell의 F.G는 potential drop 현상으로 Control Gate의 전압이 증가해야만 기존과 동일한 Channel 형성을 할 수 있어, 마치 B Cell의 문턱 전압이 증가되는 효과를 보여주고 있습니다.

F.G- F.G 간섭효과는 수식적으로  $\Delta V_t = \text{Capacitive Interference Ratio} \times \text{Neighbor Cell Vt shift}$  표현됩니다.

Capacitive Interference Ratio를 감소시키는 방법은 3종류의 Cell 구조 개선으로 ① F.G - F.G 간 Capacitance를 감소하기 위해 F.G 간 간격을 확보하는 공정 개발, ② 낮은 유전율을 갖는 매립 물질로의 변경 Nitride (7.5) → Oxide



〈그림 11〉 STI 영역의 Air-Gap 형성

Issues with 10nm FG		Solutions
X - direction	CG poly-Si void	?
	Interference	
	Bit-Line loading	
Y - direction	WL bending	
	Interference	
	WL-to-WL leakage	

10nm FG?

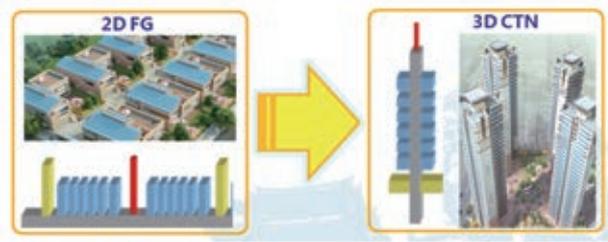


3D NAND?

〈그림 12〉 F.G의 scaling 한계 극복을 위해 3D 구조로 방향 전환<sup>[7]</sup>

(3.9) → Air-Gap(1.0) 그리고 ③ F.G 두께의 감소 방안이 필요합니다. 특히 Air-Gap은 가장 낮은 유전율을 갖고 있어 WL - WL 간 Space의 Air-Gap (〈그림 9〉)는 F.G - F.G, WL - WL, WL - F.G 간 간섭효과를 제어하기 위해 2Xnm 이하 tech에서, Shallow Trench Isolation (STI) 영역의 Air-Gap (〈그림 11〉)은 F.G-F.G, F.G-Channel, Channel - Channel 간 간섭효과를 제어하기 위해 1X nm 이하 Tech.는 필수 적용기술입니다.

전기적 및 알고리즘적 개선방법으로는 크게 3종류가 있는데 ① Neighbor Cell Vt shift에 영향을 주는 Neighbor Cell 수의 감소, ② Program 시작하는 Cell Vt를 Positive 방향으로 상향하여 Program 완료 Cell Vt와 Gap 감소, ③ Interference Vt shift를 받은 상태에서 program을 완료하는 방법 등을 적용하고 있습니다. 간섭효과에 의한 지속적인 Program Vt 분포 증가는 특히 TLC 성능에 치명적인 요소로 작용하고 있으며, 구조적, 전기적, 알고리즘적 개선 요소들을



〈그림 13〉 scaling 의 페러다임 전환 (2D FG → 3D CTN)

이미 1X nm 에서 적용한 상태에서도 각 Cell level 간 문턱전압 증첩 현상이 발생하고 있어 Strong Controller 의 도움을 받아 극복하고 있는 상황입니다.

따라서 Cell 구조의 혁신 없이 단순히 F.G Cell Scaling 은 전기적, 공정적으로 한계치에 도달하여 대부분의 NAND Flash 업체에서는 10nm F.G 대신 3D Cell 구조로 개발 방향을 전환 하였습니다(〈그림 12〉).

