한국지능정보시스템학회논문지 제9권 제1호 2003년 6월(pp.157~177)

데이터마이닝을 이용한 반도체 FAB공정의 수율개선 및 예측

백동현
천안대학교 경상학부 조교수
(estarbaek@cheonan.ac.kr)

한창희
한양대학교 디지털 경영학부 전임강사
(chan@hanyang.ac.kr)

본 논문은 반도체 FAB공정의 수율개선 및 예측을 위해 데이터마이닝 기법을 적용한 사례를 소개한다. FAB공정의 복잡성과 생산현장에서 수집되는 방대한 기술데이터로 인해 기존의 통계적 방법이나 엔지니어의 경험적 분석 방법만으로는 미처 파악하지 못하는 수율 저하 요인이 상당 수 존재한다. 본 논문은 먼저, FAB공정을 마친 웨이퍼에 불량 칩(chip)이 지리적으로 특정 위치에 집중적으로 발생하는 현상을 육안검사 대신 군집분석을 이용하여 데이터로부터 자동 판별할 수 있는 방법을 제안한다. 다음으로 연속패턴분석, 분류분석, RBF(Radial Base Function) 기법을 적용하여 수율 저하의 원인이 되는 문제 장비나 문제 파라미터를 신속, 정확하게 파악할 수 있도록 해 줄 뿐만 아니라 공정 진행 중인 제품의 미래 수율을 예측할 수 있도록 지원하는 방법을 제안한다. 또한 위 기법들을 반도체 FAB공정을 대상으로 국내 모 반도체 회사에서 정보시스템으로 구현한 Y2R-PLUS (Yield Rapid Ramp-up, Prediction, anaLysis & Up Support) 시스템을 소개한다.

논문접수일 : 2002년 12월 게재확정일 : 2003년 4월 교신저자 : 백동현

1. 서론

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

특히, "수율과의 전쟁"이라 표현할 만큼 수율이 반도체 산업에서 차지하는 의미는 매우 크다. 수율(yield)은 투입량 대비 완성된 양품의 비율로 반도체 칩(chip)이 고집적화 됨에 따라 높은 수율을 확보하기가 점점 어려워지고 있다. 또한, 공정이나 장비의 관리 미흡, 작업자의 사소한 실수, 그리고 설계나 공정 상의 파라미터 설정치 오류 등에도 수율은 상당한 영향을 받는다. 경제적 측면으로 보면 새로운 신상품이 개발되면 이를 일정 수준 이상의 수율로 하루 빨리 끌어 올려 양

* 이 논문은 2003년도 천안대학교 교내연구비 지원에 의하여 연구되었음

산체계로 이관할 수 있어야만 시장에서의 경쟁력을 확보할 수 있다. 또한 일정 수준 이상의 수율을 확보한 제품이라 하더라도 더 높은 수율을 유지할 수 있어야만 원가 절감으로 인한 가격 경쟁력이 생길 수 있는 것이다.

반도체 제조공정은 웨이퍼가 투입되어 완제품이 생산되기까지 wafer fabrication(이하 FAB공정), probe검사(probe test), 조립(assembly), 그리고 최종검사(final test) 등 4개 공정을 거쳐야 하며 총 3~4개월 정도 소요된다. 이중 특히 FAB공정은 가장 자본 집약적이고 생산 공정이 복잡하며 생산주기시간이 2~3개월 소요되는 공정으로 반도체 생산성과 수율에 가장 큰 영향을 미치는 공정이다. 64M DRAM 기준으로 볼 때 FAB공정 내의 단위공정 수가 약 350여 개, 생산장비 수가 500여 대, 계측장비 수가 250여 대이며 계측되는 변수의 종류만 해도 무려 500여 가지가 되어 한 달 평균 150 기가 바이트 이상의 방대한 기술 데이터가 수집된다. 생산현장에서 생성되는 방대한 기술 데이터를 효과적으로 활용하여 반도체 수율 향상에 도움을 줄 수 있는 정보시스템들이 개발되어 사용되고 있으나 좀 더 지능적이고 효과적인 지원시스템이 필요한 상황이다.

현재 사용되고 있는 수율관리시스템 중 대표적인 것은 통계적 품질관리(statistical quality control) 기법을 이용한 것이다. 자동으로 수집되는 다양한 계측 데이터를 이용하여 x관리도, $\bar{x}$-R관리도, 공정능력지수관리 등을 통해 단위공정의 품질을 관리하거나 장비 이상발생 이력 등을 관리하고 있다. 또한 상관분석, 회귀분석, 유의차 분석 등 다양한 통계적 기법이 사용되고 있으며 검사 데이터의 경우 그래픽을 이용한 map분석 기능을 활용하고 있다. 이밖에도 6시그마 활동, 엔지니어의 노하우를 축적, 공유, 활용하는 지식관리활동 등 다각적인 노력을 기울이고 있다.

다각적인 품질개선 활동을 전개하고 있음에도 불구하고 기존의 통계적 기법이나 map 분석 기능, 그리고 엔지니어의 경험적 분석만으로는 미처 파악하지 못하는 수율 저하 요인이 다수 존재한다. 특히, 수많은 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 1>은 기존 수

<표 1> 기존 수율관리시스템의 문제점과 개선방향

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 이루어지고 있으나, 전체 공정을 고려한 개선활동은 미비함	종합적 안목의 분석 방법 필요
엔지니어가 수많은 공정 변수를 모두 분석하기에는 너무 많은 시간이 소요되어, 실제적으로 일부 변수만이 분석되고 있음	수많은 공정 변수를 빠른 시간 내에 분석할 수 있는 방법 필요
엔지니어의 자질에 따라 분석 품질이 다름	분석의 일부를 지능적으로 지원하는 지원시스템 필요

율관리시스템의 문제점과 개선방향에 대해 정리한 것이다.

데이터마이닝(data mining)은 "대용량의 데이터로부터 이들 데이터 내에 존재하는 관계, 패턴, 규칙 등을 탐색하고 모형화하기 위해 알고리즘을 적용하는 단계"로 정의할 수 있다 (Fayyad 등, 1996). 기업들의 데이터베이스가 대형화되고 분석 변수가 많아져 기존 인간 두뇌에 의존한 자료검색 방법만으로는 데이터베이스 내에 숨겨져 있는 지식을 찾아내는 데 한계에 도달하였다. 이에 컴퓨터가 숨겨진 지식을 발견하게 하는 방법이 필요하게 되었는데 이것의 한 분야가 데이터마이닝이다. 데이터마이닝은 유통, 통신, 금융 등의 산업을 중심으로 고객관리 및 수요/판매 예측 등에 활발히 적용되고 있다. 반면, 제조 산업에도 데이터마이닝 기법의 적용이 생산성과 품질 향상 등에 기여할 수 있는 부분은 많이 있으나 아직 그 적용사례는 많지 않다.

본 논문은 반도체 FAB공정의 품질개선을 위해 데이터마이닝 기법을 적용한 사례를 소개한다. FAB공정을 마친 웨이퍼에 불량 칩(chip)이 지리적으로 특정 위치에 집중적으로 발생하는 현상 (AUF성 불량)을 시간이 많이 소요되고 엔지니어의 수준에 따라 편차가 있는 기존의 육안 검사에서 데이터마이닝 기법인 군집분석을 이용하여 데이터로부터 자동 판별할 수 있도록 구현하였다. 또한 연속패턴분석, 분류분석, RBF(Radial Base Function) 등 데이터마이닝 기법을 적용하여 수율 저하의 원인이 되는 문제 장비나 문제 파라미터를 신속, 정확하게 파악할 수 있도록 해줄 뿐만 아니라 공정 진행 중인 제품의 미래 수율을 예측할 수 있도록 지원하는 방법을 제안한다. 또한 위 기법들을 반도체 FAB공정을 대상으로 국내 모 반도체 회사에서 정보시스템으로 구현된 Y2R-PLUS (Yield Rapid Ramp-up, Prediction, anaLysis & Up Support) 시스템을 소개한다.

본 논문은 다음과 같이 구성되었다. Ⅱ장에 데이터마이닝에 대한 기본개념 및 품질개선 분야의 적용사례를 소개하며, Ⅲ장에 대상시스템인 반도체 제조공정, 특히 FAB공정에 대한 소개와 수율관리에 대해 설명한다. Ⅳ장은 반도체 FAB공정에서 데이터마이닝을 활용한 수율관리에 대해 데이터마이닝 기법 적용의 문제 상황과 적용기법을 Y2R-PLUS 시스템과 함께 설명한다. 마지막으로 Ⅴ장은 본 논문의 결론을 정리하였다.

