

반도체 메모리의 전망

유영갑(충북대학교)

I. 요약

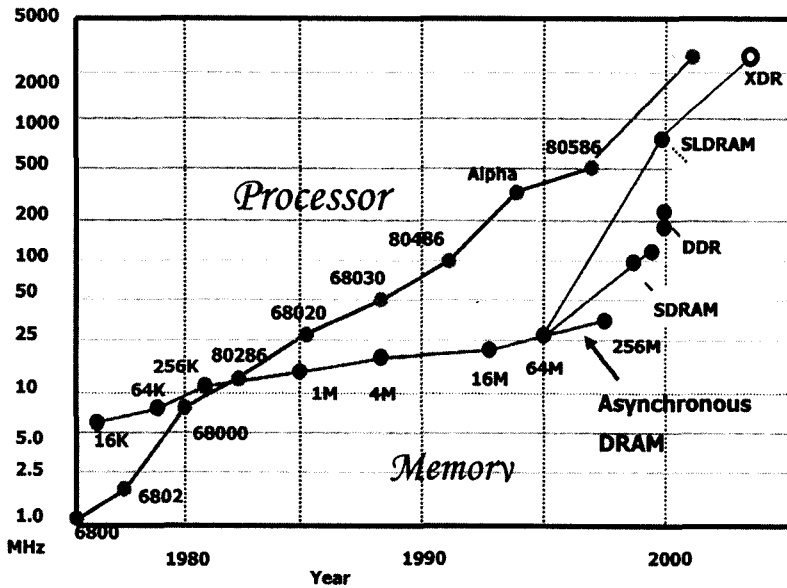
반도체 집적회로 설계 자동화 기술의 발전은 반도체 상당 수준의 시스템 구현 가능성을 비 전문가들에게 열어 주었다. 알고리즘 수준의 시스템 정의가 가능하다면 이것을 하드웨어로 만드는 것이 어렵지 않게 되었다. 시스템 설계에서 가장 핵심이 되는 메모리의 활용은 이들 비 전문가들에게 넘어서 할 큰 장애물이다. 이 글은 반도체 메모리 기술의 발전 전망을 예측하여 줌으로써 시스템 기술자들이 쉽게 어려움을 극복하도록 도와주는 것이 목적이다. 메모리 기술에 대한 접근을 쉽게 해주는 몇 가지 방법을 소개 하였다. 시스템 구성에서의 메모리, 메모리 칩의 기술 등을 요약하였다. 시스템에서의 요구 사항을 들어주는 것을 바탕으로 장래를 전망하였다.

II. 방앗간과 곳간 사이의 문제

폰 노이만은 인간의 두뇌기능을 정보저장

기능과 정보처리 기능으로 나누어 구현하였다. 이 구조는 찰스 배비지가 analytic engine에 사용하였던 방앗간(mill)과 곳간(store)의 개념을 그대로 복제한 것이다. 시스템의 구성을 데이터를 저장하는 곳간 즉 메모리와 데이터를 처리하는 방앗간 즉 처리장치로 분리하는 개념이다. 인간의 두뇌에는 특별한 부위가 정보의 저장만을 담당하거나 정보처리만을 하는 기관이 존재하지 않는다. 폰 노이만이 두뇌기능에 대한 정의에서 이 두 기능을 물리적으로 갈라놓은 것이다. 지금까지 실생활에 적용된 거의 모든 컴퓨터는 이 폰 노이만 구조에 기반을 두고 있다. 인간 두뇌를 모방하려는 시스템 설계 목표는 시작부터 단추가 이상하게 채워진 것이다.