## VI. CTD Flash & 3D Array 구조

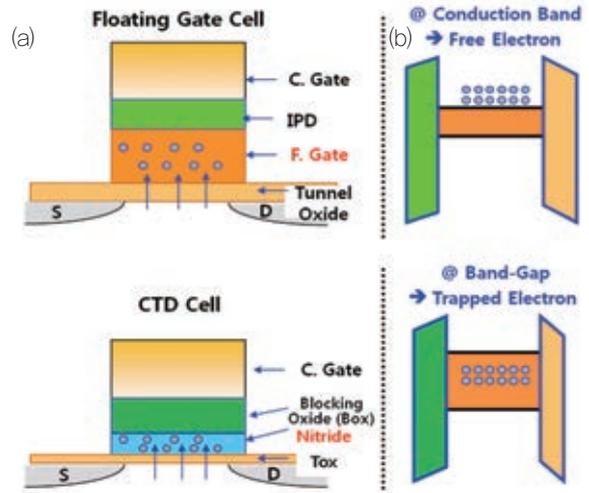
Cell array 를 수평 방향으로 형성하는 F.G Flash의 Scale down의 한계로 인한 고 집적화의 어려움을 극복하려는 돌파구로써 Cell array 를 수직 방향으로 형성하는 3D array 구조가 제안 되었습니다(〈그림 13〉).

적층 layer 수를 증가하면 Cell density 가 증가하는 구조로 Scale down 없이도 Cost down 할 수 있는 유용한 구조로 이제는 Flash 의 main array 로 자리잡게 되었습니다.

다양한 Array 구조와 F.G type, Charge Trap Nitride (CTN) type 의 단위 Cell 구조가 제안되었지만, 비교적 단순한 구조로 인해 Scaling 관점에서 좀 더 유리하여 대부분의 NAND Flash 업체가 CTN type Cell 구조를 적용하고 있습니다. 이에 따라 제 4장에서는 main stream 인 CTD (Charge Trap Device) [or CTF (Charge Trap Flash)] Cell 구조 + Vertical Channel Array 구조에 대하여 알아보겠습니다.

### 1. CTD Cell 기본 특성

Floating Gate 형 Cell 과 Charge Trap 형 Cell 간 근본 차이는 Charge storage 물질입니다. F.G Cell 은 전도체인



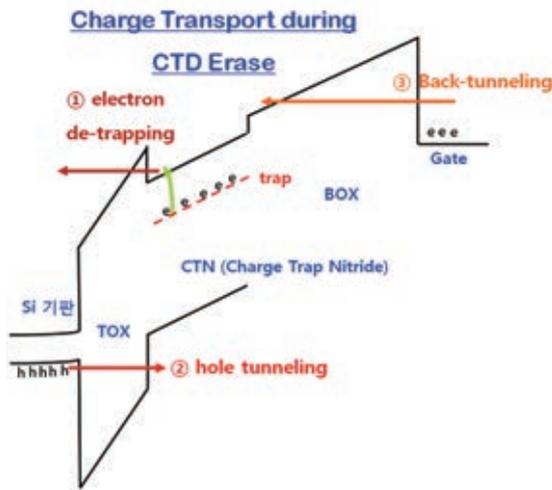
〈그림 14〉 F.G Cell 과 CTD Cell 간 (a) 구조와 (b) 전하 저장방식 비교

Doping 된 Poly Silicon (Poly) 을 사용하고 있지만, CTD Cell 은 유전물질인 Nitride 를 storage 물질로 사용하고 있습니다. Cell 의 수직 방향 구조를 보면 F.G Cell 은 Control Gate - IPD - Poly - Tunnel Oxide (Tox) - Sub Si 로 구성되고, CTD Cell 은 Control Gate - Blocking Oxide (Box) - Nitride - Tunnel Oxide (Tox) - Sub Si 로 구성되어 있습니다(〈그림 14-a〉).

Storage 물질 차이로 인해 Charge 저장 방식도 차이가 있습니다(〈그림 14-b〉). F.G Cell 은 Poly의 Conduction Band 에 Free electron 형태로 존재하고, CTD Cell 은 Nitride 의 Band Gap 내 trap site 에 Trap electron 형태로 존재 합니다. Gate Coupling Ratio 확보를 위해 F.G Poly를 높은 두께로