2. 기존 연구

가트너 그룹(Gartner group)은 데이터마이닝(data mining)을 "저장된 대용량의 데이터를 패턴인식, 통계학, 수학 등의 기법을 이용해 분석함으로써, 의미 있는 새로운 상관관계, 패턴, 그리고 경향 등을 발견하는 프로세스"로 정의하였다. 또한 Frawley 등(1991)은 "대량의 실제 데이터로부터, 이전에 잘 알려지지는 않았지만, 묵시적이고, 잠재적으로 유용한 정보를 추출하는 작업"이라고 정의하였다. 이를 정리하면, 대용량의 데이터베이스에 숨겨져 있는 데이터간의 관계, 패턴, 규칙 등을 찾아내고 모형화하여 기업의 경쟁력 확보를 위한 의사결정을 돕는 유용한 정보로 변환하는 일련의 과정으로 볼 수 있다.

정보기술의 빠른 발전은 업무의 자동화를 촉진시켜 방대한 양의 데이터를 전자적으로 수집하고 보관하는 것을 가능하게 하였다. 예를 들면, 판매시점관리(POS) 데이터, 주식거래 데이터, 병원의 환자 데이터, 은행거래 데이터, 제조업체의

품질관련 데이터 등이 기하급수적으로 쌓여가고 있다. 데이터의 양이 오늘날과 같이 방대하지 않았던 과거에는 소수의 전문가들이 통계기법이나 질의(query)를 통해 데이터를 분석하고 요약된 결과를 보고서 형식으로 제공하였다. 그러나 데이터 양이 기하급수적으로 증대함에 따라 이전과 같은 수작업에 의한 데이터 분석방법으로는 어려움이 발생하게 되었다(장남식 등, 1999). 이에 방대한 양의 데이터로부터 유용한 지식을 얻도록 사람을 지능적이고 자동적으로 지원하는 새로운 기법과 도구가 필요하게 되었다 (Fayyad 등, 1996).

<표 2>는 품질개선 분야에 데이터마이닝 기법을 적용한 사례를 정리한 것이다. 백준걸 등(2000)은 반도체 에칭(etching) 공정에서 실시간으로 수집되는 기계 이력 데이터의 분석을 통해 현재의 기계 상태를 정확히 반영한 기계 고장 정보를 추출해 낼 수 있는 점진적 의사결정나무 구축 방법을 제시하였다. 에칭율에 영향을 미치는 공정변수 (기계 내의 압력, 온도, 에칭가스 밀도, 이온 전압, 웨이퍼 위치 등)를 속성으로 정의하고 에칭율에 따라 결과 클래스 (LE: Low Etching rate, NE: Normal Etching rate, HE: High Etching rate)를 정의한 이력 데이터를 기반으로 에칭율의 이상 상황을 예측할 수 있는 의사결정나무를 작성하여 활용하였다.

안진석 등(1999)은 자동차 휠 주조공정에서 범주형 데이터와 연속형 데이터의 혼합 형태로 공정변수가 주어지며 품질특성치로서 불량률 데이터가 주어진 경우에 최적 공정조건에 대한 정보 추출을 위해서 연속형 변수의 범주화와 유사 빈도의 군집화 기법을 제시하였다. Hancock 등(1996)은 자동차 밸브 주조공정에서 ridge regression을 이용하여 공정변수와 품질특성치간의 관계를 규명하여 불량률을 감소하기 위한 공정조건을 설정하였다.

Braha 등(2002)은 반도체 클린징 공정(cleansing process)에서 미세 오염물질 제거를 위해 레이저 빔을 이용하는 advanced wafer cleansing이라 불리는 건조 클린징 기술 (dry

<표 2> 품질개선 분야 데이터마이닝 적용 사례

저자	적용분야	적용기법	목적
백준걸 등 (2000)	반도체 에칭 공정	decision tree	에칭율 이상 상황 예측
안진석 등 (1999)	자동차 휠 주조공정	연속변수 범주화, 유사빈도 군집화	최적 공정조건 추출
Hancock 등 (1996)	자동차 밸브 주조공정	ridge regression	불량률 감소 위한 공정조건 설정
Braha 등 (2002)	반도체 클린징 공정	decision tree, neural network, composite classifiers, SOM	미세물질 제거공정의 온라인 모니터링
Kusiak 등 (2001)	PCB 조립공정	rough set theory	불량원인 파악
Yamashita (2000)	Debutanizer공장 시뮬레이터	decision tree, Naive-Bayes classifier	공장 운영에 필요한 지식 추출
Lian 등 (2002)	자동차 판금공정	상관분석, 최대나무기법, 다변량 통계분석	규격변동 원인 진단

cleansing technology)의 개선을 위해 데이터마이닝 기법을 적용하였다. 이 건조 클린징 기술과 관련하여 두 가지 그룹의 입력변수들 (에너지 요인들과 가스흐름요인들)과 두 개의 목표변수가 존재한다. 에너지 요인들(energy factors)은 (1) 레이저의 강도; (2) 한 클린징 세션의 펄스 수; (3) 웨이퍼 위의 클린징 패스(pass)의 수; (4) 레이저 빔의 각도; (5) 기판의 온도; (6) pulse-train의 주파수; (7) 연속되는 pulse-train 간의 시간간격 등이다. 가스흐름요인들(gaseous flow factors)은 (1) 결합된 가스의 압력; (2) 오존 흐름; (3) 산소 흐름; (4) nitrogen tri-fluoride 흐름 등이다. 목표변수는 원래 위치로부터 이동한 미세입자(particle)의 비율(%Moval)과 웨이퍼에서 제거된 미세입자의 비율(%Removal)이다. 수집된 데이터를 대상으로 분류분석 기법을 적용하기 위하여 연속형인 목표변수를 각각 전문가의 직관과 SOM (Self Organizing Map)을 이용하여 범주형으로 변환하였다. 각각의 목표변수에 대해 decision tree induction, neural networks, 그리고 composite classifiers를 적용하였다. 저자는 데이터마이닝 기법이 데이터가 성기고(scarce), 물리·화학적 파라미터가 상당히 복잡하게 영향을 미치는 분야에 적합하다고 주장하였으며, 데이터마이닝을 이용한 클린징 프로세스의 온라인 모니터링이 상당히 효과적일 것이라고 제안하였다.

Kusiak 등(2001)은 rough set theory를 이용하여 printed-circuit board(PCB) 조립 공정 중 일부 구성 요소 밑에서 발생하는 불량의 원인을 찾고자 하였다. 기존의 통계적 방법이나 실험계획적 방법으로는 불량의 원인을 명확히 파악하지 못하였다. 이 논문은 stencil 배합, stencil 두께, 위치, 진공여부 등 14개의 속성을 이용하여 불량의 원인을 찾는 방법을 제안하였다.

Yamashita(2000)는 Debutanizer 공장의 시뮬레이터에서 생성되는 데이터에 데이터마이닝 기법들을 적용하여 공장 운영에 필요한 지식을 추출하였다. 대부분의 프로세스 공장에서는 방대한 양의 운영 데이터(operational data)가 생성된다. 이 데이터는 공정설계, 운영 및 통제에 상당히 가치 있는 정보를 줄 수 있음에도 불구하고 데이터의 방대함과 복잡성 때문에 적적히 활용되지 못하고 있다. 이 논문에서는 의사결정나무 분석을 통해 사전 지식 없이도 데이터로부터 합당한 운영규칙을 찾아낼 수 있었으며 feature-subset selection wrapper algorithm, Naive-Bayes classifier, 그리고 nearest-neighbor classifier를 결합을 통해 운영 행동(action of operation)을 성공적으로 추정할 수 있었다.

Lian 등(2002)은 판금(sheet metal) 제품의 규격 변동(dimensional variation)의 원인을 진단하기 위해 데이터마이닝 기법과 지식발견기법을 적용한 의사결정지원 방법을 제안하였다. 자동차 차체의 규격 정확도는 이후 공정에 큰 영향을 미치므로 규격변동의 원인을 신속, 정확하게 파악하는 것은 매우 중요하다. 이 논문은 우선 방대한 다차원 측정 데이터로부터 변동폭이 큰 그룹을 추출하기 위해 상관분석과 최대나무(maximal tree) 기법을 적용하였으며, 주된 변동 패턴을 발견하기 위해 다변량 통계분석을 이용하였다. 또한 변동 원인의 진단을 위해 제품과 조립 공정 지식을 기반으로 한 의사결정나무 접근을 하였다. 저자들은 이러한 접근을 통한 판금 조립공정의 변동 진단이 상당히 효과적이고 효율적이라고 언급하였다.

