

DRAM의 한계

박 영 준

(서울대 공대 전자공학과 조교수)

1. 서 론

DRAM은 본 특집호에서 여러번 강조된 바와 같이 반도체 기술의 Drive force로서 중요한 제품이다. 즉 세대별(16K→64K→256K→1M→4M→16M 등)로 개발→양산 시기가 미리 예측됨에 의해, 단위 칩당 메모리 셀의 증가에 따르는 기술적 발전이 Drive 당 하게 되었고, 이에 의해 전반적 반도체 기술 수준이 높아지게 되었다는 것이다. 또한 DRAM의 집적도의 증가에 의해 단위 byte당의 가격이 현저하게 저하됨으로써, 시스템의 설계가 Random Access 기억소자를 큰 Overhead가 없이 사용할 수 있게 되었고, 이에 의해 새로운 시스템 출현의 Drive force가 되었다고 할 수 있다. 이의 대표적인 예로 미래의 TV라고 할 수 있는 HDTV(High Definition Television)의 출현이 가능해진 것을 들 수가 있다.

여러번 인용된 바와 같이 DRAM의 세대를 특징짓는 변수로는 Cell의 면적, 최소 선폭, Chip의 면적(크기)를 들 수가 있고, 이 변수들이 년도별로 어떻게 발전되어 왔는가를 보이는 것이 흔히 사용되는 DRAM 추이표이다. 이 DRAM의 추이표는 많이 사용되는 만큼, 진부한 면도 있을 수 있으나 DRAM의 한계를 생각할 때 반복적으로 인용할 것이기 때문에 그림 1에 모아두기로 한다.

그러면 DRAM의 한계, 즉 이 세대별 집적도의 증가의 한계가 어디에서 오는가? 견해의 차이는 있겠

지만 근본적으로 4개의 칩의 가격과 이를 Printed Circuit Board(PCB)에 넣는 가격(Package 및 P.C. Board의 면적)이 4개의 칩의 집적도에 해당하는 1개의 칩과 1개의 Package 및 PCB의 면적보다 작아지게 될 때가 집적도 증가의 한계가 된다고 볼 수가 있다. 즉 4MDRAM chip 네 개를 Package하여 PCB에 꽂아 넣을 때 드는 가격이 16M DRAM 한개의 칩에 대한 그것보다 싸지게 되면 16M DRAM의 한계가 오게 된다고 할 수가 있다.

또 하나의 한계는 기술적 한계이다. 위에서 언급한 한계가 경제성에서 오는 한계라면 기술적 한계란 실제로 합리적인 칩의 면적내에 원하는 집적도를 성취하기 위해서 필요한 합리적 기술적인 부분에 의한 한계이다. 여기서 합리적이란 실험실내에서 제작할 수 있는 기술이 3-4년 정도의 노력으로 양산으로 전달될 수 있는 기술을 말한다. 즉, 어떤 기술이 수 atomic layer를 제어해야 하는 것에 관한 것이라면 이는 실험실에서는 가능하다 하더라도 양산에서는 적합하지 않으므로 이에 대한 근본적인 부대 기술이 성취되지 않고서는 합리적인 기술이라고 말할 수는 없는 것이다.

마지막으로 생각할 수 있는 한계는 이론적인 한계에 관한 것이다. 이는 기술적으로 무한히 작은 크기의 소자를 만들 수 있다고 할 때, 이론적(혹은 물리적)으로 소자의 동작에 한계가 있을 수 있다는 것이다.

물론 본 소고는 한개의 MOS 트랜지스터 스위치와 이에 의해 스위칭되는 커패시터에 의해 구성되는 1T Cell에 기초를 두고 있다. 이 DRAM Cell은 IBM의 R. H. Dennard¹⁾에 의해 발명되었는데 이에 의해 4K DRAM이 출현되었으며 그 이상의 세대에서는 이 1 트랜지스터 Cell이 DRAM의 표준이 되었다. DRAM, 혹은 대용량 RAM이 1T Cell로만 되어 있어야 한다는 이유는 없다. 그러나 점칠 수 있는 미래의 대용량 RAM의 기술이 1 트랜지스터 Cell에 머무를 것이기 때문에, 이에 대한 한계에 대해서만 다루기로 한다.

2. 일반적 DRAM Chip의 구조 및 동작원리

그림 2는 일반적인 DRAM의 Floor-Plan을 보이고 있다. 그림에서 볼 수 있듯이 크게 1 T Cell이 모여 있는 Storage Array 부분과, 이 Array에 신호를 저장하고 이를 선택하는 외부회로(Earray Interface Circuitry)로 나눌 수가 있다.

그림 2는 이 Storage Array 부분에서 1 T Cell이 어떻게 외부회로와 연결되어 있는가를 보여 주고 있다. 그림에서 X 방향, 즉 좌우로 연결되어 있는 선을 워드 선(word line)이라 하고 y방향, 즉 상하로 연결되어 있는 선을 비트 선(bit line)이라고 부른다. Cell을 선택하기 위해서는 먼저 워드 선 하나를 선택하여, 이 워드 선에 높은 전압을 가한다. 그림 2에서 이 일을 Row Decoder와 Driver가 담당하게 된다. 일단 한개의 워드 선이 선택되고 나면, 그 워드