메모리가 문제되는 것은 폰 노이만의 컴퓨터구조 제안에서 비롯된다. 이 문제는 폰 노이만 병목현상(von Neumann bottleneck)이라고 부른다. 데이터는 메모리에서 처리장치로 끊임없이 왕복하게 되었다. 인간의 두뇌에서는 저장장치와 처리장치가 분리되지 않았기 때문에 이런 문제가 발생하지 않는다. 처리장치



〈그림 1〉 메모리와 프로세서의 동작속도 차

의 성능이 좋아져서 일정한 시간 안에 처리되는 데이터가 많아지면 많아질수록 메모리-처리장치 간의 교통량 증가가 나타난다. 메모리 처리장치의 분리는 두 가지 요인에 의하여 시스템 성능에 영향을 미친다. 첫째 데이터 처리 속도가 빨라질수록 메모리 처리장치간의 데이터 교통량의 증가한다. 둘째 메모리의 동작 속도는 정체되어 있는 동안 처리장치의 처리 능력은 지수함수적인 개선을 이루어 왔다. 처리장치와 메모리의 동작속도 차이가 문제를 더욱 심각하게 만들고 있다. 최근에 이르러서 이 차이를 극복하려는 메모리 기술의 발전이 있지만 아직 근본적인 해결책으로 보기에어설픈 구석이 있다.

메모리는 차세대 반도체 집적회로 설계의 핵심으로 자리매김할 것이다. 기본 소자의 크기가 줄어들면서 메모리의 용량을 충분하게

늘릴 수 있게 되어 기능 구현에 필요한 제약이 약화된 것이다. 메모리의구성과 성능문제, 기술발전에 대한 논의 없이는 차세대 반도체 기술의 전망이 정확하지 않게 된다. 메모리가 시스템 성능에서 미치는 문제를 바탕으로 미래 메모리 기술을 전망해 보기로 하자. 기존의 폰노이만 컴퓨터 구조를 바탕으로 하는 메모리 시스템의 발전 방향을 먼저 보고 나서, 폰노이만 구조가 아닌 진보된 형태의 메모리 기술의 문제를 보기로 하자.

미래의 반도체 기술을 예측하는 ITRS (International Technology Roadmap for Semiconductor) 는 Moore의 법칙에 따라 CMOS 기술 발전이 지속 될 것으로 전제로 7년 뒤인 2014년의 메모리를 다음과 같이 보고 있다. 최소 선 폭 50nm을 기반으로 64Gbit DRAM chip이 만들어 질 것으로 본다. 트랜지

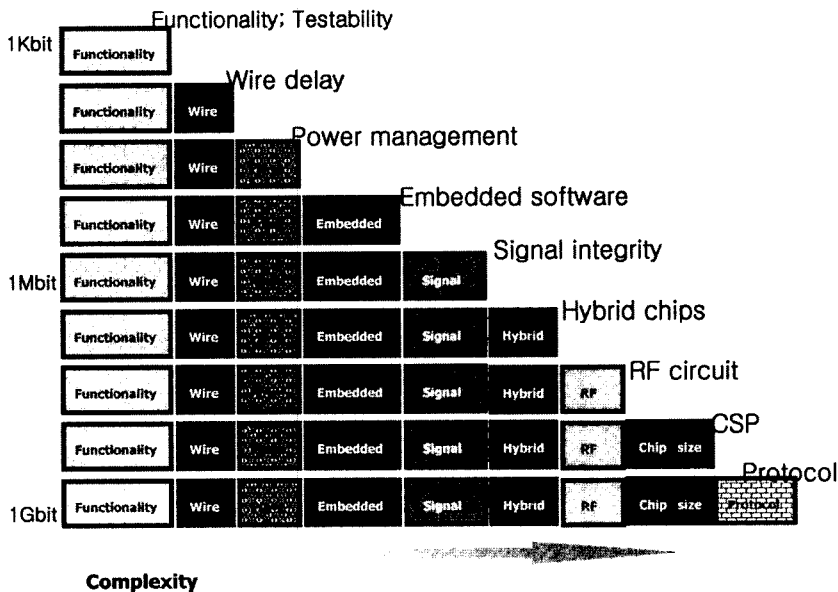
스터의 밀도는 107 TR/mm²에 해당한다. 처리장치의 동작속도는 대략 10GHz에 이르게 될 것이다. 이는 100psec의 cycle time을 의미한다.

