

## 데이터마이닝을 활용한 반도체 수율개선시스템

백동현\*, 남정곤\*\*

\* 천안대학교 경상학부 경영정보학전공, \*\* 하이닉스반도체 자동화 담당 상무

### Abstract

반도체 공정은 웨이퍼가 투입되어 완제품이 생산되기까지 수 백개의 제조공정을 수 개월에 걸쳐 진행해야 하는 매우 복잡하고 긴 공정으로 구성되어 있다. 대부분의 공정들은 먼지가 철저히 통제되는 클린 룸에서 진행되지만 아주 미세한 먼지 하나도 반도체 칩의 성능과 수율을 저하시키는 요인이 된다. 반도체 칩의 불량은 특정 생산장비에서의 이물질 발생, 생산장비의 잘못된 파라미터 값 설정 등 다양한 요인에 의해 발생할 수 있으며 불량의 원인을 요인별로 파악하여 신속하게 대처하는 것이 수율 개선의 핵심이 된다. 이를 위해 SPC 시스템, MES 그리고 6-시그마 등의 활용을 통한 다양한 수율개선 노력이 있었으나 공정의 복잡성과 대응량의 수집 데이터로 인해 기존의 통계적 방법이나 엔지니어의 경험적 분석방법으로는 미처 파악하지 못하는 수율 저하 요인이 상당 수 존재한다. 본 논문은 군집화/분류, 순차패턴 등의 데이터마이닝 기법과 다차원분석(OLAP) 도구를 활용하여 수율저하의 원인이 되는 문제공정, 문제장비, 그리고 잘못된 파라미터 값 설정 등을 신속하고 정확하게 파악하여 수율 개선을 지원하는 방법을 소개하며, 반도체 Fabrication공정을 대상으로 실제 구현된 수율개선 시스템(Y-PLUS)을 설명한다.

### 1. 서론

한국의 반도체 생산업체들은 반도체 시장의 변화에 대처하기 위해 다각적인 노력을 기울이고 있다. 메모리 제품의 우위를 유지하면서 비메모리 분야로의 제품 구조 조정, 설계 인력의 양성, 기술의 확충, 그리고 신기술 및 신상품 개발 등을 적극적으로 모색하고 있는 상황이다. 이와 더불어 주어진 생산 능력 하에서 수율(yield) 향상, 주기시간(cycle time)의 단축, 재공 재고의 감소 등 생산성 향상을 위한 관리 기술적인 노력을 기울이고 있다.

특히, 수율과의 전쟁이라 표현할 만큼 수율이 반도체 산업에서 차지하는 의미는 매우 크다. 새로운 신상품이 개발되었다 하더라도 이를 일정 수준 이상의 수율로 하루라도 빨리 끌어 올려 양산체제로 이관할 수 있어야만이 시장에서의 경쟁력을 확보할 수 있다. 또한 일정 수준 이상의 수율을 확보한 제품이라 하더라도 더 높은 수율을 유지할 수 있어야만이 원가 절감으로 인한 가격 경쟁력이 생길 수 있는 것이다.

수율의 이러한 중요성 때문에 반도체 업계에서는 관리도, 상관분석, 회귀분석, 유의차 분석 등 통계적 품질관리 기법을 적용한 품질관리시스템을 활용하거나 6 시그마 활동, 엔지니어의 노하우를 축적, 공유, 활용하는 지식관리 활동 등 다각적인 노력을 기울이고 있다. 그러나 기존의 통계적 기법이나

엔지니어의 경험적 분석만으로는 미처 파악하지 못하는 수율 저하 요인이 다수 존재한다. 특히 수 많은 공정 변수와 엄청난 양의 데이터로 인해 모든 변수를 모두 분석하기에는 너무 많은 시간이 소요될 뿐만 아니라 미처 파악하지 못하고 지나쳐 버리는 원인들이 다수 존재한다.

본 논문에서는 기존 수율관리시스템의 문제점을 개선하기 위해 데이터마이닝 기법과 다차원분석 도구를 활용한 수율 관리 방법을 제안하고 이를 정보시스템으로 구축한 Y-PLUS (Yield Prediction, anaLysis & Up Support) 시스템을 소개한다. 이 시스템은 군집화, 연속패턴분석, 분류, 신경망 등 데이터마이닝 기법을 적용하여 수율의 원인이 되는 문제 장비나 문제 파라미터를 신속, 정확하게 파악할 수 있도록 해 줄 뿐만 아니라 공정 진행 중인 제품의 미래 수율을 예측할 수 있도록 지원한다.

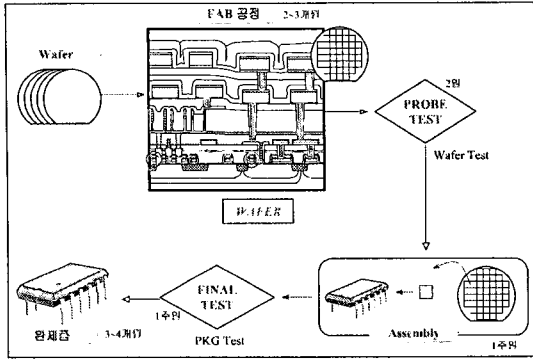
본 논문은 다음과 같이 구성되었다. 2절에 반도체 제조 공정에 대해 설명하고, 3절에서 데이터마이닝 기법과 OLAP에 대해 간략히 설명하였다. 4절은 수율개선시스템인 Y-PLUS 시스템에 대해 데이터마이닝 기법 적용의 문제 상황과 적용기법을 중심으로 설명한다. 마지막으로 5절은 본 논문의 결론을 정리하였다.

### 2. 반도체 제조공정

#### 2.1. 반도체 제조 흐름

반도체 제조공정은 크게 wafer fabrication(이하 FAB공정), probe검사(probe test), 조립(assembly), 그리고 최종검사(final test) 등 4단계로 구성되어 있다[8]. 그림 1은 반도체 공정을 설명한 것이다.

웨이퍼(wafer)는 반도체 제조를 위한 실리콘 기판으로, FAB공정을 거치면서 한 웨이퍼 위에 수 백개의 집적회로(IC, integrated circuit) 칩(chip)이 생성된다. Probe검사는 웨이퍼에 생성된 칩이 원하는 전기적인 특성을 제대로 갖추고 있는지 검사하여 양불량을 판별하는 단계이다. 조립공정은 만들어진 웨이퍼 내의 칩을 분리하여 칩의 전기적, 물리적 특성을 향상시키고 외부의 기계적, 물리적 충격으로부터 칩을 보호하기 위하여 형성시켜 준다. 마지막으로, 최종검사 단계는 조립된 칩의 전기적 특성 및 기능, 신뢰성 등을 검사하여 양품, 불량품을 판별한다. Probe검사가 웨이퍼 상태에서 각 칩을 검사하는 반면, 최종검사는 각 칩이 분리되어 조립된 상태에서 검사를 수행한다. 웨이퍼가 투입되어 완제품으로 생산되기까지는 보통 3~4개월 정도 소요된다.



