

CCA를 통한 반도체 공정 변인들의 상관성 분석 : 웨이퍼검사공정의 전압과 불량결점수와의 관계를 중심으로

김승민 · 백준걸[†]

고려대학교 산업경영공학과

Correlation Analysis on Semiconductor Process Variables Using CCA(Canonical Correlation Analysis) : Focusing on the Relationship between the Voltage Variables and Fail Bit Counts through the Wafer Process

Seung Min Kim · Jun-Geol Baek

School of Industrial Management Engineering, Korea University

Semiconductor manufacturing industry is a high density integration industry because it generates a vast number of data that takes about 300~400 processes that is supervised by numerous production parameters. It is asked of engineers to understand the correlation between different stages of the manufacturing process which is crucial in reducing production costs. With complex manufacturing processes, and defect processing time being the main cause. In the past, it was possible to grasp the correlation among manufacturing process stages through the engineer's domain knowledge. However, It is impossible to understand the correlation among manufacturing processes nowadays due to high density integration in current semiconductor manufacturing. In this paper we propose a canonical correlation analysis (CCA) using both wafer test voltage variables and fail bit counts variables. using the method we suggested, we can increase the semiconductor yield which is the result of the package test.

Keywords: Fail bit, Voltage, Wafer test process, Canonical Correlation Analysis, CCA

1. 서론

현대사회는 첨단 기술의 발전으로 인하여 스마트폰과 같은 전자 제품이 대중화 되어 있다. 전자 제품의 핵심 부품 중 하나인 반도체를 생산하는 반도체 산업은 전자 제품 사용이 대중화로 인하여 점점 더 많은 이윤을 창출하고 있다. 높은 이윤 창출을 가진 반도체 산업은 경쟁이 가장 치열한 산업 중 하나이다. 치열한 경쟁 속에서 우위를 선점하기 위해 반도체 제조업체들은 생산성 향상을 이용한 경쟁력 확보를 위한 노력을 기울이고

있다(Pieter, 2000). 또한 설계와 공정의 첨단 기술 개발을 이용하여 더욱더 고집적화된 첨단 제품의 품질 향상에 노력하고 있다(Kim *et al.*, 1998). 노력의 일환으로 제조 공정 단계 간의 상관관계를 파악하는 방법이 있다. 공정 간의 상관관계 파악을 통해 불량을 조기에 예측할 수 있고 따라서 시간과 비용을 절감할 수 있다. 공정 간의 상관관계를 이용한 기존 연구로서 공정 조건 간의 상관관계를 파악하여 최적 공정 조건을 찾는 연구가 진행되었고(Byun *et al.*, 1998), 반도체의 CMP(Chemical-Mechanical Polishing) 공정에서 사용된 변수와 웨이퍼 흡집 받

본 연구는 BK21 플러스 사업(고려대학교, 제조·물류분야에서의 빅 데이터 응용 사업팀)으로 지원된 연구임

이 연구는 2014년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2013R1A1A2010019).

[†] 연락저자 : 백준걸 교수, 02841 서울특별시 성북구 안암로 145 고려대학교 산업경영공학과, Tel : 02-3290-3396, Fax : 02-929-5888,

E-mail : jungeol@korea.ac.kr

2015년 2월 9일 접수; 2015년 6월 1일 수정본 접수; 2015년 8월 21일 게재 확정.

생과의 상관관계를 연구하였다(Jung *et al.*, 2013). 과거에는 엔지니어의 경험을 바탕으로 공정 단계 간의 대략적인 상관관계를 파악하였으나 현재는 반도체의 고집적화로 인하여 경험적 상관관계 파악도 어려운 상황이다. 따라서 본 연구는 웨이퍼 검사 공정에서 단위 공정 간의 상관성을 분석하고 최종 검사 공정을 예측하여 확인한다. 상관성을 확인하는 방안으로는 변수 군과 변수군 사이의 상관관계를 확인하는 정준 상관 분석을 이용하였다. 본 논문에서 제안된 방법으로 불량률을 조기에 탐지할 수 있고 이를 바탕으로 불량품에 대한 전처리 과정을 추가하여 수율을 향상시킬 수 있다.

반도체 제조 공정은 작게는 300~400여 개의 단위 공정으로 구성되며, 크게는 투입(Input)과 출고(Output)를 제외하고 FAB (Wafer Fabrication) 공정, 웨이퍼(Wafer Test) 검사 공정, 조립(Assembly) 공정, 최종 검사(Final Test) 공정 단계로 4단계의 공정으로 구분할 수 있다(Uzsoy *et al.*, 1992). <Figure 1>은 반도체 제조 공정을 나타낸다. FAB 공정은 고온에서 녹인 실리콘으로 만든 실리콘 기둥인 잉곳(Ingot)을 잘라 웨이퍼(Wafer)를 생성하고 패턴(Pattern) 형성 공정, 식각(Etching) 공정 등의 세부 공정을 통해 웨이퍼 표면에 여러 종류의 막을 겹쳐 하나의 웨이퍼에 수백 개의 칩(Chip) 회로를 생성하는 공정이다. 웨이퍼 검사 공정은 웨이퍼 상태의 칩들에 대해 다양한 테스트를 수행하여 웨이퍼가 전기적 특성에 따른 양품과 불량품을 판별하는 공정이다(Hsu *et al.*, 2007). 조립 공정은 양품인 웨이퍼를 칩 단위로 분리하고 칩의 전기적, 물리적 특성을 향상시킨다. 외부의 물리적, 화학적 충격으로부터 칩을 보호하기 위하여 막을 형성화하는 공정이다(Kim *et al.*, 2014). 마지막으로 패키지 테스트 공정은 조립 공정을 통과한 칩을 실제 사용되는 환경보다 더 가혹한 환경에서 검사하여 칩이 정상 작동하는지 판별하는 단계이다(Kim *et al.*, 2014).

