

역전파 신경망 공정 모델의 평가지표로서의 유클리디언 웨이트 거리

김 병 환  
세종대학교, 전자공학과

Euclidean Weight Distance as a Performance Measure for Backpropagation Neural Network Process Model

Byungwhan Kim  
Sejong University, Department of Electronic Engineering

**Abstract** - 역전파 신경망은 반도체 공정 모델링에 효과적으로 응용이 되고 있으며, 최근 선형뉴런을 비선형 함수 대신 출력층에 이용하여 모델의 예측정확도를 향상 시킨 바 있다. 본 연구에서는 그 원인을 규명하기 위한 모델의 평가지표로서의 유클리디언 웨이트 거리 (Euclidean Weight Distance)를 제안한다. 이 지표를 이용하여 신경망의 입력층과 은닉층, 그리고 은닉층과 출력층의 웨이트를 감시하였으며, 그 결과 예측정확도의 향상이 이 지표의 감소에 기인하고 있음을 알았다. 모델링에 이용한 실험데이터는 다중 유도결합형 플라즈마 장비로부터 Langmuir Probe 진단 시스템을 이용하여 수집하였다.

유클리디언 웨이트 거리를 이용했다. 초기웨이트의 불규칙성은 동일 학습 조건에서 다중 모델을 발생하는 것으로 구현하였다. 활성화함수의 종류에 따른 모델 특성의 차이를 유클리디언 웨이트 거리의 양과 비교하여 본 평가지표의 활용 가능성을 보이고자 한다. 모델링에 이용한 데이터는 반도체 소자제조에 이용되고 있는 다중 유도결합형 플라즈마 장비로부터 추출하였다. 플라즈마 특성 모델링을 위해 통계적인 실험계획표가 이용되었으며, 변화시킨 인자로는 소스 전력, 압력, 그리고 아르곤 가스의 양이다.

2. 본 론

1. 서 론

반도체 공정, 특히 플라즈마를 이용한 증착과 식각공정은 소자 제조를 위한 박막의 증착과 미세 패턴닝에 응용되고 있다. 공정개발에 소요되는 시간과 경비를 최소화하기 위해서는 공정 시뮬레이션이 요구되고 있으며, 이는 다시 정확한 예측 모델의 개발을 요구하고 있다. 지난 수십년 간에 걸쳐서 플라즈마 공정 모델링은 주로 고도의 물리적인 지식에 기초하여 개발이 되어오고 있다 [1]. 하지만, 해석적인 모델 정립과정에는 다수의 가정 (Hypothesis)이 전제되어 있어, 시뮬레이션 결과의 신뢰성이 자연 저하된다. 또한 시뮬레이션에 소요되는 시간이 상당하고, 공정변수 내지 장비 하드웨어의 변화가 있을 때, 이를 수용하기 위한 기존 프로그램 코드의 변경 내지 수정이 용이치 않다. 이러한 이유로, 실제 실험 데이터를 이용하기 위한 통계적인 반응표면 모델 [2]이 개발된 바 있다. 90년도 초반부터는 이 보다 예측효과가 더 우수한 신경망에 대한 많은 관심과 그 응용이 발표가 되고 있다 [3-6]. 주로 역전파 신경망 [7]이 이용이 되고 있으며, 최근 다른 신경망 구조의 응용이 발표된 바 있다 [8-9].

2.1 실험데이터 수집

본 연구에서 이용된 플라즈마 장비의 개략적인 모습이 그림 1에 나타나 있다. 플라즈마는 장비 상단 부에 위치하고 있는 6개의 페라이트 막대자석을 에워싸는 코일에 라디오 (Radio) 주파수의 전력을 인가함으로써 발생한다. 이후, Langmuir Probe를 장비 안으로 반경 방향으로 삼입하여 플라즈마 특성을 측정하였으며, 여기에는 전자(이온) 밀도, 전자온도, 그리고 플라즈마 전위 등이 있다.

역전파 신경망을 플라즈마 공정 모델링에 이용할 때, 그 모델의 예측정확도는 다 수의 학습인자에 의해 영향을 받는다. 이러한 인자에는 은닉층의 뉴런수, 초기웨이트의 분포범위, 그리고 활성화 함수의 종류와 그 경사 등이 포함된다. 이 들 인자들은 모델에 상당한 영향을 미치고 있으며, 따라서 이들을 동시에 최적화하기 위한 연구가 발표된 바 있다 [10-11]. 학습인자 중 최적화하기가 가장 어려운 인자는 초기웨이트 이며, 이는 모델의 성능이 초기웨이트에 따라 불규칙하게 변화하기 때문이다. 이는 다시 학습인자의 변화에 따른 모델의 성능평가를 어렵게 하는 주원인이 되고 있다. 초기웨이트의 영향을 해석하고, 나아가서 최적의 초기 웨이트를 설정하기 위한 많은 연구가 진행이 되고 있으나 [12-14], 아직 보편적으로 적용할 수 있는 평가지표가 발표되어 있지 않다. 사실 다양한 초기웨이트에 따른 모델성능을 해석적인 식으로 표현하는 것은 현실적으로 불가능하다고 사료된다.