[그림 1] 반도체 제조과정

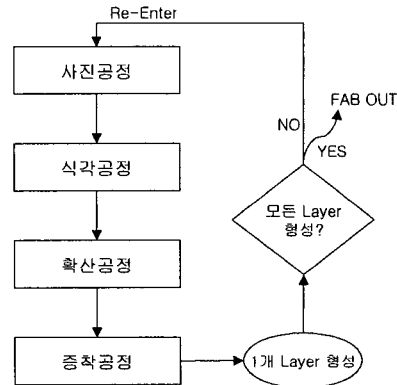
위 네 개 단계 중 FAB공정은 가장 자본 집약적이고, 생산공정이 복잡하며(200~300개 공정단계) 또한 생산 주기가 매우 긴 공정으로(2~3개월), 반도체 생산성과 수율에 가장 큰 영향을 미치는 매우 중요한 공정이다. 일정한 개수의 웨이퍼(보통 25장)가 하나의 로트(lot)를 형성하며, 로트 단위로 이동과 처리가 이루어진다. FAB공정은 그림 2와 같이 사진(photolithography), 식각(etching), 확산(diffusion), 그리고 증착(deposition) 등 4개 공정으로 구분할 수 있다.

사진공정은 패턴형성공정으로 마스크 상의 이미지를 웨이퍼 위에 전달시키는 공정이다. 피사체인 마스크에는 소자에 투사할 이미지가 그려져 있다. 감광액이 도포된 웨이퍼 위에 마스크를 정렬시키고 마스크 위에서 자외선을 비추면 감광액이 선택적으로 감광된다. 식각공정은 사진공정에서 형성된 감광막 패턴을 마스크로 하여 그 형성된 형상을 하부막에 전사시켜 가공한 후 불필요하게 된 감광막 패턴을 화학반응을 이용하여 제거하는 공정이다. 확산공정은 물에 잉크를 넣으면 퍼지는 것과 같이 웨이퍼 위에 원하는 불순물을 침적시킨 후 웨이퍼 내부로 퍼져 나가게 하는 공정이다. 증착공정은 웨이퍼 표면에 Oxide, Al, Au 등의 막을 형성하는 공정으로, 칩을 물리적 작용으로부터 보호, 전기적 배선의 형성 등을 목적으로 한다.

반도체 칩을 만들기 위해서는 보통 10-30개 정도의 층(layer)을 쌓아야 하는데, 사진·식각·확산·증착 공정을 거치면서 1개의 층이 쌓이게 된다. 따라서 하나의 칩을 완성하기 위해서는 공정을 반복해야 하는 재진입 흐름(Re-entrant flow)을 갖는다. 하지만 반드시 그림 2와 같은 순서대로 공정을 진행하는 것은 아니며, 층에 따라 일부 공정은 생략되거나 추가될 수 있으며 작업 순서도 바뀔 수 있다.

2.2 수율의 정의 및 중요성

일반적으로 수율(yield)이라 함은 투입량 대비 완성된 양품의 비율을 의미한다. 100개를 투입하여 80개의 양품을 생산하였다면 수율은 80%가 되는 것이다. '수율과의 전쟁'이라 말할 정도로 수율이 반도체 산업에서 차지하는 비중은 매우 크다. 대부분의 제조산업에서도 수율은 중요하지만, 반도체 산업에서 더욱 강조하는 이유는 반도체 산업이 가지는 몇 가지 특수성 때문이다.



[그림 2] FAB 공정 흐름

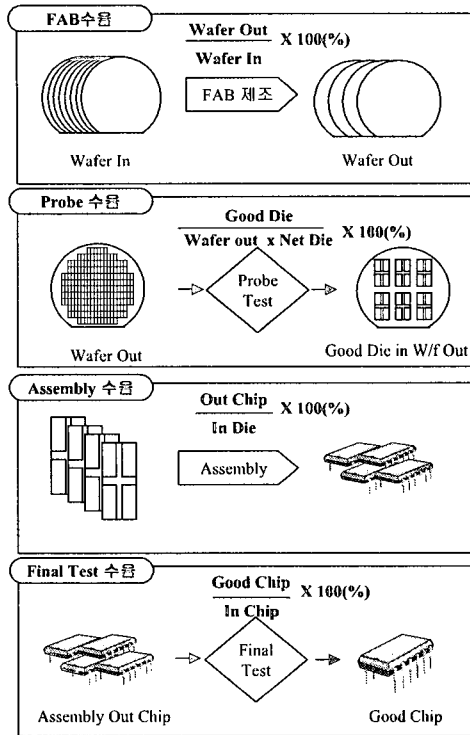
첫째, 반도체 소자가 급속하게 고집적화 됨에 따라 높은 수율을 확보하기는 점점 어려워진 반면 시장의 조기 선점과 생산성 향상을 통한 수익성 확보를 위해서는 기존 제품의 안정적인 고 수율 유지와 신제품의 안정적 수율까지 도달하는 시간(Ramp-up 기간)을 줄이는 것이 상당히 중요하다.

둘째, 원자 또는 분자 단위의 오염물질, 먼지 등이 반도체 성능 및 수율에 영향을 미친다. 그 때문에 반도체 공정은 0.1 $\mu$ m 이하의 오염물질까지 제거하는 클래스 1 수준의 클린 룸(clean room)에서 작업이 이루어진다. 클래스 1이란 1입방 피트 속에 먼지 등 오염물질이 1개 이하가 되도록 유지하는 것을 의미한다. 이런 초 청정 환경에서 작업이 이루어지지만, 공정/장비의 관리미흡이나 작업자의 사소한 실수 등으로 인한 오염물질 발생과 설계나 공정의 마진 부족이 수율에 나쁜 영향을 미치게 된다. 수율 저하 원인을 요인별로 파악하여 제거하는 것이 수율 개선을 위한 핵심이 된다.

마지막으로, 반도체 공정은 모든 공정이 끝나기 전에는 제품의 양불량을 판별할 수 없다는 특징을 가지고 있다. 제조 공정이 세분화되어 있고 투입에서 완료까지 수개월이 소요되는 긴 공정을 가지고 있으나, 공정 중간에는 발생된 불량을 판별할 수 없는 것이다. 예를 들어, 어느 공정을 담당하는 장비가 관리 소홀로 오염물질을 발생시켜 반도체 칩에 불량을 만들었다 하더라도 그 칩이 모든 공정을 마친 뒤에야 불량량이 감지된다. 따라서 불량량이 감지되고 그 원인이 되는 장비를 파악하기 전까지는 계속해서 그 장비는 불량을 발생시키는 것이다(최악의 경우 수개월 동안 계속 불량 발생). 경제적인 손실을 고려한다면 불량의 원인을 신속하게 파악하는 것이 얼마나 중요한지 짐작할 수 있다.