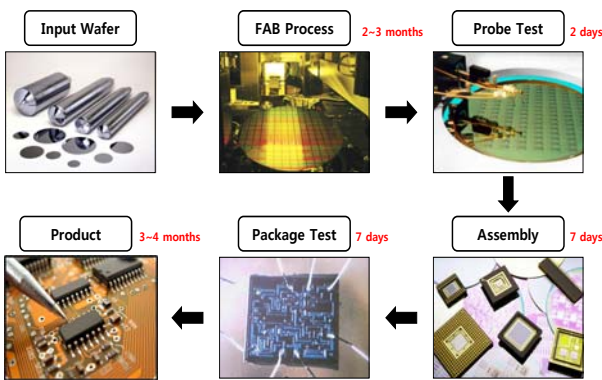


Figure 1. Semiconductor manufacturing process

반도체 업계에서는 생산 효율성을 극대화하기 위해 양품이 될 칩과 불량품이 될 칩들을 미리 구분하는 칩 품질 예측을 통해 수율 향상과 제조 시간 및 비용 단축을 추구하고 있다. 수율을 예측하는 기존 연구 방법들로 웨이퍼 검사 결과를 예측하기 위해서 FAB 공정 데이터(Chien *et al.*, 2007)와 PCM(Process Control Monitoring) 데이터를 활용(Ludwig, 2000)하였다. 패키

지 테스트의 결과를 예측하기 위해서 웨이퍼 검사 데이터(An *et al.*, 2009)를 활용하거나 웨이퍼 빈 맵(Wafer Bin Map)을 활용(Hsu *et al.*, 2007) 하는 등 다양한 연구가 진행되었다. 패키지 테스트 결과를 예측하는 연구에서는 칩에 대한 결함 유형을 설명하기 위한 코드로 빈(BIN) 정보(Quirk *et al.*, 2001), PCM 테스트 정보, 불량결점 수(Fail Bit Count), 웨이퍼의 칩 좌표 정보 등을 주로 이용하였다. 특히, 웨이퍼 검사의 불량결점수는 웨이퍼 검사의 항목별로 측정되는 결점 수를 의미하고 제품의 비정상적인 기능까지 고려된 상황에서 발생하는 정보를 포함하므로 칩의 품질 예측에 주로 사용된다(Kim *et al.*, 2014). 따라서 본 연구에서는 웨이퍼 검사 공정에서 수행되는 전압 테스트의 결과 변수군과 웨이퍼 검사 공정의 결과물인 불량결점수 변수군을 이용하여 상관관계를 밝히고자 한다.

반도체 제조 공정을 평가하는 가장 중요한 척도인 반도체 수율은 반도체 공정에서 투입된 원자재 대비 양품으로 생산된 칩의 비율을 나타낸다. 수율은 반도체 시장에서 시장 확보에 필수 조건일 뿐만 아니라, 값 비싼 생산 비용으로 생산 필요성을 결정짓는 중요한 척도이다. 반도체 제조 공정 단계에 따라 FAB 수율, 웨이퍼 검사 수율, 조립 수율, 패키지 테스트 수율로 총 4가지로 구분 가능하다(Baek *et al.*, 2002). <Figure 2>은 반도체 수율 계산을 설명하는 것이다. FAB 수율은 투입된 웨이퍼 장 수(Wafer in) 대비 불량이 발생하지 않고 양품으로 나온 웨이퍼 장 수(Wafer Out)의 비율이다(Baek *et al.*, 2002). 웨이퍼 검사 수율은 총 다이(Die) 수(Wafer Out×Dies/wafer) 대비 웨이퍼 검사 결과 양품인 다이의 비율이다. 조립 수율은 웨이퍼 테스트를 통과한 다이 대비 조립 공정을 통과한 칩의 비율이다. 패키지 테스트 수율은 최종적인 수율을 의미하며 조립 공정을 통과한 칩 중 패키지 테스트를 통과한 양품 칩의 비율이다(An *et al.*, 2009). 위의 4가지 수율의 곱이 CUM 수율(Cumulation Yield)이다(Baek *et al.*, 2002). 따라서 본 연구에서는 제조 공정 간의 파악된 상관관계를 검증하기 위한 방안으로 최종 패키지 테스트 결과를 불량결점 수와 전압을 이용한 예측 모델로 비교하여 공정 간의 상관성을 분석한다.

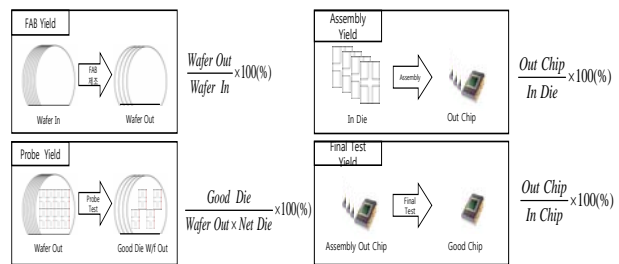


Figure 2. Semiconductor yield

2. CCA를 이용한 상관 분석

2.1 CCA(Canonical Correlation Analysis)

다변량 변수들 간의 다중 상관관계를 분석할 수 있는 모델로

서 PLS(Partial Least Square) 회귀분석, PCA(Principal Component Analysis) 등 여러 가지 방법이 있다. PLS 회귀분석은 다변량 데이터를 주성분 분석하여 종속변수와 가장 상관관계가 높은 주성분을 이용한 통계적 모델 방법으로 독립변수 간에 강한 다중공산성이 존재하고 잡음이 있을 때도 사용가능하다는 장점이 있다. 하지만 설명변수의 수가 표본의 크기보다 작으며 변수들 간의 상관관계가 높을 때 사용 가능하다. PCA는 상관이 있는 변량들의 변동을 줄이는 차원에서 상관이 없는 변량의 집합으로 기준 축을 변환하여 특징 벡터를 주축 방향으로 사영(Projection)하는 방법을 의미한다. 하지만 변수들 간의 상관관계가 존재해야 되고 정보 손실이 크다는 단점을 가지고 있다. 따라서 본 연구에서는 변수들 간의 선형조합을 이용하여 선형상관관계를 파악하는 정준 상관분석을 이용하였다.

정준 상관분석(Canonical Correlation Analysis)은 변수들로 구성된 군집 간 선형 상관관계를 파악하는 분석 방법이다. 일반적으로 두 변수군이 이용하여 각 변수군의 선형결합 간의 상관계수에 관심을 둔다(Hotelling, 1936). 변수군 간의 상관성을 가장 크게 하는 첫 번째 선형결합(Pair of combinations)을 찾고 첫 번째 선형결합을 제외한 나머지 짝들 가운데 가장 큰 상관성을 가지는 두 번째 선형결합 짝을 찾는다. 상관성이 큰 선형결합을 선택하는 과정은 선형결합 짝의 수가 한 변수군의 수가 적은 변수 수와 일치될 때까지 계속된다. 여기서 선형결합의 짝을 정준 변수(Canonical Variables)라 하고 이 변수의 상관을 정준 상관(Canonical Correlations)이라 한다. 따라서 정준 상관은 두 변수군 사이의 상관성의 정도를 나타낸다(Seok et al., 2012).