사용한 F.G Cell 과는 반대로 Storage Nitride 는 두께가 두꺼우면 상대적으로 동작 조건이 불리해 집니다. 낮은 두께와 부도체인 Storage 로 인하여 F.G Cell 의 scale-down 제약 요소로 작용한 F.G - F.G 간 Capacitive coupling 에 의한 Interference (간섭효과) 가 없어 Program Cell 의 Vt 분포를 좁게 형성할 수 있습니다. Trap 된 electron 특성상 electron 의 움직임이 제한되어 Tunnel Oxide 를 F.G 보다 낮은 두께로 형성할 수 있어 더 높은 Program Vt 가 가능하여 F.G 대비 낮은 Program Bias 사용할 수 있고, 또한 Tunnel oxide 에 trap되는 carrier 들의 절대량도 감소하여 Program /

**NAND Flash 메모리는 charge storage 물질에 따라 floating gate 형과 charge trap 형으로 나뉘어진다.**



〈그림 15〉 CTD Erase 동작 중 Charge 이동 경로

Erase 반복 횟수도 개선되는 장점이 있습니다.

## 2. Program & Erase 특성

Program 동작은 Si 기판의 electron 이 FN tunnel 방식으로 Tunnel oxide 를 통과하여 Nitride 에 trap 되는 방식으로 trap 특성은 Nitride 의 trap 밀도, trap capture cross section, Trap energy level 같이 Nitride 의 물성에 의존합니다. 또한 Nitride 와 Oxide 간 Conduction Band Barrier gap이 1.05eV 정도로 낮아 program 동작 중 Blocking Oxide (Box) 를 통하여 Control Gate 방향으로 Tunneling 방식의 electron 누출이 발생하여 Program Bias 대비 Program 효율성은 80 ~ 90% 이지만 실제 Program 특성은 F.G Cell 과 큰 차이는 없습니다. 하지만 Erase 특성은 F.G Cell 보다 다양한 Carrier path 들이 존재합니다.

〈그림 15〉에서 보듯이 Erase 동작의 주요 특성은 3가지 입니다. Erase 를 진행하는 ① Trap electron 의 de-trap 과정 및 ② Hole tunneling 과정과, Erase 를 방해하는 ③ Control Gate 로 부터 Box 를 통한 electron 의 Back tunneling 과정입니다. 각 특성 별로 살펴보면 ① Electron De-trap 은 Erase 초기 단계에서 기여하는데, Nitride 에 trap 된 electron 이 Tunnel Oxide 를 통해 Si 기판으로 이동하는 것으로 Nitride 의 trap energy level 이 작을수록 효

율성이 높아집니다. ② Si 기판으로부터 Hole tunneling 은 de-trap 이 끝나는 과정에서 Erase 에 기여합니다. 이 방식은 Si 기판에서 효율적인 Hole 생성과 Tox 의 Valance Band 의 barrier 높이가 주요 요소로 작용합니다. ③ Back tunneling 개선과정은 CTD 개발 과정 그 자체라고도 볼 수 있습니다. Back tunneling 은 High-K 물질의 Box 적용 & High work function gate material 적용으로 개선하였습니다. Gate material 은 Block layer 의 Conduction Band 와 Potential Barrier Gap 을 크게 하는 방식으로 N-type Poly → P-type Poly → Metal 로 변경 발전 하였습니다. 적용된 Block oxide 물질과 Gate 물질에 따라 붙이는 CTD 호칭도 SONOS (Si-Oxide-Nitride-Oxide-Si) 부터 SANOS (Si-Alumina-Nitride-Oxide-Si), MONOS (Metal-Oxide-Nitride-Oxide-Si), WANOS (W-Alumina-Nitride-Oxide-Si), MANOS (Metal-Alumina-Nitride-Oxide-Si) or TANOS (Titanium-Alumina-Nitride-Oxide-Si)로 다양해졌습니다.