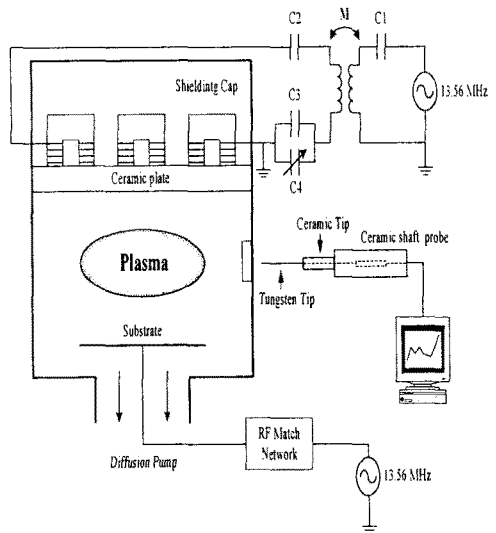


그림1: 다중 유도결합형 플라즈마 장비.

총 35개의 측정치들을 수집하였고, 이를 다시 평균하여 그 평균치를 신경망의 응답 변수로 이용하였다. 플라즈마는 공정 변수의 변화에 민감하며, 따라서 이들과 특성 인자간의 체계적인 관계 모델링을 위해 3<sup>3</sup> 인자 실험계획표 [15]이 이용되었다. 실험계획표에서 변화시킨 인자로는, 소스전력, 압력, 그리고 아르곤 가스의 양이며,

본 연구에서는 해석적 평가지표의 설정이 어렵다는 인식 하에, 신경망의 웨이트를 감시하였으며, 그 도구로

실험에 이용된 범위는 표 1에 나타나 있다.

표1: 실험 변수와 범위

공정 변수	실험 범위	단위
RF 전력	500-1000-1500	Watts
압력	2-6-10	mTorr
아르곤 가스	50-75-100	sccm

실험계획으로부터 얻어지는 27개의 데이터를 이용하여 신경망을 학습시켰다. 신경망의 학습정도를 테스트하기 위해 추가적으로 16번의 실험을 수행하여, 본 연구에서는 총 43개의 실험을 수행하였다.

### 2.2 평가지표로서의 유클리디언 웨이트 거리

신경망을 이용한 공정 모델링은 학습단계와 테스트 단계로 이루어진다. 학습을 위한 학습데이터는 전술한 통계적인 실험방식에 의해 수집하였으며, 학습된 모델의 적합성은 별도의 테스트 데이터를 만들어 평가를 한다. 학습과정에는 여러 학습인자가 관여하며, 여기에는 학습 허용도 (Tolerance), 은닉층의 뉴런수, 활성화 함수의 종류와 경사, 그리고 초기웨이트의 분포 등이 포함된다. 이러한 인자들은 개별적으로 변화시키거나, 또는 통계적인 실험방식과 최적화 기법을 연계하여 최적화하기도 한다 [10-11]. 어느 경우의 최적화든, 그 최적화의 어려움은 초기웨이트의 불규칙적인 분포에 기인한다. 초기웨이트는 일반적으로  $\{+1, -1\}$  사이에 분포되어 있으며, 그 불규칙적인 분포에 따라 각기 다른 모델성능을 보이고 있어 모델의 성능평가를 매우 어렵게 하고 있다. 임의의 하나의 고정된 웨이트에 대해서 다른 학습인자를 최적화하여 모델을 개발하기도 하지만, 이 경우 모델의 예측정확도가 크게 제약을 받게 된다. 다중 모델을 발생하여 그 평균적인 모델을 이용할 수도 있지만, 이 보다 발생한 다중 모델중 그 예측효과가 가장 우수한 모델을 추출하여 공정 모델링에 활용하는 것이 더 실제적인 방법이라 할 수 있다 [11].