본 논문은 반도체 FAB공정을 대상으로 1) 군집화 기법을 이용한 AUF성 불량 자동 판별; 2) 순차패턴(sequential pattern)분석 기법을 이용한

불량원인 장비 파악; 3) 의사결정나무 기법을 이용한 불량원인 공정 파라미터 파악; 4) RBF기법을 활용한 수율예측 등을 목적으로 하고 있다.

3. 반도체 제조공정

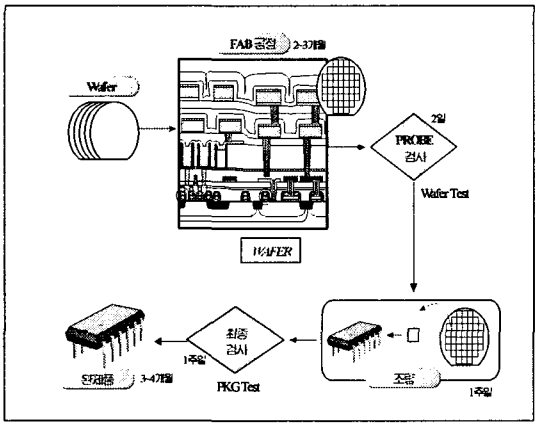
3.1 반도체 제조 흐름

반도체 제조공정은 크게 wafer fabrication (이하 FAB공정), probe검사(probe test), 조립(assembly), 그리고 최종검사(final test) 등 4단계로 구성되어 있다(한국과학기술원 산업공학과, 1996). <그림 1>은 반도체 공정을 설명한 것이다.

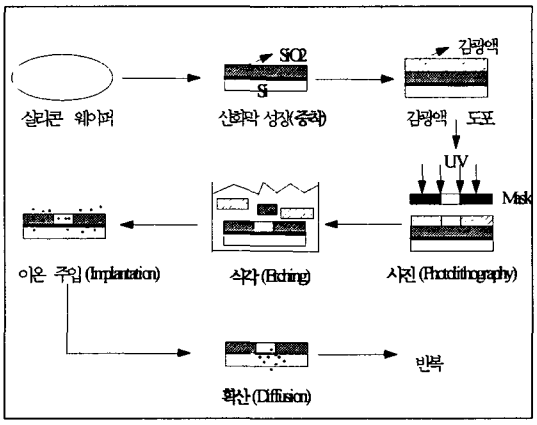
웨이퍼(wafer)는 반도체 제조를 위한 실리콘 기판으로, FAB공정을 거치면서 한 웨이퍼 위에 수 백개의 집적회로(IC, integrated circuit) 칩(chip)이 생성된다. Probe검사는 웨이퍼에 생성된 칩이 원하는 전기적인 특성을 제대로 갖추고 있는지 검사하여 양불량을 판별하는 단계이다. 조립공정은 만들어진 웨이퍼 내의 칩을 분리하여 칩의 전기적, 물리적 특성을 향상시키고 외부의 기계적, 물리적 충격으로부터 칩을 보호하기 위하여 형상화시켜 준다. 마지막으로, 최종검사 단계는 조립된 칩의 전기적 특성 및 기능, 신뢰성 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.

위 네 단계 중 FAB공정은 가장 자본 집약적이고, 생산 공정이 복잡하며(200~300개 공정단계) 또한 생산 주기시간이 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 2>와 같이 증착(deposition), 사진(photolithography), 식각(etching), 확산(diffusion) 등의 공정으로 구분된다.

증착공정은 웨이퍼 표면에 Oxide, Al, Au 등의 막을 형성하는 공정으로, 칩의 물리적 작용으



<그림 1> 반도체 제조공정



<그림 2> FAB 공정 흐름

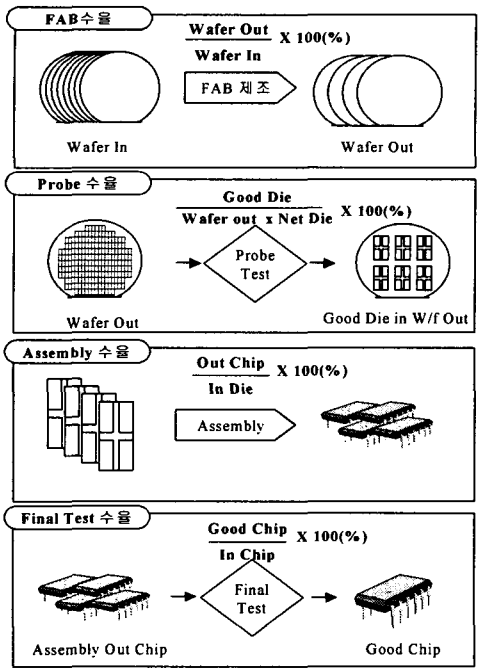
로부터 보호, 전기적 배선의 형성 등을 목적으로 한다. 사진공정은 패턴형성공정으로 마스크 상의 이미지를 웨이퍼 위에 전달시키는 공정이다. 피사체인 마스크에는 소자에 투사할 이미지가 그려져 있다. 감광액이 도포된 웨이퍼 위에 마스크를 정렬시키고 마스크 위에서 자외선(U.V)을 비추면 감광액이 선택적으로 감광된다. 식각공정은 사진공정에서 형성된 감광막 패턴을 마스크로 하여 그 형성된 형상을 하부 막에 전사시켜 가공한 후 불필요하게 된 감광막 패턴을 화학반응을 이용하여 제거하는 공정이다. 확산공정은 물에 잉크를 넣으면 퍼지는 것과 같이 웨이퍼 위에 원하는 불순물을 침적시킨 후 웨이퍼 내부로 퍼져 나가게 하는 공정이다.

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(Baek 등, 1998). 하지만 반드시 <그림 2>와 같은 순서대로 공정을 진행하는 것은 아니며, 층에 따라 일부 공정은 생략되거나 추가될 수 있으며 작업 순서가 바뀔 수도 있다.

3.2 수율관리

일반적으로 수율(yield)이라 함은 투입량 대비 완성된 양품의 비율을 의미한다. 100개를 투입하여 80개의 양품을 생산하였다면 수율은 80%가 되는 것이다. 대부분의 제조산업에서도 수율은 중요하지만, 반도체 산업에서 더욱 강조하는 이유는 반도체 소자가 급속하게 고집적화 됨에 따라 높은 수율을 확보하기는 점점 어려워지는 반면 시장의 조기 선점과 생산성 향상을 통한 수익성 확보를 위해서는 신제품의 안정적 수율까지 도달하는 시간(Ramp-up 기간)을 줄이는 것과 기존 제품의 안정적인 고 수율 유지가 매우 중요하기 때문이다. 또한, 초미세 오염물질에도 반도체 성능 및 수율이 큰 영향을 받기 때문이다. 그 때문에 반도체 공정은 0.1㎛ 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 초청정 환경에서 작업이 이루어지지만, 공정 또는 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 수율개선을 위한 핵심이 된다.

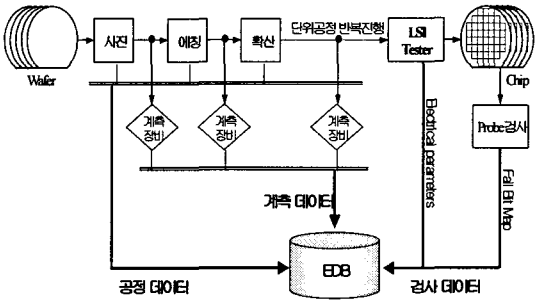
반도체 수율은 제조단계에 따라 FAB수율, probe수율, assembly수율, final test 수율로 구분되며, 이 4가지 수율을 곱하면 CUM 수율(cumulation yield)이 된다. <그림 3>은 이들 4가지 수율의 계산 방식을 설명한 것이다. FAB수율은 투입된 웨이퍼 장수(wafer in) 대비 중간에 깨지는 등의 불량이 발생하지 않고 제대로 나온 웨이퍼 장수(wafer out)의 비율이다. Probe수율은 wafer out된 총 die 수(wafer out×net die) 대비 probe검사 결과 양품인 die(good die)의 비율이다. 여기서 net die는 한 웨이퍼에 만들어지는 칩의 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

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 툴을 활용하여 효율적인 수율관리가 가능하도록 기반 시스템이 구축되어 있다.