반도체 수율은 제조단계에 따라 FAB수율, probe수율, assembly수율, final test 수율로 구분되며, 이 4가지 수율을 곱하면 CUM 수율(cumulation yield)이 된다. 그림 3은 이들 4가지 수율의 계산 방식을 설명한 것이다. FAB수율은 투입된 웨이퍼 장수(wafer in) 대비 중간에 깨지는 등의 불량이 발생하지 않고 제대로 나온 웨이퍼 장수(wafer out)의 비율이다. Probe수율은 wafer out된 총 die 수(wafer out $\times$ net die) 대비 probe검사 결과 양품인 die(good die)의 비율이다. 여기서 net die는 한 웨이퍼에 만들어지는 die의 총 수를 의미하는 것으로 제품에 따라 다르지만 보통 500-1000개 정도이다. Assembly 수율과 final test 수율은 각각 투입량 대비 완성된 양품의 비율을 나타낸다. 이 4가지를 곱한 CUM수율은 결국 FAB 투입량 대비 최종

완제품의 비율을 나타내게 된다.



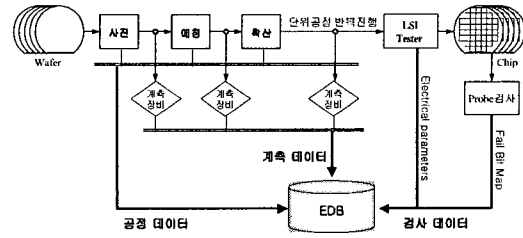
[그림 3] 수율 계산

### 2.3 기존 수율관리시스템

반도체 생산라인의 자동화 및 정보화의 정도는 다른 산업과 비교해 봐도 그 수준이 매우 높으며, 대량의 정보관리 및 물류관리, 그리고 이에 대한 실시간 활용과 모니터링을 위해 대용량의 시스템 환경과 24시간 365일 운영을 위한 안정적인 시스템 환경이 구축되어 있다. 또한 단위 공정 진행과정에서 발생하는 라인내 모든 기술 데이터가 데이터베이스로 집적되어 다양한 분석 틀을 활용하여 효율적인 수율관리가 가능하도록 기반 시스템이 구축되어 있다.

FAB공정에서 발생하는 데이터는 그림 4에 묘사된 것 같이 공정·계측·검사 데이터 등 크게 3가지 종류로 구분할 수 있다. 공정 데이터는 각 로트가 공정을 진행한 시간, 작업을 진행한 장비, 사용한 마스크 등의 생산이력과 각 장비의 이상발생, 예방보존(PM, preventive maintenance) 등의 장비이력과 관련된 데이터이다. 각 로트가 사진·식각·확산·증착 등의 단위공정을 진행한 후에는 해당 단위공정이 원하는 스펙(specification)대로 제대로 작업되었는지 검사하기 위해 계측장비로 cd, overlay, particle, thickness 등을 계측하며, 계측된 데이터는 장비로부터 자동으로 수집되어 데이터베이스에 축적된다. FAB공정을 마친 웨이퍼가 원하는 전기적인 특성을 제대로 갖췄는지 검사하기 위해 전기적인 파라미터 값이나 probe 검사 등을 하게 되며, 검사 결과 데이터도 데이터베이스에 저장된다.

64M DRAM 기준으로 볼 때 FAB공정 내의 단위공정 수가 약 200여개, 생산 장비 수가 500여대, 계측장비 수가 250여대이며 계측되는 변수의 종류만 해도 무려 500여가지가 되어 한 달 평균 100 기



[그림 4] 기술 데이터 수집

가 바이트 이상의 방대한 기술 데이터가 수집된다. 생산현장에서 생성되는 방대한 기술 데이터를 효과적으로 활용하여 반도체 수율 향상에 도움을 줄 수 있는 정보시스템들이 개발되어 사용되고 있으나 좀더 지능적이고 효과적인 지원시스템이 필요한 상황이다.

현재 사용하고 있는 수율관리시스템 중 대표적인 것은 통계적 품질관리(statistical quality control) 기법을 이용한 것이다. 자동으로 수집되는 다양한 계측 데이터를 이용한 x관리도,  $\bar{x}$ -R관리도 등의 관리도와 공정능력지수 관리 등을 통해 단위공정의 품질을 관리하거나 로트품질이력관리와 장비의 이상발생관리 등을 관리하고 있다. 또한 상관분석, 회귀분석, 유의차분석 등 다양한 통계적 기법이 사용되고 있으며 검사 데이터의 경우 그래픽을 이용한 map 분석 기능을 활용하고 있다. 이밖에도 수율관리를 위해 6-시그마 활동, 통계적 품질관리 교육, 엔지니어의 노하우를 축적, 공유, 활용하는 지식관리 활동 등을 전개하고 있다.

그러나 기존의 통계적 기법이나 map 분석 기능, 그리고 엔지니어의 경험적 분석만으로는 미처 파악하지 못하는 수율 저하 요인이 다수 존재한다. 특히, 수 많은 공정 변수와 엄청난 양의 데이터로 인해 모두 변수를 분석하기에는 너무 많은 시간이 소요되어 실제적으로 일부 변수만 분석이 되고 있다. 또한 엔지니어가 미처 인식하지 못하고 지나치는 공정 이상이 존재하고 있으며, 단위공정별 품질개선 활동은 활발히 이루어지고 있으나 전체 공정을 고려한 종합적 수율개선 노력은 부족한 상황이다. 표 1은 기존 수율관리시스템의 문제점과 개선방향에 대해 정리한 것이다.

본 연구에서는 기존 수율관리시스템의 문제점을 개선하기 위해 데이터마이닝 기법과 다차원분석(OLAP, on-line analytical processing) 도구를 활용한 수율관리의 방법을 제안하며, 실제적으로 이 방법이 구현된 수율관리시스템을 소개한다.

