

한글·한자 전산화 문제에 관한 고찰 (I)

姜 哲 熙

韓國電氣通信研究所 (工博)

I. 머릿말

세종대왕께서 한글을 제정하신지 540년이 되고 있는 최근에 와서 한글기계화 문제가 서서히 해결될 기미를 보이고 있는 것은 다행스러운 일이다. 그것을 가능케한 주역중의 하나가 다름아닌 반도체 집적회로 기술, 특히 마이크로컴퓨터 응용기술이라는 것은 주지의 사실이다.

반면, 이 기술들은 한글의 기계화에만 기여한 것이 아니라, 한자의 기계화에도 큰 영향을 미쳐, 가까운 일본에서는 발음대로 키이를 치면 단어 단위로 한자가 표시되는, 이른바 “카나 한자 변환시스템”이 개발되기에 이르렀다. 우리의 한자문제는 일본과는 판이한 입장에 있지만, 법학, 의학등 특수분야에서 처럼 한자를 꼭 필요로 하는 경우에, 그에 대응할 수 있는 한글 한자 기계화 시스템이 개발되어야 한다는 사실에는 반대할 이유는 없다고 본다.

따라서, 본 논문에서는 2회에 걸쳐, 한글을 기계화, 특히 전산화 하는데에 고려해야 할 사항들에 관하여 고찰하고, 한자와 병용하는 경우의 제반 문제들에 대하여서도 언급하기로 한다.

II. 한글 문제

1. 한글 자판 배열

영문자판 배열이 현재의 것으로 정해진 이유는, 1800년대 당시의 기계공업 기술로는 능숙한 타자수가 빨리 타자하면 타이프 바아가 서로 뒤엉키는 문제를 해결할 수 없었기 때문에 오히려 치기 어렵도록 배치한 것 때문이라고 알려져 있다. 그후 미국의 Dvorak씨가 이 문제를 해결하기 위하여 영문의 발생빈도와 손가락에 미치는 부하를 잘 고려하여 이른바 “Dvorak keyboard”를 창안해 내었으나, 이미 보급되어 있는 타자기를 전부 뜯어 고쳐야 하는 어려움등으로 인하여 몇몇 곳에

서 채용하는 정도에 그치고 말았다.^{1)~4)}

국내의 경우에도, 자판배열 문제는 논란이 거듭되어 왔던 것으로, 최근에 들어 2벌식으로 통일되어 가는 추세에 있다.

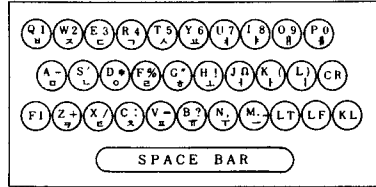
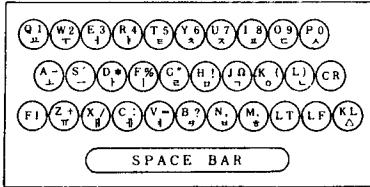
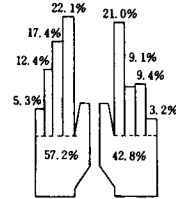
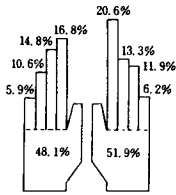
자판 배열시 고려해야 할 사항들을 열거해 보면 다음과 같다.⁵⁾

- i) 왼손보다 오른손 부하가 조금 무거울 것.
- ii) 각 손가락 힘에 알맞도록 부하가 배분되어 있을 것.
- iii) 가능한, 줄(두번째)에서 가운데 타자를 많이 할 수 있을 것. 뒷줄에 걸리는 부하는 아랫줄(밑에서 첫번째)보다 많은 편이 좋음.
- iv) 한 손 연속치기가 덜 생기도록 한다. 리드미컬한 타자를 위해서는, 한쪽 키이를 두드릴 때 다른 쪽은 다음을 계속 칠 수 있게 준비하고 있는 것이 바람직하기 때문.

