

On-Chip 학습기능을 가진 확률연산 펄스형 디지털 신경망의 구현

위재우, 이종호  
인하대학교 전기공학과

Implementation of A Pulse-mode Digital Neural Network with On-chip Learning Using Stochastic Computation

Jae-Woo Wee, Chong-Ho Lee  
Department of Electrical Engineering, Inha University

**Abstract** - In this paper, an on-chip learning pulse-mode digital neural network with a massively parallel yet compact and flexible network architecture is suggested. Algebraic neural operations are replaced by stochastic processes using pseudo-random sequences and simple logic gates are used as basic computing elements. Using Back-propagation algorithm both feed-forward and learning phases are efficiently implemented with simple logical gates. RNG architecture using LFSR and barrel shifter are adopted to avoid some correlation between pulse trains. Suggested network is designed in digital circuit and its performance is verified by computer simulation.

1. 서 론

실시간 처리가 요구되는 응용분야에 신경회로망을 적용하기 위해서는 신경회로망의 하드웨어 구현이 필요하다. 현재 급속히 발전하고 있는 VLSI 기술을 이용하여 규모가 큰 신경회로망을 집적회로화하려는 연구가 꾸준히 이루어지고 있다. 신경회로망 칩의 구현방법에는 아날로그 회로를 이용한 방법과 디지털 회로를 이용한 방법으로 크게 나눌 수 있다. 이 중 디지털 회로를 이용한 방법은 잡음에 덜 민감하고 잘 정립된 설계 방법론을 이용할 수 있다는 장점이 있는 반면에 신경회로망을 구현하는 데에 있어 기본적으로 곱셈기나 덧셈기 회로에 요구되는 칩의 소요면적이 커서 고밀도의 신경회로망 칩을 만드는데 어려움이 있다. 확률연산을 이용한 신경망은 디지털 방법의 장점을 유지하면서 적은 면적으로 덧셈기나 곱셈기와 같은 기본연산을 구현할 수 있다[1]. 또한 음성인식이나 패턴인식 등에서 요구되는 실시간 처리에 있어서 필수적인 On-chip 학습 기능을 내장하였다. Off-chip의 경우에 발생하는 호스트 컴퓨터와의 인터페이스로 인한 속도 저하나 가중치 저장을 위한 칩 면적의

증가 등의 문제를 해결할 수 있다[2]. 본 논문에서는 확률적 연산기법을 이용한 펄스형 디지털 신경망에 On-chip 학습기능을 내장하는 단순하고 효과적인 구조와 상관도가 적은 의사-난수 발생기(Pseudo-random number generator)를 제안하며 이 구조의 성능을 검증한다.

2. 본 론

2.1 확률 연산기법을 이용한 기본 신경망 소자

확률연산에서는 주어진 시구간 내에서 펄스가 1이 되는 확률로써 수를 표현하며 그림 1과 같이 신경망의 기본적인 동작을 AND 게이트와 OR 게이트 소자를 사용하여 수행할 수 있다. 만약 두 펄스열  $a_{i(n)}$ 과  $b_{j(n)}$ 에서의 펄스 발생이 통계적으로 독립적이라면 AND 게이트의 출력 펄스열에서의 펄스 발생확률은 다음과 같다.

$$c_{ij} = P(c_{ij(n)}=1) = P(a_{i(n)}=1 \wedge b_{j(n)}=1) = a_i b_j$$

입력들의 덧셈과 비선형 함수의 구현은 동시에 OR 동작에 의해서 수행된다. 단극적인 성질을 가지고 있는 펄스열은 음수값의 표현이 불가능하므로 각 연결 가중치는 흥분성 입력과 억제성 입력으로 나누어진다.

$$net_i^+ = P(c_{i1(n)}=1 \vee c_{i2(n)}=1 \vee \dots \vee c_{ij(n)}=1) = 1 - \prod_j (1 - c_{ij}^+) = 1 - \prod_j (1 - a_{ij}^+ b_j)$$