최근, 신경망 뉴런의 활성화 함수에 따른 플라즈마 모델링에의 영향을 불규칙적인 초기웨이트하에서 연구한 보고가 있었다 [16]. 이 연구에서는, 종래의 비 선형적인 시그모이드 함수에 비해 선형함수를 출력층에 이용할 경우 그 예측 정확도가 크게 향상한다는 것을 실험적으로 보여 주었으며, 초기웨이트의 불규칙성을 수용하기 위하여 주어진 함수 종류에 대해서 다중 모델을 발생하였다. 그러나, 당시의 연구에서는 선형 함수의 사용이 어떻게 정확도의 향상에 기여를 했는지 알 수 없었으며, 본 연구에서는 이를 고찰하고자 다음과 같이 유클리디언 웨이트 거리 (W)를 정의한다.

$$W = \sqrt{\sum_{i,j} (w_{ij,b} - w_{ij,i})^2}$$

여기서  $w_{ij,b}$ 은 다중 모델 중 결정된 최적의 모델에 대한 웨이트, 그리고  $w_{ij,i}$ 은 첫 번째 모델에 대한 초기웨이트를 각기 나타낸다. 이 평가지표를 이용하여 신경망의 입력층 (I)과 은닉층 (H) 그리고 은닉층과 출력층 (O)의 웨이트를 감시하는 데에 이용하였다.

### 2.3 모델링과 평가

본 연구에서는 여러 학습인자 중, 활성화 함수의 종류에 따른 모델성능만을 비교 평가하며, 다른 학습인자는 일정한 값에 고정하였다. 구체적으로 학습허용도, 은닉층의 뉴런수, 그리고 함수의 경사는 각기 0.1, 4, 1로 고정시켰다. 초기웨이트의 분포범위도  $\{+1, -1\}$ 로 제한하였다. 두 종류의 함수 유형, 즉 B/B와 B/L,에 대해서 모델성능을 평가하였으며, 여기서 "B"는 바이폴라

시그모이드 함수, 그리고 "L"은 선형함수를 지칭하고, "/"앞과 뒤의 함수는 각기 은닉층과 출력층의 위치한 함수를 지칭한다. 초기웨이트의 불규칙성을 고려하여 총 200개의 모델을 연속적으로 발생하였으며, 이 중 최적의 모델, 즉 테스트 데이터에 대한 예측 에러가 최소인 모델을 추출하여 표2에 정리하였다. 표 2에서와 같이 B/L의 함수 조합을 가진 신경망이 B/B 조합을 가진 신경망에 비해 예측정확도가 향상되었으며, 구체적으로 전자밀도, 전자온도, 그리고 플라즈마 전위에 대해서 각 17.2%, 10.8%, 그리고 36.6%의 향상된 성능을 보이고 있다. 표 2의 각 데이터는 RMSE라는 매트릭에 의해 계산된 양이며, 이는 다음과 같이 정의된다.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1}}$$

여기서  $y_i$ 와  $\hat{y}_i$ 는 실제 측정치와 모델로부터의 예측치를 지칭하고, n은 테스트 데이터의 수이다. 본 연구에서의 n은 16이다.

플라즈마 응답	B/B-신경망	B/L-신경망
전자밀도	0.609	0.504
전자온도	0.423	0.377
플라즈마 전위	2.578	1.632

표2: 함수조합에 따른 모델 예측 에러

표 2의 최적의 모델에 대한 학습후의 웨이트와 초기웨이트를 이용해 앞서 정의한 W를 계산하였으며, 그 결과가 그림 2와 3에 나타나 있다. 그림 2는 입력층과 은닉층에 대한 W이며, 그림 3은 은닉층과 출력층에 대한 W이다. 그림2에서와 같이 B/L 조합의 신경망은 B/B 조합의 신경망에 비해 감소된 W를 보이고 있다. 동일한 결과가 그림 3에 나타나 있고, 그림 2에 비교해 볼 때, 더 현격한 감소를 보이고 있다.

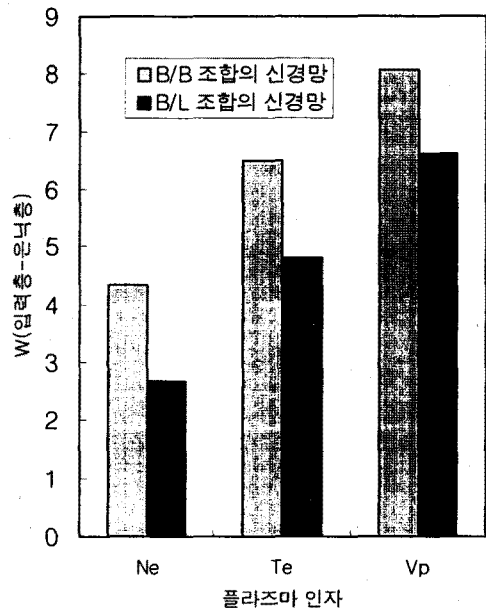


그림 2: 입력층과 은닉층에 대한 계산된 W.