이상과 같은 점들을 참작하여 한글자판을 배열해 보면, 우선 왼쪽이 모음, 오른쪽이 자음이 놓여져야 한다는 생각까지는 쉽게 할 수 있다. 그러나, 공교롭게도 한글의 모음은 “H, K”를 넣어도 12자인데 반해, 영문자리에 그대로 한글을 배치해야만 하는 경우, 왼쪽 자판의 키이 수는 15이 되어 3자가 남아 돌게 된다. 오른쪽이 11이므로, 한글 자음 14자중 3자는 왼쪽으로 넘어오게 되어, 앞에서 말한 iv)항의 문제가 필연적으로 나타난다. 즉, 한글은 자음+모음+자음(CVC) 발생빈도가 전체의 55% 정도나 되므로, 나머지 3자중 1자가 중성으로 나타나면, 오른쪽+왼쪽+왼쪽이 되어 한 손 연속치기가 다발하게 되기 때문이다.

그림 1 (a)에 앞의 낭법을 적용한 컴퓨터시뮬레이션에 의해 얻어진 자판 배열중 하나를 소개하고 있다.

그림 1 (b)는 현용 2벌식(특히, 텔렉스)자판에 관해 조사한 결과이다. 이것을 보면, 오른쪽에 모음을 배치



(a) 모음을 왼쪽에 배열해 본 자판 예

(b) 한글 텔렉스 자판

그림 1. 한글 자판 배열과 각 손가락에 미치는 부담율⁽⁸⁾

했기 때문에 부하가 조금 가벼운 것으로(42.8%) 나타났으며, 손가락에 대한 부하 비율은 왼쪽은 제대로 되었으나 오른쪽은 그렇지 못하게 나타나 있다.

그림 1 (a)와 (b)를 비교해 보면, 손가락 부하 배분에 관해서는 비교적 이상적이라고 할 수 있는 (a)가, 한 손 연속치기 문제에서는 앞서 설명한 이유 때문에 좌, 우가 16%, 3% (b)가 1%, 6% 정도나 되어 바람직하지 못한 결과를 보이고 있다. 단, 이 데이터에는 “각가”를 칠 경우의 “각”의 “ㄱ”받침 다음에 “가”의 “ㄱ”이 음으로 인한 한 손 연속치기 발생율은 포함되어 있지 않다. 또한, 쌍자음의 경우는 문헌[6]의 “△”개념을 도입하여, 예를 들어 “쌍씨”를 “스△스△스△”로 타자하는 것으로 하였다.

결국, 한글 특성과 기존 타자의 전통적인 손가락 배치법에 따르면, 모음을 좌측에 배열하여도 그렇게 큰 효과를 얻지 못한다는 사실을 알 수 있게 된다. 반면, 그림 1 (b)에 보인 현용 2벌식 자판배열은 현재 정해져 쓰이고 있다는 강점과 최근 경향이 자판 우측에 숫자 키만 모아 놓은 별도의 숫자 키 그룹이 채용되는 경우가 많아지고 있으므로 우측 부하가 보다 무거워지게 되어 균형이 잡히게 될 것이라는 점을 감안하면 표준자판으로서 충분히 통용될 수 있는 것이라 생각된다.

2. 한글 입력 방식

한글의 입력은 기본적으로 풀어쓰기 순서대로 쳐넣으면 입력부의 제어기능(통상, 마이크로프로세서에 의해 실현)에 의해 한 글자로 조합되어 처리부로 넘겨지게 된다. 쌍자음의 경우는 쉬프트 키 사용법, △

사용법, 스페이스 키 사용법 등에 의해 입력된다. 앞 절에서 언급한 △사용법이 풀어쓰기대로 입력한다는 개념에는 가장 합치한다고 볼 수 있으나, 과거치 표준인 쉬프트 키를 눌러 쌍자음 키를 넣는 방법이 이미 표준으로 권고되고 있으므로 구태어 △사용법을 주장할 근거는 미약하다고 본다.

출력과도 관련이 있는 영문 혼용시의 입력법이 문제가 된다. 즉, 한글 1자에 대해 영어 1자를 대응시키면 영어 글자가 너무 커지는 문제이다. 이에 대처하기 위해, 한글 1자에 영어 2자를 칠 수 있게 하는 “반각” 모드 키를 설정해야 하는데, 그것을 어디에 위치시키고, 어떻게 이용하느냐를 결정해야 한다. 한글과 같은 크기의 영문자를 선택하는 기능을 “반각” 키가 겸용한다면, “반각·전각” 모드 키가 될 것이다. “한글” 모드에서 “영문” 모드로 바뀔 때, 앞의 상태를 기억했다가 “전각” 또는 “반각” 상태로 들어갈 것인지, 아니면 항상 “영문”에 들어갈 때에는 “반각” 상태로 들어가도록 할 것인지에 대해 고려해 보면, 후자쪽이 보다 일반적인 것 같다. 그 이유는 한·영 혼용문에서도 영어는 “반각” 모드에서 쓰는 경우가 많을 것으로 예측되기 때문이다.