### 3. 데이터마이닝과 OLAP

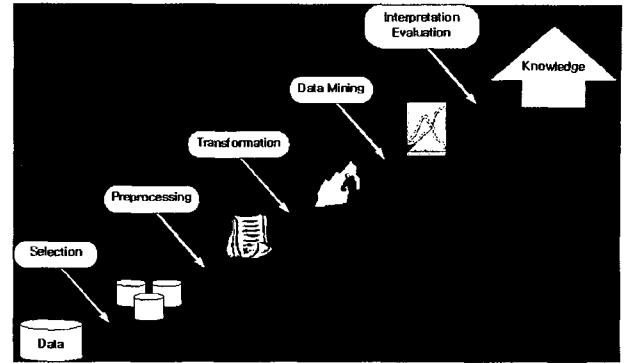
기업들의 데이터베이스가 대형화되고, 분석변수가 많아져 기존 인간의 두뇌에 의존한 자료검색 방법만으로는 데이터베이스내에 숨겨져 있는 정보를 찾아내는데 한계에 도달하였다. 이에 컴퓨터가 숨겨진 정보를 발견하게 하는 방법이 필요하게 되었는데 이것의 한 분야가 데이터마이닝이다. 데이터마이닝은 "대량의 실제 데이터로부터, 이전에 잘 알려지지 않는 것, 묵시적이고, 잠재적으로 유용한 정보를 추출하는 작업"이라고 정의한다[1]. 즉, 대용량의 데이터베이스에 숨겨져 있는 데이터간의 관계, 패턴, 규칙 등을 찾아내고 모형화하여 기업의 경쟁력 확보를 위한 의사결정을 돕는 유용한 정보로 변환하는 일련의 과정이다.

기존 시스템의 문제점	개선방향
기존의 통계적 방법이나 엔지니어의 경험적 분석방법으로도 미처 파악하지 못하는 수율저하요인 존재함	장비요인, 장비외적요인별 수율저하요인을 지능적으로 식별하여 제공하는 방법론 필요
단위공정별 품질개선활동은 활발히 이루어 지고 있으나, 전체 공정을 고려한 개선활동은 미비함	종합적 안목의 분석 방법 필요
엔지니어가 수 많은 공정 변수를 모두 분석하기에는 너무 많은 시간이 소요되어, 실제적으로 일부 변수만이 분석되고 있음	수 많은 공정 변수를 빠른 시간 내에 분석할 수 있는 방법 필요
엔지니어가 인식하지 못하고 지나치는 공정이상 존재함	공정 이상을 자동적으로 피드백하는 메커니즘 필요
엔지니어의 자질에 따라 분석 품질이 다름	분석의 일부를 지능적으로 지원하는 지원시스템 필요

[표 1] 기존 수율관리시스템의 문제점과 개선방향

종종 데이터마이닝과 지식발견(KDD, knowledge discovery in database)이라는 용어를 혼용해서 사용하고 있다. 데이터마이닝은 통계학자, 데이터분석가, 그리고 데이터베이스 분야에서 많이 사용되는 용어이고 지식발견은 인공지능이나 기계학습(machine learning) 분야에서 자주 사용되는 용어이다. Fayyad 등 [2]은 지식발견은 데이터로부터 유용한 정보를 발견하는 전체 프로세스이고 데이터마이닝은 지식발견 프로세스 중에서 데이터로부터 정보를 추출하기 위하여 기법을 적용하는 특정 단계로 정의하고 있다. 본 논문에서는 Fayyad 등의 구분에 따르기로 한다.

일반적인 지식발견 프로세스는 그림 5와 같다 [2, 9]. 먼저 대상 문제에 대한 명세화 또는 지식발견의 목표를 명확하게 정의한 후, 지식발견의 대상이 되는 목표 데이터 집합 또는 변수 집합을 정의하고 만드는 일이 수행된다(selection). 사전처리(preprocessing) 단계는 데이터의 적재, 변환, 클린징을 하는 단계이다. 특히, 데이터 클린징은 데이터 내의 잡음 제거, 망실 데이터 필드에 대한 처리 전략 등에 관한 기본적인 작업을 수행하는 것으로 지식발견 프로세스 중에서 많은 시간과 노력을 요구하는 단계이다. 데이터 변형(transformation) 단계에서는 고려 대상이 되는 변수의 수를 줄이거나 좀더 효과적인 표현을 발견하여 변형시키는 단계이다. 다음은 지식발견의 목표를 가장 효과적으로 달성할 수 있는 데이터마이닝 방법 또는 알고리즘이 무엇인지 선택하여 데이터마이닝을 수행하는 단계이다. 분류, 군집, 회귀분석 등 다양한 방법을 적용하여 데이터 속에 숨어 있는 패턴을 발견한다(data mining). 마지막으로, 발견된 패턴을 사용자가 해석 가능하게 시각화하고 정의된 평가 기준에 의해 데이터마이닝 결과를 평가한다(interpretation & evaluation). 평가가 성공적으로 이루어지지 못한 경우에는 이전 단계로의 회귀와 반복이 필요하다. 성공적인 데이터마이닝 결과는 의사결정 문제의 해결을 위해 사용되어 진다.



[그림 5] 지식발견 프로세스

데이터마이닝은 기계학습(machine learning), 패턴인식(pattern recognition), 데이터베이스, 통계학, 인공지능, 지식획득(knowledge acquisition), 데이터 시각화, 고성능 컴퓨팅 등 다양한 연구분야와 관련을 가지면서 많은 기법들이 개발되었다[1]. 그림 6은 데이터마이닝의 분석유형, 분석기법, 그리고 적용 예를 정리한 것이다. Goebel과 Gruenwald[3]는 그림 6의 분석유형에 대해 다음과 같은 정의하였다.

classification(분류)

미리 정의된 범주형 클래스가 있을 때, 어떤 특정 데이터 아이템이 어느 클래스에 속하는지 결정하는 것. 예를 들어 고객을 우량고객, 불량고객, 일반고객의 세 클래스로 구분할 때, 특정 고객이 어느 클래스에 속하는지 분류하는 것이다. 판별분석(discrimination analysis), 회귀분석, 의사결정나무(decision tree), 신경망(neural network) 등의 분석 기법이 사용된다.

clustering(군집화)

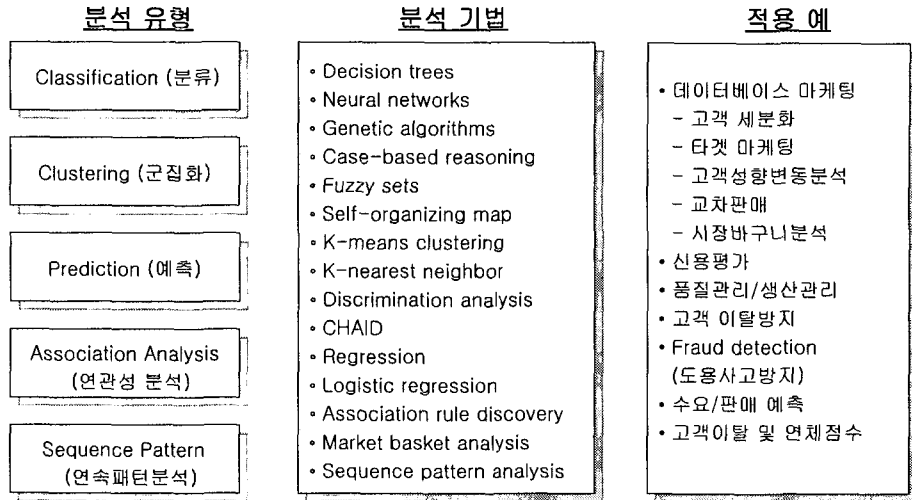
데이터 아이템의 집합이 주어졌을 때, 비슷한 특징을 가진 아이템들이 함께 그룹지어지도록 데이터 아이템의 집합을 몇 개의 소 집단으로 분할하는 것. 고객 집단을 구매 행태가 비슷한 몇 개의 소집단으로 구분하는 것이 군집화의 예가 될 수 있다. 군집화에 사용되는 분석기법으로는 self organizing map, k-means clustering, k-nearest neighbor 등이 있다.

