

반도체에 적합한 복합 학습곡선 모형

하정훈[†]

홍익대학교 정보컴퓨터공학부

Compound Learning Curve Model for Semiconductor Manufacturing

Chunghun Ha

School of Information and Computer Engineering, Hongik University, Seoul, Korea, 121-791

The learning curve model is a mathematical form which represents the relationship between the manufacturing experience and its effectiveness. The semiconductor manufacturing is widely known as an appropriate example for the learning effect due to its complicated manufacturing processes. In this paper, I propose a new compound learning curve model for semiconductor products in which the general learning curve model and the growth curve are composed. The dependent variable and the effective independent variables of the model were abstracted from the existing learning curve models and selected according to multiple regression processes. The simulation results using the historical DRAM data show that the proposed compound learning curve model is one of adequate models for describing learning effect of semiconductor products.

Keyword: learning curve model, semiconductor manufacturing, yield learning, growth curve

1. 서론

반도체의 주요 제조공정은 크게 웨이퍼 가공(wafer fabrication), 웨이퍼 검사(wafer probe), 패키지 조립(assembly), 그리고 최종기능검사(final functional test) 등 4가지로 구분할 수 있다. 이 중에서 웨이퍼 가공 및 검사 공정은 반도체 생산에 있어서 대부분의 공정을 담당하며 막대한 투자비를 요구하므로 수익에 지대한 영향을 미친다(Cunningham *et al.*, 1995). 최근 반도체 칩의 복잡도와 집적도가 급속도로 증가하면서 웨이퍼 가공설비(FAB)의 구축비용 또한 현저히 증가하고 있다. 이에 따라 반도체 산업업체에서는 비용절감을 위하여 웨이퍼 FAB의 가동률, 수율, 최대 생산량, 다이(die) 크기의 감소 등을 생산관리의 중요한 관리대상으로 하고 있다.

반도체 생산에서 가장 중요한 관리요소 중 하나는 수율

(yield)로서, 수율은 총 투입량 대비 판매 가능한 제품의 비율로 정의할 수 있다. 반도체 생산에서 수율은 가공공정의 순서에 따라 라인수율(line yield), 다이수율(die yield), 조립수율(assembly yield), 그리고 최종검사수율(final test yield)로 분류할 수 있다. 라인수율은 웨이퍼 투입량 대비 웨이퍼 검사의 통과량을 기준으로 산출하고, 다이수율은 검사를 통과한 웨이퍼에서 분리된 다이(die)를 기준으로 산출한 수율이다. 조립수율과 최종검사수율은 패키지 조립을 거쳐 최종기능검사를 마친 반도체 완성품에 대한 수율이다. 최종수율 또는 누적수율은 각 공정수율의 곱으로 표현할 수 있다(Cunningham, 1990).

수율손실(yield loss)의 80%는 <Table 1>과 같이 오염(contamination)에 기인한 결함(defect)으로 알려져 있다(ICE 1997). 오염은 크린룸(cleanroom), 공정(processes), 장비, 사람으로부터 발생 할 수 있는데, 1985년경에는 장비와 사람으로부터 기인한 오염

이 논문은 2007학년도 홍익대학교 학술연구진흥비에 의하여 지원되었음.

[†]연락처 : 하정훈 교수, 121-791 서울 마포구 상수동 72-1, 홍익대학교 정보컴퓨터공학부 산업공학전공, Fax : 02-336-1130,

E-mail : chungun.ha@hongik.ac.kr

투고일(2009년 06월 04일), 심사일(1차 : 2009년 09월 23일, 2차 : 2010년 03월 02일), 게재확정일(2010년 03월 22일).

Table 1. VLSI chip 생산에 있어서 수율손실에 관련된 주요 인자들(ICE 1997)

원인	수율손실 기여도(%)
오염	80
설계 마진	10
프로세스 변동	6
노광에러	2
소재 결함	2
총 손실	100

이 각각 30% 정도로서 가장 많은 비중을 차지하였으나, 2000년경에는 약 80%가 프로세스에서 기인한 결함으로 보고되고 있다. 결함의 발생 원인을 분석하고 그 원인을 제거하는 혁신 활동은 수율의 향상에 기여한다. 그러나 이러한 정태적인 분석은 현재의 상황을 파악하고 개선하는 데는 유용하지만 생산 초기로부터 안정기간까지 시간에 따른 수율변화를 분석하는 데는 한계가 존재한다.

초기생산 단계에서의 낮은 수율은 공정이 안정화되면서 점차 상승하고 최종적으로 정상수율로 수렴한다. 시간 또는 생산량에 따라 수율이 개선되는 이러한 동태적 현상을 수율학습(yield learning)이라고 한다. 정상수율은 FAB에서 생산이 안정화된 이후의 수율을 결정하므로 반도체 생산기업에서 가장 중요시하는 관리요소 중 하나이다. 한편, 반도체 제품의 라이프 사이클은 매우 짧아 신제품 출시단계에는 수요가 공급보다 많아 프리미엄을 받고 판매가 가능하지만, 시간이 지남에 따라 공급이 수요를 초과하게 되면 제품의 가격이 급격히 하락하며 수익 또한 급격히 감소한다. 따라서 초기수율에서 정상수율에 도달할 때까지의 기간인 수율 램프업(ramp-up) 기간 또한 수익을 결정하는 중요한 요소가 된다. 수율 램프업 기간을 6개월 단축하면 투자수익의 2배가 증가하고, 램프업 기간의 6개월 지연은 투자수익의 3분의 2를 감소시킨다고 알려져 있다(Weber, 2004). 그러므로 반도체 생산기업에서는 조기에 수율을 극대화시킬 수 있도록 수율 램프업 기간을 줄이는 것이 필수적인 생산활동 중의 하나이다.