FAB공정에서 발생되는 데이터는 <그림 4>에 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 단위 공정이 원하는 스펙(specification)대로 제대로 작업되었는지 검사하기 위해 계측장비로 cd, overlay, particle, thickness 등의 계측변수를 측정하며, 측정된 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.



<그림 4> 기술 데이터 수집

4. 반도체 FAB공정의 데이터마이닝을 활용한 수율관리

본 장에서는 우선 국내 모 반도체 회사에서 FAB공정을 대상으로 수율 저하의 원인분석 및 수율 예측을 위해 데이터마이닝과 OLAP 기법을 적용하여 개발한 수율관리시스템인 Y2R-PLUS에 대해 간단히 소개한다. 다음으로 Y2R-PLUS 시스템 내에 구현된 기능 중 데이터마이닝 기법을 적용한 경우를 해당 문제상황과 적용한 데이

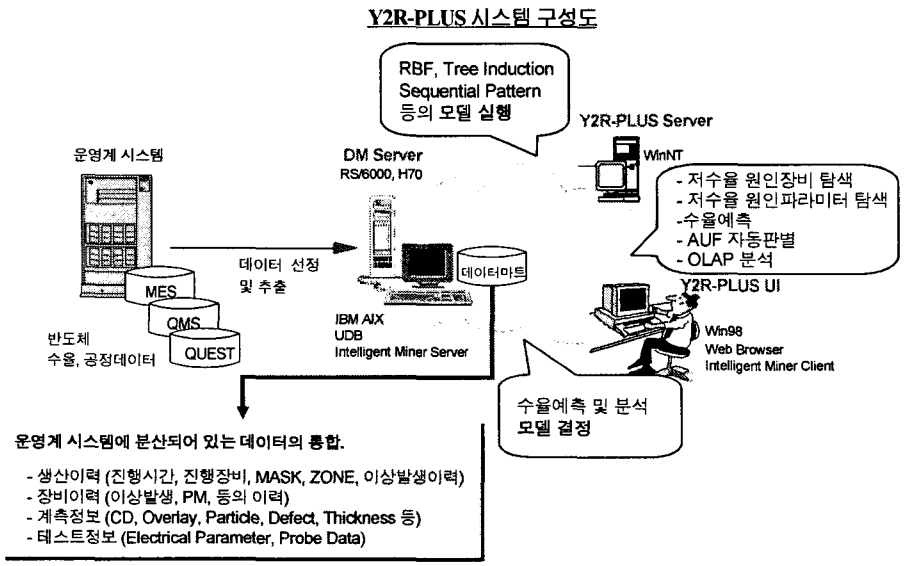
터마이닝 기법을 중심으로 설명하기로 한다.

4.1 Y2R-PLUS 시스템 개요

Y2R-PLUS(Yield Rapid Ramp-up, Prediction, anaLysis & Up Support) 시스템은 과학적이고 종합적인 수율 분석을 통해 최적의 공정 조건을 산출하고 수율 문제원인을 조기에 해결함으로써 신제품의 수율을 빠른 시간 내에 일정 수준 이상으로 끌어올리고 기존 제품의 수율 향상 및 안정화를 목적으로 개발되었다.

<그림 5>는 Y2R-PLUS 시스템의 구성도이다. Y2R-PLUS 시스템은 데이터마이닝 서버(DM Server), Y2R-PLUS서버, Y2R-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 Y2R-PLUS 서버에 요구할 수 있다. 모델의 실행을 통해 수율 예측이나 수율저하요인을 분석할 수 있으며, 또한 OLAP 기능을 이용하여 다양한 레포트를 생성시킬 수 있다.

<그림 6>은 Y2R-PLUS 시스템의 기능에 대해 정리한 것이다. 수율 저하의 원인이 되는 잘못된 공정 parameter를 찾거나(원인공정 parameter탐색) 오염물질 발생 등 문제를 발생시킨 공정 장비의 탐색이 가능하며(원인장비탐색),

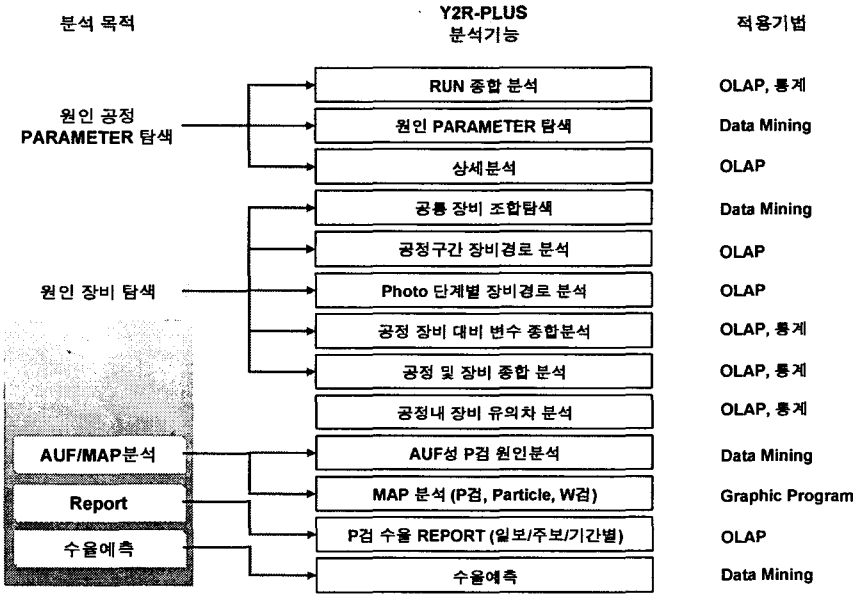


<그림 5> Y2R-PLUS 시스템 구성도

웨이퍼에 불량 칩(chip)이 지리적으로 특정 위치에 집중적으로 발생하는 현상을 기존의 육안 검사에서 군집화 기법을 이용하여 데이터로부터 자동 인식할 수 있도록 구현하였다 (AUF/MAP 분석). 또한 각종 레포팅 기능과 수율예측 기능을 내장하고 있다. 이를 위해 간단한 통계 기능, 데이터마이닝 기법, OLAP, 그리고 그래픽 프로그램을 활용하였다. 데이터마이닝을 적용한 기능에 대해서는 다음절에 자세히 설명하기로 하고, 여기서는 OLAP 적용의 중요성을 간단히 언급하기로 한다. 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 전산 부서에 필요한 데이터를 요청하면 상당한 시일이 흐른 뒤에야 데이터를 받을 수 있었고 이를 엑셀 등을 이용하여 분석하였다. 하지만 데이터마트가 구축되고 다차원분석이 가능한 OLAP 도구가 제공됨에 따라 현업의 엔지니어는 즉각 수율의 문제를 분석할 수 있게 되었다. 이 분석 과정에 엔지니어의 과거 분석 경험이 충분히 반영되어 상당히 효율적이고 효과적으로 수율 문제를 분석할 수 있게 되었다. 데이터마이닝 기법을 활용한 자동화된 지능적 분석과 더불어 OLAP도구를 활용한 경험적 분석 기능 통로를 제공함으로써 효율적이고 효과적이며 또한 입체적으로 문제의 원인을 파악할 수 있게 되었다.

4.2 데이터마이닝 기법 적용

본 절에서는 Y2R-PLUS 시스템 내 구현된 기능 중 데이터마이닝 기법을 적용한 경우를 해당 문제상황과 적용한 데이터마이닝 기법을 중심으로 설명한다. <표 3>은 이를 요약한 것이다.