3. 한글 출력 방식

한글을 디스플레이나 프린터에 출력할 때의 방식에는 크게 나누어 조합방식과 패턴방식이 있다.

조합방식은 문자 그대로 한정된 한글 자소패턴을 입력정보에 따라 적절히 배치 조합하여 한글을 합성하는 방식으로, 패턴 메모리 용량을 수KB~십수 KB (글자패턴 13×13~24×24 dots)로 줄일 수 있는 장

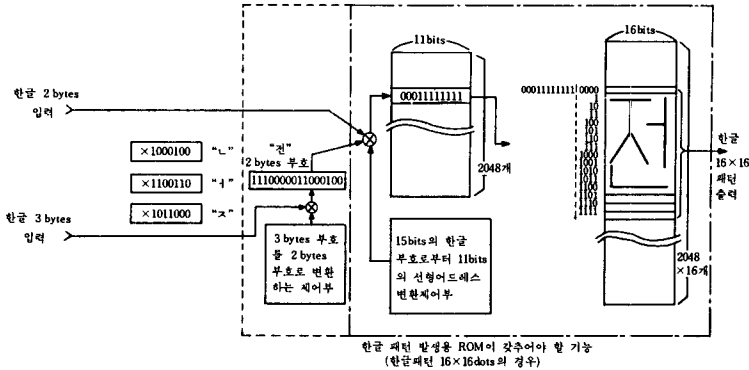


그림 2. 지능형 한글 패턴 발생용 ROM의 기능 설명도

점이 있는 반면 어느 수준이상의 품질좋은 글자 모양을 표시할 수 없다.

패턴방식은 처음부터 합성된 한글 패턴을 갖고 있으면서, 필요시에는 합성하는 절차없이 바로 패턴을 불러내어 표시하는 방식이다. 문자패턴 디자인에 따라 dot수가 갖는 해상도의 범위까지는 얼마든지 품질을 높일 수 있는 반면, 패턴 메모리 용량이 64KB~144KB(한글 글자수를 2000자로 가정하여, 16×16, 24×24일 때)정도나 필요하게 되며, 그림 2와 같은 선형어드레스 변환기능을 위해 수KB의 테이블이 필요하게 되는 결점을 갖는다. 이 기능은 2 bytes 한글 부호가 2¹⁶ 개의 어드레스 공간 전체에 흩어져 위치하는 것을 2¹¹ 개(한글자수를 2000개 이내로 가정)의 어드레스 공간에 선형적으로 배치되도록 위치(어드레스) 변환 역할을 수행하게 된다.

여기서, 한글 패턴발생용 ROM 개발의 필요성을 제안하고 싶다. 그림 2에서 보인 바와 같은 기능들, 즉 3, 또는 2 bytes 한글부호가 입력되었을 때, 선형어드레스변환부를 거쳐, 선정된 어드레스의 패턴을 읽어내어 출력하는 기능까지가 하나의 칩(chip)에 집적된 지능형 ROM을 말한다. 영문 알파벳 및 특수기호 등을 합한 256종의 패턴에, 한글패턴 1800여종을 넣으면 2048종이 될 것이며, 16×16, 24×24, 32×32 dots의 패턴을 기억시킨다면, 그 용량이 각각 512Kbits, 1Mbits, 2Mbits이 되어, 현재의 반도체 기술로서도 충분히 몇 개 이내의 칩으로 구성할 수 있는 규모가 될 것이다.

국내 기술을 총동원한 한글패턴발생용 국산 ROM의 조속한 개발은 한글에 의한 정보입출력 기술 발전에

커다란 도움이 될 것으로 기대한다.

문헌[7]의 조사에 의하면, 한글사용도수 누적백분율이, 그림 3과 같이, 270번째 글자에서 90%를 차지하고 있다. 일본에서의 한자 사용도수 누적백분율이 512자에서 겨우 80%가 되는 것과 비교해 보면, 한글이 얼마나 국한된 글자를 자주 쓰고 있는가를 알 수 있을 것이다. 따라서, 이 특성을 활용하여 글자품질을 높이는 아이디어에 관해 소개한다.