수율학습에 가장 필수적인 것은 수율에 영향을 주는 요소를 신속히 파악하고 이를 조기에 해결하는 학습 능력이다. 이와 더불어 수율의 동적인 변화를 예측하고 이를 반영한 적절한 생산 및 판매계획을 수립하여 수익을 최대화 하도록 노력해야 한다. 이를 위해서는 수율의 동적인 변화를 정량적으로 분석하고 해석할 수 있는 도구의 확립이 필요하다. 학습곡선 모형은 이러한 과정에 대한 이론적 근거 및 모형을 제시한다는 점에서 중요한 역할을 한다.

학습곡선(learning curve)이란 반복 학습(또는 시간)과 효율성의 관계를 표현하는 함수로서 비용-생산량 관계(cost-quantity relationship), 생산가속곡선(product acceleration curve), 경험곡선(experience curve), 효율곡선(efficiency curve) 등 다양한 이름으로

불리고 있다(Badiru, 1992). 생산에서의 학습은 동기와 원인에 따라 반복적인 생산활동의 과정에서 획득하는 학습(learning by doing), R&D와 같이 탐구에 의해 획득하는 학습(learning by searching), 관계자와의 지식의 교환 및 공유에 의해 이루어지는 학습(learning by interacting), 사용자에 적합하도록 개선하려는 과정에서 생성되는 학습(learning by using)으로 분류할 수 있다. 이중에서 반복적인 실행에 의한 학습(learning by doing)은 기술적으로 복잡한 프로세스를 가진 산업에서 두드러지게 나타난다. 반도체 산업은 매우 복잡한 프로세스로 구성되어 있으므로 실행에 의한 학습이 대표적으로 나타나는 산업이라고 할 수 있다(Gruber, 1994).

학습효과는 경험과 효율성의 관계를 나타내고 학습곡선 모형은 이러한 관계를 수학적으로 표현하는 것이다. 모형의 구성을 위해서는 독립변수와 종속변수를 구분하여 정의해야 하는데, 종속변수로는 누적 평균(또는 단위) 직접 노무비, 판매 가격, 또는 단위생산시간 등이 사용이 되고, 종속변수에 영향을 미치는 독립변수로는 누적 생산량, 초기 경험치, 총생산기간 등이 사용된다.

일반적으로 적용하는 평균 생산비용, 노무비용과 같은 독립변수는 직관적으로 쉽게 이해가 가능하나 기업의 비용정보의 비공개성으로 인하여 이를 직접 이용하기는 매우 어렵다. 이를 해결하기 위하여 제품의 가격을 비용의 대리변수로서 사용하는 연구가 진행되었다(Lieberman, 1984; Gruber, 1992). 제품의 가격으로서 비용 대신 가격을 대리변수로 사용하기 위해서는 분석기간 동안 가격 대비 비용 마진이 일정해야 하며 가격 대비 비용 마진의 변화가 생산비용의 변화보다 상대적으로 작아야 하고 필요시 이를 조정해 주어야 한다.

Gruber(1992)는 대리변수로서 평균 실제 판매가격을 적용하여 메모리 반도체 종류별 학습곡선의 특성을 분석하였다. Gruber의 연구에 따르면, EPROM의 경우 학습곡선의 영향이 생산비용에 영향을 미치나 대량생산에 따른 규모의 경제나 제품의 생산기간은 생산비용에 영향을 미치지 못하며, DRAM의 경우는 규모의 경제와 생산기간, SRAM의 경우 생산기간만 생산비용에 영향을 주며 학습에 의한 효과는 거의 없는 것으로 주장하였다. 그러나 Gruber가 대리변수로 이용한 판매 가격은 공급과 수요의 변화에 따라 그리고 반도체 가격주기에 따라 가격의 변동이 심하게 나타나는 것이 주지의 사실이므로 대리변수로서 적합하지 않다고 판단된다. 반도체 산업의 경우 가격 대비 비용의 마진의 변화가 크므로 가격을 대리변수로 적용한 분석은 결과가 왜곡될 가능성이 많다. 따라서 반도체 산업의 경우 비용의 대리변수로서 수율(Gruber, 1994; Chung, 2001) 또는 결함 밀도(Hatch, Mowery, 1998)를 적용하는 연구가 진행되었다.

학습곡선 모형에서는 다양한 종속 및 독립변수의 설정이 가능하므로 학습곡선 모형 구성 시 해당 산업의 특성에 따른 학습효과를 가장 잘 설명할 수 있는 변수를 적용하는 것은 필수적이다. 본 논문에서는 반도체에 관련된 기존의 학습곡선 모형을 분석하여 새로운 학습곡선 모형을 제시하고 반도체 산업

에 효과적인 변수를 설정하고자 한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 일반적인 학습곡선 모형에 대하여 설명하고 제 3장에서는 반도체에 관련된 기존의 대표적인 학습곡선 모형을 소개한다. 제 4장에서는 기존의 연구의 분석을 통한 반도체에 적합한 학습곡선을 도출하기 위한 주안점과 새로운 모형을 제시하고 이를 검증한다. 마지막으로 제 5장에서는 결론과 향후 연구방향에 대하여 논한다.

2. 학습곡선 모형

생산의 한계비용은 현재의 생산량, 누적 생산량, 그리고 생산기간에 따라 감소하며, 실행에 의한 학습은 누적되고 이렇게 누적된 지식은 새로운 제품과 기술의 혁신의 원천으로서 작용한다(Alchian, 1959). 이러한 기술적 혁신의 과정은 학습곡선 모형으로 모델링이 가능한데, 일반적으로 다음과 같이 표현할 수 있다(Gruber, 1992).

$$Y = f(C, N, Q, T)$$

여기서, Y 는 종속변수로서 학습의 결과에 대한 대표 변수(누적 평균 생산비용, 단위 생산비용, 단위 생산시간 등)를 나타내며, C, N, Q, T 는 각각 독립변수로서 C 는 학습이 진행되기 이전의 초기치, N 은 누적생산량, Q 는 해당기간의 생산량, T 는 제품이 생산된 기간을 의미한다. 각 독립변수는 생산되는 제품 또는 산업의 특성에 따라 다르게 설정이 가능하며, 독립변수가 한 개인 경우 단일요인 모형(univariate model)이라 하며, 두 개 이상인 경우 다중요인 모형(multivariate model)이라고 한다(Badiru, 1992).