같은 방법으로,

$$net_i^- = 1 - \prod_j (1 + a_{ij}^- b_j)$$

따라서, 만약  $net_i^{+(n)}$ 과  $net_i^{-(n)}$ 이 서로 통계적으로 독립적이라고 가정하면 뉴런 몸체 i내의 AND 게이트의 출력열  $o_i(n)$ 에서의 펄스 발생확률은 다음과 같다.

$$o_i = P(net_i^+ = 1 \wedge net_i^- = 0) = net_i^+ \cdot (1 - net_i^-) = [1 - \prod_j (1 - a_{ij}^+ b_j)] \prod_j (1 + a_{ij}^- b_j)$$

2.2 역전파 알고리즘을 적용한 신경망 모델

가장 널리 사용되어지는 모델인 역전파 알고리즘을 적용한 다계층 신경망 구조는 입력층과 출력층 사이에 하나 또는 그 이상의 은닉층으로 구성된다. 각 뉴런은 연결 가중치에 의하여 두 인접한 층들의 모든 뉴런과 연결되어진다. 이 구조는 입력값으로부터 출력값을 계산하는 순방향단계와 가중치를 학습하는 역방향 단계로 나뉜다.

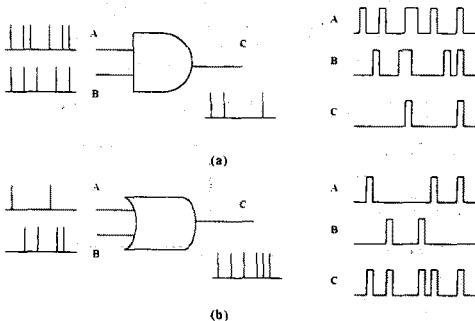


그림 1. (a) AND 게이트를 이용한 곱셈기  
(b) OR 게이트를 이용한 덧셈기

## 2.3 시스템의 구현

### 2.3.1 전체 구조

확률연산 방법을 이용하여 36개의 입력층, 16개의 은닉층, 4개의 출력층으로 구성된 패턴인식에 적합한 구조로 구현한다. 시스템의 분해능은 확률연산에서 양질의 결과를 얻을 수 있는 7비트로 하였다. 그림 3은 확률연산 필스형 디지털 신경망칩의 출력층 구조이다. 각 뉴런은 하나의 단위뉴런과 16개의 시냅스를 포함하고 모든 시냅스부와 뉴런부는 동시에 처리된다. 하나의 입력 패턴은 순방향(재현)단계와 역방향(학습)단계에서  $2^7-1=127$ 비트의 필스열의 길이를 갖는다. 단위 뉴런은 시냅스들로부터 전달되는 가중치와 입력이 곱해진 값들의 덧셈과 비선형 함수의 역할을 수행한다. 각각의 시냅스부는 입력 필스열과 가중치 필스열의 곱셈을 수행한다. 시냅스부는 가중치 레지스터(WR), 펄스 발생부(PG), 가중치 갱신부(WU), 오차 계산부(EG) 등으로 구성된다. 펄스 발생부는 뒤에 기술할 난수 발생부(RNG)에서 생성된 난수열을 사용하여 가중치를 확률적 필스열로 변환한다. 가중치 갱신부는 수정된 가중치 필스열을 계산하는 부분이고 다음 학습 단계 이전에 새로운 가중치 값을 다시 저장하게 된다. 오차 계산부는 이전 층으로부터의 오차값( $\delta$ )을 계산한다.