III. 설계의 문제

메모리 시스템 구성이나 설계의 문제는 주로 이 폰 노이만 병목현상 방지 내지는 그 데이터 트래픽에 의한 성능저하 문제를 완화시키는데 집중되어 왔다. 처리장치 가까이에서 소규모 고속 메모리를 두고 여기에 자주 사용하는 데이터나 프로그램의 일부를 두는 것이다. 이런 종류의 메모리는 캐쉬(cache)라고 부른다. 또한 여러 개의 독립적인 메모리 뱅크를 두고 연속적인 메모리 접근을 가능하도록 하

는 것도 있다. 메모리 인터리빙(memory interleaving)이라고 부르는 기법이다. 메모리와 처리장치 간의 버斯拉인의 수를 늘려서 한꺼번에 많은 양의 데이터가 전달되도록 하는 방법 등이다.

메모리 칩과 시스템 구현에 통신 시스템에서 사용하고 있는 다양한 기술이 도입되었다. 동기화 기술, 프로토콜 처리 기능, 임피던스 매칭 기능, 그리고 가변 signaling 방식 등이다. 메모리에 클럭으로 사용하여 동기 기능이 강화되었다. 과거 상당기간 동안 메모리는 주소 스트로빙 신호에 의하여 메모리 기능이 가동되었다. 메모리 동작의 전 과정이 비 동기 식이었던 것이다. 지금은 클럭에 의하여 내부 동작은 물론 데이터 출력도 외부 시스템과 동기되고 있다.



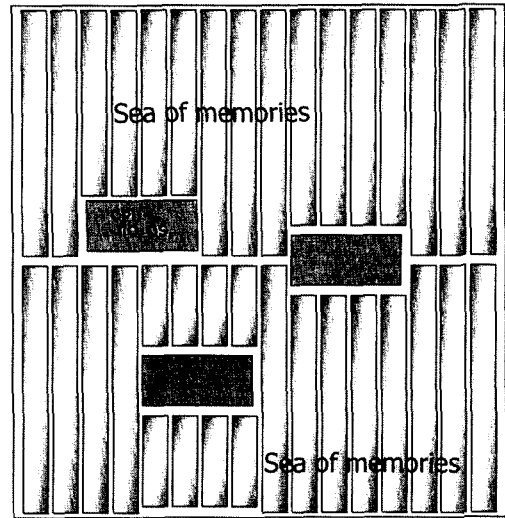
〈그림 2〉 메모리 설계 변수의 복잡화 추세 [2]

처리장치와 사이에 데이터 전송이 bit stream 방식이 아닌 패킷 전송 방식이 되면서 데이터 전송량을 획기적으로 높일 수 있었다. 패킷 크기는 처리장치의 데이터 버스의 폭에 따라 달라진다. 패킷 전송에 필요한 외부와의 클럭 동기화를 위하여 도입된 대표적인 회로가 DLL(delay locked loop) 회로이다. 이것은 기존의 PLL(phase locked loop) 방식을 디지털 신호의 특성에 맞추어 간소화한 것이다.

메모리 설계에 필요한 기술의 범위는 지속적으로 넓어져 왔다. 그림 2는 이런 추세를 요약한 것이다. 초창기에는 기능적 동작 유무만을 고려하면 되었다. 메모리의 용량이 커지고 고속화 과정을 거치면서 신호의 무결성(signal integrity)이 중요한 설계 변수가 되었다. 고속 회로에서나 볼 수 있었던 문제가 메모리 설계의 변수가 된 것이다. 저전력 설계의 문제, 칩 사이즈 패키지, 프로토콜 등이 모두 같이 고려되어야 한다. System의 논리 기능 일부가 메모리 안에 수용되면서 검증의 문제가 더욱 중요하게 되었다. 논리회로만으로 구성된 칩 설계보다 더 많은 설계 변수를 가지게 되었다.