prediction(예측)

데이터 아이템과 예측모형이 주어졌을 때, 그 데이터 아이템의 특정 속성 값을 예측하는 것. 신용카드의 경우, 남의 카드를 훔치거나 주워서 몰래 불법적으로 사용하는 것을 방지하는 것이 매우 중요하다 (fraud detection). 이를 위해 신용카드의 거래당일 누적 회수, 거래간 시간차이, 평균거래금액, 사용장소 등을 이용하여 신용카드 불법사용 예측모형을 만든 후, 특정 카드거래가 불법일 가능성을 예측할 수 있다.

association analysis(연관성분석)

데이터 아이템에 어떤 패턴이 존재하는 것이 다른 패턴이 존재하는 것을 암시하는 것과 같은 관계를 찾아 내는 것. “맥주를 구입한 사람의 70%는 콜라도 같이 구입한다”와 같이 하나의 거래나 사건에 포함되어 있는 항목들간의 연관성을 파악하는 것이 그 예가 될 수 있다.



[그림 6] 데이터마이닝 기법

sequence pattern(연속패턴분석)

연관성 분석에 시간정보가 포함된 형태로서, 시간 상에 순차적으로 나타나는 데이터 아이템 간의 관계를 찾아 내는 것이다. 주식 투자에서 “지난 주에 X 주식이 5일 동안 10% 오르고 Y주식이 같은 기간 동안 10~20% 올랐을 때, 금주에 Z주식이 오를 확률은 70%이다” 와 같은 패턴을 찾아내는 것이 연속패턴분석의 예이다[4].

이와 같은 분석유형을 지원하기 위해 다양한 분석기법들이 개발되었다. 많은 경우 통계학과 기계 학습 분야에서 기존에 연구되었던 방법들이 데이터 마이닝에 적용되고 있다. 본 논문에서는 각각의 기법에 대한 자세한 설명은 생략하기로 한다. 그림 6의 맨 우측은 데이터마이닝의 적용 사례를 정리한 것이다. 데이터베이스 마케팅, 고객의 신용평가, 고객 이탈방지, 수요 및 판매 예측 등에 활용되고 있으며, 금융·통신·유통업체에 적용된 사례는 많은 반면 제조업체에 적용된 사례는 별로 없는 상황이다.

대규모의 다차원 데이터를 동적으로 온라인에서 분석하고, 통합하고 보고서를 만드는 과정을 OLAP(Online analytical processing)이라고 하며, 이를 지원하는 도구를 OLAP 도구라 한다. OLAP 도구를 이용하면 반도체 제조공정 내의 수많은 변수들을 다양한 차원을 서로 교차하면서 분석할 수 있으므로, 현장 전문가의 경험을 십분 발휘하여 문제의 원인을 빠른 시간 내에 발견할 수 있도록 지원할 수 있을 것이다.

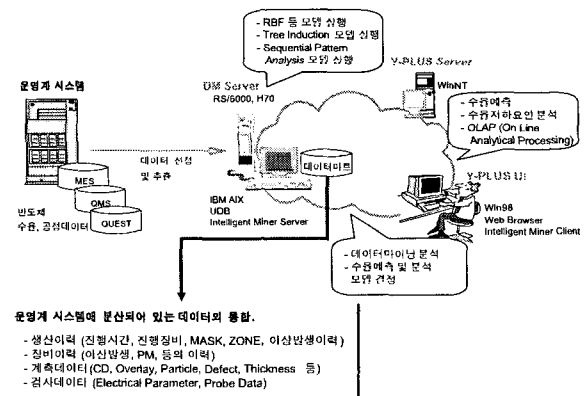
4. Y-PLUS 시스템

4.1. Y-PLUS 시스템 개요

Y-PLUS(Yield Prediction, anaLysis & Up Support) 시스템은 과학적이고 종합적인 수율분석을 통해 최적의 공정 조건을 산출하고 수율 문제원인을 조기에 해결함으로써 신제품의 수율을 빠른 시간 내에 일정 수준 이상으로 끌어 올리고 기존 제품의 수율 향상 및 안정화를 목적으로 개발되었다.

주요 개발내용은 저수율의 원인분석, 목표수율 달성을 위한 주요 공정 파라미터 관리치 제공, 관리치 이탈 항목에 대한 문제공정 또는 장비의 규명, 그리고 수율 예측 등이다.

그림 7은 Y-PLUS 시스템의 구성도이다. Y-PLUS 시스템은 데이터마이닝 서버(DM Server), Y-PLUS서버, Y-PLUS UI 등 3 tier 구조로 되어 있으며, 데이터마이닝 서버에는 데이터마이닝 툴인 IBM의 intelligent miner가 설치되어 있다. 분석에 필요한 데이터를 위해 MES(manufacturing execution system), QMS(Quality Management System), QUEST 등 기존의 운영계 시스템에 흩어져 있는 생산이력, 장비이력, 계측데이터, 검사데이터 등 필요한 데이터를 자동으로 추출/변형/적재하여 데이터마트를 구성하였다. 사용자는 데이터마이닝 서버에 직접 접근하여 데이터마이닝 기법을 이용한 수율 예측 및 분석 모델을 설정하거나, 웹브라우저를 이용해 이미 만들어진 RBF, tree induction, sequential pattern analysis 등 모델을 실행토록 Y-PLUS 서버에게 요구하여 수율 예측이나 수율저하요인을 분석할 수 있으며, 또한 OLAP분석을 할 수도 있다.



[그림 7] Y-PLUS시스템 구성도

4.2. 데이터마이닝 기법 적용 기능

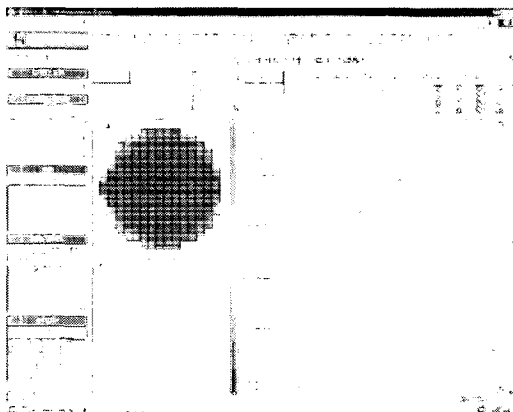
본절에서는 Y-PLUS 시스템 내 구현된 기능 중 데이터마이닝 기법을 적용한 경우를 해당 문제상황과 적용한 데이터마이닝 기법을 중심으로 설명한다.