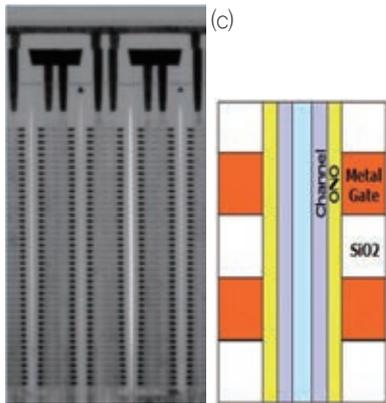
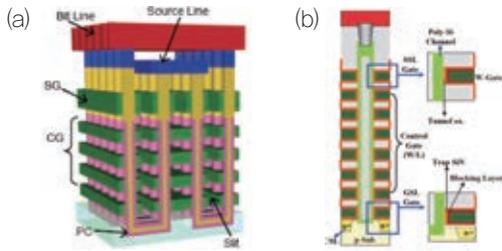
## 3. Vertical Channel 3D NAND Flash Array 구조

대표적인 Vertical Channel 3D array 구조는 U자형 Cell array 구조인 P-BiCS (Pipe connected Bit Cost Scalable)<sup>[2]</sup>, SMArT (Stacked Memory Array Transistor)<sup>[4]</sup> 와 I 자형 Cell array 구조인 TCAT (Terabit Cell Array Transistor)<sup>[3]</sup> 입니다. 상기 array 구조의 장점은 Gate가 Channel을 완전히 감싸는 Gate All-Around (GAA) 구조이기 때문에 Gate에 의한 channel의 지배력이 크다는 장점이 있는 반면에 ONO와 WL을 포함하는 stack 전체가 단위 cell의 X,Y 방향 dimension에 포함되기 때문에 단위 cell size가 커서 scaling이 어렵고, 단위 cell size가 큰 만큼 높은 층수를 필요로 하기 때문에 공정 난이도가 높은 특징이 있습니다.

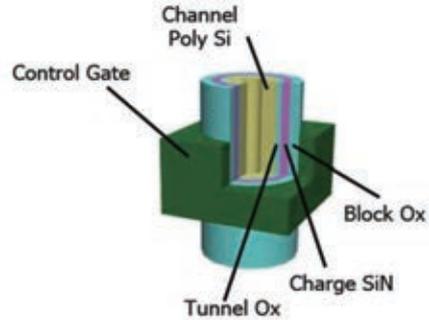
### (1) Channel 구조

기본 Cell 구조는 〈그림 16〉에서 보듯이 Gate 구조와 Gate 간 space 영역을 형성할 material 을 반복하여 적

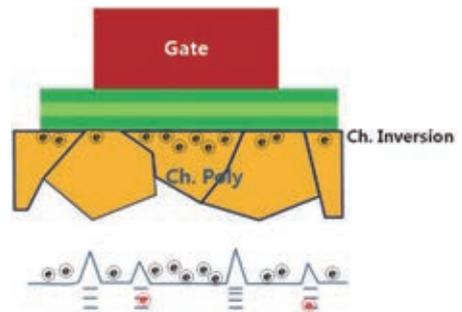
**Vertical channel 3D 구조는 게이트에 의한 채널의 지배력이 크다는 장점과 scaling이 어렵고 공정 난이도가 높은 단점을 갖고 있다.**