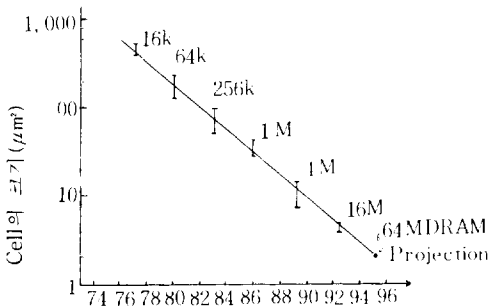
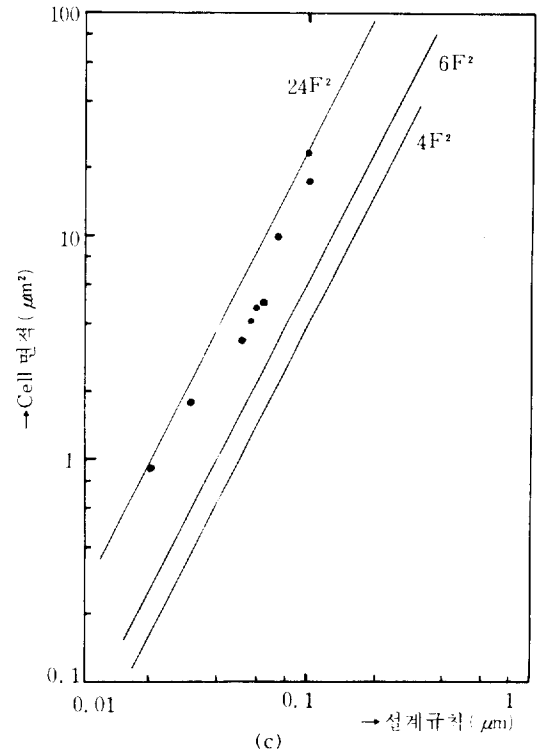
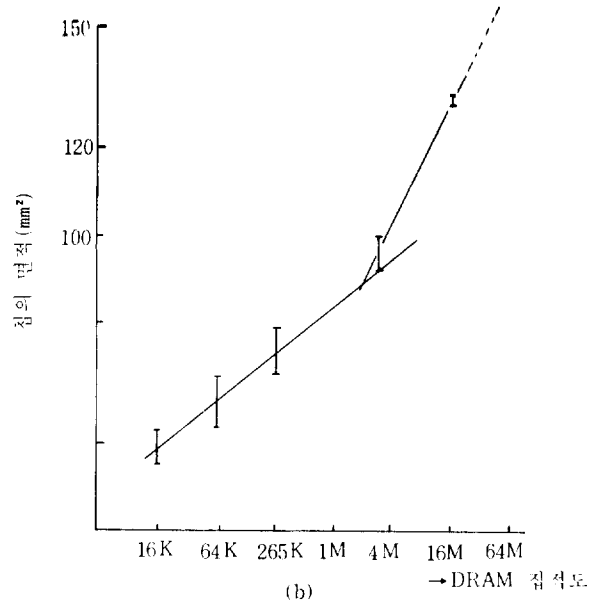


그림 1 (a) 년도별 Cell 크기의 추세
 (b) DRAM 세대별 칩 면적의 추이
 (c) 설계 규칙에 따른 Cell의 면적

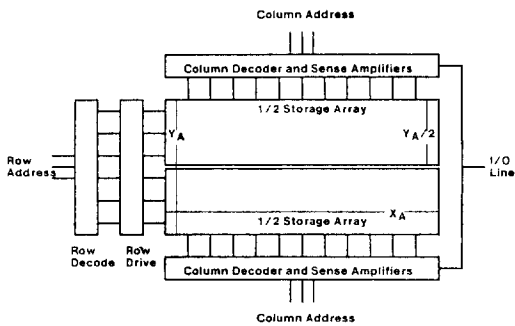


그림 2. 일반적인 DRAM 칩의 구조

선에 물려 있는 모든 Cell들에 저장되어 있는 정보(DRAM인 경우 전하의 있고 없음)가 비트 선을 통하여 감지 증폭기(Sense Amplifier)로 들어가게 되는데 이 증폭기에 의해 Cell들에 저장되어 있던 미세한 정보(전하량)가 바깥 세계로 나올 수 있는 충분한 세기로 증폭되는 것이다. 이렇게 증폭된 신호는 Column Decoder에 의해 원하는 Cell만이 선택되어 출력선(보통 입력선과 공통으로 물려 있다)으로 나가게 되어 칩의 출력단 buffer를 구동시키게 되는 것이다. 여기서 설명한 동작은 읽어내기(read) 동작에 관한 것인데 거꾸로 써넣기(write) 동작도 이와 비슷한 원리로 이루어 지게 된다. 일단 워드 선이 선택되어, 선택된 워드 선에 물려 있던 커패시터의 정보가 비트선으로 나오게 되면, 일단 모든 Cell들은 정보를 잃게 된다. 그러나 감지 증폭기가 이 Cell들의 작은 정보를 가지고 증폭을 하게 되면, 그림 3에서 보듯이 감지 증폭기의 출력이 입력으로 연결되어 있어서 읽어 내었던 정보를 다시 그 Cell에 즉시 써넣어 주게 되는 것이다.

여기서 DRAM 칩의 속도에 대해서 생각해 보자. 먼저 읽어 내기시의 속도란 그림 1에서 보이지 않았지만 buffer register가 외부의 번지 신호를 받아, 워드 선 Driver를 구동하여, 워드 선을 선택하여, 선택된 Cell의 정보가 감지 증폭기를 통하여 Column Address Decoder에 의해 I/O 선, I/O Data Butter로 나오게 되는 시간에 의해 결정되는데, 이를 Access time이라고 부른다. 기본적으로 Access time은 Row Decoder, 워드 선 그리고 감지 증폭기의 지연 시간에 의해 결정되는데 써넣기 시에는 감지 증

폭기에서의 지연 시간이 필요없기 때문에 읽어내기 시 보다는 Access time이 빠르다.

DRAM의 동작 원리에 대한 설명을 마치기 전에 몇가지 중요한 기하학적 변수에 대해 생각해 보기로 하자. 첫째, 그림 2의 Storage Array에서 보는 바와 같이 단위 Cell의 세로크기는 워드 선과 워드 선 사이의 길이(이를 워드 선 Pitch라 부른다)에 의해 결정되고, 가로 크기는 비트선 사이의 길이(비트선 Pitch)에 의해 결정된다. 또한 그림 2에서와 같이 Storage Array가 두 부분으로 나누어져 있는 경우, 이 Array의 세로를 Y_A , 가로를 X_A 라고 하면 비트선의 길이는 $0.5Y_A$, 워드선의 길이는 X_A 가 될 것이다.