식 (1) 과 식 (2)로 정의된 다수의 변수들로 구성된 두 변수군 벡터 X와 Y에 대한 각각의 선형 조합을 식 (3) 과 식 (4)와 같이 정의하는데, 이는 X와 Y를 W_x 와 W_y 방향으로 투영한 새로운 축을 정의하게 된다.

$$X = \{x_1, x_2, \dots, x_m\} \quad (1)$$

$$Y = \{y_1, y_2, \dots, y_n\} \quad (2)$$

$$S_{x, W_x} = \{(W_x, x_1), \dots, (W_x, x_m)\} \quad (3)$$

$$S_{y, W_y} = \{(W_y, y_1), \dots, (W_y, y_n)\} \quad (4)$$

식 (5)와 같이 선형 조합 S_x 와 S_y 의 상관성이 최대가 되게 하는 W_x 와 W_y 를 구하고 이를 제 1 정준 변수(First canonical variate)라 한다. 그리고 이 때 ρ 를 제1 정준 상관 계수(First canonical correlation)라 한다. ρ 의 계산 방법인 식 (5)의 상관계수가 최대가 되기 위해서는 조건으로 식 (6)과 식 (7)을 만족해야 한다. 또한 식 (8)을 최댓값을 구함으로써 식 (5)를 최대화할 수 있다(Seok et al., 2012).

$$\rho = \max(S_{x, W_x}, S_{y, W_y}) \\ = \max \frac{W_x' Cov(X, Y) W_y}{\sqrt{W_x' Var(X) W_x W_y' Var(Y) W_y}} \quad (5)$$

$$W_x' Var(X) W_x = 1 \quad (6)$$

$$W_y' Var(Y) W_y = 1 \quad (7)$$

$$\max W_x' Cov(X, Y) W_y \quad (8)$$

W_x 와 W_y 에 대해 최대가 되는 ρ 를 찾기 위해 식 (9)와 같이 Lagrange Multiplier를 적용한다.

$$L(\lambda, W_x, W_y) = W_x' Cov(X, Y) W_y \\ - \frac{\lambda_x}{2} (W_x' Var(X) W_x - 1) \\ - \frac{\lambda_y}{2} (W_y' Var(Y) W_y - 1) \quad (9)$$

Lagrange Multiplier를 적용하여 얻은 식 (9)을 W_x 와 W_y 에 대해 각각 미분하여 식 (10) 과 식 (11)을 얻을 수 있고 식 (10)과 식 (11)에 각각 W_x 와 W_y 곱하여 식 (12)와 식 (13)을 얻은 뒤 식 (13)에서 식 (12)와의 차를 이용하여 최종적으로 식 (14)을 얻을 수 있다(Seok et al., 2012).

$$\frac{dL}{dW_x} = Cov(X, Y) W_y - \lambda_x Var(X) W_x = 0 \quad (10)$$

$$\frac{dL}{dW_y} = Cov(Y, X) W_x - \lambda_y Var(Y) W_y = 0 \quad (11)$$

$$W_x' Cov(X, Y) W_y = \lambda_x W_x' Var(X) W_x \quad (12)$$

$$W_x' Cov(X, Y) W_y = \lambda_y W_y' Var(Y) W_y \quad (13)$$

$$\lambda_y W_y' Var(Y) W_y - \lambda_x W_x' Var(X) W_x = 0 \quad (14)$$

식 (14)에서 도출된 $\lambda_x - \lambda_y = 0$ 즉 $\lambda = \lambda_x = \lambda_y$ 를 이용하여, 식 (11)의 W_y 를 계산할 수 있다.

$$W_y = \frac{Var(Y)^{-1} Cov(Y, X) W_x}{\lambda} \quad (15)$$

식 (15)에서 구해진 W_y 를 식 (10)에 대입하면 식 (16)를 구할 수 있다. 식 (16)의 λ 는 일반화된 고윳값(Eigen Value)을 의미하며 λ 를 이용하여 상관계수의 최대치인 정준상관계수를 계산할 수 있다(Lee, 2002).

$$Cov(X, Y) Var(Y)^{-1} Cov(Y, X) W_x \\ = \lambda^2 Var(X) W_x \quad (16)$$

2.2 유의성 검정통계량

정준 상관계수나 고유 값은 모두 도출된 정준 상관 함수의 중요성을 알려주는 통계량이지만 이 값에 대해 일관성 있게 판단할 수 있는 기준은 아직까지 마련되어 있지 않다. 따라서 상관관계 여부를 파악하기 위해 일반적으로 유의성 검정 통계량을 이용하여 두 변수군의 상관관계를 검정한다. 유의성 검정통계량은 두 변수군의 변수들이 다변량 정규분포를 이룰 때, “모집단에서의 정준 상관계수의 값이 0이다.”라는 귀무가설을 검정할 수 있다. 이러한 정준 상관계수에 대한 가설을 검