가장 일반적으로 적용하는 학습곡선 모형은 식 (1)과 같은 Cobb-Douglas 생산함수의 형태이다.

$$Y = CN^\alpha Q^\beta T^\gamma \quad (1)$$

양변에 로그를 취하면 식 (2)와 같이 선형화가 가능하다.

$$\ln Y = \ln C + \alpha \ln N + \beta \ln Q + \gamma \ln T \quad (2)$$

위의 선형모형을 기본으로 과거의 데이터를 적용하여 회귀방정식을 구하면, 각 독립변수의 탄력성 α, β, γ 의 추정이 가능하다. 도출된 각 요소에 대한 탄력성은 종속변수 Y 에 대한 각 독립변수 N, P, Q 의 영향의 정도이므로 이를 통하여 효율성을 향상시킬 수 있는 방법을 도출하거나 미래의 효율성을 예측하는 데 사용이 가능하다.

3. 반도체에서의 학습 곡선 모형

반도체에 관련한 학습곡선은 여러 연구자에 의해 다루어 졌지

만 대표적인 모형은 다음과 같다.

3.1 Gruber(1992) 모형

Gruber(1992)는 메모리 반도체에 대하여 다음과 같은 학습곡선 모형을 제시하였다.

$$P_{ijt} = CN_{ijt}^\alpha Q_{ijt}^\beta T_{ijt}^\gamma$$

여기서 P 는 실제 평균 판매가격, C 는 학습이 진행되기 이전의 초기치, N 은 누적생산량, Q 는 해당기간의 생산량, T 는 해당 세대의 제품이 생산된 기간을 의미한다. 또한, t 는 연도, j 는 반도체 세대(256K, 1M, 4M 등), i 는 메모리 반도체의 형태(EPROM, DRAM, SRAM 등)를 나타내는 인덱스이다. 여기서 누적생산량 N 은 학습효과, 해당기간 생산량 Q 는 규모의 경제(economies of scale)를 대변하는 독립변수이다.

Gruber의 연구에 따르면 반도체의 형태에 따라 가격에 미치는 영향은 다르며, EPROM은 학습효과, DRAM은 규모의 경제와 해당 세대의 생산기간, SRAM은 해당 세대의 생산기간만 판매가격에 영향을 미친다. 또한, 반도체 산업의 학습효과에 있어서 제품 간 변화보다 생산 라인 간 변화가 더욱 크다고 주장하였다.

Gruber(1992) 모형은 모형이 단순하고 데이터의 확보가 용이하여, 적용이 간편하다는 장점이 있다. 그러나 종속변수로 사용한 반도체의 실제 판매가격의 변동이 매우 심하며, 판매가격은 주로 수요와 공급의 차이에 영향을 가장 많이 받으므로 적합한 종속변수로 보기 어렵다. 또한, 메모리 반도체는 반도체의 용량, 다이의 크기, line width, mask layer의 수 등에 따라 수율의 변화가 심하고 수율을 가격에 영향을 주므로 판매가격은 학습의 결과에 대한 대리변수로 보기에는 무리가 있다.

3.2 Gruber(1994) 모형

Gruber는 과거의 모형을 개선하여 수율을 종속변수로 하는 새로운 학습곡선 모형을 제시하였다. 수율 학습 모형은 다음과 같이 표현할 수 있다.

$$Y = Y_D(D_0, A, \theta) \cdot Y_M(X)$$

여기서, Y_D 는 die 수율에 대한 결함율 이용한 수율모형(Poisson 수율 모형, Negative Binomial 수율 모형 등)을 의미하며 인자로서 단위면적 당 결함밀도(D_0), die 면적(A), 수율 모형에 따른 파라미터(θ , 예를 들어, Poisson Model의 경우는 평균 결함의 수, Negative Binomial Model의 경우는 평균 결함의 수와 클러스터링 파라미터)를 갖는다. $Y_M(X)$ 은 학습 효과에 의해 지속적으로 증가하는 최대 가능 수율(정상수율)이며, 인자 X 는 학습의 독립변수로서 반도체 세대 또는 누적생산량 등을 의미한다.

$Y_M(X)$ 을 학습에 따른 정상수율이라고 하고 $Y_M(0) = 0$,

$Y_M'(X) > 0$, $Y_M''(X) < 0$, $\lim_{X \rightarrow \infty} Y_M(X) = 1$ 의 특성을 가진다고 가정하면, 세대 j 에서의 수율은 다음과 같이 표현할 수 있다.

$$Y_{jt} = Y_{Dj} \cdot \exp\left\{\frac{\delta_j}{H_{jt}}\right\}$$

여기서 H_{jt} 는 시간 t 에서 세대 j 를 생산한 기간, δ_j 는 탄력성을 나타내는 계수로서 음수이다. δ_j 가 음수이므로 H_{jt} 가 증가함에 따라 정상수율 $\exp\{\delta_j/H_{jt}\}$ 는 최대값 1에 수렴하는 오목 함수가 된다.