#### Fail category의 AUF성 자동판별

2절에서 설명한 바와 같이, FAB 공정을 마친 웨이퍼는 생성된 칩이 원하는 전기적인 특성을 제대로 갖추고 있는지 검사하는 probe검사를 받게 된다. probe검사를 통해 웨이퍼 내의 각 칩은 양품 또는 불량품의 판정을 받게 되는데, 불량이 발생한 경우 probe검사를 담당하는 장비는 그 불량 형태에 따라 불량을 몇 가지 종류로 구분한다. 예를 들어 배선계열에 문제가 있어 불량이 발생하는 경우 그 불량 정도에 따라 '3불량', '2불량' 등과 같이 분류하고 그 외에도 'W불량', 'R불량' 등 약 10가지 종류의 불량으로 구분하고 있다. 이러한 불량의 종류를 현업에서는 Fail category라 부르고 있다.

많은 경우에 특정 Fail category (예, 3불량)가 웨이퍼 상에서 지리적으로 특정 위치에 집중적으로 발생하는 현상이 일어나곤 한다. 이를 현업에서는 AUF(area usage factor)성 불량이라고 한다. 예를 들어 3불량이 웨이퍼의 중앙 부분에 집중적으로 발생하면, 그 웨이퍼는 '3불량'-'중앙' AUF성 불량이라고 한다. AUF성 불량 타입은 그 발생 위치에 따라 중앙, 상, 하, 좌, 우, 좌상, 좌하, 우상, 우하, 주변 등으로 구분하고 있다. 그림 8의 우측 부분은 R불량이 '주변' AUF성 불량 타입으로 발생한 웨이퍼들을 보여 주고 있다. 웨이퍼 상에서 검게 나타난 부분이 R불량이 발생한 칩을 나타낸다.

AUF성 불량이 발생하는 원인은 다양하다. 특정



[그림 8] AUF 성 불량 자동 판별

공정을 담당하는 장비가 오염물질을 웨이퍼 상의 특정 부분에 발생시킨다거나, 식각공정의 경우 웨이퍼의 위치에 따라 온도 차이가 발생하여 식각정도가 달라 불량이 발생하는 등 다양한 원인이 존재한다. 그런데, AUF성 불량을 현업에서 중요하게 생각하는 이유는 동일한 AUF성 불량이 발생한 웨이퍼들의 경우 대부분 동일한 원인 때문에 불량이 발생했다는 인과관계 때문이다. 즉, 동일한 AUF성 불량이 발생한 웨이퍼들을 모아서 그것들이 공통적으로 거쳐온 장비를 찾아내거나 공정의 파라미터를 조사해 보면 쉽게 그 원인을 찾을 수 있다는 것이다.

지금까지는 현업의 엔지니어가 컴퓨터 화면상에 보여지는 웨이퍼 map을 보고서 AUF성 불량이 발생했는지 일일이 육안 검사를 하였다. 육안검사로

인해 많은 시간과 노력이 들게 되고, 엔지니어의 숙련도에 따라 검사의 질이 달라지는 문제가 있었다.

AUF성 불량 타입을 판별하는 문제는 군집화 문제에 해당한다고 볼 수 있다. 웨이퍼 상에 발생한 특정 Fail category에 대해 지리적으로 가깝게 위치한 것들이 하나의 군집에 포함되도록 군집화시킨 후, 그 군집이 존재하는 위치에 따라 AUF성 불량 타입을 결정할 수 있을 것이다. 본 연구에서는 군집화를 위해 k-means clustering 알고리즘을 응용한 알고리즘을 개발하였다. Distance measure는  $\max(|X_i - X_j|, |Y_i - Y_j|)$ 를 사용하였는데, 여기서  $(X_i, Y_i)$ 는 칩의 웨이퍼 상의 위치를 나타내는 것으로 좌측 상단에 위치한 칩을 기준점 (0,0)으로 하여 설정한 상대적 위치이다.

AUF성 불량을 군집화 알고리즘을 이용해 빠른 시간 내에 정확하게 판별할 수 있게 됨에 따라 수율 저하의 원인을 신속하게 파악할 수 있게 되었다. 앞서서도 말했듯이 동일한 불량 형태를 보이는 웨이퍼들을 모으고, 그 웨이퍼들을 갖는 공정 상의 공통점을 찾으면 불량의 원인을 신속, 정확하게 파악할 수 있기 때문이다.

#### 원인장비 탐색

Probe검사가 끝난 로트 중, 분석을 위해 몇 개의 로트를 모아 놓은 것을 '분석그룹'이라고 하자. 분석그룹을 설정하는 주된 이유는 불량이 발생한 하나의 로트에 대해 그 불량의 원인을 찾기 보다는 비슷한 형태의 불량이 발생한 로트의 집합을 대상으로 그 불량 원인을 찾는 것이 더 확실하게 원인을 파악할 수 있기 때문이다. 분석그룹을 설정하는 방법은 앞의 Fail category의 AUF성 자동판별을 이용하여 동일한 AUF성 불량 타입을 갖는 로트들의 집합을 이용할 수도 있으며, 현업의 엔지니어가 직접 로트들을 지정하여 그룹을 지을 수도 있다.

일단, 불량의 원인을 파악하고자 하는 분석그룹이 만들어지면 '원인 장비 탐색'에서는 어떤 장비가 불량을 발생시켰는지와 같이 생산장비 측면에서 불량의 원인을 파악하게 된다. 즉, 분석그룹에 포함된 로트들이 공통적으로 진행한 장비 또는 장비들의 sequence를 발견함으로써 원인이 되는 장비를 찾아내는 것이다.

이를 위해 Y-PLUS 시스템에서는 Intelligent miner의 '연속패턴 마이닝 함수'를 사용하였다. 입력되는 데이터의 형태는 분석그룹 내의 각 로트가 FAB공정을 시작할 때부터 끝날 때까지 거쳐간 장비들의 리스트이다. 데이터마이닝의 결과는 분석그룹 내의 로트들이 공통적으로 거쳐간 장비나 장비들의 sequence를 발견해 주는 것이다.

표 2는 어떤 분석그룹에 대한 불량 원인장비 탐색의 결과를 비교그룹과 대비하여 보여 주는 것이다. 표에서 support가 의미하는 바는 분석그룹 (또는 비교그룹) 내에 포함된 로트 중 해당 장비를 거쳐간 로트의 비율을 의미한다. 표 2에서 분석그룹에 있는 로트들의 90%가 1430번 L-ASHER공정의 E-9A-B 장비를 거쳐서 공정을 진행했음을 알 수 있다.