단위 Cell의 면적 A_c 를 칩의 제조 공정에서 제작할 수 있는 최소 선폭 F 로 다음과 같이

$$A_c = N_c F^2$$

로 표시할 때, N_c 는 기본 면적인 F^2 몇개가 Cell을 형성하는가 하는 소위 Number of squares를 나타내게 된다. 그림 2의 Storage Array에서 Y_A 라 X_A 의 비를 Aspect Ration R_A 라 하면

$$R_A = \frac{Y_A}{X_A}$$

로 표시할 수가 있다.

표1은 역사적으로 이러한 상수들이 어떻게 변화해 왔는가를 보이고 있다. 예를 들어 Number of squares N_c 가 12인 경우(여기서 4×3 F^2 로 가정하자)에 대해 몇 가지의 계산을 해 보자.

우선 비트 선 당 Cell의 수는

$$N_c / \text{비트선} = \frac{0.5Y_A}{4F}$$

로 표시할 수 있고 $M = \frac{A_A}{12F^2}$ 혹은 $F = \sqrt{\frac{A_A}{12M}}$

이므로

$$N_c / \text{비트선} = \frac{0.5\sqrt{A_A/R_A}}{3\sqrt{A_A/(12M)}} = \sqrt{\frac{M}{12R_A}}$$

로 표시할 수가 있다. 그리고 비트 선의 수를 N_{BL} 이라 하면

$$N_{BL} = \frac{X_A}{3F} \text{ 이고 } F = \sqrt{\frac{A_A}{12M}} \text{ 이므로}$$

$$N_{BL} = \frac{\sqrt{A_A/R_A}}{2\sqrt{A_A/12M}} = \frac{1}{2}\sqrt{\frac{12M}{R_A}} = 1.7\sqrt{\frac{M}{R_A}}$$

로 표시된다. 한 개의 비트선 당 한개의 감지 증폭기가 달려 있다면 전체 감지 증폭기의 수도 N_{BL} 과 같을 것이다.

이상과 같이 DRAM의 기하학은 전체 집적도 M과 Storage Array의 Aspect Ratio R_A 로 표시할 수가 있다. 그러면 이러한 기초적 DRAM의 기하학을 배경으로 하여 DRAM의 한계에 대해 생각해 보기로 하자.

3. DRAM의 기술적 한계

DRAM의 기술적 한계는 크게 다음과 같이 세 가지로 나눌 수가 있다. 즉 DRAM의 집적도가 세 배별로 증가할 때 Cell의 면적이 줄어 들게 되고, 이 면적이 줄어들게 될 때 이 작은 면적안에 그림 2에서 보듯이 안정되게 동작할 스위칭 소자와 정보를 저장할 수 있는 커패시터를 만들 수 있는가가 그 첫째 문제가 될 것이다. 두 번째 문제는 윗 절에서 언급한 바와 같이 집적도가 증가할수록 감지 증폭기의 수가 증가하게 되는데 이에 따라 정보를 읽어 낼 때 증폭기가 동시에 동작하게 되어 전력을 소모하게 되는데 있다.

즉, 단위 면적당 전력 소모량의 증가가 한계를 줄 수가 있다는 것이다. 세번째는 비트선 Pitch와 워드 선 Pitch가 작아짐에 따라 감지 증폭기를 비트선 Pitch 안에, 워드 선 Driver를 워드 선 Pitch 내에 만들어 넣을 수 있는가 하는 문제이다. 각각에 대하여 간략하게 생각해 보기로 하자.

3.1 Cell 커패시터에 의한 한계

한개의 트랜지스터로 구성된 DRAM의 한계는 결

국 감지 증폭기가 실수 없이 Cell에 저장된 전하에 따라 동작하는가(Latch가 되는가)에서 온다. 감지 증폭기는 그림 3에서 보듯이 기본적으로 두개의 MOSFET에 의한 차동 증폭기의 모양을 가지게 되는데, 만약 한쪽 비트선의 전압이 다른 쪽 전압 보다 조금이라도 크면, Q_1 트랜지스터는 Q_2 보다 더 전류를 잘 통하게 되고, 이에 의해 Q_1 의 드레인의 전압이 낮아지는데, 이 전압이 Q_2 의 게이트에 물려 있으므로 Q_2 는 전류를 더욱 통하지 않게 되어 Q_1 의 게이트 전압은 더욱 올라가게 된다. 이런 식으로 감지 증폭기는 두 비트선의 미소한 전압의 차이에 의해, 확대 재생되게 된다. 그러나 실제로는 두개의 MOS

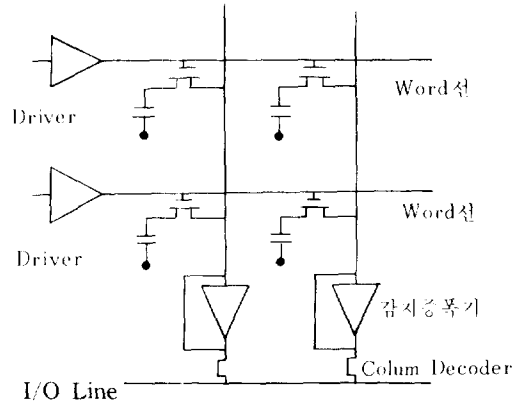


그림 3. 「DRAM의 storage Array Cell은 워드선과 비트선에 의해 선택된다」

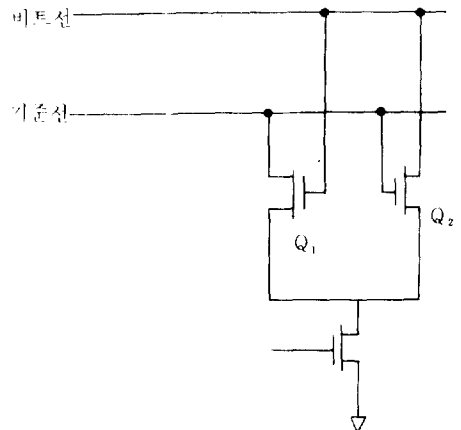


그림 4. 차동 증폭기의 원리

소자가 크기 및 문턱 전압이 공정에 의해 조금씩 다르므로 확실한 감지 증폭기의 동작을 위해서는 두 비트선의 전압 차이가 보통 $\pm 20\text{mV}$ 는 되어야 하는데 이를 offset 전압이라고 한다.