학습효과를 반영한 수율(Y)과 반도체 chip 당 가격(C)간의 수율과의 관계는 다음과 같이 설정할 수 있다.

$$C = \frac{W}{Y \cdot N^P} + Z \quad (3)$$

여기서 W 는 웨이퍼의 가격, N^P 는 한 장의 웨이퍼에서 생산 가능한 다이의 총 개수, Z 는 어셈블리와 패키징에 관련된 다이 당 고정비이다. 반도체 기업은 이익을 최대화하기 위하여 chip 당 가격을 지속적으로 낮추려고 한다. 기술이 동일하다면 웨이퍼 가공비용 W 는 고정적이므로 기업은 다이 수율 Y 를 증가시켜 chip 당 가격 C 를 감소시키려고 한다. 만약 Y 가 정상수율에 도달하여 개선이 어려우면 기업은 N^P 를 증가(다이 면적의 감소)시켜 C 를 낮추려는 노력을 하게 된다.

본 모형은 기존의 변동성이 심한 반도체 가격에서 보다 안정적인 수율로 종속변수를 개선하고 학습효과는 정상수율의 크기로 나타나도록 한 것이 특징이다. 그러나 수율은 die의 크기, line width 등 반도체 설계와 매우 관련이 높아 세대 간 반도체의 설계 또는 기술이 급격히 변화할 경우, 정상수율만 학습에 의하여 증가한다고 주장하기 어려운 단점이 존재한다.

3.3 Hatch and Mowery의 모형

Hatch and Mowery(1998)(이하 HM 모형)는 반도체 산업에서 새로운 프로세스 도입에 따른 학습곡선의 변화를 모델링하였다. HM 모형은 다른 모형과 달리 몇 가지 특징이 존재한다. 첫째, 기존의 학습곡선모형의 경우 비용, 가격, 또는 수율 등의 종속변수를 사용하였으나, HM 모형은 결함밀도(defect density)를 종속변수로 갖는다. 학습효과는 FAB 간의 차이가 분명히 존재하며 결함밀도는 FAB 간 차이를 대표하는 대표적인 변수이므로 이는 타당한 설정이라고 할 수 있다. 둘째, HM 모형은 FAB의 지식수준을 반도체 기업의 지식수준과 FAB에 특화된 지식으로 구분하여 모형을 설정하였다. 반도체 기업의 지식수준(K_t)은 반도체 설계기술, 공정의 운영지식 등과 같이 FAB에 특화되지 않고 다른 제품 또는 공정의 생산을 통하여 획득한 지식을 의미하며, FAB에 특화된 지식수준은 해당 FAB의 설비에 관련한 지식, 청정도 등 해당 FAB에만 국한된 지식을 의미한다. 새로운 공정을 FAB에 도입할 경우 기업의 지식수준은 그

대로 보전되며, FAB의 지식수준은 과거의 경험을 토대로 초기치가 결정되고 학습에 의해 개선된다. 셋째, HM 모형은 기존 모형과 달리 학습의 독립변수로서 line width, mask layer의 개수, 크린룸의 청정도 등을 고려함으로써 반도체 설계 및 공정의 변동성을 고려하였다.

HM 모형에서 t 시점에서 FAB의 지식수준 L_t 는 반도체 기업의 지식수준과 FAB에 특화된 지식의 합으로서 다음과 같이 모델링 된다.

$$L_t = aCV_t + bCE_t + npi \cdot xfer + L_0$$

여기서 a 와 b 는 추정해야 하는 계수, CV_t 는 누적생산량, CE_t 는 누적 개발경험, L_0 는 FAB의 초기 지식수준을 나타낸다. npi 는 이진변수로서 $npi = 1$ 은 새로운 프로세스도입을, $npi = 0$ 은 기존프로세스의 진행을 의미한다. $xfer$ 는 새로운 프로세스가 도입될 때 FAB의 특화된 지식수준을 의미하며 다음과 같이 모델링 된다.

$$xfer = x_0 + x_1dev + x_2prox + x_3dup + x_4fab_1$$

여기서 x_i 는 추정 계수, dev 는 별도의 개발 FAB에서 개발된 프로세스의 개수, $prox$ 는 동일한 개발 FAB에서 생산 FAB으로 이전한 프로세스의 개수, dup 는 개발 FAB과 동일한 설비를 가진 생산 FAB으로부터 이전된 프로세스의 개수, fab_1 은 새로운 프로세스가 생산 FAB으로 이전된 프로세스의 개수를 의미한다.

t 시점에서 반도체 공정의 결함밀도(DD_t)는 FAB의 지식수준과 설계 및 공정환경에 의하여 다음과 같이 결정된다.

$$DD_t = [\gamma + \phi e^{-\delta L_t}] + [cCR + lLW + mML + vVin_t]$$

여기서 CR 은 크린룸의 청정도, LW 는 line width, ML 은 mask layer의 수, Vin 은 장비의 노후정도를 의미하며, γ , ϕ , δ , c , l , m , v 는 추정해야 하는 계수이다.

HM 모형은 상당히 정교하고 논리적으로 구성된 모형으로써 공정이 안정되고 신뢰도가 높은 데이터가 확보가 가능하면 적용해 볼 수 있는 모형이다. 그러나 모형에서 사용하는 중간종속변수가 K_t , L_t 와 $xfer$ 와 같은 지식수준으로서 이는 측정이 불가능한 변수이고 또한 적절한 대리변수를 도출하는 것도 용이하지 않다. 게다가 모형의 복잡도가 높고 독립변수의 수가 많아 수집해야 하는 정보량이 많고 하나의 종속변수에 대하여 추정해야 하는 계수가 많아 독립변수 간에 상관관계가 존재할 가능성이 높아 다수의 데이터를 확보해야 하므로 적용에 주의할 기술여야 한다.

4. 새로운 반도체 학습모형

4.1 학습의 주체와 학습의 결과

학습모형을 구현하기 위해서는 우선 학습의 주체와 학습의

결과가 정립되어야 한다. 학습의 주체는 학습을 통하여 개선되어지는 대상으로서 학습곡선 모형의 독립변수를 결정하는 주요요소가 된다. 학습의 결과는 학습의 주체가 학습을 통하여 개선되고 그 결과로서 나타나는 기업의 성과로서 학습곡선 모형의 종속변수로 사용할 수 있다.