〈그림 16〉 3D NAND array 구조 (a) P-BICS, (b) TCAT (c) SMART Cell 구조<sup>[2-4]</sup>



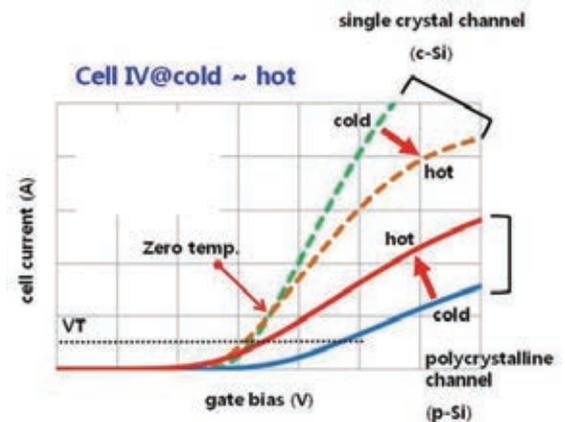
〈그림 17〉 GAA & 마카로니 형 Cell 구조



〈그림 18〉 Poly Channel 에서 전자 이동도 및 trap 모델

층 한 후 Channel Plug 을 형성하고서 측벽에서부터 중심을 향해 Block Oxide / Storage Nitride / Tunnel Oxide / Channel Poly / Oxide 를 증착하여 Cell 을 구성합니다(〈그림 17〉). Plug 중심부는 Poly 가 아닌 Oxide 로 매립하는 일명 마카로니 구조를 적용하고 있습니다. 근본 이유는 Grain Boundary(G.B)가 형성된 Poly Channel 구조적 특성 때문입니다(〈그림 18〉). Read 동작 시 Channel 에서 발생하는 Depletion layer 폭보다 Poly 두께를 작게 할 수 있어 Grain Boundary 의 charge trap 영향을 감소시킬 수 있고 그에 따른 transport 특성을 획기적으로 개선할 수 있어<sup>[10]</sup> 3D Vertical channel 방식에서는 모두 적용하고 있습니다.

3D Cell은 GAA 구조로 Channel width 가 넓지만, 증착된 Poly Silicon 에서 Channel 이 형성되고 Junction free 구조까지 더해져 F,G Cell 에 비해 작은 Cell Current 특성을 갖고 있습니다. (〈그림 19〉) Cell Current 의 온도 특성도 Single Si Channel 과 차이가 있습니다. Grain Boundary 에서 온도가 높을수록 Potential Barrier 가 낮아짐에 따라 ON / OFF Current 가 같이 증가하는 현상이 발생되고 있습니다. 또한 온도간 문턱 전압 shift 도 크게 나타나서 F,G Cell 보다 더 강화된 온도 보상 알고리즘을 적용하고 있습니다.

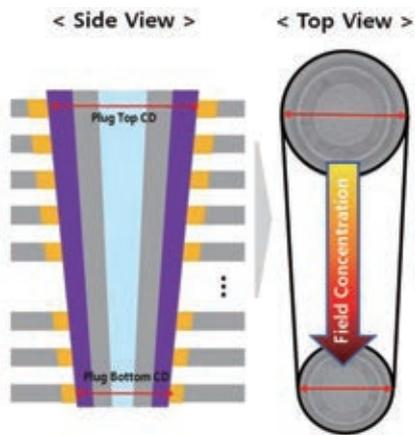


〈그림 19〉 Single Crystal channel 과 Poly Channel 간 온도 별 Vg-Id Curve 비교

이와 같은 사항은 Channel Poly 의 Grain Boundary 에 기인하고 있어 grain boundary 에서 낮은 potential barrier 형성 방법, Large Grain size 형성하는 요소 기술과 Single crystal-like 증착 요소 기술 확보가 핵심입니다.

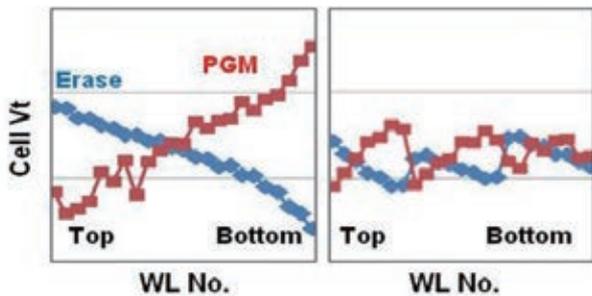
## (2) Plug 구조

GAA 구조에서는 electric flux 밀도 차이로 인해 곡률 반경이 작을수록 Electrical Field 증가합니다. 이러한 특성의

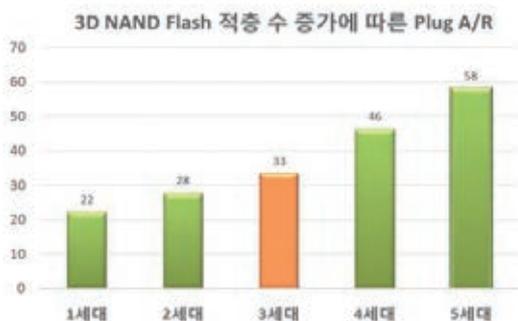


〈그림 20〉 3D Plug 의 단면 과 Top view

로 인해 Tunnel Oxide 영역의 E-Field 값이 Block Oxide 영역의 E-field 값보다 큰 값을 갖고 있어 Planar 구조보다는 Program / Erase / Read 동작 모두 도움이 되고 있습니다. 그러나 〈그림 20〉에서 보듯이 Plug 형성하는 식각 공정의 특성상 시작 부분에서 끝 부분으로 갈수록 직경이 좁아져서 상단부의 Cell 특성과 하단부의 Cell 특성간 차이가 발생하고 있습니다.