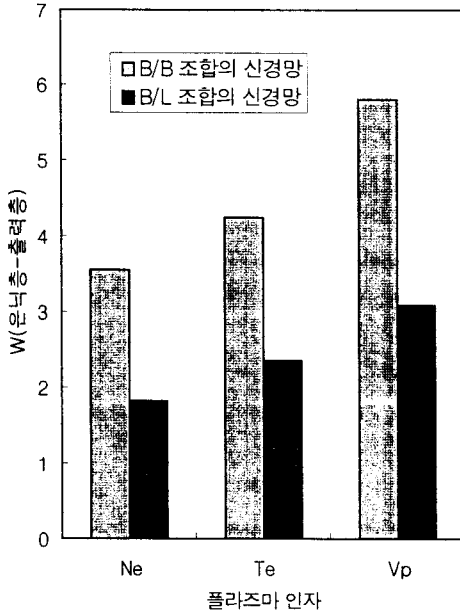


그림3: 은닉층과 출력층에 대한 계산된 W.

이는 은닉층과 출력층의 W가 함수 조합에 따른 모델 성능 변화에 더 민감하다는 것을 암시한다. 이상의 결과로부터 선형함수의 사용에 따른 성능향상은 W의 감소에 기인하고 있음을 알 수 있었다.

### 3. 결 론

신경망 성능 평가지표로서의 유클리디언 웨이트거리 (W)를 제안하였다. 이 지표를 이용하여 활성화 함수의 종류에 따른 모델 예측능력을 평가하였다. 그 결과 출력층에서의 선형함수 이용에 따른 예측성능의 향상이 W의 감소에 기인 하다는 것을 알 수 있으며, 이러한 현상은 입력층과 은닉층, 그리고 은닉층과 출력층의 웨이트에서 모두 관찰되었다. 현재 다른 학습인자들에 대한 이 지표의 응용가능성을 연구 중에 있다.

### (참 고 문 헌)

[1] R. A. Stewart, P. Vitello, and D. B. Graves, Two-dimensional fluid model of high density inductively coupled plasma sources, *J. Vac. Sci. Technol., B*, vol. 12, no. 1, pp. 478-485, 1994.

[2] G. S. May, J. Hung, and C. J. Spanos, Statistical experimental design in plasma etch modeling, *IEEE Trans. Semicond. Manufact.*, vol. 4, pp. 83-98, 1991.

[3] B. Kim and G. S. May, Reactive ion etch modeling using neural networks and simulated annealing, *IEEE Trans. Comp. Pack. and Manufact. Technol., C*, vol. 19, no. 1, pp. 3-8, 1996.

[4] B. Kim, K. H. Kwon, and S. H. Park, Characterizing metal-masked silica etch process in a CHF<sub>3</sub>/CF<sub>4</sub> inductively coupled plasma, *J. Vac. Sci. and Technol., A*, vol. 17, no. 5, pp. 2593-2597, 1999.

[5] B. Kim, J. Sun, C. Choi, D. Lee, and Y. Seol, Use of neural networks to model low temperature tungsten etch characteristics in high density SF<sub>6</sub> plasma, *J. Vac. Sci. and Technol., A*, vol. 18, no. 2, pp. 417-422, 2000.

[6] B. Kim and G. T. Park, Modeling equipment plasma using neural networks, *IEEE Plasma Sci.*, vol. 29, no.

2, 2001.

[7] D. E. Rummelhart and J. L. McClelland, *Parallel Distributed Processing*, M.I.T. Press, Cambridge, 1986.

[8] 박성진, 김병환, "반도체 공정 최적화를 위한 일반화된 희귀 신경망 플라즈마 모델", *전기학회 하계 학술회의*, pp. 2744-2746, 2000.

[9] B. Kim and S. Park, "Modeling of process plasma using a radial basis function network: a case study", *Trans. Cont. Autom. Syst. Eng.*, pp. 268-273, 2000.

[10] B. Kim and G. S. May, An optimal neural network process model for plasma etching, *IEEE Trans. Semicond. Manufact.*, vol. 7, no. 1, pp. 12-21, 1994.

[11] B. Kim, S. Park, An optimal neural network plasma model: a case study, accepted for publication to *Chemom. Intell. Lab. Syst.*, vol. 56, no.1, pp. 39-50, 2001.

[12] D. Sarkar, Randomness in generalization ability: A source to improve it, *IEEE Trans. Neural Networks* vol. 7, no. 3, pp. 676-685, 1996.

[13] P. Bartlett, The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network, *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 525-536, 1998.

[14] A. Atiya and C. Ji, How initial conditions affect generalization performance in large networks, *IEEE Trans. Neural Networks*, vol. 8, no. 2, pp. 448-451, 1997.

[15] Montgomery, *Design and Analysis of Experiments*, John Wiley & Sons, New York, 1991.

[16] 김병환, 박성진, 뉴런 활성화 경사최적화를 이용한 개선된 플라즈마 모델, *한국 자동제어 학술회의*, 2000.