<그림 6> Y2R-PLUS 시스템 분석 기능

<표 3> 데이터마이닝 기법 적용

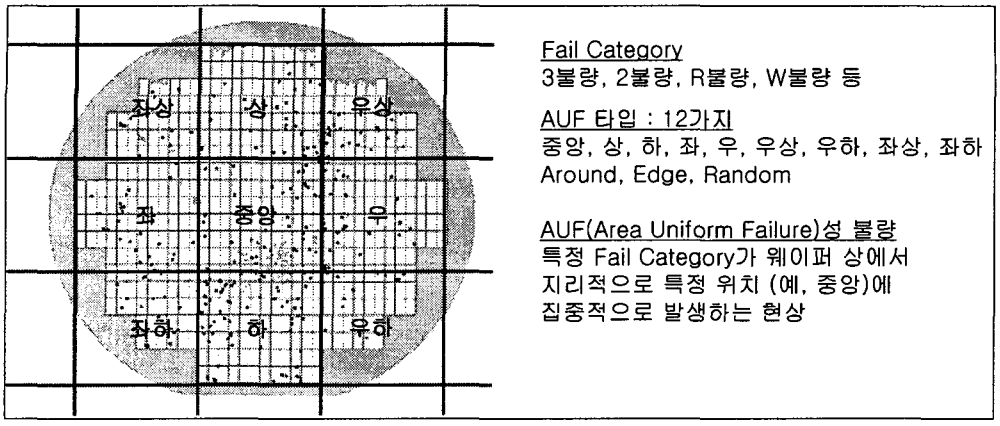
기법	적용내용	적용효과
Clustering (군집)	AUF성 불량 자동 판별 Wafer상에 나타난 특정 불량에 대해 중앙, 주변, 좌상 불량 등 12가지 패턴 발생 유무를 군집화 알고리즘을 이용해 자동 판별	· 엔지니어의 육안검사를 대체하여 AUF성 불량을 자동판별 · 시간/노력 절약 및 정확도 증가 · 동일 AUF성 불량 Wafer를 그룹핑함으로써 불량원인 분석의 기초제공
Sequential Pattern (순차패턴)	원인장비탐색 유사불량이 발생한 Wafer들이 공통적으로 진행한 장비 또는 장비들의 Sequence를 발견	· 불량의 원인이 되는 장비를 빠른 시간 내에 찾아 조치를 취할 수 있도록 함
Classification (분류)	원인 Parameter탐색 문제가 되는 Wafer Group과 문제가 없는 Wafer Group간의 차이를 발생하게 한 공정 Parameter를 Decision Tree방법을 사용하여 규명	· 불량의 원인이 되는 Parameter를 판별함으로써 수율 개선 Point도출
Prediction (예측)	수율예측 공정 진행 중에 수집한 Parameter값을 이용해 LOT의 수율을 사전에 예측함. (RBF기법 사용)	· 미래의 수율을 사전예측함으로써 Flexible 생산관리의 기초 제공

4.2.1 AUF성 불량 자동 판별

앞에서 설명한 바와 같이, FAB공정을 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 불량의 형태에 따라 불량을 몇 가지 종류로 구분한다. 예를 들어 배선계열에 문제가 있어 불량이 발생하는 경우 그 불량의 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 uniform failure)성 불량이라고 한다. <그림 7>의 좌측 부분은 웨이퍼를 9개 영역으로 구분한 것을 보여주는 것이다. AUF 타입은 그 발생 위치에 따라 중앙, 상, 하, 좌, 우, 좌상, 우상, 좌하, 우하, Edge, Around, Random 등 12가지로 구분한다. 예를 들어 3불량이 웨이퍼의 중앙 부분에 집중적으로 발생하면, 그 웨이퍼는 '3불량'-'중앙' AUF성 불량이 발생했다고 한다. <그림 7>에서 검은 점이 표시된 칩은 불량이 발생한 칩을 의미한다.

AUF성 불량이 발생되는 원인은 다양하다. 특정 공정을 담당한 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 불량



<그림 7> AUF성 불량

이 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 통해 AUF성 불량이 발생했는지 일일이 육안 검사를 하였다. 육안검사로 인해 많은 시간과 노력이 들게 되고, 엔지니어의 숙련도에 따라 검사의 질이 달라지는 문제가 있었다.

AUF 타입을 판별하는 문제는 군집화 문제에 해당한다고 볼 수 있다. 군집화 문제는 레코드의 집합이 주어졌을 때, 비슷한 속성을 가진 레코드들이 함께 그룹 지어지도록 레코드의 집합을 몇 개의 소집합으로 분할하는 분석유형이다. 웨이퍼 상에 발생한 특정 Fail category에 대해 지리적으로 가깝게 위치한 것들이 하나의 군집에 포함되도록 군집화시킨 후, 그 군집이 존재하는 위치에 따라 AUF 타입을 결정할 수 있을 것이다. 본 연구에서는 군집화를 위해 계층적 군집화 기법 중 하나인 단일결합법(single linkage method)을 약간 변형한 알고리즘을 개발하여 적용하였다. 계층적 군집화는 처음 각 대상이 독립군집으로 출발하는데(예를 들면 대상이 10개이면 군집이 10개), 거리가 가장 가까운 어느 두 대상이 군집을 이루기 시작하여 가까운 군집들끼리 계속적인 군집화가 이루어지는 방법이다(이학식 등, 2002). 계층적 군집화는 군집들간의 거리계산 방식에 따라 단일결합법, 완전결합법, 그리고 평균결합법으로 구분한다. 군집들간의 거리를 계산할 때 단일결합법은 최소거리, 완전결합법은 최장거리, 그리고 평균결합법은 평균거리를 사용한다.

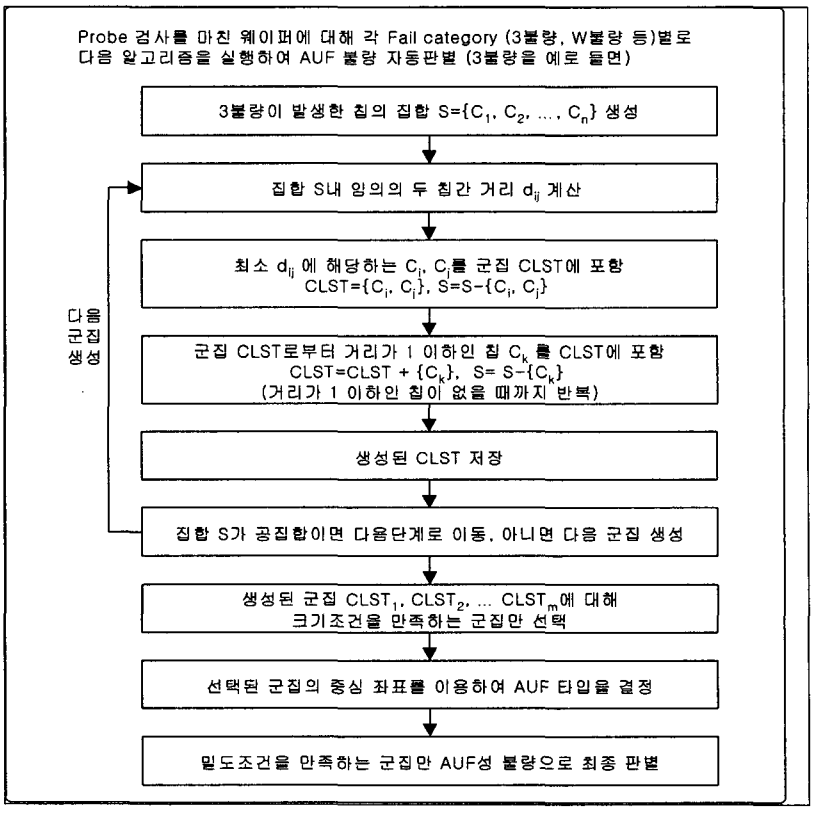
<그림 8>은 Probe 검사를 마친 웨이퍼에 대해 각 Fail Category별(3불량, W불량 등)로 AUF성 불량이 발생했는지, 발생했다면 어떤 AUF 타입(중앙, 좌상 등)이 발생했는지 자동으로 판별하는 알고리즘을 도식적으로 보여주는 것이다. 우선 Probe 검사 결과로부터 해당 웨이퍼 상에 특정 Fail Category (<그림 8>에서는 3불량)가 발생한 칩의 집합 $S=\{C1, C2, \ldots\ldots, Cn\}$를 생

성한다. 칩 C_i는 좌표 (X_i, Y_i)로 표현할 수 있는데, 여기서 (X_i, Y_i)는 칩의 웨이퍼 상의 위치를 나타내는 것으로 좌측 상단에 위치한 칩을 기준점 (0, 0)으로 하여 설정한 상대적 위치이다 (X_i, Y_i는 정수).