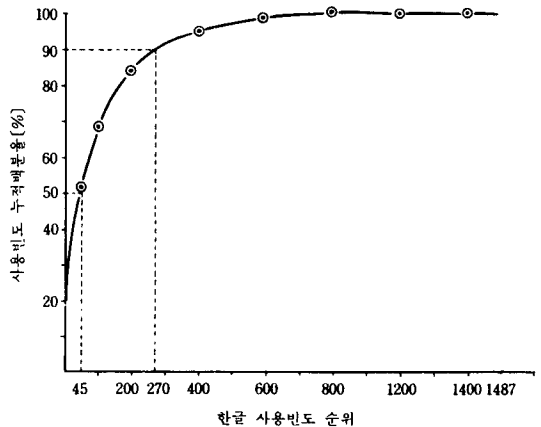


그림 3. 한글 사용도수 누적백분율과 빈도순위⁽⁵⁾

그림 4에 보인 방식개념도를 보면서 설명하기로 하자. 출력을 원하면, 우선 한글코드 메모리에 저장된 뒤, 하나씩 읽어내어 그것이 한글 패턴 메모리(예를 들어, 그림 3에서 얻은, 90%의 누적백분율을 갖는 270자의 패턴)에 속하는지, 아니면 여러 종류의 한글

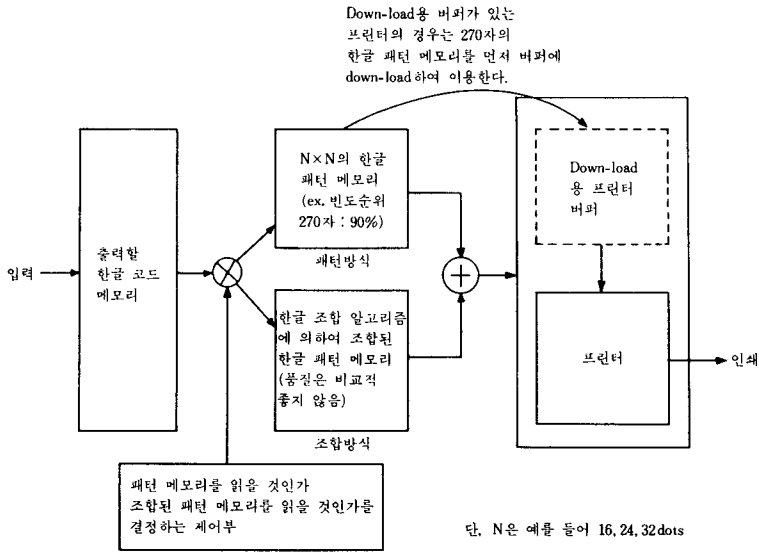


그림 4. 한글 사용 빈도수에 따른 새로운 한글 인쇄방식 개념도

자소 패턴으로 부터 미리 정해 놓은 알고리즘에 따라 조합하여 패턴(조합된 한글 패턴)을 출력할 것인지를 결정하게 된다. 만일, 프린터가 호스트로부터의 문자 패턴 정보를 받아 저장하고 있다가 정해진 코드에 의해 프린트 될 수 있는 전용 하드웨어를 갖고 있다면, 저장용 버퍼 메모리에 시스템 초기화시 패턴 메모리에 저장될 내용이 공급되도록 함으로써 패턴 메모리를 없앨 수 있다. 대신, 패턴 메모리에 존재하지 않는 글자 패턴은 조합된 한글 패턴 메모리로부터 공급받아 프린터 버퍼의 한부분을 채우도록 하여 프린터가 이부분을 읽어 낼 수 있게 한다.

이 방법은, 메모리가 한정된 시스템 설계시, 또는 멀티 폰트(고딕체, 명조체 등의 혼용)의 한글을 출력하고자 할 때에 유효한 방법이 될 수 있을 것이다. 앞으로, 사용빈도 순위를 몇으로 하면 실질적인 프린터 구동속도가 최대가 될지, 패턴 메모리로부터 인쇄된 글자 품질과 조합하여 만든 패턴으로부터 인쇄된 것과의 혼용 비율이 어느 정도일 때 사람의 눈이 그것을 못느끼게 되는지 등에 관해 정량적인 연구가 필요하다고 생각한다.