장비	Support1 (분석그룹)	Support2 (비교그룹)	차이
1430 L-ASHER E-9A-B	90%	89%	1%
1415 L-DRY-ET E-2B-05B	89%	25%	64%
1290 L-SOX-CL D-03-04	50%	45%	5%
1330 L-NIT-DE D-5A-06	30%	40%	-10%
2230 LNW ASHER E-9B-06	20%	22%	-2%

[표 2] 원인 장비 탐색 결과

이 분석에서는 분석그룹 뿐만 아니라 비교그룹에 대해서도 분석을 수행한다. 비교그룹은 분석그룹과 비슷한 시기에 probe검사를 하였지만 불량이 없거나 불량률이 있더라도 서로 상이한 불량타입을 보이는 로트들의 집합으로 구성된다. 비교그룹을 설정한 이유는 분석 결과의 잘못된 해석을 방지하기 위해서이다. 비교그룹을 설정하지 않았다면 "분석그룹의 90%가 1430번 L-ASHER공정의 E-9A-B를 공통적으로 거쳐 갔기 때문에 분석그룹의 불량원인은 이 장비일 가능성이 높다" 라는 첫부른 결론을 내릴 수 있다. 하지만 비교그룹에 대한 support가 89%라는 수치를 보면 그러한 결론을 내릴 수 없을 것이다. 두 그룹 모두 이렇게 높은 support가 나온 이유는 1430번 공정의 장비 중 E-9A-B의 가동률이 원래 그렇게 높거나 다른 장비의 고장 등으로 로트들이 어쩔 수 없이 E-9A-B를 진행한 경우라고 해석할 수 있을 것이다. 1415 L-DRY-ET공정의 경우, 분석그룹의 89%가 E-2B-05B에서 작업을 진행한 반면 비교그룹은 단지 25%만이 작업을 진행했음을 알 수 있다. 동일한 불량 타입을 갖는 분석그룹은 E-2B-05B에서 작업한 비율이 높은 반면 정상 또는 이질적인 불량을 갖는 로트들로 구성된 비교그룹은 작업비율이 낮은 것으로 보아, E-2B-05B 장비가 분석그룹의 불량 원인 장비일 가능성이 높다는 결론을 내릴 수 있을 것이다.

#### 원인파라미터 탐색

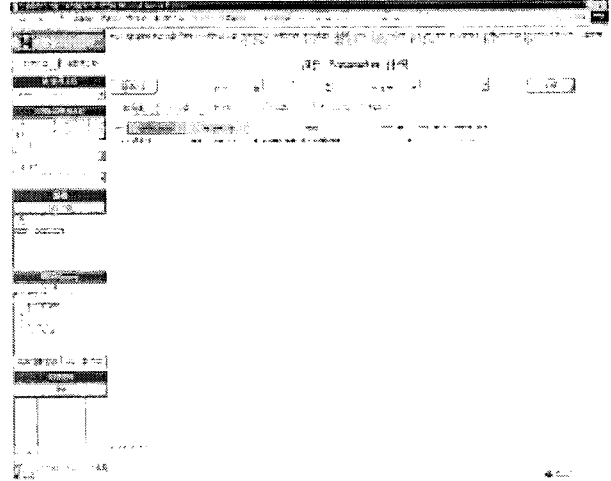
각 생산 공정에서는 여러 가지 공정 파라미터를 설정하여 작업을 하게 되는데, 파라미터의 잘못된 값 설정이 저수율의 원인이 되는 경우가 많이 있다. 각 단위공정별로 통계적 품질관리 기법을 이용하여 파라미터가 관리치를 벗어나는 이상발생 여부를 실시간으로 감지하지만, 여기서 문제가 되는 것은 해당 파라미터의 관리치가 잘못 설정되어 있으면 수율에 악영향을 미친다는 점이다.

FAB공정의 공정 단계 수가 200~300개이고 각 공정단계에서 설정해야 할 파라미터 수가 십 여개씩 있다는 것을 생각하면, 어떤 파라미터의 잘못된 관리치 설정이 수율에 나쁜 영향을 미친다는 것을 파악하는 것이 결코 쉬운 일이 아님을 알 수 있다.

불량 원인파라미터 탐색은 분류(classification) 문제에 해당한다. 문제가 되는 로트들인 분석대상 그룹과 문제가 없는 반대특성그룹을 설정하고, 두 그룹간의 차이를 발생하게 한 공정 파라미터를 decision tree 기법을 적용하면 찾아낼 수 있다. 불량 원인 파라미터 탐색을 위해 intelligent miner의 '트리분류 마이닝 함수'를 이용하였으며, 입력 데이터의 형태는 각 로트가 공정을 진행할 때 해당 공정 장비에서 설정한 파라미터 값들의 리스트이다.

그림 9는 불량 원인 파라미터 탐색의 결과를 보여 주는 화면이다. 동일한 수로 이루어진 분석대

상그룹 F-2-8SHOT과 반대특성그룹 F2-4의 차이를 알아 보기 위해 원인 파라미터 탐색을 한 결과, L-건식각 공정에서 64SLE-L/64SLE-L recipe로 작업한 경우 C1415MP0358STD 파라미터가 0.00916 미만일 경우에 F-2-8SHOT 그룹은 9개(75%)가 포함되고 F2-4 그룹은 3개(25%)가 포함된다. F-2-8SHOT그룹이 대다수인 것으로 보아 반대특성 그룹에 대한 F-2-8SHOT그룹의 특성은 C1415MP0358STD가 0.00916보다 작다는 것이다.



[그림 9] 불량 원인 파라미터 탐색

이 때 오차(ERROR)는 25%이다. 따라서 이 변수의 설정치를 분석대상그룹의 문제 원인 파라미터로 의심해 볼 수 있다.

#### 수율예측

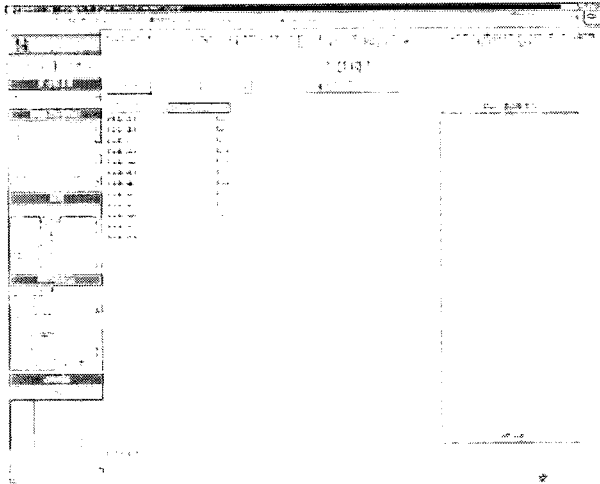
반도체 공정의 특성 상 모든 공정이 끝나야 제품의 양품/불량품을 판정할 수 있다. 그런데 제조공정이 수 개월인걸 감안해 본다면 각 로트가 모든 공정을 마쳤을 때의 대략적인 수율을 공정 진행 중에 미리 예측할 수 있다면 생산관리 측면에서 상당히 바람직한 효과를 거둘 수 있을 것이다.