반도체의 생산에 있어서 학습의 주체는 크게 반도체 설계, 반도체 공정, 생산운영으로 분류할 수 있다. 반도체 설계에서는 설계자의 능력 및 경험, 설계 장비의 능력, 설계 관련 데이터 등이 학습을 통하여 개선이 되며, 반도체 설계의 학습정도를 표현할 수 있는 측정치로는 누적 설계 개수, 설계 경험자의 수 또는 비중, 설계 관련 특허의 수, 설계 장비 및 S/W의 버전, line width, mask layer의 수 등이 있다. 반도체 공정에서는 작업자의 숙련도 및 경험, 공정 장비의 능력, 공정의 청정도, 공정 프로세스, 결함에 대한 분석 및 개선 능력 등이 학습을 통하여 개선이 된다. 공정의 학습정도를 측정할 수 있는 파라미터는 누적 생산량, 작업자의 근무 연수, 크린룸 청정도 수준, 재료의 순도, 해당 제품의 생산량, 총 생산기간, 설비의 배치 등을 고려할 수 있다. 학습을 통하여 생산운영은 생산계획 능력, 스케줄링 능력, 품질관리 능력, 검사능력 등에 대한 개선이 이루어진다. 생산운영의 학습 측정 파라미터로는 단위기간 당 생산능력, 불량 검출률, 운영관리자의 근무기간 등을 이용할 수 있다.

학습의 결과로 나타나는 것은 생산비용의 감소, 생산능력(capacity)의 증가, 산출량(throughput)의 증대, 그리고 사이클 타임의 감소(cycle time) 등이다. 생산비용은 일반적으로 비공개이며 정확하게 산정하기 어려워 판매가격, 수율, 결함밀도와 같은 대리변수를 적용할 수 있다. 생산능력, 산출량, 사이클 타임 등은 생산운영의 학습결과로서 본 연구에서는 고려 대상에서 제외하기로 한다.

학습곡선 모형은 학습의 주체와 학습의 결과에 따라 독립변수와 종속변수를 결정하고 변수 간의 관계를 구성하면 된다. 제 3장의 기존 학습곡선모형의 분석과 학습의 주체와 결과에 대한 분석을 바탕으로 반도체에 적합한 학습곡선 모형을 구성하기 위한 기준을 제시하면 다음과 같다.

첫째, 독립변수는 학습의 주체 중에서 생산운영을 제외한 반도체 설계와 반도체 공정에 관련된 변수들이 적합하다고 판단된다. 다양한 변수를 선택할 수 있으나 본 연구에서는 데이터 확보의 한계로 인하여 반도체 설계에 관련한 변수로서 transistor 개수, 적용된 line with, mask layer의 수, metal layer의 수, poly layer의 수, die의 크기, transistor의 크기를 고려하며, 반도체 공정에 관련한 독립변수로는 생산연도, 분기 생산량, 해당 제품의 누적 생산량, 해당 제품의 생산기간을 고려한다.

둘째, 반도체 산업에 가장 적합한 종속변수는 결함밀도(defect density)로 판단된다. 결함밀도는 외부변화의 영향이 적으며, 반도체 설계의 경험과 반도체 공정의 학습의 결과를 반영할 수 있다. 수율은 결함밀도로부터 수율모형(Poisson, Negative Binomial 등)을 통하여 계산이 가능하다. 메모리 반도체의 경우는 리던던시(redundancy) 또한 수율에 지대한 영향을 미치므로

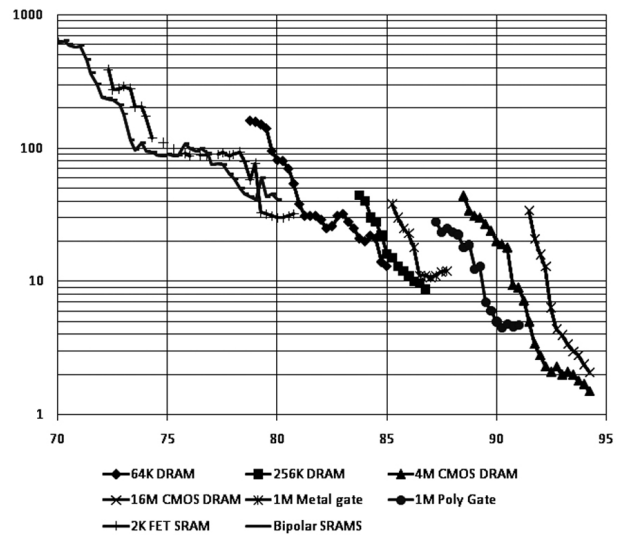


Figure 1. 메모리 반도체의 상대적 결함밀도(Stapper 1995)

이를 반영해 주어야 한다. chip 당 가격은 수율로부터 추정이 가능하다.

셋째, 종속변수와 독립변수 간의 관계는 Cobb-Douglas 생산함수(CD 함수)와 성장곡선의 복합형태가 적합하다. CD 함수는 다양한 실증 연구결과를 가지고 있는 고전적 학습곡선 모형과도 일치하며, 단순하므로 적용이 용이하다. 또한 로그를 취하면 선형화가 가능하므로 파라미터 추정에도 용이하다. <Figure 1>의 결함밀도 그래프를 보면 결함밀도는 로그를 취하여도 선형화되지 않고 성장곡선의 형태를 유지하고 있다. 따라서 결함밀도를 종속변수로 정할 경우에는 성장곡선 또한 반영해 주어야 한다.

