

난수발생기를 이용한 뉴런경사 제어와 플라즈마 식각공정 데이터 모델링에의 응용

김성모, 김병환
 세종대학교 전자공학과

Neuron gradient control by random generator and application to modeling a plasma etch process data

Sungmo Kim, Byungwhan Kim
 Sejong University, Electronic Engineering

Abstract - 역전파 신경망 (BPNN)은 반도체 공정 모델링에 효과적으로 응용되고 있다. 뉴런의 활성화 함수는 동일한 값을 가지며, 이로 인해 예측정확도를 증진하는 데에는 한계가 있었다. 본 연구에서는 난수발생기(Random generator-RG)를 이용하여 뉴런 경사들이 다중값을 가지도록 최적화하였다. 본 기법은 은닉층의 뉴런수의 함수로 고찰하였으며, 종래의 고정된 경사를 갖는 모델과 그 성능을 비교 평가하였다. 평가에 이용된 데이터는 플라즈마 식각 공정데이터이며, 모델에 이용된 응답은 식각률과 프로파일 각이다. 비교결과 종래의 모델에 비해 예측정확도가, 식각률의 경우 19%-43%, 프로파일의 경우 10%-56% 정도 향상하였으며, 이는 제안된 기법이 모델개발에 매우 효과적으로 적용될 수 있음을 보여준다.

인 데이터 수집을 위해, 통계적인 실험계획법이 이용되었다. 식각응답의 측정에는 scanning electron microscopy (SEM)이 이용되었다.

2. 본 론

2.1 실험데이터 수집

식각공정은 유도결합형 플라즈마를 이용한 식각장비에서 수행이 되었으며, 장비와 테스트 패턴 제작공정은 기존 발표된 보고에서 이용된 것과 동일하다 [2]. 단 Ni-마스크 층을 이용한 것과, 플라즈마 유형이 NF₃/CH₄라는 점에서 이전의 연구와 상이하다. 식각에 이용된 공정 변수는 4종류이며, 여기에는 소스전력, 바이어스 전력, 압력, 그리고 가스비가 포함이 된다. 통계적인 실험계획 표 작성에 이용된 각 변수별 실험범위는 표 1에 나타나 있다.

1. 서 론

반도체 공정, 특히 플라즈마를 이용한 증착과 식각공정은 소자 제조를 위한 박막의 증착과 미세 패턴닝에 응용되고 있다. 또한 공정 모델은 경제적인 공정개발, 즉 공정 최적화에 소요되는 시간과 경비를 줄이기 위해 절실히 요구되고 있다. 반도체 공정 모델은 주로 물리적 모델에 의해 개발되어오고 있으나, 계산에 소요되는 시간이 엄청나고, 또한 모델개발 과정에 수반되는 사정들의 존재로 그 예측의 신뢰도가 높지 않다는 단점이 있다. 이에 반해, 신경망은 실험데이터를 기초로 용이하게 개발될 수 있으며 그 예측정확도는 매우 우수해 공정 최적화에 효과적으로 응용할 수 있다. 신경망중 역전파 신경망(BPNN)이 가장 많이 응용이 되었으며, 플라즈마 방전[1]과 식각공정의 모델링 [2]에 효과적으로 응용된 바가 있다. 역전파 신경망 모델의 예측성능은 다수의 학습인자에 의해 결정이 되며, 여기에는 학습허용도, 초기 웨이트의 크기, 활성화함수의 종류와 경사, 그리고 은닉층의 뉴런수가 포함된다. 대부분의 응용에 있어 마지막 인자인 뉴런수를 조정하는데 그치고 있으며, 인자 상호간의 영향을 고려한 모델 개발 기법이 발표된 바 있다 [3-4]. 종래의 모델에서, 뉴런 활성화함수 경사는 고정된 값 (대개 1로 고정)을 이용하고 있다. 모델성능의 경사에서의 의존도는 공정데이터 모델링중에 고찰한 바 있으며, 상당한 영향을 주고 있음이 발표된 바 있다 [2]. 이 연구에서는, 경사값을 일정한 범위에서 실험적으로 변화시켜 그 영향을 최적화하였다. 최적화한 모델의 경사값은 모든 뉴런에 대해서 동일한 값을 가지게 된다. 한편, 뉴런 경사값이 고정된 값이 아닌, 다중값을 갖게 결정할 경우 예측정확도의 향상을 기대할 수 있으며, 이 같은 연구사례는 아직 보고된 바가 없다.