정할 수 있는 다변량 통계량에는 네 가지가 있는데 도출된 정준 상관 함수의 고유 값을 이용한다. <Table 1>은 다변량 검정 통계량의 의미이다. 식에서 λ_i 는 i 번째 도출된 고유치를 의미한다. λ_{max} 는 가장 큰 고유치인 첫 번째 고유치 λ_1 이고 m 은 연구자가 정한 정준 상관함수의 수를 나타낸다(Lee, 2002). 첫 번째로 윌크스의 람다(Wilk's Lambda)에서 $1/(1+\lambda_i)$ 은 i 번째 정준 상관 함수로서 설명되지 않는 분산의 비율을 나타내 주기 때문에 윌크스의 람다 Λ_i 는 각 정준변량으로 설명되지 않는 분산의 누적이다. 두 번째로 Pillai's trace에서 $\lambda_i/(1+\lambda_i)$ 는 i 번째 정준변량에 의해 설명되는 분산의 비율이다. 따라서 이 값들의 합인 Pillai trace V_i 는 정준상관함수들로서 설명할 수 있는 분산의 비율을 의미하는 통계량이다. 세 번째 Hotelling-Lawley trace 식에서 λ_i 는 설명되는 분산을 설명되지 않는 분산으로 나눈 값과 같다. Hotelling-Lawley trace T_i 는 각 정준 상관 함수의 설명되는 분산의 설명되지 않는 분산에 대한 비율의 합을 의미하는 통계량이다. 마지막으로 최대 특성근 R(Roy's greatest root)은 가장 설명력이 큰 첫 번째 정준 상관 함수로서 설명될 수 있는 분산의 비율을 나타낸다. 검정 통계량의 검정 방법으로는 유의수준(p-value)과 유의 확률을 비교하여 검정할 수 있다. 유의 확률이 설정된 유의 수준보다 작을 경우 귀무가설을 기각할 수 있다(Lee, 2002). 본 연구에서는 통계량 평균(μ) 차이를 가장 잘 나타내는 윌크스의 람다를 사용하여 검정을 실시한다.

Table 1. Significant test statistic

Wilks' Lambda	$A_i = \prod_{i=1}^m \frac{1}{1+\lambda_i}$
Pillai's trace	$V_i = \sum_{i=1}^m \frac{\lambda_i}{1+\lambda_i}$
Hotelling-Lawley trace	$T_i = \sum_{i=1}^m \lambda_i$
Roy's greatest root	$R = \frac{\lambda_{max}}{1+\lambda_{max}}$

3. 실험

3.1 실험 데이터

본 연구의 실험은 국내 반도체 기업의 산업 현장에서 발생하는 실무 데이터를 사용하였다. 웨이퍼 검사 결과 데이터는 불량률점수 변수 29개(HF_1~24, CF_1~5)와 웨이퍼 검사에서 실행한 전압 관련 변수 13개(HV_1~10, CV_1~3)를 가지고 있다. 전압 관련 변수는 웨이퍼 검사 공정에서 프로브 카드(Probe Card)에 의해 측정된 값이며, 전압 관련 변수의 값이 관리한계선 값보다 높거나 낮을 경우 불량률점수의 값이 증가한다. 그러므로 이는 반도체 품질과 직결된다. 본 연구는 위의 변수군을 이용하여 정준 상관 분석을 실시하였다.

3.2 정준 상관 함수 선택

불량률점수로 구성된 종속 변수군과 전압 변수로 구성된 설명 변수군에 대해 정준 상관 분석을 실시하였고 최소 변수군과 동일한 수인 13개의 정준 상관 함수가 생성되었다. 정준 상관 분석의 특성상 최대 상관성을 갖는 값에서부터 제 1정준 상관 함수를 생성하고 다음으로 높은 상관성을 가지는 값으로부터 상관 함수를 생성한다. 따라서 제 1정준 상관 함수에서 제 2정준 상관 함수로 갈수록 상관성이 낮아지는 특성을 보인다. <Figure 3>은 정준 상관 함수와 상관계수의 그래프이며, 제 3정준 상관 함수를 넘어가면서 계수의 차이가 크게 감소하는 상태를 보인다. 타 연구에서는 상관성이 제일 큰 제 1정준 상관 함수만을 살펴보지만 본 연구에서는 변수 간의 다양한 해석을 위해 상관성이 가장 큰 3개의 함수를 이용하여 실험을 실시하였다. <Figure 4>는 제 1정준 상관 함수의 전압 변수와 불량률점수 변수의 정준 부하량을 나타낸 그림으로서, 약 0.825의 상관성을 보이고 있다. 정준 부하량의 상관성을 통하여 각 변수의 선형조합이 상관성을 있음을 확인할 수 있다.

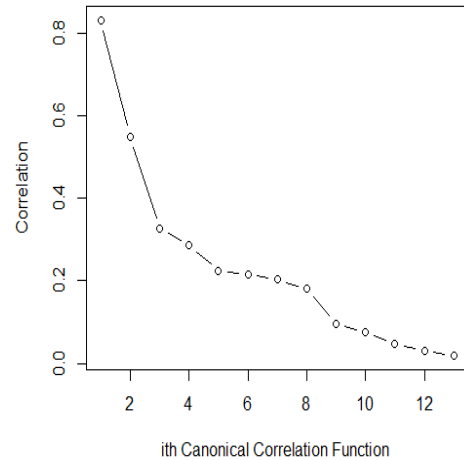


Figure 3. Correlation of canonical correlation functions

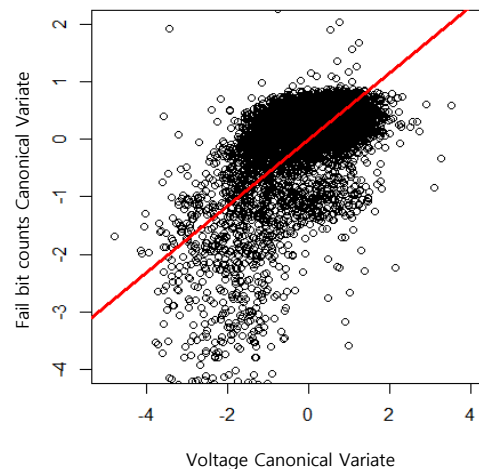


Figure 4. Canonical variate of first canonical correlation function

3.3 정준상관 분석 결과

본 연구의 목적은 공정 간의 상관관계를 파악하는 것이다. 즉, 웨이퍼 검사 공정의 불량결점수에 영향을 미치는 전압 변수들을 확인하는 것이다. 정준 상관계수를 통하여 두 변수군 간의 관계 정도를 알 수 있기 때문에 불량 결점수에 미치는 영향을 파악하기 위해서는 정준 상관계수를 확인해야 한다(Lee, 2002). <Table 2>는 웨이퍼 검사의 결과인 불량결점수 변수와 웨이퍼 검사의 전압 변수와의 정준변형계수이다. 정준 상관 분

석의 결과로 나온 정준변형계수에 대하여 살펴보면, 각 변수들의 측정 척도가 다를 뿐만 아니라, 각 변수의 평균과 표준편차가 서로 다르기 때문에 표준화된 정준변형계수를 해석해야 한다(Lee, 2002). <Table 3>은 표준화된 정준 변형계수를 나타내는 표이며 표준화된 정준변형계수표의 정준변형계수의 절댓값이 1 이상인 값에 대해 변수간의 상관관계를 파악하였다.