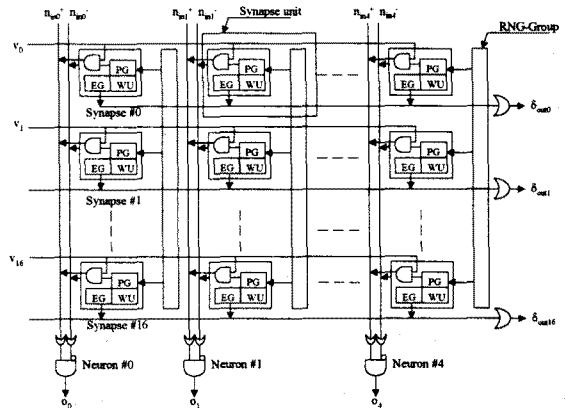


그림 3. 전체 구조(출력층)

### 2.3.2 의사-난수 발생부(RNG) 구조

확률연산 구조에서 필스열들은 확률적으로 서로 독립적이어야 한다. 만약 필스열 사이에 상관(correlation)이 발생한다면 확률연산을 이용한 필스열의 계산 정확도가 상당히 저하된다. 의사-난수 발생부는 LFSR (Linear Feedback Shift Register)와 디지털 비교기로 구성된다. 여기서 n비트 LFSR은 최대길이 주기  $P=2^n-1$  을 갖는다. 기존의 방법에서는 n비트 LFSR 사이의 상관을 제거하기 위해서 서로 다른 n개의 LFSR, 다른 LFSR의 초기값, 또는 다른 XOR tap구조의 임의적인 선택 등의 방법을 사용하였다. 그러나 이 방법들은 회로가 복잡해질 뿐만 아니라 구현이 어렵다. 본 논문에서는 하나의 LFSR로부터의 의사-난수열을 barrel shifter를 사용하여 주어진 주기 안에서 난수열을 회전(rotate)시킴으로써 상관을 제거하는 방법을 사용하였다. Mux의 조합으로 이루어지는 barrel shifter는 한 클럭 주기 안에 데이터를 하나 또는 그 이상의 비트로 이동 또는 회전시킨다. 이 방법을 사용하여 상관도가 적은 의사-난수열을 간단하고 용이하게 구현할 수 있다. 또한 한 클럭 주기 안에 수행하기 때문에 속도가 빠르다. 그림 4와 같이 난수 발생부 회로를 구성하였다.

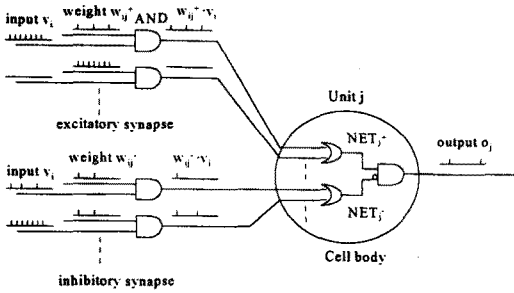


그림 2. 다수 입력 뉴런의 순방향 단계 연산 회로

확률연산의 단순한 논리적 게이트를 사용하여 역전과 알고리즘을 효과적으로 수행한다.

### 2.2.1 순방향 단계

그림 2는 확률연산을 하는 칩의 순방향 단계를 나타낸다. 입력  $v_i$ 와 연결 가중치  $w_{ij}$ 가 확률적 디지털 필스열로 나타내어지고, 연결의 곱셈연산  $w_{ij} \times v_i$ 는 하나의 AND 게이트에 의하여 수행된다. 이 가중치를 가진 흥분성 입력과 억제성 입력들이 OR 게이트를 통해 각각 더해지게 되고 필스열  $O_j$ 는  $NET_j^+$ 신호와 반전된  $NET_j^-$ 신호의 AND 연산으로 계산된다. 필스들이 임의로 분포되었다고 가정하면 신호의 값은 한 주기 이내에 펄스 비트가 ON이 될 확률로 표현된다.