본 연구에서, RG를 이용하여 활성화 함수경사값을 최적화하는 기법을 제안한다. 제안된 기법은 난수발생 범위와 뉴런수의 인자로 고찰하였다. 기법의 우수성을 확인하기 위해 종래의 모델과 비교하였다. 평가에 이용된 데이터는 silicon carbide (SiC)의 플라즈마 식각공정을 통해 수집한 식각률과 프로파일 각이다. 체계적

표1. 실험변수와 범위

| 공정변수 | 범위 | 단위 |
|---------|---------|-------|
| 소스전력 | 700-900 | Watts |
| 바이어스 전력 | 50-150 | Watts |
| 압력 | 6-12 | mTorr |
| 가스비 | 0.2-1.0 | |

표 1에 2⁴ 전인자 실험계획을 적용하였으며, 그 결과 수집된 16개의 실험과 중심점에 해당하는 실험을 포함한 17번의 실험이 모델개발을 위한 학습인자로 이용되었다. 모델예측능력을 평가하기 위한 테스트 데이터는 총 15개로 구성이 되었으며, 따라서 모델개발에 이용된 전체 실험수는 32개이었다.

2.2 역전파 신경망

BPNN은 플라즈마 공정데이터 해석에 가장 많이 응용되고 있는 구조이다. BPNN은 주로 은닉층이 하나인 3개의 층으로 구성되며, 학습은 일반화된 델타 규칙 [6]을 이용하여 진행이 된다. 신경망을 이용한 공정 모델링은 학습단계와 테스트 단계로 이루어진다. 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 실험 방식과 최적화 기법을 연계하여 최적화하기도 한다. 뉴런 활성화 함수는 다음과 같이 표현된다.

$$f(x) = \frac{1 - e^{-x/S}}{1 + e^{-x/S}} \quad (1)$$

식 (1)에서 S는 함수의 경사를 의미하며, 종래의 모델에서는 고정된 값, 즉 1.0에 고정하고 있다. 전술한 바와 같이, 경사값의 변화에 따라 예측성능은 상당히 변할 수 있으며, 이는 실제적으로 확인된 바 있다 [2]. 이 경

우, 경사값을 일정한 범위에서 실험적으로 변화시켜 최적화하였으며, 예측정확도를 크게 증진할 수 있었다. 한편, 최적화된 경사는 모든 뉴런에 대해서 동일한 값을 가지고 있어, 예측정확도를 더 증진하는데에는 한계가 있었다. 경사값을 다중값을 갖게 할 경우, 예측정확도의 증진을 기대할 수 있으며, 이를 본 연구에서 고찰하고자 한다.

2.3 모델링과 평가

본 연구에서는 RG를 이용하여 모델예측성능이 뉴런 경사가 단일 값이 아닌 다중값에서 최적화가 되게 하는 기법을 제안한다. 출력층의 뉴런은 선형함수로 구성이 되었으며, 그 경사는 1로 고정하였다. 은닉층의 뉴런에 대해서만 본 기법을 적용한다. 난수범위는 $\pm 0.2 \sim \pm 2.0$ 까지 0.2 간격으로 발생하였으며, 뉴런수는 2-6까지 변화시켰다. 주어진 난수 범위와 뉴런수에 대해서 200개의 모델을 발생하였다. 각 모델에 대해서 초기웨이트의 값은 동일하게 설정하였다. 본 기법은 종래의 모델기법과 비교하여, 그 우수성을 확인한다.