<Table 4>는 변수간의 상관관계를 나타낸 표로서 같은 행에 있는 변수 간에는 양의 상관관계를 가지며 다른 행과는 음의

Table 2. Canonical coefficients

DC voltage variable	Canonical coefficient		
	First CCF	Second CCF	Third CCF
HV_1	3.862220	23.725271	-176.1019206
HV_2	3.5345228	-18.626461	68.8126594
HV_3	-3.5301373	-6.403897	2.9403665
HV_4	-4.2767865	-1.829754	-1.624945
HV_5	-9.9610514	-8.06303	-18.4906575
HV_6	-0.260741	0.670621	-0.1165229
HV_7	4.9500092	-124.173452	107.7539
HV_8	-0.9243469	-1.665309	-63.9567095
HV_9	-7.0356879	112.948638	118.6534829
HV_10	-3.6504063	20.85367	-27.8886183
CV_1	-6.2055626	-3.834513	-5.0854194
CV_2	4.7513816	1.994459	0.7302153
CV_3	10.8226894	5.506297	-1.1275807

Fail bit count	Canonical coefficient		
	First CCF	Second CCF	Third CCF
HF_1	1.29E-05	1.07E-05	1.60E-04
HF_2	-8.83E-03	1.54E-02	-5.39E-03
HF_3	2.90E-04	-9.83E-03	-3.29E-03
HF_4	7.53E-05	-1.86E-03	-7.12E-04
HF_5	-1.27E-03	8.52E-03	1.83E-02
HF_6	1.83E-03	4.09E-05	-2.07E-03
HF_7	1.92E-08	1.22E-07	3.26E-07
HF_8	-9.55E-08	-2.06E-07	-2.66E-07
HF_9	1.99E-07	-7.32E-07	2.38E-07
HF_10	-2.97E-07	7.59E-07	6.41E-07
HF_11	2.17E-07	1.68E-07	-1.20E-06
HF_12	7.58E-07	-1.69E-06	5.32E-06
HF_13	-1.00E-06	1.56E-06	-7.69E-06
HF_14	3.29E-07	2.25E-07	3.50E-06
HF_15	-5.99E-07	1.19E-07	1.28E-06
HF_16	5.33E-07	-4.81E-07	-2.50E-06
HF_17	-3.98E-07	6.27E-07	1.75E-06
HF_18	1.52E-06	3.62E-08	-4.86E-06
HF_19	-1.36E-06	-6.76E-07	3.51E-06
HF_20	9.66E-07	-8.57E-07	1.36E-06
HF_21	6.34E-07	2.73E-06	8.78E-06
HF_22	-1.57E-06	-1.81E-06	-1.01E-05
HF_23	7.13E-07	-1.06E-06	-5.35E-06
HF_24	-7.46E-07	1.12E-06	5.31E-06
CF_1	3.24E-04	-6.47E-03	2.67E-03
CF_2	3.33E-04	-5.64E-03	3.52E-03
CF_3	-2.47E-04	-7.78E-03	-5.57E-03
CF_4	-3.66E-07	-4.37E-08	1.66E-07
CF_5	5.58E-07	1.03E-07	-1.45E-07

Table 3. Standardized canonical coefficients

DC voltage variable	Canonical coefficient		
	First CCF	Second CCF	Third CCF
HV_1	0.534	3.279	-24.342
HV_2	12.080	-6.367	23.512
HV_3	-0.474	-0.859	0.394
HV_4	-0.387	-0.166	-0.147
HV_5	-0.215	-0.174	-0.400
HV_6	0.049	0.126	-0.022
HV_7	0.866	-21.727	18.854
HV_8	-0.395	-0.712	-27.345
HV_9	-1.140	18.316	19.241
HV_10	-1.392	7.949	-0.632
CV_1	-0.245	-0.151	-0.201
CV_2	0.442	0.186	0.068
CV_3	0.392	0.189	-0.039

Fail bit count	Canonical coefficient		
	First CCF	Second CCF	Third CCF
HF_1	0.004	0.003	0.050
HF_2	-0.161	0.282	-0.098
HF_3	0.002	-0.063	-0.021
HF_4	0.007	-0.184	-0.071
HF_5	-0.024	0.160	0.344
HF_6	0.008	0.000	-0.008
HF_7	0.024	0.154	0.412
HF_8	-0.123	-0.265	0.342
HF_9	0.257	-0.944	0.307
HF_10	-0.386	0.987	0.833
HF_11	0.286	0.222	-1.590
HF_12	1.000	-2.230	7.020
HF_13	-1.320	2.060	-10.200
HF_14	0.446	0.306	4.750
HF_15	-0.814	0.162	1.750
HF_16	0.728	-0.657	-3.420
HF_17	-0.555	0.874	2.440
HF_18	2.140	0.051	-6.870
HF_19	-1.940	-0.097	5.000
HF_20	1.420	-1.260	1.990
HF_21	0.930	4.000	12.900
HF_22	-2.310	-2.660	-14.800
HF_23	1.070	-1.600	-8.050
HF_24	-1.120	1.680	7.990
CF_1	0.032	-0.641	0.265
CF_2	0.002	-0.031	0.020
CF_3	-0.005	-1.430	-0.102
CF_4	-1.750	-0.209	-0.794
CF_5	2.700	0.497	-0.703

상관관계를 나타낸다. 예를 들면 <Table 4>에서 전압 변수 HV_2는 불량결점수 변수 HF_18/20/23/CF_5와 양의 상관관계를 가지고 HF_13/19/22/24/CF_4와 음의 상관관계를 갖는다고 해석할 수 있다.

Table 4. Correlation between variables

First Canonical Correlation		
Variable	DC Voltage	Fail bit count
Correlation	HV_2	HF_18/20/23, CF_5
	HV_9/10	HF_13/19/22/24, CF_4
Second Canonical Correlation		
Variable	DC Voltage	Fail bit count
Correlation	HV_1/9/10	HF_13/21/24
	HV_2/7	HF_12/20/22/23
Third Canonical Correlation		
Variable	DC Voltage	Fail bit count
Correlation	HV_2/7/9	HF_12/14/15/17/19/20/21/24
	HV_1/8	HF_11/13/16/18/22/23

파악된 상관관계에 대해 정확한 검정을 통해 상관성의 유무를 파악하기 위하여 정준 상관계수가 0인지에 대해 확인해야 한다. 정준 상관계수가 0인지를 판별하기 위해서는 <Table 1>에서 제시한 유의성 검정 통계량 중 윌크스의 람다 검정 통계량을 이용하여 설정된 가설($H_0: \rho=0$ vs. $H_1: \rho \neq 0$)에 검정을 실시하였다.