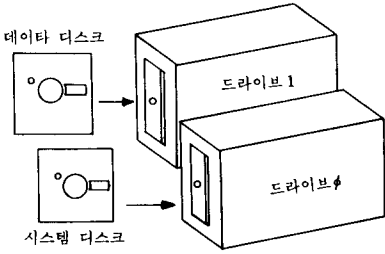
4. 문서 교환용 플렉시블 디스크 카트리리지 (flexible disk cartridges)의 파일(file) 규격

최근 한글 워드프로세서(이후 WP라 약함.)의 개발이 활발해지고 있음은 매우 다행한 일이다. 그러나,

선진국이 걸은 과정을 잘 살펴서 선진국의 과오는 다시 반복함이 없도록, 주의 깊게 WP개발을 추진해 나가야 할 것이다. WP보급 대수가 10만대에 육박하고 있는 일본의 경우, 각 회사간에 플렉시블 디스크(이후 디스크라 약함.)파일의 호환성이 보장되어 있지 않아 타사 제품의 WP에서 만든 문서 파일을 이용할 수가 없다. 이것은 미국의 경우도 마찬가지이며, 최근 문서교환용 디스크 파일 규격의 표준화 작업이 이 문제를 해결하기 위해 끝났거나 진행중에 있다(ISO DIS 7665, JEIDA-35-1983등).^(*)

우리 나라의 경우는 WP개발의 초창기에 해당되는 지금 이 문제를 충분히 고려하여 우리 고유의 규격을 만들어 놓는다면 선진국의 전철을 밟지 않을뿐더러 WP의 조속한 보급에 크게 기여할 것이다.

우선, 디스크의 파일 포맷이 규격화되었을 때, (규격 디스크라 약함), 어떻게 하여 비규격의 WP가 그것을 해석하느냐에 관해 설명한다. 규격 디스크는 데이터 디스크로서 드라이브 1에 삽입시키고 그것을 구동하는 것은 드라이브 0에 넣어 놓은 비규격의 회사 고유 시스템 디스크에 의하고 있다. 그림 5에서 보는 바와 같이, 시스템 디스크를 읽은 호스트 컴퓨터는 규격 디스크를 읽거나 쓰는 프로그램을 실행하여, 데이터 드라이브로부터 또는 드라이브에 규격화된 문서 정보를 옮겨 오거나 또는 비규격의 회사 고유 포맷의 문서정보를 옮겨가게 해준다. 이 방법은 WP 시스



- 시스템 디스크
회사 고유의 OS를 포함한 소프트웨어가 고유의 파일 포맷 방법에 의해 기록.
- 데이터 디스크
규격화된 문서교환용 디스크는 여기에 삽입되어 시스템 디스크의 유틸리티 프로그램에 의해 회사 고유의 디스크로부터 그 내용이 옮겨진다.

그림 5. 규격화된 파일 디스크의 이용 방법

템의 파일 포맷에는 영향을 미침이 없이 규격화된 파일 정보를 취급할 수 있다는 장점이 있는 반면, 기본적으로 드라이브가 항상 2대 있어야 하는 점이 문제가 될 수 있다.

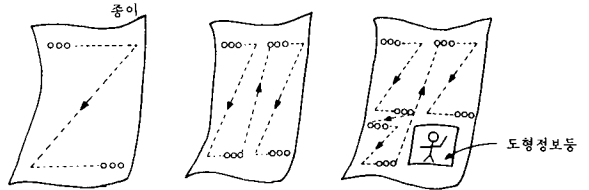
한편, WP 전용 디스크를 의식한다면, 1대의 드라이브 만을 이용하는 방법도 생각될 수 있다. 특히 우리 나라와 같이 드라이브의 순수 국산화를 시키지 못하고 있는 실정을 감안하면, 드라이브 1대만의 WP도 현실적일 수 있을 것이다. 이런 경우, 1장의 디스크에 시스템 프로그램(예를 들어, OS에 최소한의 유틸리티 프로그램)과 문서 교환용 데이터 디스크가 공존하게 된다. 따라서, 데이터 영역이 좁아지며, 특히 문서내용을 읽어내어 다른 프로그램에 의해 그것을 어떤 형태로 처리하려고 하는 경우에는 프로그램(예를 들어, language 프로그램인 BASIC, COBOL 등) 영역이 부족한 경우가 발생하거나, 기존 제품의 시스템 소프트웨어 파일 포맷을 바꾸어야 하는 결점들을 들 수 있다. 드라이브 장치 가격이 비교적 비싸고, WP의 보급 초기단계인 것을 감안하면 적합한 방법이라 생각할 수 있으나 장래의 확장성등을 생각하여 함부로 정할 수는 없을 것이며, 따라서 충분한 연구가 필요할 것이다.