2.3 모델 성능별 분석

2.3.1 고정경사 모델

그림 1은 종래의 모델기법을 이용하여 개발한 식각률 모델의 예측정확도를 경사와 뉴런수의 함수로 도시하고 있다. 주어진 뉴런수에 대해서 초기웨이트의 값은 동일하다. 주어진 뉴런수와 경사값에 대해서 발생한 모델의 수는 1이다. 뉴런수가 2인 경우 경사가 0.2에서 0.4로 변화할 때, 예측정확도는 급격히 감소하고 있으며, 경사값을 더 증가시켰을 때, 예측정확도의 변화는 미미하다. 이는 뉴런수가 5와 6인 경우에도 나타난다. 뉴런수가 3과 4인 경우, 경사값에 따라, 예측정확도가 불규칙적으로 변화하고 있다. 일반적으로 작은 경사값에 대해서 예측정확도의 변화가 컸다. 그림 1에서 알 수 있듯이, 뉴런수 2에서, 경사값이 1.8일 때 가장 좋은 모델을 결정하였으며, 그 RMSE는 41.0 nm/min이었다. 여기서 RMSE는 root mean-squared error의 약자이다.

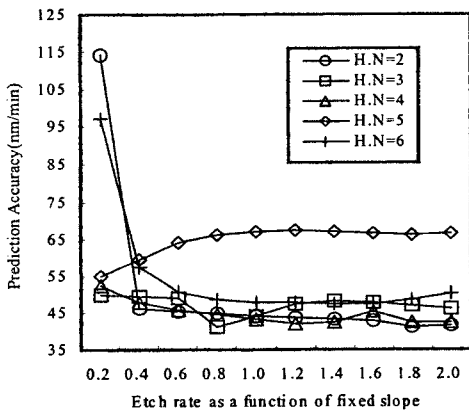


그림 1. 예측정확도의 경사와 뉴런수에 따른 변화

그림 2는 종래의 모델기법을 이용하여 개발한 프로파일 모델의 예측정확도를 경사와 뉴런수의 함수로 도시하고 있다. 그림 2에서와 같이, 뉴런수 2와 3인 경우 경사값의 변화에 따른 그 경향이 비슷하다. 뉴런수 4,5,6에 대한 경향도 서로 비슷하며, 작은 경사값에 대해서 예측정확도의 변화가 크고, 경사값의 증가에 따라 감소하고 있다. 그림 2에서와 같이, 뉴런수 2와 경사값 0.6에서 최적의 모델을 결정하였으며, 그 RMSE는 2.78°이다.