그림 3의 회로에서 보듯이 이 offset 전압이 감지 증폭기에 나타나기 위해서 Cell에 저장되어야 하는 전하량 Q_C 는

$$Q_{ST} = 40\text{mV} \times (C_C + C_B)$$

여기서 C_{BC} , C_B 는 각각 Cell과 비트선의 전체 커패시턴스이다. 만약 비트선의 커패시턴스가 800fF 라면 Q_C 는 약 30fC 이 될 것이다. 즉 이상적으로 Cell에서 30fF 만 저장하고 있으면, 이 전하가 비트선을 충전하고 나서도 40mV 이상의 전압이 감지 증폭기에 나타나게 된다는 결론이다. 그러나 Package 등에서 방출되는 방사선인 α 입자등에 의해 비트 선에 전하가 들어오게 되는데 이 전하의 크기를 이겨내고, 감지 증폭기에 올바른 정보를 주기 위해서, 이 α 입자에 의한 전하만큼이 더 Cell에 저장되어야 한다. 즉 Q_{CRIT} 는

$$Q_C - Q_{ST} = Q_{CRIT}$$

로 표시되는데 여기서 Q_C 는 Cell에서 저장된 전하이 고 Q_{ST} 는 감지 증폭기를 실 수 없이 동작시키기 위해 필요한 전하이다. 따라서 Q_{CRIT} 가 α 입자등에 의한 전하 upset을 방지할 만큼 크지 못하면, DRAM의 저장된 정보에 실수(Soft Error라고 한다)가 발생하게 되는데 이것이 DRAM의 한계가 되는 것이

표 1. Cell과 Array 파라미터

M	F(μm)	$A_A(\text{cm}^2)$	N_C	Cell
4 Kb	7	0.07	36	MOS+Cap.
16 Kb	5	0.1	32	CC
64 Kb	3	0.15	24	HI-C
256 Kb	2	0.3	24	HI-C
1 Mb	1.3	0.2	12	Planar Cell
4 Mb	0.9	0.5	15	3D Cell
16 Mb	0.6	0.8	15	3D Cell
64 Mb	0.3	1	15	3D Cell

다.

Mohne 등²⁾은 이 실수 발생율(SER)이 Q_{CRIT} 이 줄어 듬에 따라 증가함을 다음과 같은 식으로 나타내었다.

$$F_{rsa} \exp(-Q_{CRIT}/13\text{fc})$$

즉 Q_{CRIT} 이 30fc 줄어들면 SER이 10배 정도 커지게 된다는 것이다. 4MeV 에너지를 가진 α 입자에 의해 약 130fc 정도의 전하가 유지된다면 Q_C 는 적어도 160fc 정도는 되어야 한다는 결론이 나오게 된다.

그러면 이 160fc 정도의 전하량이 결국은 Cell의 면적 A_C 내에 들어가야 한다는 결론이 나오게 되는 것이다. 의미있는 DRAM이 되기 위해서 Cell의 크기 A_C 를 보면 4M DRAM 이 $13\mu\text{m}^2$, 16M DRAM 이 $6\mu\text{m}^2$, 64M DRAM 이 $3\mu\text{m}^2$ 가 되어야 한다. 그런데 A_C 내에는 Q_C 외에는 스위칭 소자 및 Cell간의 격리를 위한 면적도 포함되어야 한다. 보통 전체 A_C 중 커패시터의 면적이 40% 정도 되므로 결국 160fc 가 4 , 16 , 64M DRAM 각각에 대해 $5.2\mu\text{m}^2$, $2.4\mu\text{m}^2$, $1.2\mu\text{m}^2$ 정도가 될 것이다. 즉 160fc 을 이 면적 내에 넣어야 하는 것이다! 혹은 단위 면적당의 전하량이 각각 $31\text{fc}/\mu\text{m}^2$, $67\text{fc}/\mu\text{m}^2$, $133\text{fc}/\mu\text{m}^2$ 이 된다. Gauss의 법칙에 의해 이만한 전하밀도를 보장하기 위한 전계는 SiO_2 인 경우 $8.9\text{MV}/\text{cm}$, $19.4\text{MV}/\text{cm}$, $38.5\text{MV}/\text{cm}$ 가 된다. 여기서 주의할 것은 만약 유전체로서 SiO_2 보다 큰 유전율을 가지는 물질을 사용하는 경우, 이 전계가 줄어들 것이다. 따라서 유전율이 큰 물질을 사용하려는 노력이 당연히 이루어져야 한다. 이는 작은 전계로서도 160fc 의 전하를 저장하기 위해서이다. 문제는 유전율이 큰 물질을 절연성이 좋게 만들 수 있는가에 있다. 유전체에서의 전류의 도전 특성은 전계가 증가하면 증가하게 되는데 보통

$$J = J_0 \exp(B/E)$$

의 관계를 가진다. 따라서 커패시터에 사용되게 될 유전체는 적어도 전류가 반도체의 PN 접합의 누설 전류 보다도 작아야 한다. Noble³⁾ 등은 1985년도

표 2. Insulator의 누설 전류에 의한 최대 전하밀도의 한계

구분	J_0 (A/cm)	B (cm/M v) ^{1/2}	E (M v / cm)	K_{ins}	Q/A (fC/ μ m ²)
SiO ₂	5.1×10^{-28}	17.7	7.9	3.9	27
Si ₃ N ₄	9.2×10^{-18}	11.7	5.0	7.0	31
Ta ₂ O ₅	4.1×10^{-15}	23.4	0.7	23.0	14