Table 5. Wilk's Lambda test result

	Wilk's Lambda	F	df1	df2	P
First CCF	0.1450935	245.636208	377	527713.37	0.00E+00
Second CCF	0.4713375	102.3735225	336	490950.89	0.00E+00
Third CCF	0.6754719	59.519525	297	453740.39	0.00E+00

<Table 5>는 불량결점수 변수와 전압 변수의 정준 상관분석에 대한 유의성 검정 결과이다. 검정 결과 3개의 정준 상관에 대한 모든 정준 상관계수의 유의 확률(P-value)이 0.001 이하의 값을 나타낸다. 이는 유의 수준(Significant level)이 0.01이라 할 때, 유의 확률이 유의 수준보다 현저히 낮으므로 귀무가설(H_0)을 기각하게 된다. 즉, 모든 정준 상관계수는 0이 아니라고 할 수 있다. 상관계수가 0이 아니므로 불량결점수에서 추출된 주요 변수와 웨이퍼 검사의 전압 변수 간에는 상관성이 존재함을 알 수 있다. 또한 연구 결과를 통해서 본 연구에서 목적으로 하는 웨이퍼 검사 공정 간 상관성을 확인할 수 있었다.

다음으로 정준 상관분석의 결과로 정준변량에 대해 <Figure 5>와 <Figure 6>를 살펴볼 것이다. <Figure 5>와 <Figure 6>의 그림은 전압 변수와 불량결점수 변수의 2차원 정준 변량 공간에 나타난 그림으로 <Figure 5>와 <Figure 6>을 비교하여 제2축

에 있어 전압 변수 HV_10 과 HF_13의 대응성이 뛰어난 것을 알 수 있다. <Figure 5>를 이용하여 HV_11과 HV_13이 가깝게 위치하고 있어 정준 변량의 결정에 있어 비슷한 역할을 하였음을 알 수 있지만 그렇다고 두 변수의 상관도가 높은지는 상관계수를 이용하여 직접 해석해야 된다. 본 연구에서는 각 변수 간의 상관성을 이용하여 예측 모델을 생성한 뒤 결과를 확인한다.

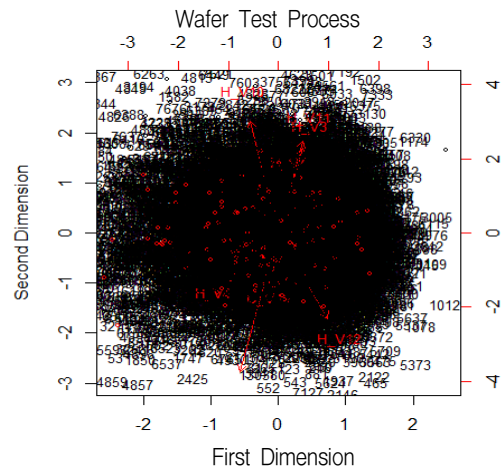


Figure 5. Second dimension canonical space of DC voltage variable

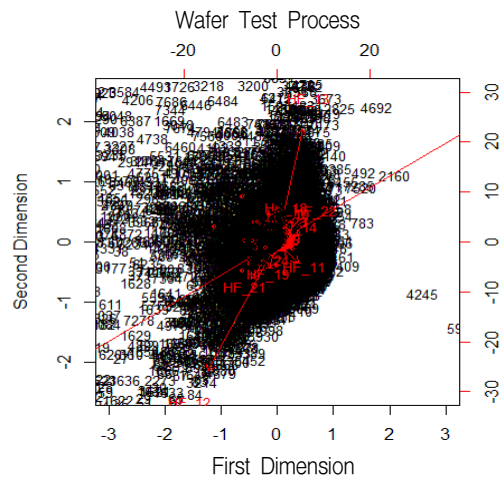


Figure 6. Second dimension canonical space of Fail bit count variable

3.4 실험 성능 평가

정준 상관분석을 이용하여 전압과 불량결점수의 상관관계를 확인하였다. 파악된 상관관계를 이용하여 패키지 테스트의 결과를 예측하고 전압 변수를 이용하여 불량결점수를 측정하는 공정 전에 패키지 테스트의 결과를 예측하여 전처리를 통한 수율을 향상시키는 방법을 제안한다. 예측 모델의 예측 정확도를 판단하기 위해서는 민감도(Sensitivity)와 특이도(Specifi-

city)를 이용하였다. 식 (15)은 민감도를 나타낸 식이고 식 (16)은 특이도를 나타낸다. 양품을 양품으로 예측하는 것이 가장 중요하지만, 공정에서 불량품 또한 양품으로 진행되는 이원화 공정을 고려할 경우, 불량품에 대한 예측 정확도 즉, 특이도 또한 중요한 요소이다(Kim *et al.*, 2014).

$$Sensitivity = \frac{TP}{TP+FN} \quad (15)$$

$$Specificity = \frac{TN}{TN+FP} \quad (16)$$

- TP : True Positive (불량을 불량으로 예측)
- FN : False Negative (불량을 양품으로 예측)
- TN : True Negative (양품을 양품으로 예측)
- FP : False Positive (양품을 불량으로 예측)

하지만 민감도와 특이도만으로는 분류 모델의 성능을 파악하기가 어려움이 있으며 실제 공정에서 발생하는 데이터의 형태는 불균형 데이터이기 때문에 민감도만을 성능 척도로 이용하면 소수 범주에 과적합(Over-fitting)된 분류 모델이 생성될 수 있다. 또한 과적합으로 인하여 다수 범주의 분류 성능이 감소하여 전체적인 분류 성능의 저하가 발생한다. 불균형 데이터의 분류 성능에 대한 문제점을 해결하기 위해서는 다수 범주와 소수 범주의 분류 성능을 고려한 기하 평균(Geometric mean)을 이용하였다(Kubat *et al.*, 1997). 기하 평균은 소수 범주와 다수 범주의 분류 성능을 고려할 수 있는 성능 척도이며 기하 평균은 식 (17)과 같이 계산한다(Lee *et al.*, 2015). 본 논문에서는 실험 결과를 평가하기 위해 민감도와 특이도 그리고 기하평균을 살펴본다.