IV. System on Memory

메모리의 용량의 제약을 극복하기 위하여 도입된 가상 메모리(virtual memory)는 운영체제의 메모리 관리 체계의 핵심요소가 되었다. 메모리의 가격이 하락하면서 메모리의 용량 확대가 부담이 줄어들게 되었다. 저가의 컴퓨터까지도 충분한 양의 메모리를 가질 수 있게 되었다. 가상메모리의 메모리 관리 체계는 이제 캐쉬 메모리 관리용으로 용도가 전환되고

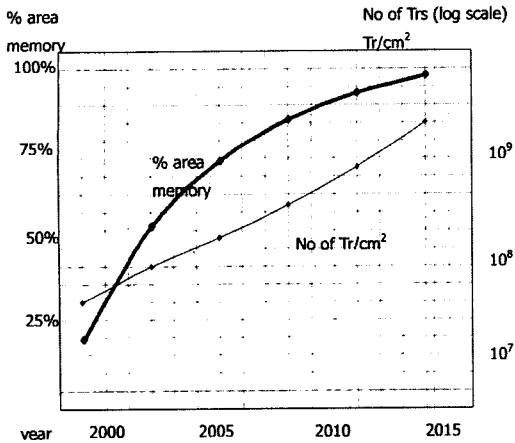


〈그림 3〉 Sea of memory 구조^[2]

있다. 새로 캐쉬로 불러 올 페이지를 위하여 사용 가능성이 낮은 페이지를 캐쉬 메모리에서 내보내는 기능 등이다. 이 시스템은 하드웨어로 구현되지만 실리콘 면적의 증가는 크지 않다. 대부분 운영체제기능은 하드웨어 논리 회로로 구현되어도 칩 크기에 미치는 영향이 미미한 것이다.

주변회로의 크기가 획기적으로 늘지 않는 한 시스템 구성에 필요한 논리소자의 양은 큰 변동이 없게 된다. 그러나 메모리 수요는 과거 하드 디스크에 저장하였던 데이터와 명령어를 처리장치 가까이에서 두기 위한 노력을 반영하여 급격하게 커질 것이다. 단일 칩화 함으로써 얻어지는 제조비용의 절감과 성능 개선이 이루어 지기 때문이다. 칩 면적의 대부분은 메모리가 차지하고 논리소자는 드문드문 자리잡는 구조가 될 것이다^[2].

반도체 메모리가 SoC 칩 안에서 차지하는 면적 비율은 계속 증가하고 있다. 수년 전에



〈그림 4〉 SoC 칩 안에서 메모리가 차지하는 면적 비율의 증가추세^[2]

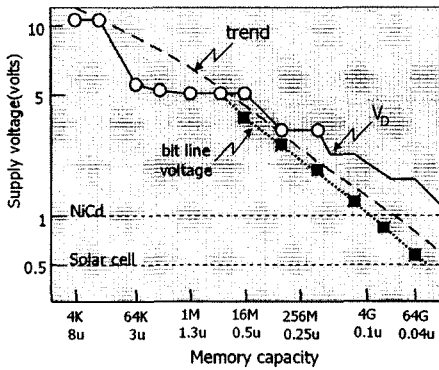
발표된 Intel 사의 Itanium processor도 칩 면적의 반 이상을 메모리로 채우고 있다. 2014년 경이 되면 SoC 칩 면적의 95%를 메모리가 차지하게 된다는 보고가 있다. 이것은 System on Chip(SoC)의 개념이라기 보다는 System on Memory(SoM)의 개념이 될 것 같다. 사용하게 될 처리장치나 주변 회로 등을 모두 합쳐도 전체 면적의 5% 내외로 충분할 것이라는 예상이다. 나머지 사용 가능한 면적은 칩 외부에 두었던 메모리를 칩 안에 내장 시키는 것이다.

SoM 방식의 도입이 가져오는 좋은 점은 칩 또는 인쇄회로 제조 단가의 절감을 통하여 시스템 제조 비용이 줄어 든다는 것이다. 여기에 이어서 저전력 고속이 이루어 진다. 칩 밖에 있는 인쇄회로 기판의 거대한 저항, 캐패시터를 구동하기 위한 전력 소모를 현저하게 줄일 수 있다. (칩 소모 전력의 상당 부분이 출력 구동회로에서 소진된다.) 또한 물리적 거리를 단축시켜서 전송 거리의 제공에 비례하는 지연 시간을 줄이게 된다. 구동해야 하는 용량성 부

하(capacitive load)의 크기를 획기적으로 줄일 수 있어서 부하의 총 방전에 소요되는 시간이 짧아지게 되고 따라서 속도개선 효과가 커지게 된다.