까지의 SiO₂, Si₃N₄, Ta₂O₅ 등의 유전 물질에 대한 실험적 Data를 근거로 이 누설 전류의 한계를 넘지 않을 전계를 구하여 이 전계가 유지할 수 있는 전하를 구했는데 이를 보인 것이 표2이다. 이 표에서 보면 Si₃N₄ > SiO₂ > Ta₂O₅의 순으로 전하밀도가 크다. 여기에서 보면 Ta₂O₅의 유전율이 비록 Si₃N₄, SiO₂의 그것보다 우수하나, 누설 전류 때문에 단위 면적당 저장할 수 있는 전하는 오히려 떨어진다. Ta₂O₅의 특성은 1988년 Sputtering/Anodization의 방법에 의해 개선되었으나⁴⁾ SiO₂나 Si₃N₄의 수준에는 아직 요원하다. 따라서 특별한 물질의 발명이 없는 한 유전체가 단위 유지할 수 있는 전하량은 30 fC/ μ m²이다. 따라서 5.4 μ m²의 커패시터 면적이 필요하다는 결론이다. 그러면 이 전하량을 4M, 16M, 64M DRAM에 넣을 수 있는 방법은 무엇인가? 즉 각각 5.2, 2.4, 1.2 μ m²의 면적에 160fC를 넣을 수 있는 방법은 무엇인가? 전하밀도의 한계가 30fC/ μ m²이므로 필요한 면적은 5.3 μ m² 이므로 4M DRAM까지는 이론상은 평면상의 면적이 보장된다. 그러나 차세대인 16M DRAM 부터는 평면상의 면적이 이를 보장하지 못하므로, 3차원 구조의 커패시터

를 사용해야만 한다. 흔히 사용되는 방법으로 폴리실리콘 사이의 면적을 이용하거나(Stacked Capacitor), Cell의 커패시터 면적에 소위 Trench라고 하는 구멍을 뚫어 면적을 늘리는 기술을 들 수가 있다. 그림 5는 이러한 두개의 Cell을 보이고 있다. 5.3 μ m²의 면적을 보장하기 위한 수직 방향의 높이를 H라고 하면 H는 16.64M DRAM의 경우 각각 약 0.4 μ m, 1.2 μ m가 된다. 따라서 64M DRAM이 되면 Cell 커패시터의 수직대 수평 Aspect Ratio가 1이 된다. 따라서 이런 추세로 간다면 256M DRAM Cell의 커패시터의 Aspect Ratio는 4이상이 될 것이다. 실리콘 기술에서 수직대 수평의 Aspect Ratio가 2이상이 되면 이미 양산성을 잃을 수 있으므로, DRAM Cell의 한계는 이러한 Aspect Ratio 때문에 64M DRAM이 한계가 될 것이다.

3.2 스위칭 소자에 의한 한계

기본적으로 1 T Cell의 스위칭 트랜지스터는 워드 선을 통해서 높은 전압이 들어오면, ON되어 커패시터의 정보가 비트선을 통하여 나올 수 있고 그렇지 않은 경우 OFF 되어 커패시터의 정보가 보존될 수가 있어야 한다. 이러한 이상적인 ON-OFF 동작은 트랜지스터의 채널 길이가 작아지면 작아질수록 어려워지는데 이는 트랜지스터의 소스-드레인간의 전계가 증가하여 상호 간섭을 일으키기 때문이다. 이러한 문제를 해결하기 위하여 소자의 채널 도핑 농도가 커져야 한다. 그림 6은 확실한 ON-OFF 동작을 위한 도핑 농도를 채널 길이에 따라 예측해 본