$$\begin{aligned} \text{기하 평균} &= \sqrt{Sensitivity \times Specificity} \quad (17) \\ &= \sqrt{\frac{TP}{TP+FN} \times \frac{TN}{TN+FP}} \end{aligned}$$

3.5 통계적 예측 모형(Statistical Predictive model)

본 연구의 목표는 전압 관련 변수와 불량결점수 변수를 이용하여 공정 간의 상관관계를 파악함으로써 품질을 향상시키는 것이며 최종 칩 품질을 예측함으로써 결과를 확인하고자 한다. 품질 예측을 위하여 통계적 예측 모형을 적용하였다. 예측 모형 선택에 고려된 방법들로서 ANN(Artificial Neural Network)와 k-NN(k-Nearest Neighbor Classification) 그리고 SVM(Support Vector Machine)을 이용하였다. 동일 변수를 가지고 구현된 모델들의 기하 평균을 이용하여 예측 정확도를 비교하였다. <Table 6>에서 나타난 것처럼 3가지의 모델 중 SVM의 기하 평균이 0.4555로 가장 우수한 성능을 보였다. 본 연구에서는 상관성 확인을 위한 예측 모델로서 SVM(Support Vector Machine)을 선택하였고 종속변수로 패키지 테스트 결과를 독립변수로 각각 전압 변수와 불량결점 수를 사용하였다. SVM은 두 범주를 갖는 객체들

을 분류하는 방법으로서 두 범주의 여백을 최대화하는 초평면을 찾는 방법이다. SVM은 구조적 위험도를 최소화하여 일반적인 오류를 줄이는 장점을 가지고 있다(Park *et al.*, 2014).

Table 6. Geometric mean by predictive model

Predictive Model	Geometric Mean
ANN	0.3137
k-NN	0.0991
SVM	0.4555

3.6 실험 결과

본 연구에서 정준상관분석을 이용하여 공정 간의 상관관계를 파악하였고 상관관계로 인한 결과의 영향을 확인하기 위해 전압을 사용한 예측 모델과 불량결점수를 SVM을 이용한 예측 모델을 사용하였다. 실험 결과의 일반성을 확보하기 위해 5-Fold Cross Validation을 이용하였다. <Figure 7> 불량결점수와 전압 변수를 사용한 모델의 민감도를 나타내고 <Figure 8>는 위의 모델의 특이도를 나타낸다. 그리고 <Figure 9>는 위 두 모델의 기하 평균을 나타낸 상자그림이다. 민감도의 상자그림과 특이도의 상자 그림을 살펴보면 민감도 상자그림에서 변수 별 분산의 차이가 발생한 것을 알 수 있다. 이는 불량결점 수를 변수

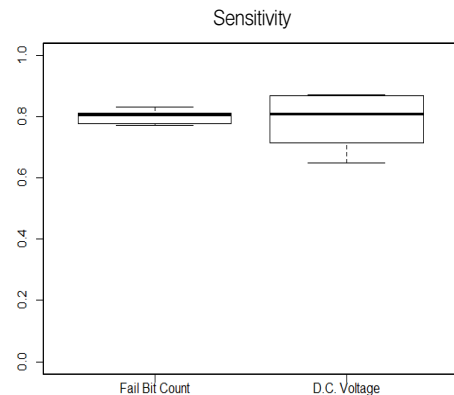


Figure 7. Sensitivity of predictive model

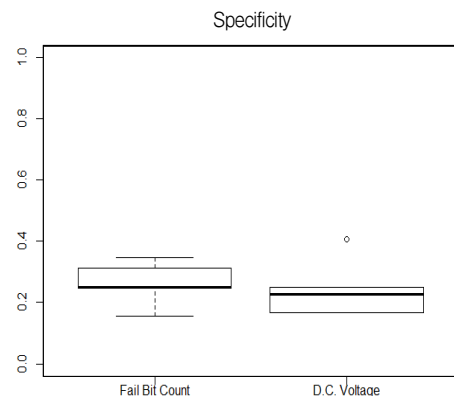


Figure 8. Specificity of Predictive model

로 사용한 모델이 전압을 변수로 사용한 모델에 비해 양품을 더 잘 예측 가능하다고 판단할 수 있다. 또한 특이도 상자그림의 경우 전압변수로 그려진 상자그림에 이상치가 존재하기는 하지만 불량결점수 변수에 비해 상대적으로 분산이 낮은 것을 확인할 수 있다. 따라서 전압 변수를 사용한 예측 모델이 불량결점수 변수를 사용한 예측 모델보다 더 불량을 잘 예측함을 확인할 수 있다. <Table 7>은 실험결과와의 평균을 나타낸 표이다.

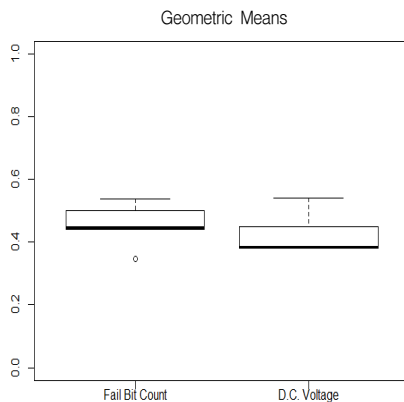


Figure 9. Geometric Mean of Predictive model

Table 7. The average of test result

	Input Variables	
	Fail Bit Count	D.C. Voltage
Sensitivity	0.8003	0.7833
Specificity	0.2629	0.2436
Geometric Mean	0.4587	0.4368

<Table 7.>에서 보는 바와 같이, 특이도의 경우 민감도에 비해 낮게 나타난다. 이는 공정에서 발생하는 양품 대비 불량품의 비율이 100 : 2로 매우 극심한 불균형 데이터이기 때문에 발생한다. 기하평균 또한 특이도와 민감도의 곱으로 표현되기 때문에 낮게 나타난다. 불량결점수를 사용한 모델이 민감도와 특이도, 기하 평균에서 약 0.02 차이로 높은 값을 지니고 있다. 이는 근소한 차이이므로 전압 변수를 사용한 모델과의 차이가 없다고 할 수 있다. 불량결점수는 칩 레벨에서 패키지 테스트의 결과를 예측하는 중요한 변수이다. 본 연구의 결과로 불량결점수와 상관성이 높은 전압 변수를 이용하여 패키지 테스트 결과인 칩의 품질을 예측한 결과 비슷한 결과를 나타낸다. 따라서 전압을 테스트하는 공정과 불량결점수를 측정하는 공정 사이에 불량에 대한 전처리 과정을 진행한다면 최종 패키지 테스트의 수율을 향상시킬 수 있다.