다음으로 집합 S내에 있는 두 칩 C_i, C_j 간의 거리 $d_{ij} = \max(|X_i - X_j|, |Y_i - Y_j|)$를 모든 가능한 쌍에 대해 계산한다. 계산 결과를 이용하여 최소의 거리를 갖는 두 칩 C_i, C_j를 군집 CLST에 포함시키고 집합 S에서는 제외시킨다. 다음 단계는 이렇게 생긴 군집 CLST로부터 거리가 1 이하인 칩들을 CLST에 포함시킨다. CLST로부터의 임의의 칩 C_k까지의 거리는 최소거리를 이용하였다. 즉, CLST에서 칩 C_k까지 거리= $\mathrm{Min}\{d_{ik} : C_i \in \mathrm{CLST}\}$이다. CLST로부터 거리가 1 이하인 칩 C_k는 CLST에 포함시키고 집합 S에서 제외시킨다. 새롭게 확장된 CLST에 대해 위 과정을 거리가 1 이하인 칩이 없을 때까지 반복한다. 이것은 군집 CLST에 인접해 있는 칩들을 군집에 포함시켜 나가는 과정이다.

군집 CLST로부터 거리가 1 이하인 칩이 더 이상 없으면 위 과정을 멈추고 생성된 군집 CLST를 저장한다. 만약 집합 S가 공집합이면 다음 단계로 넘어가고 그렇지 않으면 앞 단계로 이동하여 다음 군집을 생성한다.

위의 과정을 통해 군집 CLST_1, CLST_2, ...



<그림 8> AUF성 불량 자동판별 알고리즘

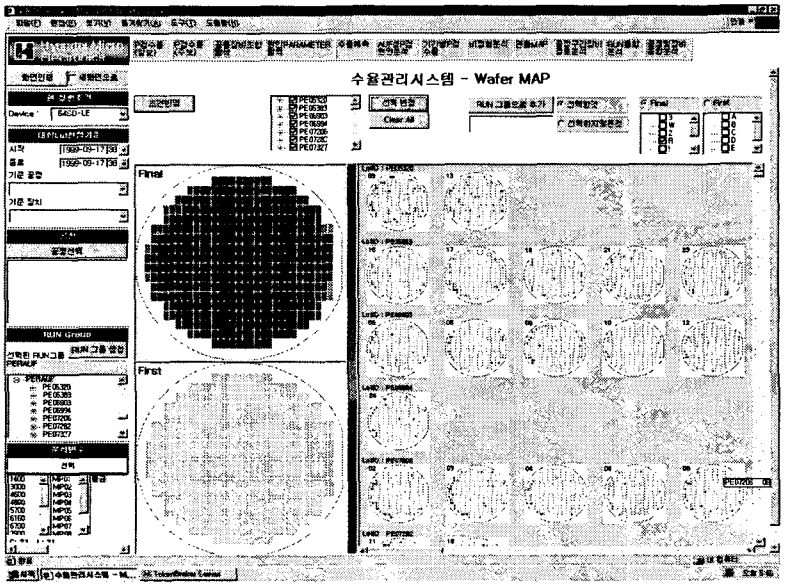
$CLST_m$ 등 m개의 군집이 생성되었다고 가정하자. m개의 군집 중 크기조건을 만족하는 군집만을 선택하는 과정이 필요하다. 왜냐하면 군집 내에 포함된 칩의 수가 일정 수 이하로 너무 작아 군집으로서의 의미가 없는 경우를 배제하고자 하는 것이다.

다음으로 크기조건을 만족하는 군집들에 대해 각 군집의 중심좌표를 구하고 그 중심좌표가 어디에 위치해 있느냐에 따라 AUF 타입을 결정한다. 군집의 중심좌표는 군집 내 속해 있는 칩들의 (x, y) 좌표의 산술평균으로 구하였다. 이 결과로서 웨이퍼 내에 복수 개의 군집이 형성될 수 있는데, 예를 들어 3불량이 '좌상'과 '중앙'에 군집을 형성하여 2개의 AUF 타입을 보일 수 있다.

마지막은 밀도조건을 만족하는 군집만을 AUF성 불량으로 최종 판별하는 단계이다. 밀도조건이란 Fail category의 전체 밀도에 비해 AUF 밀도가 일정 배수 (본 연구에서는 2배) 이상 되어야만 AUF성 불량으로 판별하는 것을 의미한다. 여기서 fail category의 전체 밀도는 웨이퍼 상에 해당 fail category가 발생한 칩의 수를 웨이퍼 상의 전체 칩 수 (net die)로 나눈 것이다 (fail category 발생 칩의 수÷net_die). 그리고 AUF 밀도는 AUF 타입이 결정된 군집 내 칩의 수를 AUF 타입 영역에 존재하는 전체 칩의 수로 나눈 것이다. 예를 들어 임의의 군집 $CLST_k$에 속한 칩의 수가 10이고 그 중심좌표가 '좌상'에 위치한 경우(좌상 영역에 해당하는 전체 칩의 수를 50이라 가정하자), 군집 $CLST_k$의 AUF 밀도는 10/50=0.2가 되는 것이다.

<그림 9>의 우측 부분은 R불량이 'Edge' AUF 타입으로 발생한 웨이퍼들을 보여 주고 있다. 웨이퍼 상에서 검게 나타난 부분이 R불량이 발생한 칩을 나타낸다.

AUF성 불량을 군집화 알고리즘을 이용해 빠른 시간 내에 정확하게 판별할 수 있게 됨에 따



<그림 9> AUF 성 불량 자동 판별

라 수율 저하의 원인을 신속하게 파악할 수 있게 되었다. 앞에서도 말했듯이 동일한 불량 형태를 보이는 웨이퍼들을 모으고, 그 웨이퍼들이 가지는 공정 상의 공통점을 찾으면 불량의 원인을 신속, 정확하게 파악할 수 있기 때문이다. 데이터마이닝 기법을 적용하기 전에는 여러 명의 엔지니어가 일일이 육안검사를 통해 AUF성 불량 여부를 판별하였는데 이는 엔지니어의 숙련도에 따라 그 정확도가 달라지는 문제뿐만 아니라 무엇보다도 시간이 많이 소요되었다. 하루에 수 만장의 웨이퍼가 생산되는 상황을 감안하면 육안검사를 통해 AUF성 불량여부를 파악하는 것은 시간적으로 한계가 있다. 반도체 공정의 특성상 FAB공정이 모두 끝나기 전에 제품의 양불량을 판별할 수 없다. 제조 공정이 세분화되어 있고 투입에서 완료까지 수개월이 소요되는 긴 공정을 가지고 있는 점을 감안하면 Probe검사 단계에서 불량이 감지되고 그 원인이 되는 장비나 공정 파라미터를 파악할 때까지 소요되는 시간을 단축하는 것이야말로 수율 개선 측면에서 너무나도 중요하다. 왜냐하면 원인을 찾고 있는 동안에도 원인 장비나 공정 파라미터에 의해 불량품은 계속 발생하기 때문이다. 불량의 원인을 파악하는 데 이전에는 1~2일 소요되던 것이 AUF성 불량 자동판별과 뒤에서 설명할 원인장비탐색과 원인파라미터 탐색 방법을 적용하게 됨에 따라 1~2시간 이내로 단축할 수 있게 되었고 이에 따라 상당한 수율개선 효과를 볼 수 있었다.

4.2.2 원인장비 탐색

Probe 검사가 끝난 웨이퍼 중, 불량 발생의 원인을 분석하기 위해 몇 개의 웨이퍼를 모아 놓은 것을 '분석그룹'이라고 하자. 분석그룹을 설정하는 주된 이유는 하나의 웨이퍼에 대해 그 불량의 원인을 찾기보다는 비슷한 형태의 불량이 발생한 웨이퍼의 집합을 대상으로 그 불량 원인을 찾는 것이 더 확실하게 원인을 파악할 수 있기 때문이다. 분석그룹은 (1) Fail category의 AUF성 자동판별을 이용하여 동일한 AUF성 불량 타입을 갖는 웨이퍼들의 집합, (2) 수율이 낮은 웨이퍼들의 집합, (3) 현업 엔지니어가 직접 지정한 웨이퍼들의 집합 등 다양한 방법에 의해 설정할 수 있다.

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

이를 위해 Y2R-PLUS 시스템에서는 Intelligent miner의 '연속패턴 마이닝 함수'를 사용하였다. 입력되는 데이터의 형태는 분석그룹 내의 각 웨이퍼가 FAB공정을 시작할 때부터 끝날 때까지 거쳐간 장비들의 리스트이다. 데이터마이닝의 결과는 분석그룹 내의 웨이퍼들이 공통적으로 거쳐간 장비나 장비들의 sequence를 발견해 주는 것이다.