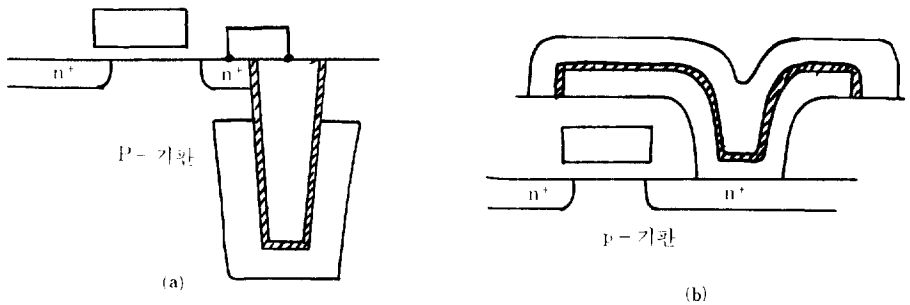


그림 5. 3차원 Cell의 모양

(a)는 Trench cell (b)는 stacked cell 빗금 친 부분이 유전체이다

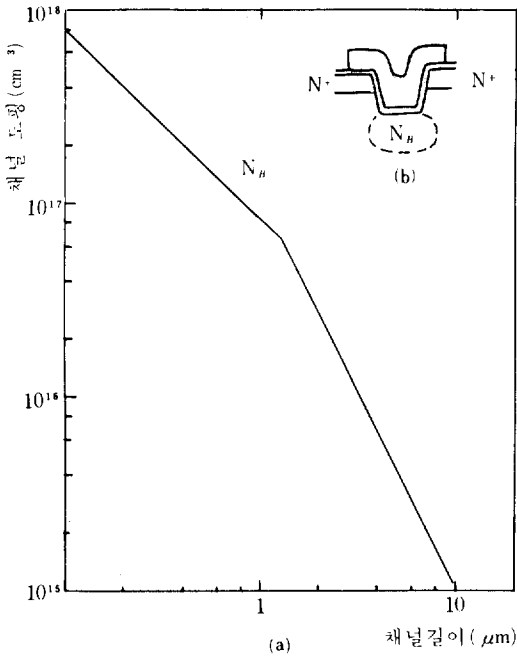


그림 6. (a) 스위칭 역할을 보장해 주기 위한 채널 도핑과 채널 길이와의 관계
(b) N⁺P 누설 전류를 개선해 주기 위한 이상적 구조

것이다. 그림에서 보듯이 채널 길이가 $0.5\mu\text{m}$ 인 경우 약 $1 \times 10^{17}/\text{cm}^3$, 채널 길이가 $0.3\mu\text{m}$ 인 경우 약 $5 \times 10^{17}/\text{cm}^3$, $0.1\mu\text{m}$ 인 경우 약 $8 \times 10^{17}/\text{cm}^3$ 가 된다. 즉 채널 길이가 짧아지더라도 도핑 농도를 높여 줌으로써 ON-OFF 동작을 보장해 줄 수가 있다. 이는 최근의 연구 결과⁴⁾로도 밝혀져 있다. 문제는 이러한 도핑 농도가 Cell 커패시터의 N⁺ 접합과 닿아 있다는데 있다. 즉 다시 말하면 누설 전류가 적은 N⁺P 접합을 만들어 줄 수 있느냐 하는 문제이다. 만약 채널 도핑 농도가 커지면 커패시터의 정보가 이 누설 전류를 통하여 손실될 것이기 때문이다.

El-Mansy 등⁵⁾은 Hi-C Cell에서 보론의 농도를 $1.5 \times 10^{17}/\text{cm}^3$ 로 제한하였다. 이는 기본적으로 항복 전압을 증가시키기 위한 것이다. 또한 N⁺P 접합에서 Tunneling 현상이 생기기 시작하는 농도가 $5 \times 10^{17}/\text{cm}^3$ 정도이므로 이 농도에 의해 채널 길이의 한계가 존재하게 될 것이다. DRAM의 전하량을 주어

진 전계에서 보장해 주기 위해서 비트선의 전압이 3V 정도에서 그 이하로 내려 가지 않는다고 하면 이 전압에서 Tunneling이 생기는 농도가 채널 농도의 한계가 될 것이고 이 채널 농도에서 스위칭 소자의 ON, OFF가 보장되는 채널 길이가 스위칭 소자의 한계가 될 것이다. 위의 계산에 의하면 농도 한계는 약 $5 \times 10^{17}/\text{cm}^2$, 채널 길이는 $0.2\mu\text{m}$ 가 DRAM Cell에서 사용되는 스위칭 소자의 한계가 될 것이고 이것이 64M DRAM의 최소 선폭이 된다(그림 1). 그러나 이는 표준적인 평면 기술을 사용했을 경우이고, 좀 더 이상적인 스위칭 소자의 구조를 그림 5(b)와 같이 생각할 수가 있다⁶⁾. 이 구조는 채널의 도핑이 소스/드레인의 N⁺ 영역과 물리적으로 닿지 않음으로 해서 위에서 설명한 문제점을 해결해 줄 수가 있을 것이다.

3.3 최소 선폭의 한계

최소 선폭이란 반도체 칩에서 트랜지스터를 제작하고 이들을 결선하기 위해서 필요로 하는 최소한의 크기를 말한다. 기억 소자가 아닌 다른 칩에서는 회로의 속도를 증가시키기 위해서 다른 부분은 그대로 두고 게이트 부분(즉 MOS 소자의 채널 길이)만 줄여 나가는 경향이 있으나 DRAM 칩의 경우는 Cell의 집적도 문제 때문에 어느 한부분만 줄이는 것이 아니라, 트랜지스터, Cell의 커패시터, Cell간의 격리, 접촉점(Contact)등을 모두 줄여나가야 한다. DRAM 칩의 공정이 어려운 것은 바로 여기에 기인하는 것이다.

그림 1에서 보듯이 DRAM의 세대별 최소선폭이 30%/세대로 줄어들음을 알 수가 있다. 따라서 16M DRAM인 경우 $0.6\mu\text{m}$, 64M DRAM인 경우가 $0.3 \sim 0.2\mu\text{m}$ 정도의 최소선폭을 요구할 것이 예측된다. 그러면 이러한 추세에 대한 실제 제작상의 한계는 무엇인가?

첫째, Lithography의 문제점을 들 수가 있다. Lithography는 크게 Mask, 광원에서 광 에너지를 받아 Mask에 의해 선별적으로 상이 전달되는 Photoresist, 그리고 광원 및 Lens 그리고 정렬을 할 수 있는 Alignment 시스템을 갖춘 Aligner(혹은 stepper 등)로 나눌 수 있다. 회절 현상에 의한 분

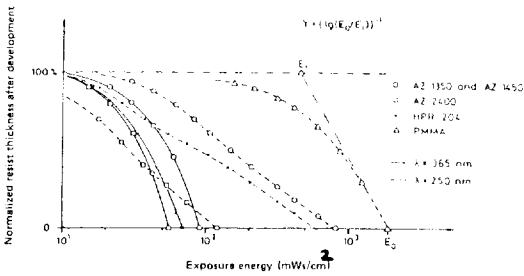


그림 7. 상용 Resist의 Deep UV Sensitivity

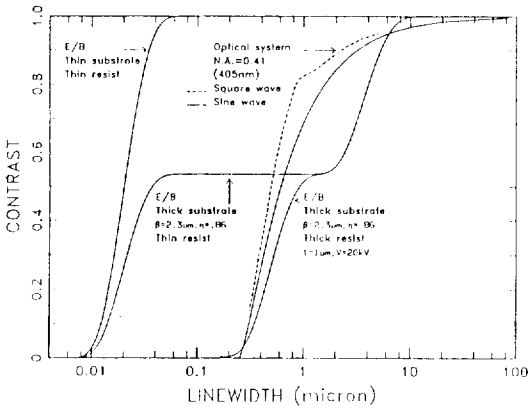


그림 8. E-beam과 광학 Lithography 시스템에서 선폭과 대비율의 관계

해능(resolution)의 감소를 막기 위해 Lithography 시스템의 광원이 Deep U. V., X-Ray, E-beam으로 변천해 나갈 것이라는 것은 자연스러운 현상이다. 따라서 Resist 물질은 다음과 같은 성질을 가져야 한다. 즉 광원의 에너지 스펙트럼과 Resist의 흡수율이 맞아 떨어져야 한다. 또한 원하는 스펙트럼의 빛에 대해 Resist의 대비율(contrast)이 좋아야 한다는 것이다. 예를 들어 Positive Resist인 경우 에너지를 받게 되면 Polymer가 형성되지 않아 현상되어 없어지므로, 원하는 주파수에서 에너지의 Dose에 따라 그 분기점이 sharp 하면 할수록 좋다. 이렇게 되어야만 흡수되지 않는 빛이나 난반사에 의한 패턴의 간섭이 작아질 것이다. 따라서 파장이 짧은 광원에 대한 감광율과 대비율이 좋은 Resist의 개발이 당연히 요구된다. 예를 들어 그림 7)은 현재 사용되는