〈그림 21〉 WL 보상 알고리즘적용 (a) 전 / (b) 후<sup>[11]</sup>



〈그림 22〉 3D NAND Flash 적층 수 증가에 따른 Plug구조의 aspect ratio

이러한 Plug 직경 차이에 기인한 Program & Erase Vt 차이는 각 WL에 동일한 전압을 인가하지 않고 전압차이를 주는 방식으로 보상을 하는 알고리즘을 적용하고 있습니다(〈그림 21〉).

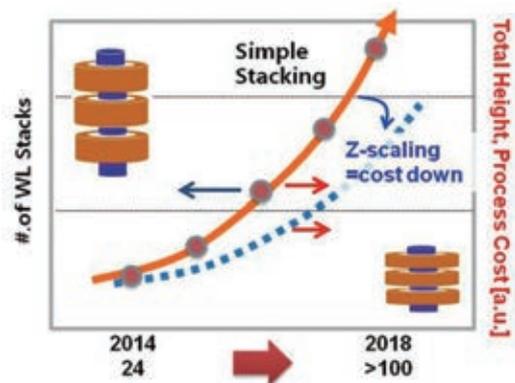
하단부 직경 대비 Plug 높이 비율을 표현하는 Aspect Ratio (A/R) 을 비교하면 (〈그림 22〉) 적층 layer 수가 증가할수록 이 현상은 심화되고 있습니다. 하단부 직경 확보를 위해서는 상단부 직경을 크게 해야 함으로 chip size 증가를 유발하는 문제로 나타납니다.

이에 대한 대응책으로 적층 layer 의 폭을 감소하여 Plug 높이를 하향하는 요소 기술 개발로 Plug A/R 을 개선하여 Cell 간 위치 별 특성 variation 증가 및 Cost 증가를 억제할 수 있습니다.(〈그림 23〉)

### (3) Cycling & Retention 특성

Cycling 특성은 F.G 대비 우수한 특성을 보이고 있습니다. 주 요인은 2 가지로 MANOS 구조에서 ① 얇은 두께의 Tunnel Oxide 적용에 따른 Oxide 내 Trap density 감소와 ② 초기 Program Vt 분포가 F.G 대비 월등하게 좁아서 Cycling 후 분포가 넓어져도 Read 판별이 가능한 점입니다.

Retention 특성은 F.G 대비 열화 된 특성을 보입니다. 주 요인은 4 가지로 ① Tunnel Oxide 를 산화 방식 대신 증착 방식으로 형성하여 Oxide quality 열화, ② Poly Si 적용에 의한 Channel interface 특성 열화, ③ Storage Nitride 와 Box 에서의 복잡한 Charge trap 현상, ④ 이웃한 Cell 과 Storage Nitride 가 서로 연결된 구조로 인해 발생하는 Lateral



〈그림 23〉 Plug aspect ratio 개선 & Cost down 을 위한 적층 layer 폭 감소<sup>[11]</sup>

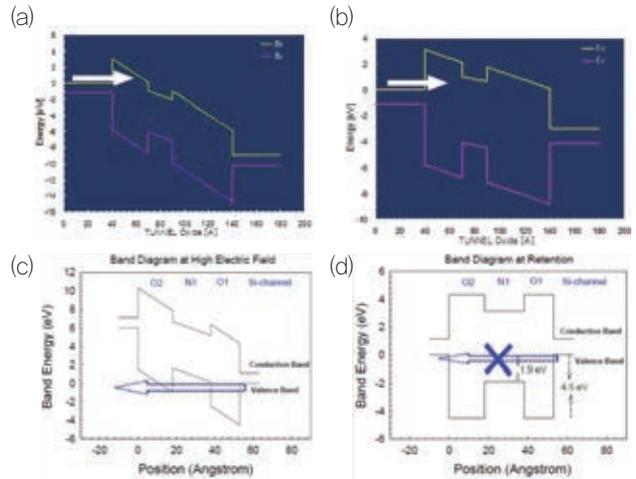


Charge Spreading 현상입니다. 상기 요인은 3D CTD 의 구조적인 문제점으로 현재는 알고리즘과 우수한 cycling 특성으로 대응하고 있지만 개선을 위한 요소 기술 개발이 필요한 상황입니다.