다음에는 규격화되어야 할 파일의 내용에 관해 간단히 설명해 둔다.

i) 적용범위

그림 6 (a)에서 보는 바와 같이, 어떤 문서를 편집할 때, 그 구성을 단순히 종이 왼쪽 및 구석에서부터 오른쪽 아래 구석으로 글자를 써내려가는 방법, 즉 단순 구성 형태와, 문서를 반으로 나누어 구성하는 (b) 와

같은 2단 블록 구성형태, (c)와 같이 여러 블록으로 나누어 구성하는 복수 블록 구성 형태등으로 문서 구성방법을 달리 할 수 있다. 초기에는 단순구성에 관해서만 규격화가 이루어지는 것이 바람직하나 문서에 도형정보를 포함하는 보다 진보된 WP(CCITT에서는 mixed mode terminal이라 부르고 있음.)의 문서에 대해서도((C) 형태)추후에 규격화가 이루어져야 할 것이다.



(a) 단순구성, (b) 2단블록구성, (c) 복수블록구성

그림 6. 문서 편집시의 문서블록의 구성 방법⁽⁸⁾

ii) 적용항목

공통화를 위하여, 문서 파일의 구조도 다음과 같이 분류되어 규격화되어야 할 것이다. 그림 7에서 보는 바와 같이, ISO 표준 레벨(label)은 볼륨 인덱스 레벨과 파일 인덱스 레벨로 구성되며, 파일 인덱스 레벨은 파일 디렉터리(directory)가 존재하는 위치를 지정해 준다. 또한 파일 디렉터리에는 문서명과 문서 저장 장소를 지정해 주는 정보가 들어 있다. 문서에는 문서 인식부(header)와 문서 내용부로 나누어 존재하며, 문서 인식부는 문서의 속성 및 서식을 규정해주고, 내용부는 문자 부호 및 제어문자 부호에 의해 문서 구성

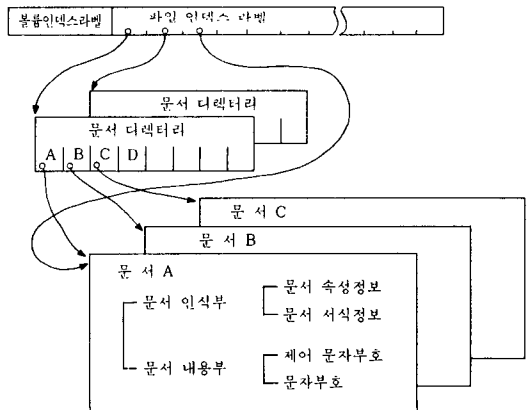


그림 7. 파일 구조 개념도⁽⁸⁾

형태를 규정해주고 있다.

이 부분에 관해서는 보다 많은 연구를 하여 다음 기회에 상세히 보고하기로 하겠다.

이제까지 한글 자판 배열 문제, 한글 입력력방식 문제, 한글 문서 교환용 파일 규격에 관한 문제에 대해 고찰해 왔다. 다음호에서는 계속하여 범용 영문용 OS에의 한글문자 도입문제, 한자 시스템에 관한 고찰등에 관해 논할 것이다.

參 考 文 獻

[1] Parkinson, R.: "The drorak simplified keyboard: forty years of frustration", *Computers and Automation* Nov. 1972.
 [2] Ferguson, D. and Duncan, J.: "Keyboard design and operating posture", *Ergonomics* vol.17, no.6, pp.731-744, 1974.