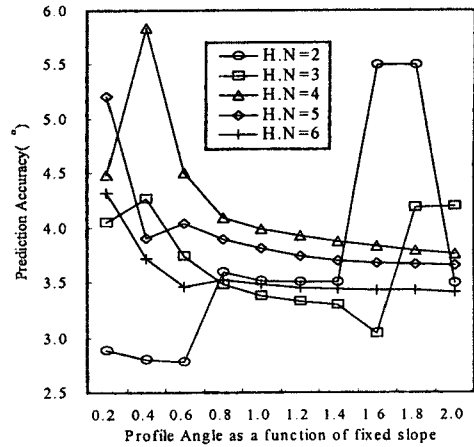


그림 2. 예측정확도의 경사와 뉴런수에 따른 변화

2.3.2 RG-기초 다중경사 모델

그림 3은 RG를 이용하여 개발한 식각률 모델의 예측정확도를 난수의범위와 뉴런수의 함수로 도시하고 있다. 그림 3에 도시된 각 데이터 요소는 주어진 난수와 뉴런수에 대해서 200개의 모델을 발생하고, 그 중 가장 우수한 모델의 RMSE를 나타내고 있다. 그림1의 고정경사 모델에 비해 그림3의 다중경사 모델은 뉴런수와 난수범위에 대해 훨씬 복잡하게 변화하고 있다. 경사값의 변화에 따른 예측정확도는 모든 뉴런수에 대해서 불규칙적이다. 그림1과 비슷한 점은 작은 경사값에서 예측정확도의 변화가 비교적 크다는 점이다. 그림에서와 같이 뉴런수 2, 그리고 난수범위 ± 1.6 에서 최적의 모델을 얻을 수 있었으며, 그 RMSE는 29.3 nm/min이고, 결정된 경사는 0.16과 1.38이다. 그림 1에서 결정된 모델과 비교할 때 28.5%정도 예측정확도가 향상되었다.

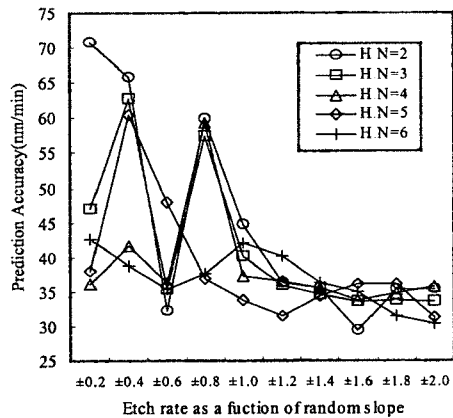


그림 3. 예측정확도의 난수범위와 뉴런수에 따른 변화

그림 4는 RG를 이용하여 개발한 프로파일 모델의 예측정확도를 난수의범위와 뉴런수의 함수로 도시하고 있다. 그림 2의 고정 경사모델에 비해, 다중경사 모델이 뉴런수와 난수범위의 변화에 따라 더 복잡하게 변화하고 있다. 이는 식각률 모델링에서 이미 관측된 바 있다. 일반적으로 뉴런수 2의 경우를 제외하고는, 경사값이 큰 범위에서 경사변화에 따른 예측정확도의 변화가 상대적으로 크다. 그림에서와 같이, 뉴런수 5 그리고 난수범위 ± 1.8 에서 최적의 모델이 발생하였으며, 그 RMSE는

1.59°이다. 결정된 다중 경사값은 1.12 -1.07 0.83 -1.66 1.21이다. 한편, 그림2에서 결정된 모델과 비교할 때 예측정확도는 42.8% 증진되었다. 결국 다중 경사모델이 예측정확도를 향상하는데 더 효과적임을 알 수 있었다.

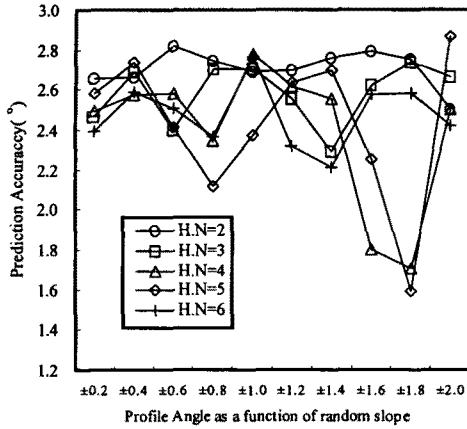


그림 4. 예측정확도의 난수범위와 뉴런수에 따른 변화

2.3.3 고정경사와 다중경사 모델의 비교

제안된 기법의 우수성을 입증하기 위해, 각 뉴런에 대해서 결정된 최적의 모델을 상호 비교한다. 그림 5는 식각률의 경우, 그 비교를 보이고 있다. 그림 5에서와 같이 모든 뉴런수에 대해서 제안된 모델의 RMSE는 작으며, 가장 큰 향상은 뉴런수 5에서 얻었으며, 이는 수치적으로 43.2%이다.