(4) Tunnel Oxide 의 Band engineering

Tunnel Oxide 영역에서의 High Electric Field 는 쉬운 Program 과 Erase 동작 동작 구현을 위한 필수 조건입니다. 그러나 Oxide 두께 감소만으로 High Electric Field 를 구현하면 유사한 원리로 Low Electric Field 에서 비 선택 Cell 에서 Storage Nitride 에 electron 유입으로 인한 문턱전압이 shift 되는 Disturb 현상과 Storage Nitride 부터 de-tarp 에 의한 Data 보존 능력 저하 현상과 같은 부작용이 발생합니다. 이를 극복할 수 있는 방법으로 High E-Field 에서는 Tunneling 특성을 강화하여 높은 Program  $V_t$ 와 더 낮은 Negative Erase  $V_t$ 를 구현하면서도 Low E-Field 에서는 Tunneling 특성을 제어하여 Disturb 와 Data 보존능력을 개선하는 win-win 전략을 위한 Tunnel Oxide 구조 변경을 진행하고 있습니다(〈그림 24〉).

Tunnel Oxide 구조를 기존의 Oxide 단일막 구조에서 3중막 구조로 변경하는데, Oxide와 Oxide 사이에 Oxide 보다 Band Gap 이 작은 material (주로, Nitride)을 형성하는 구조입니다. High Field 조건에서는 Si 의 Conduction Band energy level 에서 Tunnel Oxide 의 중간에 Conduction Band energy 감소 영역이 존재하여 Tunneling Barrier width 가 감소하므로 FN tunneling electron 이 증가하여 높은 Program  $V_t$  특성을 갖게 됩니다. 하지만 Read bias or Pass Bias 와 같이 상대적으로 Low field 에서는 Si 의 Conduction Band energy level 의 Barrier width 가 동일하므로 Tunneling 특성도 변하지 않아 문턱 전압 shift 가 증가하지는 않습니다. 동일한 개념으로 Erase 와 Retention 특성을 비교 할 수 있습니다. High Field 가 인가되는 Erase mode에서는 Band bending 으로 3-layer 에서 단지 Si-Channel 방향의 Oxide 만 Barrier로 작용하여 Barrier 폭이 감소함으로 Channel 로부터 Hole tunneling 이 증가하여 더 낮은 Negative Erase  $V_t$  를 갖지만, Low field 인 Retention mode 에서는 Band Bending 이 없어 3-layer 모두 barrier

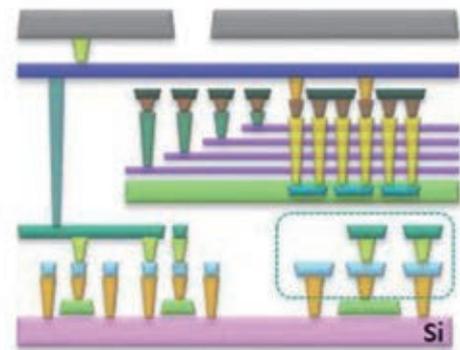


〈그림 24〉 Band engineering Tunnel Oxide 에서 Pgm / Erase 특성 (a) High Field (Program mode), (b) Low Field (Disturb mode), (c) High Field (Erase mode)<sup>[12]</sup>, (d) Low Field (Retention)<sup>[12]</sup>

로 작용하여 기존과 동일한 특성을 갖게 됩니다<sup>[12]</sup>. Tunnel Oxide 의 Band engineering 기술은 CTD Cell, F.G Cell 에서 도약의 기반을 제공하는 기술입니다.