### 2.2.2 역방향 단계

표 1에서 역전과 알고리즘과 비교한 확률연산 모델의 알고리즘을 보인다. 역전과 신경망 알고리즘에서 오차 신호들은 양수 또는 음수일 수 있는데 필스열은 항상 양수값만을 표현 가능하기 때문에 그림 2에서 NET 값을 두 성분으로 나눈 것과 같이 오차신호를  $\delta^+$ 와  $\delta^-$ 의 두 성분으로 나누어 부호를 보존하였다. 이전 층으로부터의 오차신호를 발생시키는 활성화함수의 미분 계산은 근사된 회로를 사용하였다(3). 은닉층의 오차신호는 출력층으로부터의 각 오차신호를 가중치와 곱한 합으로 나타내어진다. 생성된 오차신호에 입력값과 학습율을 곱한 후 이전 가중치와 더하여 가중치를 갱신하게 된다.

표1. 확률연산을 이용한 모델

	확률연산을 이용한 모델	역전과 알고리즘
순방향 단계	$net_j^+ = OR(v_i \text{ AND } w_{ij})$ $net_j^- = OR(v_i \text{ AND } w_{ji})$ $o_j = net_j^+ \text{ AND NOT } net_j^-$	$net = \sum w_{ij} \cdot v_i$ $o_j = f(net_j)$ $= 1 / (1 + e^{-net_j})$
출력층의 오차 (EU)	$\delta_k^+ = t_k \text{ AND NOT } o_k \text{ AND } f'$ $\delta_k^- = \text{NOT } t_k \text{ AND } o_k \text{ AND } f'$ $f' = o_k \text{ AND NOT } o_{k-1}$	$\delta_k = (t_k - o_k) \cdot f'(net_k)$ $f'(net_k) = f_k \cdot (1 - f_{k-1})$
	$\delta_j^+ = f' \text{ AND } \{ (OR(\delta_k^+ \text{ AND } w_{kj})) \}$ $OR(OR \delta_k^- \text{ AND } w_{kj}) \}$ $\delta_j^- = f' \text{ AND } \{ (OR(\delta_k^- \text{ AND } w_{kj})) \}$ $OR(OR \delta_k^+ \text{ AND } w_{kj}) \}$ $f' = o_j \text{ AND NOT } o_{j-1}$	$\delta_j = f'(net_j) \sum \delta_k \cdot w_{kj}$
가중치 갱신 (WU)	$\Delta w^+ = \eta \text{ AND } \delta^+ \text{ AND } v$ $\Delta w^- = \eta \text{ AND } \delta^- \text{ AND } v$ $w^+ = w^+ \text{ AND NOT } \Delta w^+ \text{ OR } \Delta w^+$ $w^- = w^- \text{ AND NOT } \Delta w^- \text{ OR } \Delta w^-$	$\Delta w = \eta \cdot \delta \cdot v$ $w_n = w_{n-1} + \Delta w$