4.2 반도체 학습곡선모형 제안

제 4.1절의 검토사항에 따라 본 논문에서 제시하는 반도체 학습곡선 모형은 다음과 같다.

$$D = D_c(X_c, \theta_c) D_g(X_g, \theta_g) \quad (4)$$

여기서 D 는 결함밀도, $D_c(\cdot)$ 는 CD 함수, X_c 는 CD 함수에 사용되는 독립변수 벡터, θ_c 는 파라미터 벡터, $D_g(\cdot)$ 는 성장곡선 함수, X_g 는 성장곡선 함수에 사용되는 독립변수 벡터, θ_g 는 파라미터 벡터이다. 본 연구에서는 대표적인 성장곡선 중의 하나인 Gompertz 성장곡선을 적용하였다.

수율은 결함을 이용한 수율 모형을 이용하여 계산이 가능하다. 가장 많이 사용하는 Negative Binomial 수율모형을 적용하면 다음과 같이 예측수율을 구할 수 있다.

$$\hat{Y} = Y_0 \left(1 + \frac{A_c D}{\alpha} \right)^{-\alpha} \quad (5)$$

여기서 \hat{Y} 는 예측수율, Y_0 는 defect외의 원인에 의해 결정된

수율, A_c 는 해당 제품의 critical area, α 는 클러스터링 파라미터이다. 예측 수율로부터 식 (3)을 이용하면 chip 당 가격 또한 예측이 가능하다.

본 논문에서 제안한 학습곡선 모형의 검증은 위해서는 관련 변수에 대한 데이터 확보가 필수적이다. 그러나 반도체에 관한 데이터, 특히 공정에 관련한 데이터는 반도체 생산기업의 기밀사항에 속하므로 접근하기 매우 어렵다. 따라서 본 논문에서는 다양한 자료를 통하여 공개된 데이터와 과거의 연구에 포함된 데이터를 이용하여 검증을 수행하였다.

우선 종속변수인 결합밀도와 독립변수인 생산연도는 Stapper(1995)에 제시된 데이터<Figure 1>를 사용하였다. 설계 및 공정 데이터인 line width, transistor의 개수, mask layer의 수, poly 및 metal layer의 수, die size는 IC Knowledge(2008)의 데이터를 적용하였다. 이때 결합밀도는 분기별 데이터이고 공정 데이터는 연도별 데이터 이므로 연초에 공정이 적용되는 것으로 가정하였다. 제품의 생산량에 대한 데이터는 Victor and Ausubel(2002)의 연도별 누적 출하량 데이터를 다음과 같이 분기별 데이터로 변환하여 사용하였다. 우선, 연도별 누적 출하량에 대하여 Piecewise cubic Hermite interpolation을 이용하여 분기별 누적 데이터를 생성하고 각 분기별 누적 데이터를 차감하여 분기별 생산량을 도출하였다.

모형 검증을 위한 데이터는 '78년 3분기부터 '94년 1분기까지의 64K, 256K, 1M, 4M, 16M DRAM의 데이터를 사용하였으며 데이터의 총 개수는 102개이다.

4.3 1단계 : 성장곡선의 추정

제안한 학습곡선 모형이 복합함수로 구성되어있고 이 함수는 로그를 취해도 선형화되지 않으므로 단계별 추정을 해야 한다. Gompertz 성장곡선은 일반적으로 다음과 같이 표현된다.

$$D_g = L e^{be^{ax}}$$

여기서 L 은 D_g 의 최대 값이고 α 와 β 는 추정해야하는 모수이다. 양변을 L 로 나누고 양변에 로그를 취하면

$$\ln \ln \left(\frac{D_g}{L} \right) = \ln b + aX_g$$

의 선형함수가 된다.

성장곡선의 모수 a 와 b 를 추정하기 위해서 성장곡선의 종속변수 D_g 를 결정해야 한다. 제안한 학습곡선 모형의 종속변수는 결합밀도 D 이지만 이는 DRAM의 모델에 따라 생산기간과 확보한 데이터의 개수가 다르므로 성장곡선의 모수를 추정하는 데 바로 적용할 수 없다. 따라서 본 논문에서는 다음과 같이 결합밀도를 각 모델 별로 표준화하여 성장곡선을 추정하는 방법을 사용한다.

우선 각 모델 i 와 해당 모델의 생산기간 M_i 에 대하여 $\ln D_{i,M_i}$

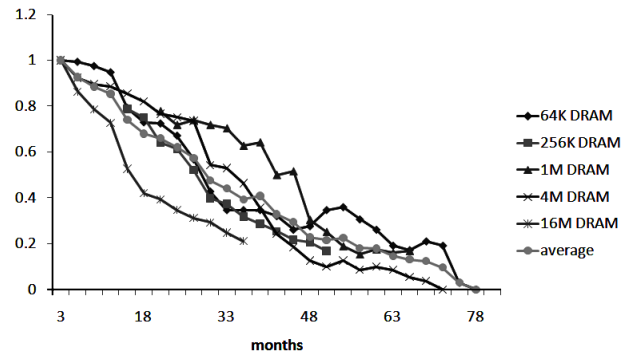


Figure 2. DRAM 모델별 표준화된 결합밀도

를 구하고 각 모델별 최대값과 최소값을 다음의 식을 이용하여 구한다.

$$d_{\max,i} = \max \{ f_t(T_{s,i}), \max \{ \ln D_{i,M_i}, \forall M_i \} \}$$

$$d_{\min,i} = \max \{ \min \{ f_b(T_{e,i}), \min \{ \ln D_{i,M_i}, \forall M_i \} \}, 0 \}$$

여기서 $f_t(T_{s,i})$ 은 모델 별 초기 생산 연도 $T_{s,i}$ 와 결합밀도 $\ln D_{i,T_{s,i}}$ 에 대한 회귀방정식이고, $f_b(T_{e,i})$ 은 모델별 최종 생산 연도 $T_{e,i}$ 와 결합밀도 $\ln D_{i,T_{e,i}}$ 에 대한 회귀방정식이다. 종속변수인 표준화된 로그결합밀도 L_{d,i,M_i} 는 다음과 같이 계산한다 (<Figure 2> 참조).