Positive Photoresist를 두개의 다른 파장($\lambda=365\text{nm}$, 250nm)의 광원에 대해 Sensitivity를 보인 것으로 λ 가 365nm 경우(i-line) 보다도 λ 가 250nm 인 경우가 대비율이 나빠짐을 알 수가 있다.

광학에서 빛의 회절 현상에 의해 Pattern이 작아짐에 따라 대비율이 또 나빠지는데 이 경우 대비율은 광원의 파장이 작아질 수록, 대비율의 감소가 더디게 나타난다. 그림 8은 광학 시스템과 Electron beam시스템 대비율이 Mask의 선폭의 감소에 따라 어떻게 나빠지는가를 보인 것이다. 그림에서 보듯이 광학을 사용한 경우는 $0.5\mu\text{m}$ 정도부터는 급격하게 열화된다. 자연스러운 방향은 에너지의 원을 광원(optical source)에서 파장이 짧은 Laser나 X-Ray, E-beam으로 사용하게 하는 것이다. X-Ray 시스템에서나 E-beam 시스템에서의 해상도를 제한하는 요소는 Mask와 Wafer에서의 회절 현상과 Resist를 통과하는 경우, Resist에서 생기는 Photoelectron 혹은 전자의 충돌에 의한 대비율의 감소이다. 그림 7에서 E-beam시스템에서 Resist의 두께를 얇게 한 경우 대비율이 $0.1\mu\text{m}$ 이하까지도 감소하지 않음을 알 수가 있다. 결국 Lithography에서의 문제점은 무엇인가? 이는 대비율이 좋은 Mask와 Resist의 개발로 초점을 맞출 수가 있다. Resist의 두께는 좋은 해상도 및 대비율을 얻기 위해서 얇아져야 하는데, 위에서 언급하였듯이 수직 스케일링에 수평 스케일링 속도(Aspect Ratio가 증가)를 따라가지 못하기 때문에 Topography가 나빠지므로 Resist가 두껍게 공정을 해야 한다는 문제점에 봉착하게 되는 것이다. 이 문제를 해결하기 위하여 그림 9와 같은 1층 이상의 Resist를 이용한 공정을 생각할 수가 있다. 가장 이상적인 경우는 Photon이든 전자이든 Back Scattering이 없고, 평면화가 잘 되는 물질 위에 얇은 Resist를 코팅한 경우가 될 것이다. Mask

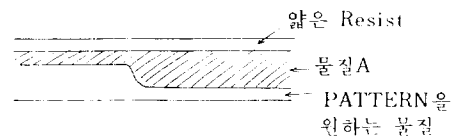


그림 9. 물질 A는 평탄화가 잘 되고 Photon 혹은 전자를 100% 흡수하면 이상적이다.

의 경우는 X-Ray Lithography 시스템의 경우 문제점으로 대두되는 데 이 역시 패턴의 정확성을 보장하기 위한 물질의 개발과 패턴이 아닌 부분이 X-Ray를 잘 통과하기 위한(대비율이 높기 위한) 물질의 개발에 관심이 달려있다. 양산을 고려하면 E-beam의 기술이 사용될 가능성은 매우 작다. 앞서도 언급하였듯이 기억 소자가 아닌 논리회로 칩인 경우는 어느 한층(예를 들어 게이트)에만 E-beam을 사용할 수가 있으나 모든 층에서 모두 E-beam을 적용하는 것은 불가능하리라 생각된다. 또한 수직 방향과 수평 방향의 Aspect Ratio의 증가에 의해 평면화가 문제가 되는데 이 평면화 기술이 완벽해지고 그림 8과 같이 back scattering이 없는 중간 물질이 안정성있게 증착되어 사용될 수 있다면 양산성이 있는 Lithography의 한계는 0.3-0.4 μ m 까지는 가리라 생각한다. X-Ray 기술은 Mask를 X-Ray가 흡수되지 않고 통과할 수 있는 물질이 얇은 막이기 때문에, 진동 등에 의해 안정된 패턴의 형성이 어렵다. 또한 X-Ray 역시 Resist에서의 2차 광 전자의 방출 때문에 Resist가 얇아야만 하는 문제가 위에서 언급한 바와 같이 상존하게 된다.

X-Ray 기술이 Mask나 Resist 등의 문제로 양산에 실용화 되기가 힘들다고 가정하고 KrF Laser를 사용한 Deep U. V.의 파장을 사용한 연구 결과를 보았을 때⁹⁾, 위의 결론과 같이 0.3 μ m 정도가 기술적인 한계로 볼 수가 있다.

4. 그 외의 문제점들

3절에서는 양산성을 고려한 기술적 한계에 대해 언급하였다. 이상에서 언급한 항목 이외에도 회로의 밀도가 증가함에 따라 칩의 Power 밀도가 증가하게 되어 회로의 PN 접합에서의 온도가 증가하게 된다. 이 온도의 증가는 Cell의 커패시터의 Storage Node의 누설 전류의 증가를 가져오고 열잡음의 증가로 감지 증폭기등의 동작을 오동작이 되도록 할 수가 있다. 전체 메모리 칩에서의 Power의 증가를 보면 1M, 4M, 16M DRAM 세대별로 Cycle 시간에 따라 다르지만 약 200n sec일 때 250MW, 300MW, 350MW 정도로 증가함을 알 수가 있다. 표준적인 DIP (Dual in-line Package)인 경우 열저항이 약 80°C/

W 정도 되므로 각각에 대해 주위 온도보다 칩내의 소자 온도가

$$P = \frac{1}{R} (J_s - T_A)$$

에서 각각 20°C, 24°C, 28°C 정도가 된다. DRAM이 최악의 경우 70°C 정도의 주위 온도에서도 동작해야 한다면 접합에서의 온도는 90°C, 94°C, 98°C 그리고 64M DRAM이 되면 100°C를 넘을 것이라는 것을 알 수가 있다.