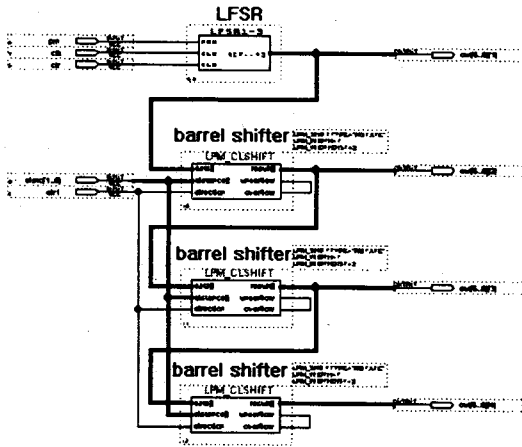


그림 4. LFSR과 barrel shifter로 구성된 상관도가 적은 의사-난수 발생부

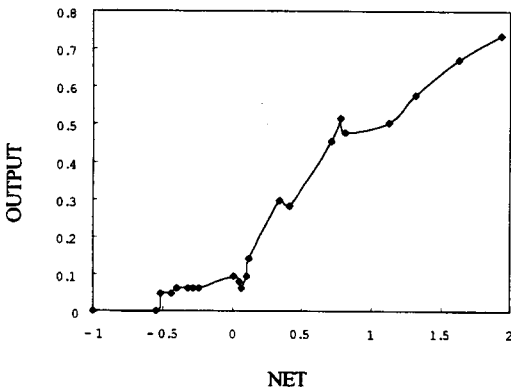


그림 5. 두 입력값에 대한 비선형 활성화함수의 파형

분해능은 7비트이며 다수의 barrel shifter를 연속으로 위치하면 지연시간이 상당히 길어지므로 하나의 LFSR에 3개의 barrel shifter를 연결하였다. 제안한 RNG 구조를 이용하여 펄스열을 발생하고 두 개의 입력을 갖는 그림 2의 회로를 구성하였다. 그림 5는 이 회로의 시뮬레이션 결과 제안한 RNG 구조를 통해 입력을 받은 하나의 뉴런이 비선형 활성화함수의 형태로 출력을 내보내는 것을 나타내었다. 그림 6은 barrel shifter에 의하여 회전된 의사-난수열들 간의 교차상관을 나타낸다. 회전한 비트의 간격이 증가할수록 상관성이 감소함을 알 수 있다.

### 3. 결 론

본 논문에서는 on-chip 학습능력을 가진 펄스형 신경망의 구조를 제안하고 성능을 검증하였다. 확률연산의 장점은 대량의 병렬처리를 위한 신경회로망의 하드웨어 구현이 적은 면적에 가능하다는 점이다. 이와 같은 장점은 많은 면적을 필요로 하는 실시간 처리를 위한 on-chip 학습기능을 내장한 칩에 적합하다. 반면에 확률연산은 모든 연산을 확률적으로 수행하기 때문에 정확도 측면에서 단점이 있다. 앞으로 연구해야 할 과제는 제안한 방법을 이용하여 충분한 정확도를 확보한 칩을 제작하는 것이다. 그리고 현재 제안한 신경회로망을 FPGA칩으로 구현하는 연구를 진행하고 있다.

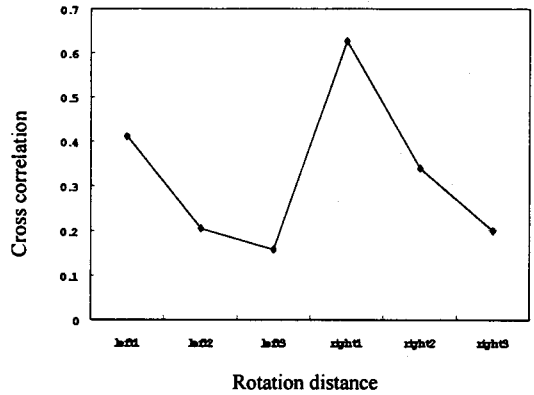


그림 6. 회전 간격에 따른 난수열 간의 상관도

(참 고 문 헌)

- [1] M.S. Tomlinson Jr., D.J. Walker and M.A. Sivilotti, "A digital neural network architecture for VLSI", Proc. IJCNN, Vol. 2, pp. 545-556, 1990.
- [2] D. Hammerstrom, "A VLSI architecture for high-performance, low-cost, on-chip learning", Proc. IJCNN, Vol. 2, pp. 537-544, 1990.
- [3] S. Oteki, A. Hashimoto, T. Furuta, and S. Motomura, "A digital neural network VLSI with on-chip learning using stochastic pulse encoding", Proc. IJCNN, Vol. 3, pp. 3039-3045, 1993.
- [4] Y.C. Kim and M.A. Shanblatt, "Architecture and statistical model of a pulse-mode digital multilayer neural network", IEEE Tran. on neural networks, Vol. 6, No. 5, pp. 1109-1118, 1995.
- [5] J. M. Zurada, "Introduction to artificial neural systems", Info access distribution pte ltd, pp. 185-206, 1992.