<표 4>는 어떤 분석그룹에 대한 불량 원인장비 탐색의 결과를 비교그룹과 대비하여 보여 주는 것이다. 표에서 support가 의미하는 바는 분석그룹 (또는 비교그룹) 내에 포함된 웨이퍼 중 해당 장비를 거쳐간 웨이퍼의 비율을 의미한다. <표 4>의 2행으로부터 분석그룹에 있는 웨이퍼들의 90%가 1430번 L-ASHER공정의 E-9A-B 장비를 거쳐서 공정을 진행했음을 알 수 있다.

<표 4> 원인 장비 탐색 결과

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%

이 분석에서는 분석그룹 뿐만 아니라 비교그룹에 대해서도 분석을 수행한다. 비교그룹은 분석그룹과 비슷한 시기에 probe검사를 하였지만 불량이 없거나 불량이 있더라도 서로 상이한 불량타입을 보이는 웨이퍼들의 집합으로 구성된다. 비교그룹을 설정한 이유는 분석 결과의 잘못된 해석을 방지하기 위해서이다. 비교그룹을 설정하지 않았다면 <표 4>의 2행에서 "분석그룹의 90%가 1430번 L-ASHER공정의 E-9A-B를 공통적으로 거쳐갔기 때문에 분석그룹의 불량 원인은 이 장비일 가능성이 높다"라는 섣부른 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 웨이퍼들이 어쩔 수 없이 E-9A-B 장비를 진행한 경우라고 해석할 수 있을 것이다. 1415 L-DRY-ET 공정의 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 웨이퍼들로 구성된 비교그룹은 작업비율이 낮은 것으로 보아, E-2B-05B 장비가 분석그룹의 불량 원인 장비일 가능성이 높다는 결론을 내릴 수 있을 것이다.

4.2.3 원인파라미터 탐색

각 단위 공정에서는 장비와 관련된 여러 가지 공정 파라미터에 대해 관리치를 설정하여 관리치 이내가 되도록 관리하고 있지만, 잘못된 관리치 설정이 저수율의 원인이 되는 경우가 종종 있다. 웨이퍼가 사진, 증착 등 각 단위공정을 마치면 작업이 제대로 진행됐는지 검사하기 위해 여러 가지 파라미터를 계측하게 되는데, 보통 통계적 품질관리 기법을 이용하여 관리치를 벗어나는 이상발생 여부를 실시간으로 파악하고 있다. 그러나 여기서 문제가 되는 것은 해당 파라미터의 관리치 자체가 잘못 설정되어 있다면 아무리 단위공정에서 관리범위 내로 관리한다하더라도 수율에 악영향을 미친다는 점이다. 이러한 현상은 연구개발이 끝나고 생산라인으로 이관된 지 얼마 되지 않은 제품의 경우에 많이 발생한다. 왜냐하면 연구개발 단계에서의 장비와 청정도 등 조건이 생산라인에서의 조건과 다르기 때문이다.

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

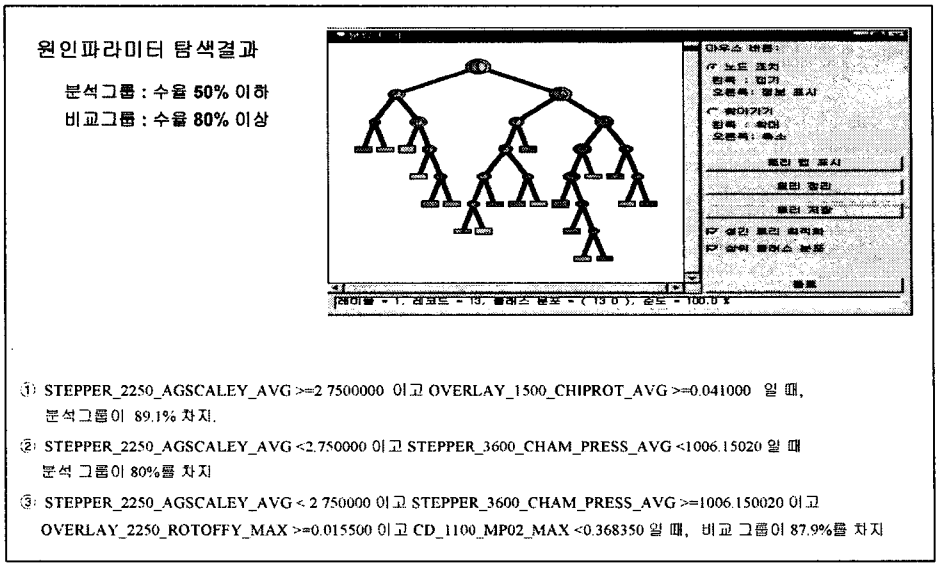
원인파라미터 탐색은 분류(classification) 문제에 해당한다. 문제가 있는 웨이퍼들의 집합인 분석그룹과 문제가 없는 비교그룹을 설정하고, 두 그룹간의 차이를 발생하게 한 공정 파라미터를 decision tree 기법을 적용하면 찾아낼 수 있다.

예를 들면 분석그룹은 수율이 50% 이하인 웨이퍼들의 집합이고 비교그룹은 수율이 90% 이상이 웨이퍼들의 집합으로 설정하여 두 그룹간 차이의 원인이 되는 공정 파라미터를 찾아 볼 수 있다.

<표 5>는 분류기법 적용을 위한 데이터의 형태를 보여 주고 있다. 입력변수는 각 웨이퍼가 단위공정을 마친 뒤 계측된 파라미터이며 목표변수는 '분석그룹' 또는 '비교그룹' 중 하나의 값을 갖는다. 여기서 입력변수로 사용하는 파라미터는 500여 개 정도 되는 모든 계측 파라미터를 이용할 수도 있지만 보통의 경우 현업 전문가들은 불량의 패턴을 보면 원인이 되는 파라미터의 종류를 추정할 수 있으므로 현업 전문가가 입력변수를 지정하여 그 수를 가급적 줄이는 것이 더 효과적이다. 원인 파라미터 탐색을 위해 intelligent miner의 '트리분류 마이닝 함수'를 이용하였다. <그림 10>은 불량 원인 파라미터 탐색의 결과를 보여 주는 화면이다.

<표 5> 원인 파라미터 분석을 위한 데이터 형태

Wafer ID	입력변수						목표변수
	CD_3716_MP01_AV G	CD_4400_MP02_Max	OVERLAY_2250_MAX_X_AV G	STEPPER_2250_AGSCALEY_AV G	.....	OVERLAY_2350_ROTOFFY_MAX	
PW4144	0.2354	0.4323	0.7863	2.7863		0.0159	분석그룹
PW4298	0.4746	0.5632	0.3243	2.1102		0.0023	비교그룹
PW4436	0.312	0.4432	0.6545	2.7921		0.1632	분석그룹



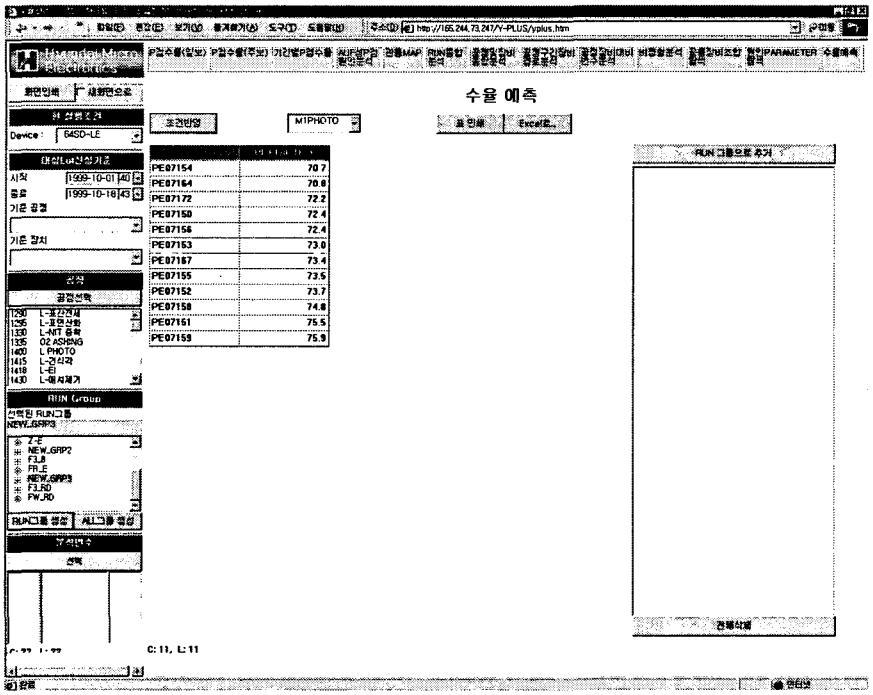
<그림 10> 원인 파라미터 탐색 결과

4.2.4 수율예측

반도체 공정의 특성 상 Probe 검사가 끝나야 웨이퍼 상의 각 칩이 양품인지 불량품인지 판정할 수 있다. 그런데 FAB공정이 2~3개월인 점을 감안해 본다면 각 웨이퍼가 모든 공정을 마쳤을 때의 대략적인 수율을 공정 진행 중에 미리 예측할 수 있다면 생산관리 측면에서 상당히 바람직한 효과를 거둘 수 있을 것이다.