4. 결론

반도체 제조 공정에서는 생산성 향상과 품질 개선을 위해 제

조 공정 간의 상관관계 파악에 힘쓰고 있다. 본 연구에서는 고 집적화된 반도체 제조 공정 간의 상관성을 파악하기 위해 웨이퍼 검사의 공정 변수인 전압 변수와 웨이퍼 검사의 결과인 불량결점수의 정준 상관 분석을 사용하였다. 또한 불량결점수와 전압 변수를 SVM을 이용하여 최종 패키지 테스트 결과를 예측 모델을 구축하였고 예측 모델의 결과로 민감도와 특이도 그리고 기하평균을 비교하였다 3가지 척도에 대해 비슷한 결과를 나타내었다. 따라서 전압 변수만으로도 불량결점수와 유사한 결과를 도출할 수 있으며, 본 연구를 이용하여 전압을 측정하는 공정과 불량결점수를 측정하는 공정 사이에 전처리 과정을 추가한다면 최종 패키지 테스트의 수율을 향상시킬 수 있다. 공정 간의 상관성을 파악하는 방법을 이용한 추후 연구 과제로 정준 상관 분석을 이용하여 상관성을 확인하는 것이 아니라 빅데이터 속에서 중요 변수를 선택하는 데 활용하는 것이다. 또한 웨이퍼 검사 단계뿐만 아니라 다른 공정과의 상관관계를 파악하는 연구를 통해 최종 패키지 테스트의 결과 즉, 수율과의 상관성에 있어 연구가 진행된다면 품질 예측 알고리즘의 성능을 향상시킬 수 있을 것이다.

참고문헌

- An, D.-W., Ko, H.-H., Baek, J.-G., and Kim, J.-Y. (2009), A Yield Prediction in the Semiconductor Manufacturing Process Using Stepwise Support Vector Machine, *IE Interfaces*, **22**(3), 252-262.
- Baek, D.-H. and Nam, J.-G. (2002), Semiconductor yield improvement system using the data mining, *IE Interfaces*, 293-300.
- Byun, J.-H. and Kim, Y.-G. (1998), An EVOP Procedure Using the Relationship Between Quality Characteristics and Process Factor Conditions, *IE interfaces*, **1**, 1-5.
- Chien, C. F., Wang, W. C., and Cheng, J. C. (2007), Data mining for yield enhancement in semiconductor manufacturing and an empirical study, *Expert Systems with Applications*, **33**(1), 192-198.
- Hsu, S. C. and Chien, C. F. (2007), Hybrid data mining approach for pattern extraction from wafer bin map to improve yield in semiconductor manufacturing, *International Journal of Production Economics*, **107**(1), 88-103.
- Hotelling, H. (1936), Relations between two sets of variates, *Biometrika*, **28**, 321-377.
- Jung, S.-Y. and Sung, I.-H. (2013), Investigation of the Relationship Between Various Process Parameters and Chatter Scratch Formation in Chemical-Mechanical Polishing, *Journal of the KSTLE*, **57**, 133-134.
- Kim, K.-N., Hwang, C.-G., and Lee, J.-G. (1998), DRAM technology perspective for gigabit era. *Electron Devices, IEEE Transactions on*, **45**(3), 598-608.
- Kim, K.-H. and Baek, J.-G. (2014), A Prediction of Chip Quality using OPTICS (Ordering Points To Identify the Clustering Structure)-based Feature Extraction at the Cell Level, *Journal of the Korean Institute of Industrial Engineering*, **40**(4), 404-414.
- Kubat, M., Holte, R., and Matwin, S. (1997), Learning when negative examples abound, *Proceedings of the 9th European Conference on Machine Learning, ECML-97*, 146-153.
- Lee, C.-J., Park, C.-S., Kim, J.-S., and Baek, J.-G. (2015), A Study on

- Improving Classification Performance for Manufacturing Process Data with Multicollinearity and Imbalanced Distribution, *Journal of the Korean Institute of Industrial Engineering*, **41**(1), 25-33.
- Lee, Y.-J. (2002), *Understanding of the Canonical correlation analysis*, Seok-Jung Moonhwasa, 27-54.
- Ludwig, L., Sapozhnikova, E., Lunin, V., and Rosenstiel, W. (2000), Error classification and yield prediction of chips in semiconductor industry applications, *Neural Computing and Applications*, **9**(3), 202-210.
- Park, S.-R., Kim, J.-S., Park, C.-S., Park, S.-H., and Baek, J.-G. (2014), Under Sampling for Imbalanced Data using Minor Class based SVM (MCSVM) in Semiconductor Process, *Journal of the Korean Institute of Industrial Engineering*, **40**(4), 404-414.
- Pieter, P. B. (2000), 2000 begins with a revised industry roadmap, *Solid State Technology*, 31-44.
- Quirk, M. and Serda, J. (2001), *Semiconductor manufacturing technology, USA : Prentice Hall*, **1**, 49-50.
- Seok, J.-W., Kim, T.-H., and Bae, G.-S. (2012), Underwater Target Analysis Using Canonical Correlation Analysis, *Journal of Korean Institute of Information and Communication*, **16**(9) 1878-1883.
- Sohn, S.-Y. and Lee, S.-G. (2012), Probe test yield optimization based on canonical correlation analysis between process control monitoring variables and probe bin variables, *Expert Systems with Applications*, **33**(1) 192-198.
- Uzsoy, R., Lee, C. Y., and Martin-Vega, L. A. (1992), A review of production planning and scheduling models in the semiconductor industry part I : system characteristics, performance evaluation and production planning, *IE transactions*, **24**(4), 47-60.