최근의 Packaging 기술과 Cooling 기술의 발전을 보면 열저항이 DIP의 경우 보다 훨씬 낮게 만들 수가 있으므로, Power가 DRAM 집적도의 한계를 줄 것으로는 생각되지 않는다.

또한 감지 증폭기가 동작할 때 생기는 전류의 크기가 시간적으로 너무 빠르게 변화하기 때문에 이에 의한 금속 저항과 리액티브 성분 등에 의한 전압강하로 회로의 각부분의 전압 전압과 접지 전압의 흔들리게 되는 경향이 집적도가 커질수록 증가하게 된다. 이 또한 여러가지 회로기술로 최대 전류를 제한함으로써 전압의 불안정과 금속선의 열화를 막을 수가 있다.

5. 경제적인 한계

여태까지 DRAM 칩의 집적도 증가에 따른 기술적인 한계에 대해서 논하였다. 기술적으로는 64M DRAM 정도가 양산 기술을 생각할 때의 한계가 될 것이라고 하는 것은 전술한 바와 같다.

그러나 기술적인 한계 이외에도 경제적인 한계가 존재한다. 즉 서론에서도 지적하였듯이 1개의 칩의 가격이 그 전 세대의 4개의 칩의 가격보다 더 비싸지게 되고 패키징의 기술의 발전에 의해 4개의 칩을 그 Overhead 없이 PCB에 꽂을 수 있다면 기술적인 한계 이전에 다음 세대를 생산할 필요성을 잃게 될 것이다. 이에 대한 정확한 연구는 대단히 많은 시간과 자료를 요구하리라 생각된다. 따라서 본 소고에서는 이에 대한 근거 있는 추측을 유보하기로 한다. 그러나 우리가 적어도 생각해 볼 수 있는 내용은 -세대별 집적도 증가시 양산에 필요한 장비의 단

가의 증가

—세대별 면적의 증가에 따른 수율의 감소와 이를 만회하기 위한 청정도 유지비

—Hybrid Package 방법등을 이용한 Memory Module의 생산가격

6. 결 론

본 소고에서는 1T Cell을 이용한 DRAM 집적도의 향상에 따른 몇 가지의 한계 요인을 생각해 보았다. 특별한 물질의 획기적 방법이 없는 한, Cell의 수직 대 수평 Aspect Ratio가 2이상인 되고, 스위칭 소자의 채널 도핑이 $5 \times 10^{17}/\text{cm}^3$ 이상이 되는 64M DRAM이 양산기술을 고려한 한계가 될 것이다. 64M DRAM에 필요한 최소 선폭은 $0.3 \sim 0.4 \mu\text{m}$ 정도로 예측되는데, 실제로 최소 선폭에 관한 한, 한계는 이 보다 훨씬 더 작아질 것이다. 실제로 그림 8에서 보듯이 얇은 Resist를 사용하는 경우 최소 선폭이 $0.1 \mu\text{m}$ 이하로 가능하다.

실험실에서는 이러한 방향으로 Cell의 축소가 진행될 것은 물론이다. 그러나 다른 측면에서 한계가 어디서 왔는지를 보는 것은 우리들에게 DRAM에 대한 새로운 시각을 보여 주리라 생각한다. 즉 왜 DRAM Cell이 적어도 160fC 정도의 전하량을 필요로 하는가에 있다. 이는 결국 Cell의 전하를 저장하는 Capacitive Node가 실리콘 내부에 존재하기 때문에, α 입자나 열 누설 전류가 문제가 된다. 이는 스위칭 소자가 반도체 Bulk에서 만들어 지기 때문에 생기는 문제로 만약 스위칭 소자를 SOI 구조 등으로 Bulk와 전기적으로 절연되어 있다면 이 문제가 해결될 수가 있을 것이다.

또 하나의 시각은 2절에서 설명한 바와 같이 X-Y addressing을 하지 않고 마치 자기 Core 기억 소자와 같이 X-Y addressing을 동시에 하여 비트 선을 실리콘 bulk에 닿지 않고 3차원 적으로 연결하는 방법을 생각할 수가 있다. 이러한 개념의 변화가 DRAM의 한계를 극복하는 방법이 될 수도 있을 것이다.

참 고 문 헌

- 1) R. H. Dennard, "Field Effect transistor Memory," U. S. Patent No. 3, 387, 286, 4 June 1968
- 2) A. Mohsen, et. al., "A High Density, High Performance 1T DRAM Cell," IEDM, pp. 616—619, Dec., 1982
- 3) W. P. Noble and W. W. Walker, "Fundamental Limitations on DRAM Storage Capacitors," IEEE Circuit and Device magazine, pp. 45—51, Jan. 1985
- 4) Berkeley의 연구결과, Private Communication
- 5) Y. A. El-mansy and R. A. Burghard, "Design Parameters of Hi-C DRAM Cell," IEEE J. SSC, pp. 951—956, Oct. 1982
- 6) K. Sunouchi et. al., "DLC Structure for Sub-half micron MOSFET," IEDM, pp. 226—229, 1988
- 7) P. Van Pelt, SPIE, Vol. 275, p. 150, 1981
- 8) A. N. Boers, IEDM, p. 2, 1980
- 9) M. Sasago et. al., "New High Transparent Resist and Process Technology for KrF Excimer Laser Technology," IEDM, pp. 88—91, 1988