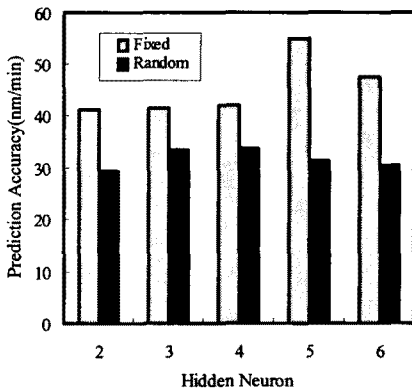


그림 5. 뉴런수에 따른 식각률 모델의 비교

식각 프로파일에 대한 비교는 그림 6에 도시되어 있으며, 식각률에서와 같이 모든 뉴런수에 대해서 다중경사 모델이 더 작은 RMSE를 보이고 있다. 특히, 그 향상도는 뉴런수 4와 5에서 상당하며, 이는 각각 54.6%과 56.5%에 해당한다. 경사와 뉴런수가 같이 변할 때 얻는 최적의 모델간의 비교는 이전에서 행한 바 있으며, 그 수치는 식각률과 프로파일의 경우, 각기 28.5%와 42.8%에 해당한다. 이 두가지 비교에서 RG-기초 다중경사 모델이 고정경사 모델에 비해 더 우수한 모델임을 입증하였다.

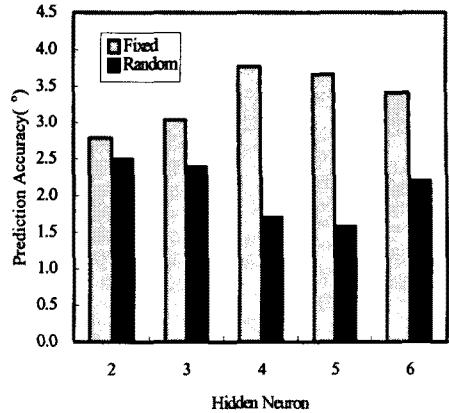


그림 6. 뉴런수에 따른 프로파일 모델의 비교

3. 결 론

RG를 이용하여 다중경사값을 갖는 BPNN 모델을 개발하였다. BPNN의 성능을 난수의 범위와 뉴런수의 함수로 고찰하였으며, 결정된 모델은 종래의 고정경사모델과 비교하였다. 두 종류의 식각공정데이터에 적용하여 비교한 결과, 다중경사 모델이 고정 경사모델에 비해 더 우수한 예측성능을 보였다. 이는 제안된 기법이 BPNN의 예측성능을 크게 증진하는데 기여하고 있음을 의미한다.

감사의 글

본 연구는 IMT 2000 연구비에 의해 지원 받았으며, 한국보건산업진흥원에게 감사사를 드립니다.

[참 고 문 헌]

- [1] B. Kim and G. T. Park, "Modeling plasma equipment using neural networks," *IEEE Trans. Plasma Sci.*, vol. 29, no. 1, 8-12, 2001.
- [2] B. Kim, S. M. Kong, and B. T. Lee, "Modeling SiC etching in C₂F₆/O₂ inductively coupled plasma etching," *J. Vac. Sci. Technol. A*, vol. 20, no. 1, 146-152, 2002.
- [3] B. Kim and G. S. May, "An optimal neural network process model for plasma etching," *IEEE Trans. Semicond. Manufact.*, vol. 7, 12-21, 1994.
- [4] B. Kim and S. Park, "An optimal neural network plasma model: a case study," *Chemom. Intell. Lab. Syst.* vol. 65, 39-50, 2001.
- [5] B. Kim, J. H. Sun, C. J. Choi, D. D. Lee, and Y. S. S., "Use of neural networks to model low temperature tungsten etch characteristics in SF₆ plasma," *J. Vac. Sci. Technol. A*, vol. 18, no.2, 417-422, 2000.
- [6] D. E. Rummelhart and J. L. McClelland, *Parallel Distributed Processing*, Cambridge, M.I.T. Press, 1986.