(5) Peripheral Under Cell (PUC)

Si 기판에는 Cell 동작을 위한 Peripheral Transistor 들을 배열하고, 그 위로 증착 공정을 통해 Cell array 를 구성하는 방안 입니다. 적절한 Metal 배선 구성과 함께 Peripheral 부면적을 큰 폭으로 줄일 수 있는데, 3D array 에서만 구현할 수 있는 구조로써 큰 폭의 Chip size 개선이 가능하여 Cost down 면에서 2D 대비 3D array의 큰 요소 기술입니다(〈그림 25〉).



〈그림 25〉Peripheral Under Cell (PUC) 개략도<sup>[10]</sup>



## V. Summary

고집적화를 위한 Floating Gate NAND 개발과정에서 몇 차례 기술적 한계상황에 직면하였었지만, Air-Gap, Double patterning, Multi-level Cell, Error Correction Code 과 같은 breakthrough idea 을 활용하여 1Xnm까지 성공적인 scale-down 을 하였고 10nm 까지도 바라보고 있지만, 10nm 미만으로는 적절한 방안을 찾지 못한 상황입니다. CTD 의 3D NAND Flash는 Aspect Ratio, Poly channel의 intrinsic 특성, Data 보존 능력 등 해결 해야 할 issue 들이 남아 있지만, F.G Flash 의 지난 20년간 Lesson-learn 과 Band engineering, Channel Si, PUC 의 요소기술 개발 및 System algorithm 개발, QLC 개발 등을 통하여 F.G Flash 를 넘어 지속적인 Cost-down 이 가능할 것입니다.

### 참고 문헌

[1] S. Hong, "Memory Technology Trend and Future Challenges", IEEE IEDM Tech. Dig., pp 292-295, 2010.

[2] H. Tanaka et al., "Pipe-shaped BiCS Flash Memory with 16 Stacked Layers and Multi-Level-Cell Operation for Ultra High Density Storage Devices", SOVT, pp 136-137, 2009.

[3] J.H. Jang et al., "Vertical Cell Array using TCAT(Terabit Cell Array Transistor) Technology for Ultra High Density NAND Flash Memory TCAT", SOVT, pp 192-193, 2009.

[4] E.S. Choi, "Device Considerations for High Density and Highly Reliable 3D NAND Flash Cell in Near Future" IEDM, p211, 2012

[5] S.K. Lee, "Scaling Challenges in NAND Flash Device toward 10nm Technology" IMW, p6,2012

[6] J.D. Choi, 2010 IMW Short course

[7] S.W. Park, 2012 Flash Memory Summit, keynote2

[8] D. Kang et al., "The Air Spacer Technology for Improving the Cell Distribution in 1 Giga Bit NAND Flash Memory", NVSMW, pp. 36-37, 2006

[9] J.-D. Lee, S.-H. Hur, and J.-D. Choi, "Effects of floating-gate interference on NAND flash memory cell operation," IEEE Electron Device Letters, vol. 23, no. 5, pp. 264-266, 2002

[10] Y. Fukuzumi et al. "Optimal integration and characteristics of

vertical array devices for ultra-high density, bit-cost scalable Flash memory" IEDM Tech. Dig., 449-52 (2007)

[11] S.K Park, "Technology scaling challenge and Future prospects of DRAM and NAND flash memory" IMW 2015

[12] H.T. Lue et al, "BE-SONOS: A bandgap engineered SONOS with excellent performance and reliability", IEDM Tec. Dig., pp 547 - 550(2005)



이희열

- 1987년 2월 고려대학교 이학사 (물리학)
- 1990년 2월 고려대학교 이학석사 (고체물리)
- 1993년 8월~현재 SK하이닉스 수석연구원  
FG NOR Flash, FG NAND Flash, 3D CTN NAND Flash 개발

<관심분야>  
NAND Flash Cell 개발, CTN Cell 특성, NAND Cell 동작 Algorithm



박성계

- 1988년 02월 경북 대학교 전자공학과 졸업
- 1990년 02월 한국과학기술원 전기전자공학과 석사 졸업
- 1994년 08월 한국과학기술원 전기전자공학과 박사 졸업
- 1994년~현재 SK 하이닉스 R&D 상무

<관심분야>  
NAND Flash 메모리 소자/제품 개발