$$L_{d,i,M_i} = (\ln D_{i,M_i} - d_{\min,i}) / (d_{\max,i} - d_{\min,i})$$

독립변수로서 각 모델별 해당기간의 생산량 Q_{M_i} , 누적생산량 C_{M_i} , 생산기간 M_i , 생산연도 T 와 각 변수의 로그를 사용하여 L_{d,i,M_i} 에 대하여 다중회귀분석을 실시한 결과 $\ln Q_{M_i}$, $\ln M_i$, T 의 p-value가 모두 0.01미만으로 독립변수로서 유의미한 것으로 판정되었다. 이때 성장곡선의 회귀방정식은

$$\begin{aligned} \ln \ln(L_{d,i,M_i}) \\ = -13.11 - 0.21 \ln Q_{M_i} + 2.22 \ln M_i + 0.065 T \end{aligned}$$

이고 추정된 모수와 선택된 독립변수는 각각 $\hat{b} = -2.023 \times 10^{-6}$, $\hat{a} = (-0.21, 2.22, 0.065)$, $L = 1$, $X_g = (\ln C_{M_i}, \ln M_i, T)$ 이다. 모형의 설명력과 적합도는 R-Sq(adj)가 84.2%, 잔차평균제곱(MSE)이 0.189, F-statistics가 151.63으로 매우 높은 수준임을 알 수 있다.

4.4 2단계 : 학습곡선 모형의 추정

1단계에서 추정된 성장곡선을 이용하여 2단계에서는 학습곡선모형을 추정한다. 이를 위한 독립변수로서 transistor 개수(\pm), 적용된 line width(-), mask layer의 수(+), metal layer의 수(+), poly layer의 수(+), die의 크기(\pm), transistor의 크기(\pm), 생산연도

Table 2. 회귀분석을 통한 파라미터 추정결과

Predictor	Coef	SE Coef	t	p
Constant	17.436	1.886	9.24	0.000
$\ln W$	-3.2723	0.3515	-9.31	0.000
$\ln C$	-0.1485	0.0276	-5.39	0.000
$\ln S$	1.9458	0.1535	12.67	0.000
D_g	1.8175	0.2574	7.06	0.000
S	0.2353			
R-Sq	96.1%			
R-Sq(adj)	95.9%			
MSE	0.055			
F	524.34			

(-), 분기 생산량(-), 해당 제품의 누적 생산량(-), 해당 제품의 생산기간(-), 성장곡선(+))을 사용한다. 각 독립변수는 물리적 의미를 가지고 있으며 결합밀도에의 영향에 따라 논리적으로 계수의 부호가 결정된다. 예를 들면, 해당 제품의 누적 생산량이 증가하면 학습효과에 의해 결합밀도는 감소하게 되므로 계수의 부호는 음이 되어야 한다. 각 변수에 대한 물리적으로 타당한 계수의 부호는 괄호로 표기하였다.

제시한 독립변수를 이용하여 결합밀도 D 에 대하여 다중회귀분석을 실시하였다. 독립변수 중에서 타당한 계수의 부호와 유의성 기준($p\text{-value} < 0.01$)을 만족하는 변수는 line with W , 해당 제품의 누적 생산량 C , transistor의 크기 S , 성장곡선 D_g 로 판정되었다. 이때 선택된 모든 변수의 $p\text{-value}$ 는 0.000으로 유의수준 1% 이내에서 유의함을 알 수 있다.

1단계와 2단계에서 선택된 독립변수를 고려한 최종적인 학습곡선모형은 다음과 같다.

$$D = AW^\alpha C^\beta S^\gamma \exp(\delta \exp(a_1 \ln C + a_2 \ln M + a_3 T))$$

최종모형에 대하여 다시 다중회귀분석을 재실시한 결과는 <Table 2>에 요약하였다. 모형의 설명력을 나타내는 R-Sq(adj)는 95.8%로서 모형의 적합도가 상당히 높음을 알 수 있다. 이때 MSE는 0.055이고 F-statistics는 524.34로써 본 논문에서 사용한 데이터의 정확도를 고려할 때, 이는 매우 높은 수준으로 판단된다.

학습곡선 모형을 통한 예측 결합밀도와 실제 결합밀도는 <Figure 3>에 표시하였다. 그림에서 예측 결합밀도는 실제를 잘 대변하고 있음을 알 수 있다. 식 (5)의 Negative Binomial 수율 모델을 이용하여 예측 결합밀도와 실제 결합밀도로부터 추정된 수율의 비교는 <Figure 4>이다. 여기서, Y_0 는 0.8, A_c 는 5, α 는 3을 임의로 적용하였다.

Gruber(1992)가 언급하였듯이 반도체 산업의 학습효과는 제품 간 변화보다 생산 라인 간 변화가 더욱 크다. 본 논문에서 적용한 데이터는 생산라인에 관한 데이터가 아니고 제품별 데이