수율예측을 위해 Y2R-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 웨이퍼의 향후 수율을 예측하는 모델을 만들기 위해 probe검사까지 마친 웨이퍼들의 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 너무 앞부분에 위치한 공정이라면 k 공정 이후에 너무 많은 수율 변동 요인이 존재하므로 정확한 수율예측은 어려운 일이다. 하지만 공정 k가 상대적으로 뒷부분에 위치한 공정이라면 비교적 정확한 수율예측이 가능해 진다. <그림 11>은 RBF 예측 함수를 이용해 M1 Photo 공정에 있는 로트들의 향후 수율을 예측한 결과를 보여 주는 화면이다.



<그림 11> 수율예측

5. 결론

반도체 공정에서의 수율관리는 매우 많은 공정 단계, 긴 생산 주기시간, 그리고 엄청난 수의 관련 변수가 복잡하게 얽혀 있는 매우 복잡한 문제이기는 하지만 초기 시장 선점과 원가경쟁력 확보라는 측면에서 반드시 해결해야만 하는 매우 중요한 문제이다. 국내외 반도체 회사는 수율관리를 위해 다각적인 노력을 기울이고 있는 상황이나, 기존의 통계적인 방법이나 엔지니어의 경험적 분석방법만으로는 해결하지 못하는 복잡한 수율 저하 원인들이 존재한다. 특히, 수많은 공정 변수와 엄청난 양의 데이터로 인해 모두 변수를 분석하기에는 너무 많은 시간이 소요되어 실제적으로 일부 변수만 분석이 되고 있다. 생산현장에서 생성되는 방대한 기술 데이터를 효과적으로 활용하여 반도체 수율 향상에 도움을 줄 수 있는 지능적이고 효과적인 지원시스템이 필요한 상황이다.

본 논문은 반도체 FAB공정을 대상으로 기존 수율관리시스템의 문제점을 개선하기 위해 데이터마이닝 기법을 활용한 수율관리 방법을 제안하고 이를 국내 모 반도체 회사에서 정보시스템으로 구축한 Y2R-PLUS (Yield Rapid Ramp-up, Prediction, anaLysis & Up Support) 시스템을 소개하였다. 데이터마이닝 적용 기능으로서는 우선, 군집화 기법을 이용하여 FAB공정을 마친 웨이퍼에 불량 칩(chip)이 지리적으로 특정 위치에 집중적으로 발생하는 현상 (AUF성 불량)을 기존의 육안 검사에서 데이터로부터 자동 인식할 수 있도록 구현하였다. 또한 연속패턴분석, 분류분석, RBF(Radial Base Function) 등의 기법을 적용하여 수율 저하의 원인이 되는 문제 장비나 문제 파라미터를 신속, 정확하게 파악할 수 있도록 해 줄 뿐만 아니라 공정 진행 중인 제품의 미래 수율을 예측할 수 있도록 지원하는 방법을 제안하였다. Y2R-PLUS 시스템은 현업에 적용하여 매우 큰 효과를 보았으며 계속해서 그 적용범위를 확대해 나가고 있는 상황이다.

참고문헌

강현철, 한상태, 최종후, 김은석, 김미경, SAS Enterprise Miner 4.0을 이용한 데이터마이닝, 자유 아카데미, 2001

백준걸, 김강호, 김성식, 김창욱, "실시간 기계상태 데이터베이스에서 데이터마이닝을 위한 적응형 의사결정트리 알고리듬," *대한산업공학회지*, Vol. 26, No. 2, pp171-182, 2000.

안진석, 고용민, 장중순, "데이터마이닝을 이용한 최적 공정조건 탐색," *대한설비관리학회지*, Vol. 4, No. 2, pp129-144, 1999.

이학식, 김영, SPSS 10.0 매뉴얼, 법문사, 2002

장남식, 홍성완, 장재호, 데이터마이닝, 대청, 1999

지원철, 김민용, "데이터마이닝과 의사결정 지원 시스템," *정보과학회지*, 16권(9호), pp. 24-36, 1998.

한국과학기술원 산업공학과, 반도체 공정의 생산 및 공정관리 시스템 개발 최종보고서, 과학기술처, 1996

Baek, D. H., Yoon, W. C. and Park, S. C., "A spatial rule adaptation procedure for reliable production control in a wafer fabrication system," *International journal of production research*, Vol. 36, No. 6, pp1475-1491, 1998.

Braha, D., and Shmilovici, A., "Data mining for improving a cleansing process in the semiconductor industry," *IEEE transactions on semiconductor manufacturing*, Vol. 15, No. 1, pp91-101, 2002.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P., "From data mining to knowledge discovery

in databases," Advances in knowledge discovery and data mining, AAAI Press/MIT Press, pp. 1-34, 1996.

Frawley, W. J., Piatetsky-Shapiro, G., Matheus, C. J., "Knowledge discovery in databases: An overview, Knowledge discovery in databases," AAAI Press/MIT Press, pp. 1-27, 1991.

Goebel, M. and Gruenwald, L., "A survey of data mining and knowledge discovery software tools," *1999 ACM SIGKDD*, Vol. 1, No. 1, pp. 20-33, 1999.

Hancock, W. M, Yoon, J. W., and Plont, R., "Use of ridge regression in the improved control of casting process," Quality engineering, Vol. 8, No. 3, pp. 395-403, 1996.

IBM, Inteligent miner for data 사용자 매뉴얼 버전 6 릴리스 1, IBM, 1999.

Kusiak, A. and Kurasek, C., "Data mining of printed-circuit board defects," *IEEE transactions on robotics and automation*, Vol. 17, No. 2, pp. 191-196, 2001.

Lian, J., Lai, X. M., Lin, Z. Q., and Yao, F. S., "Application of data mining and process knowledge discovering in sheet metal assembly dimensional variation diagnosis", *Journal of materials processing technology*, Vol. 129, pp315-320, 2002.

Yamashita, Y., "Supervised learning for the analysis of process operational data", *Computer and chemical engineering*, Vol. 24, N0. 2, pp471-474, 2000.

Abstract

Application of Data mining for improving and predicting yield in wafer fabrication system

Dong-Hyun Baek* · Chang-Hee Han**

This paper presents a comprehensive and successful application of data mining methodologies to improve and predict wafer yield in a semiconductor wafer fabrication system. As the wafer fabrication process is getting more complex and the volume of technological data gathered continues to be vast, it is difficult to analyze the cause of yield deterioration effectively by means of statistical or heuristic approaches. To begin with, this paper applies a clustering method to automatically identify AUF (Area Uniform Failure) phenomenon from data instead of naked eye that bad chips occurs in a specific area of wafer. Next, sequential pattern analysis and classification methods are applied to find out machines and parameters that are cause of low yield, respectively. Furthermore, radial bases function method is used to predict yield of wafers that are in process. Finally, this paper demonstrates an information system, Y2R-PLUS (Yield Rapid Ramp-up, Prediction, anaLysis & Up Support), that is developed in order to analyze and predict wafer yield in a korea semiconductor manufacturer.

Key words: 데이터마이닝(data mining), 반도체 FAB공정(wafer fabrication), 수율(yield), Y2R-PLUS

* Division of Business & Commerce, Cheonan University
** Department of Business Administration, Hanyang University