메모리 안에서 정보가 저장된 위치를 찾는 방식이 바뀌게 될 것이다. 기존의 메모리에서는 정보의 저장 위치는 주소에 의하여 고정된다. 정보가 저장된 위치정보 즉 주소를 잃으면 정보는 찾을 수 없다. 오류정정 부호 없이는 손상된 정보의 복구가 지극히 어려운 것이 지금의 메모리 기술이다. 앞으로는 인간의 두뇌처럼 주소 없이 정보를 저장하는 방식의 도입이 광범위해 질 것이다. 지금도 연상메모리(associative memory)에서는 정보의 일부를 주소로 활용하는 방식을 사용하고 있다. 메모리 안에 주소를 다루는 decoder 대신에 저장된 정보와 찾고자 하는 일부 정보를 맞추어 보는 매칭 회로(matching circuit)가 사용되고 있다. 이 연상 메모리의 개념은 더욱 확대되어 정보의 추상화나 여러 정보의 연관을 통하여 접근하게 되는 기능을 가지게 될 것이다. 정보자체의 종류나 특성에 따라 관련 정보를 같이 저장하는 방식의 도입이 가능해 질 것이다. 관련되는 정보가 여러 군데 분산되어서 저장될 것이다. 한번 손상된 정보는 오류정정 부호가 들어 있지 않아도 관련 정보의 처리에 의하여 어느 정도 복구 가능하게 된다. 인간의 기억이 뇌 세포의 파괴에도 불구하고 상당 부분 되살릴 수 있는 것과 같은 개념이다.

대형 시스템의 구현에서 다루게 될 중요한 과제가 저전력화이다. 메모리를 사용하지 않을 때의 소모 전력 절감과 동작 시의 전력 감소를 목표로 꾸준한 노력이 기울여졌다. 메모리를 사용하지 않을 때의 전력 소모는 궁극적



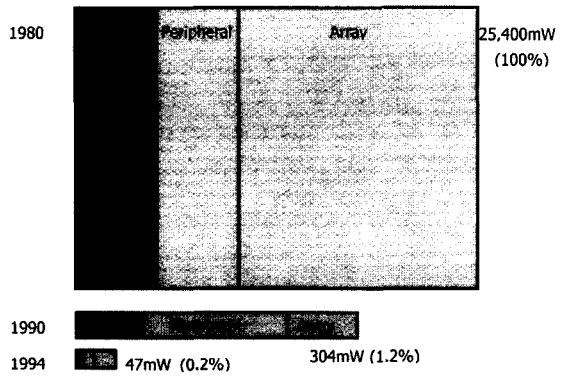
〈그림 5〉 메모리 동작 전압의 변화^[2]

으로 MRAM, FRAM 또는 PRAM 같은 비 휘발성 메모리 기술이 활용될 것이다. 동작의 전원은 태양전지의 출력 전압을 목표로 진화하고 있다.

소모전력의 절감도 획기적으로 이루어져 왔다. 내부 동작 전압의 감소, 메모리 array의 부분 동작 등이 주로 활용되었다. 다음 그림은 64Mbit 로 환산하였을 때의 소모 전력 감소 과정을 그림으로 보이고 있다. 그림에서 절대면적의 비율이 곧 소모 전력 비율을 나타낸다. 1980년의 DRAM에 비하여 1994년의 DRAM은 0.2%의 전력만을 소모하고 있다 약 500배 소모전력 절감을 이룩한 것이다. 이 경향은 그 이후에도 지속적으로 이루어지고 있다