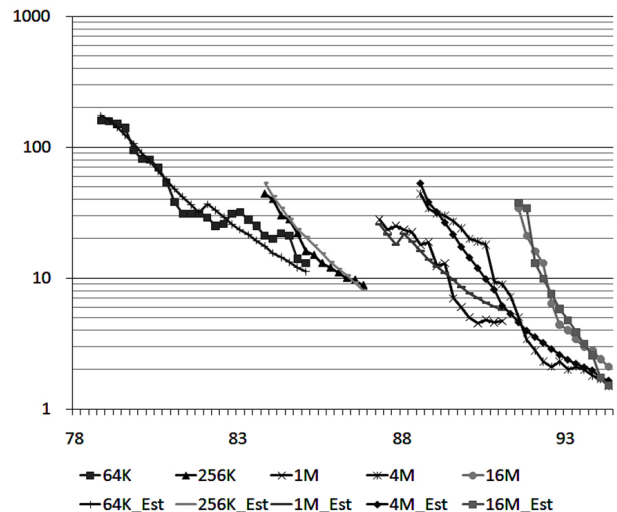


Figure 3. 실제 결합밀도와 예측 결합밀도의 비교

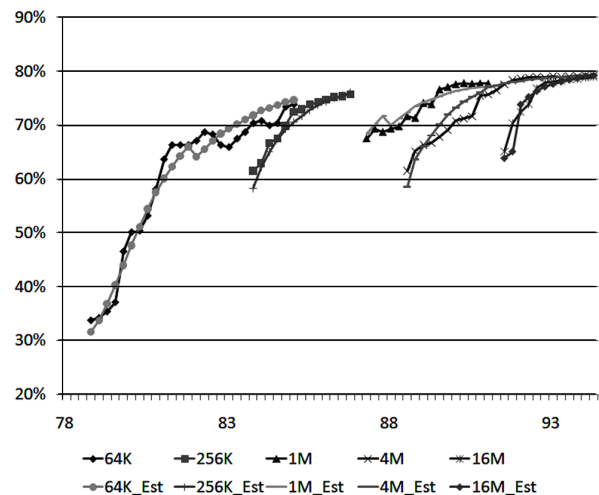


Figure 4. 실제 결합밀도와 예측 결합밀도로부터 추정된 수율의 비교

터이므로 생산라인 간 변화를 관측할 수 없는 단점이 존재한다. 또한, 제시한 검증결과는 데이터 수집의 어려움으로 인하여 실제 생산현장의 데이터를 사용한 것이 아니므로 분명히 한계가 존재한다. 그러나 본 논문의 목적이 정확한 파라미터의 추정이 아니고 반도체 산업에 적합한 새로운 학습곡선 모형의 제안과 이의 적용 방법이라는 점에서는 만족할 만한 수준의 결과라고 판단된다.

5. 결론

본 연구를 통하여 기존의 학습곡선 모형을 검토하고 이를 바탕으로 반도체에 적합한 새로운 학습곡선 모형을 제시하고 이를 검증하였다. 다중회귀분석 결과 본 논문에서 제안한 다단계 학습곡선 모형의 R_Sq(adj)가 95.8%로서 적합도가 매우 우수한 수

준으로 나타났다. 그러나 실제 생산현장의 데이터 수집의 한계로 인하여 기존의 연구와의 직접적인 비교 검토가 충분히 이루어지지 않은 점이 아쉬움으로 남는다. 반도체에 학습곡선 모델을 적용함에 있어서 제 4.1절에 제시된 바와 같이 다양한 독립변수 및 종속변수의 적용이 가능하며, 각 기업에 따라 다른 선택과 다른 결과를 초래할 수 있으므로 본 연구결과의 적용에 있어서는 실증적으로 검증하는 연구가 추가되어야 한다.

참고문헌

- Alchain, A. (1959), Cost and output, *Allocation of Economic Resources*, Abramoviz, M. (ed.), Stanford University Press, Stanford, 23-40.
- Badiru, A. B. (1992), Computational survey of univariate and multivariate learning curve models, *IEEE Transactions on Engineering Management*, 39(2), 176-188.
- Chou, S. K. (1933), What Is the Curve of Forgetting?, *The American Journal of Psychology*, 45(2), 348-350.
- Chung, S. (2001), The learning curve and the yield factor : the case of Korea's semiconductor industry, *Applied Economics*, 33, 473-483.
- Cunningham, J. A. (1990), The use of evaluation of yield models in integrated circuit manufacturing, *IEEE Transactions on Semiconductor Manufacturing*, 3(2), 60-71.
- Cunningham, S. P., Spanos, C. J., and Voros, K. (1995), Semiconductor yield improvement results and best practices, *IEEE Transactions on Semiconductor Manufacturing*, 8(2), 103-109.
- Gruber, H. (1992), The learning curve in the production of semiconductor memory chips, *Applied Economics*, 24, 885-894.
- Gruber, H. (1994), The yield factor and the learning curve in semiconductor production, *Applied Economics*, 26, 837-843.
- Hatch, N. W. and Mowery, D. C. (1998), Process innovation and learning by doing in semiconductor manufacturing, *Management Science*, 44(11), 1461-1477.
- IC Knowledge (2008), Technology Trends, www.icknowledge.com.
- ICE (1997), Yield and Yield Management, Chapter 3 in *Cost Effective IC Manufacturing 1998-1999*, Integrated Circuit Engineering Corporation.
- Kara, S. and Kayis, B. (2005), The effect of the learning process in concurrent engineering projects, *Concurrent Engineering : Research and Applications*, 13(3), 209-217.
- Lieberman, M. B. (1984), The learning curve pricing in the chemical industries, *Rand Journal of Economics*, 15, 213-228.
- Victor, N. M. and Ausubel, J. H. (2002), DRAMs as Model Organisms for Study of Technological Evolution, *Technological Forecasting and Social Change*, 69(3), 243-262.
- Weber, C. (2004), Yield learning and the source of profitability in semiconductor manufacturing and process development, *IEEE Transactions on Semiconductor Manufacturing*, 17(4), 590-596.
- Wright, T. P. (1936), Factors affecting the cost of airplanes, *Journal of Aeronautical Science*, 3(2), 122-128.



하정훈

연세대학교 전자공학과 학사
 Texas A&M University 산업공학 석사
 Texas A&M University 산업공학 박사
 현재 : 홍익대학교 정보컴퓨터공학부 조교수
 관심분야 : 최적화, 생산경영, SCM