[3] Kawakami, A. et al: "Human factors in rainputto keyboard for kanji input", 1st UJCC, 1972.
 [4] 山田, 他: "漢字入力法の人間工学的検討", 情報処理学会 19回 全国大会, パネル討論會資料, 1978.
 [5] 姜哲熙: "文字図形情報処理方式に関する研究", WASEDA 대학 박사학위청구논문, 1980.
 [6] 이주근 외, "Some structural analysis of Hangeul information source and its application to the improved coding method, *JKIEE*, vol.15, no.2, pp.1-7, (May, 1978).
 [7] 한글기계화연구소: "한글기계화연구 I", pp. 5-28, Aug. 1975.
 [8] 日本電子工業振興協會: "日本語文書交換用 ファイル仕様 (基本形) JEIDA-35-1983", Mar. 1983.*

알아봅시다

高性能 ZnO 薄膜表面波 變換素子

非壓電性 基板上에서 音響表面波를 發生 및 檢出하는 가장 좋은 方法은 壓電性 薄膜變換素子에 의한 것이 있으나 이 중에서도 ZnO膜을 사용하는 것이 壓電效果가 크고 結晶性이 良好한 膜의 製作이 可能하다고 한다.

美國 Motorola社의 Fred. S. Hickernell氏는 5年 間에 걸쳐서 研究한 結果 特性이 良好한 表面波 變換素子로서 ZnO膜의 製作條件과 膜特性 등을 明確히 했다. 從來 컴파운드스퍼터링에서는 다이오드方式을 사용했으나 3極方式보다 스퍼터링壓力이 낮게 되며 結晶粒을 적게(약 100Å) 할 수 있어 良好한 特性의 變換素子를 얻었다. 그런데 良好한 變換素子 特性을 갖는 ZnO膜의 蒸着條件을 들면 다음과 같다.

첫째, 基板溫度는 基板材料에 따라서 다르나 보통 200~350℃의 範圍에서 에피택셜成長의 경우에는 300~500℃의 範圍이며, 蒸着速度는 1.0μm/h로서 에피택셜膜의 경우에는 400Å/h 以下이다.

둘째, 最適배 그라운드 가스壓力은 直流다이오드, RF 다이오드, 直流 3極方式에 대하여 10~100μmHg, 5~20μmHg 및 0.5~5μmHg로서 使用하는 가스는 보통 알콘과 酸素를 混合한 것이며 알콘은 酸素의 10~20%이다.

셋째, 良好한 C軸 垂直方位를 얻기 위하여는 基板表面과 같이 眞空系를 淸淨하여야 하며, 基板表面의 結晶方位가 良好해야 된다.

네째, 基板의 溫度特性에 의해서 C軸의 極性이

上向 또는 下向하므로 溫度特性에 留意하고, 소스材料의 純度는 높아야 되나 예를 들면 Al, Cu를 意識적으로 添加함으로써 C軸의 平行成長과 垂直成長을 促進시킬 수 있다. 그리고 結晶粒의 C軸은 原子流의 方向으로 成長하므로 剪斷모우드의 變換素子인 경우 소스에 대하여 基板을 비스듬히 둘 것들을 들 수 있다.

ZnO膜의 表面波 特性을 評價하는 實用的인 方法은 表面波速度, 結合係數 및 C軸과 垂直方位膜의 減衰量을 調査하여 單結晶膜의 것과 比較하면 良好하다.

Motorola社에서는 蒸着條件으로 直流 3極方式을 사용해서 良好한 變換素子特性을 갖는 ZnO膜을 製作하고 있으나 測定結合係數가 理論值의 90%로서 表面波速度와 減衰量이 單結晶의 것과 거의 同一한 ZnO膜을 分析하였다. 또한 表面粗는 300Å以下에서 密度가 높았으며 C軸의 垂直方位는 ±5°內의 良好한 方位를 表示하였고, 結晶粒의 크기는 약 100Å(X線測定)였다. 이와 같은 優秀한 表面波 變換特性을 갖는 ZnO膜은 結晶粒이 微細하고 高密度로서 結晶方位가 良好하고 膜成長表面이 平滑하다는 特徵을 갖고 있다.

ZnO膜의 두께가 5μm以上이 되면 上表面의 粗가 顯著하여 微細結晶粒, 良好한 結晶方位임에도 不拘하고 結合係數가 理論值의 75% 以下였다. 또한 高周波 變換素子の 경우에는 結晶粒의 微細가 매우 重要하다고 한다.