V. 대용량화 추세

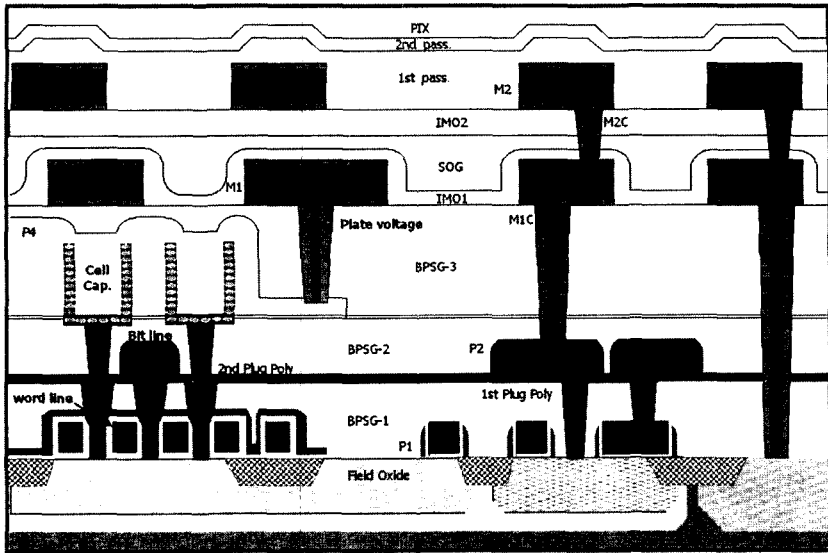
단위 면적 당 정보저장 능력을 확대하기 위한 방안으로 기본 소자의 크기를 줄이는 것과 함께 회로를 단순화 하는 기술 혁신이 개입된



〈그림 6〉 DRAM의 소모전력 절감^[2]

다. 회로 소자의 크기는 최소 선 폭의 축소에 따라 이루어진다. 최소 선 폭 축소는 ITRS 자료에 잘 나타나 있다. 이것은 제조 업체가 시장 경쟁력을 유지하기 위하여 기울이는 부단한 노력을 반영한 것이다.

반도체 집적회로의 구현은 실리콘 표면에서 이루어지기 때문에 구성요소가 차지하는 면적을 줄이는 것이 중요하다. 최소 선 폭의 크기를 F 라고 하면 현재의 4F²의 크기를 가지는 메모리 셀과 구조가 사용되고 있다. 이 표면적 셀의 축소는 3차원 구조를 효과적으로 만들어 왔다. DRAM의 경우 cell capacitor의 형태에 따라 trench cell 이나 stacked cell 형태로 구현되어 왔다. Trench cell의 경우 trench의 aspect ratio 즉 trench의 구멍의 직경과 깊이의 비율이 수십 대 일까지 높아지고 있다. 과도한 aspect ratio는 칩 하나에 수십억 개의 구멍을 파야 하는 공정 부담으로 인하여 반도체 제조 후에 얻어지는 칩 수율 향상에 과도한 부담으로 나타난다. 대부분의 DRAM은 trench cell이 아닌 stacked cell을 채택하는 이



〈그림 7〉 전형적인 3차원 구조의 COB 형 DRAM 셀^[2]

유가 여기에 있다.

DRAM cell 에서 trench cell 구조가 지하 100층 건물을 건설하는 것이 비유한다면 stacked cell은 지상 100층 건물에 비유할 수 있다. 당연히 지하 건물 보다 지상 고층 건물이 건축하기 쉽기 때문에 도시들은 고층 건물로 채워지고 있다. Stacked cell 구조에서는 cell 의 구성 요소들을 칩 표면 위에 차곡차곡 쌓아두는 방식이다. 특히 bit line과 cell capacitor의 위치에 따라 CUB(capacitors under bit lines)와 COB(capacitors over bit lines)로 나누어 진다. 상부 층의 평탄화 문제 등으로 COB 구조가 널리 채택되고 있다. Stacked cell 구조가 거의 모든 상업목적 대용량 DRAM 칩에 사용되는 것은 trench cell 구조에 비하여 제조가 쉽다는 것이다.