수율예측을 위해 Y-PLUS 시스템에서는 intelligent miner의 RBF(radial basis function) 예측함수를 사용하였다. RBF 예측함수는 multi-layer perceptron 신경망 구조와 유사하나 은닉층이 하나만 있고 각 노드의 결합함수를 비선형함수를 사용한다는 점이 다른 점이다.

어느 특정 공정 단계 k에 있는 로트의 향후 수율을 예측하는 모델을 만들기 위해 probe검사까지 마친 로트들의 계측데이터와 수율 데이터를 이용하여 학습을 시켰다. 입력 데이터의 형태는 ((공정 k-1 단계까지의 주요 계측 데이터), 수율)이다. 즉, k-1 단계까지 수집된 계측데이터를 이용해 모든 공정 단계를 마친 뒤의 수율을 예측하는 것이다. 물론, 공정 단계 k가 전체 공정에서 앞 부분에 위치한 공정이라면 k 공정 이후에 너무 많은 수율 변동요인이 존재하므로 정확한 수율예측은 어려운 일이다. 하지만 공정 k가 상대적으로 뒷부분에 위치한 공정이라면 비교적 정확한 수율예측이 가능해진다. 그림 10은 RBF 예측 함수를 이용해 M1 Photo 공정에 있는 로트들의 향후 수율을 예측한 결과를 보여 주는 화면이다.

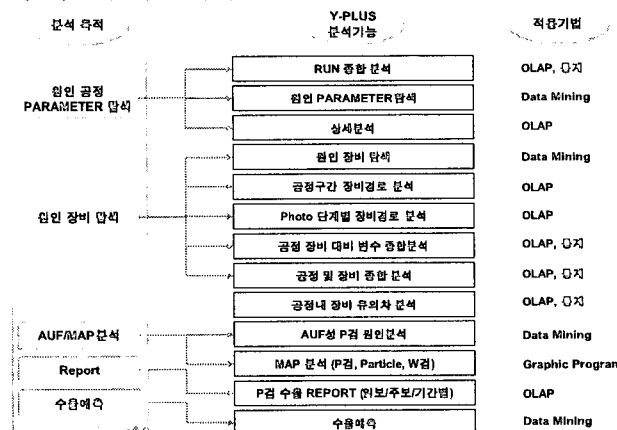
#### 4.3. Y-PLUS 시스템의 기타 기능

Y-PLUS 시스템에는 데이터마이닝을 이용한



[그림 10] 수율예측

기능 외에 OLAP을 활용한 다양한 기능이 존재한다. 여기 저기 흩어져서 존재하던 수율분석에 필요한 데이터를 한 곳에 모아 데이터마트를 구성함으로써 현업의 엔지니어는 다차원적으로 수율의 문제를 분석할 수 있게 되었다. 이 시스템이 구축되기 전에는 수율문제를 분석하기 위해 전산부서에 필요한 데이터에 요청하면 상당한 시일이 흐른 뒤에야 데이터를 받을 수 있었고 이를 엑셀 등을 이용하여 분석하였다. 하지만 데이터마트가 구축되고 다차원 분석이 가능한 OLAP 도구가 제공됨에 따라 현업의 엔지니어는 즉각 수율의 문제를 분석할 수 있게 되었다. 이 분석과정에 엔지니어의 과거 분석 경험이 충분히 반영되어 상당히 효율적이고 효과적으로 수율 문제를 분석할 수 있게 되었다. 그림 11은 Y-PLUS 시스템의 분석기능을 설명하는 그림이다. 각 기능에 대한 자세한 설명을 생략하기로 한다.



[그림 11] Y-PLUS 시스템 분석기능

## 5. 결론

반도체 공정에서의 수율관리는 매우 많은 공정 단계, 긴 생산 주기시간, 그리고 엄청난 수의 관련 변수가 복잡하게 얽혀 있는 매우 복잡한 문제이지만 초기 시장 선점과 원가경쟁력 확보라는 측면에서 반드시 해결해야만 하는 매우 중요한 문제이다. 수율관리를 위해 다각적인 노력을 기울이고 있는 상황이나, 기존의 통계적인 방법이나 엔지니어의 경험적 분석방법만으로는 해결하지 못하는

복잡한 수율 저해 원인들이 존재한다.

본 논문은 기존 수율관리시스템의 문제점을 개선하기 위해 데이터마이닝 기법과 다차원분석 도구를 활용한 수율 관리 방법을 제안하고 이를 정보 시스템으로 구축한 Y-PLUS (Yield Prediction, anaLysis & Up Support) 시스템을 소개하였다. 이 시스템은 군집화, 연속패턴분석, 분류, 신경망 등 데이터마이닝 기법을 적용하여 수율의 원인이 되는 문제 장비나 문제 파라미터를 신속, 정확하게 파악할 수 있도록 해 줄 뿐만 아니라 공정 진행 중인 제품의 미래 수율을 예측할 수 있도록 지원한다. Y-PLUS 시스템은 현업에 적용하여 매우 큰 효과를 보았으며 계속해서 그 적용범위를 확대해 나가고 있는 상황이다.

## 참 고 문 헌

- [1] W. J. Frawley, G. Piatetsky-Shapiro, C. J. Matheus, Knowledge discovery in databases: An overview, Knowledge discovery in databases, AAAI Press/MIT Press, pp. 1-27, 1991.
- [2] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, From data mining to knowledge discovery in databases, Advances in knowledge discovery and data mining, AAAI Press/MIT Press, pp. 1-34, 1996.
- [3] M. Goebel, L. Gruenwald, A survey of data mining and knowledge discovery software tools, 1999 ACM SIGKDD, vol. 1, Issue 1, pp. 20-33, 1999.
- [4] 장남식, 홍성완, 장재호, 데이터마이닝, 대청, 1999.
- [6] 강현철, 한상태, 최종후, 김은석, 김미경, SAS Enterprise Miner 4.0을 이용한 데이터마이닝, 자유 아카데미, 2001.
- [7] Intelligent miner for data 사용자 매뉴얼 버전 6 릴리스 1, IBM, 1999.
- [8] 반도체 공정의 생산 및 공정관리 시스템 개발 최종보고서, 과학기술처 (한국과학기술원 위탁 과제), 1996.
- [9] 지원철, 김민용, 데이터마이닝과 의사결정 지원 시스템, 정보과학회지, 제 16권, 제 9호, 1998.