회로 크기의 축소가 지속적으로 이루어져 왔다. DRAM의 경우 SRAM의 6 트랜지스터 셀에서 출발하여, 4TR-3TR-2TR-1TR 셀로 진화하여 왔다. 현재는 1TR 셀이 모든 DRAM에 적용되고 있다. 이 진화는 0.5TR 이하의 셀 가능성이 지속적으로 탐색되었다. 또한 cell에 정보저장을 위한 capacitor가 없는 capacitor-less cell의 가능성도 추구되었다^[3]. 예를 들면 자유공간에서 파의 이동시간을 이용하는 방식이 실험되었다. 정보는 공중에 떠 있는 동안 기억 되어 있다고 본 것이다.

Capacitor가 없는 셀 구조로서 정보저장에 사용되는 capacitance 대신에 물질의 상 변화(phase change), 자성의 방향(magnetic orientation), magnetic hysteresis, 저항 차 등을 이용할 수 있다. 대표적인 메모리로서는 FRAM, MRAM,

PRAM 그리고 RRAM 등이 있다. 이들 물질의 사용은 만들기 까다로운 capacitor를 보다 만들기 쉬운 구조체로 바꾸어 준다. 메모리 셀의 크기가 줄어들어서 칩의 크기가 작아지게 된다. 대부분 전원이 꺼져도 정보가 살아 있는 비 휘발성 특징을 가진다. 사용하지 않을 때는 전원을 아예 꺼 두는 저전력 운용에 도움을 주는 좋은 점이 있다. 아직은 제조 기술이 성숙하지 못한 것으로 보이고 기대하는 만큼의 수율이 가능한지 분명하지 않다. 또한 대부분 동작속도가 상대적으로 느리다.

VI. Non-von

메모리와 처리 기능이 통합된 컴퓨터 구조라면 인간의 두뇌의 모습에 가까워 진다. 이 구조는 von Neumann bottleneck 의 문제에서 자유롭게 해 준다. 더 이상 데이터는 메모리에 저장되었다 처리장치로 나들이를 할 필요가 없어지게 되는 것이다. Non-von 구조라고 부르는 일련의 컴퓨터 구조가 이들을 대표한다. 이들은 기존 컴퓨터의 명령어 형태가 동작코드(opcode)와 연산자(operand)로 구성되는 명령어 형태와는 완전히 다른 명령어를 가져야만 한다.

메모리 셀에 간단한 정보처리 기능을 가지는 구조체를 다룰 수 있다. 여기에 첨가되는 처리 기능은 조합회로의 기능적 완결성(functional completeness)을 유지하는 최소의 형태로 NAND 나 NOR gate를 들 수 있다. 이들 논리 기능은 가급적 다른 셀과 연결통로의 형성과정에 자연스럽게 만들어 지면 회로 단순화에 기여하게 된다. 특수한 기능의 구현을

위하여 보다 복잡한 조합회로를 추가할 수도 있을 것이다. 모든 셀에 같은 조합회로를 갖도록 하는 homogeneous cell array가 사용될 수 있다. 두 개 이상의 조합회로 기능을 가지는 heterogeneous cell array도 가능하다. 이들 셀은 학습효과를 저장하기 위한 non-volatile 특성을 가지는 메모리 요소가 들어 갈 수 있다.

참고문헌

- [1] ITRS, Int'l Tech. Roadmap for Semiconductors 2006 Update, ITRS 2006
- [2] Y. You, Fundamentals of DRAM Design and Analysis, Chungbuk National University Press, 2004
- [3] S. Okhonin, M. Nagoga, J. M. Sallese, and P. Fazan, "A capacitor-less 1T-DRAM cell," IEEE Electron Device Letters, vol. 23, no. 2, Feb. 2002

저자소개



유 영 갑

1975년 서강대학교 전자공학과 학사 졸업
 1975년-1979년 국방과학연구소 연구원
 1981년 Univ. of Michigan, Ann Arbor 전기전산학과 석사 졸업
 1986년 Univ. of Michigan, Ann Arbor 전기전산학과 공학박사 졸업
 1986년-1988년 금성반도체 (주) 책임 연구원
 1993년-1994년 아리조나 대학교 객원교수
 1998년-2000년 오레곤 주립대학교 교환교수
 1988년-현재 충북대학교 정보통신공학과 교수

주관심 분야 : VLSI 설계 및 테스트, 고속 인쇄회로 설계, 암호학