

《主 題》

신경망 VLSI 기술의 발달과 현재

한 일 송
(한국통신 연구개발단)

■ 차	■ 례 ■
I. 개 요	Ⅲ. 신경망 VLSI의 고속화 그리고 고집적화
Ⅱ. 신경망 VLSI화 요소	Ⅳ. 결 언

〈요 약〉

신경망 실용화에 기본이 되는 신경망 VLSI 기술의 최근 발전 추세에 관하여 검토하였다. 대규모 고속 신경망 VLSI 구현 방법들인 디지털, 아날로그, 하이브리드 신경망 칩들을 비교하였으며, 십 수만 단위의 하이브리드 신경망 칩기술을 제시하였다.

I. 개 요

신경망은 복잡한 정보 환경속에서 적응 원리에 따라 자동적으로 처리하는 방대한 정보시스템을 의미한다. 이러한 신경망은 지난 반세기동안 “programmed computing”식 정보처리 흐름과는 다른 선택이다. 정보처리란 수학적으로 대수분석적인 면과 기하구조적인 면 모두를 가지고 있다는 점을 감안하면, 1940년대 이후로 현재까지 대부분의 정보처리-컴퓨터는 대수 분석위주이였으나 신경망 컴퓨터는 기하구조적 정보처리인 것이다. 신경망은 기존의 정보처리 방법과는 근본적으로 다르므로, 실용적 구현을 위하여는 특별한 성능의 하드웨어 즉 신경망 VLSI가 필요하다.

현재 신경망 VLSI는 기존의 VLSI기술인 실리콘 즉 CMOS 기술이 가장 널리 쓰이고 있다. 초기의 신경망 칩은 소규모 즉 1개의 처리소자를 가지며 연결고리는 시분할 방식으로 사용되는 것에 불과하였으나, 최근

의 연구는 CMOS나 변형된 CMOS 공정으로 단위칩당 수천 연결고리의 용량에서 제한적이지만 수만 연결고리 까지 구현시키고 있으며 실리콘 웨이퍼를 통째로 1개의 칩처럼 이용하기에 이르렀다. 국내에서도 여러가지 종류의 신경망칩이 연구되고 있으며, 특히 현재 최대 규모의 용량이라 할 수 있는 십수만 연결고리를 집적시킨 범용 하이브리드 신경망 칩이 제작중이다. 이러한 점에서, 제 6세대 신경망 컴퓨터의 실현은 현실화되고 있으며, 기존의 컴퓨터와 비유하여 하드웨어의 골격인 CPU와 MEMORY 역할을 동시에 하는 신경망 VLSI의 특성과 그 발전 추세에 대한 분석은 앞으로 통신과 컴퓨터가 융합되는 정보화 사회에서 중요한 역할이 예상된다.

Ⅱ. 신경망 VLSI화 요소

신경망의 VLSI화 즉 하드웨어 구현은 여러가지 필요성과 문제점을 가지고 있으나, 그 주된 필요성은 기

존의 컴퓨터 하드웨어상에서 현실적인 적용을 목적으로 소프트웨어 시뮬레이션 할 때 생기는 처리 속도와 규모의 제한을 하드웨어로 극복하고자 함에 있다. 하드웨어 구현을 위하여 신경망 VLSI를 구현할 때 생기는 문제점은 실현 가능한 규모에 있다 - 즉, 그 지능이 증가적으로 지렁이에 해당하는가 벌에 해당하는가 라는 문제를 생각할 수 있다. 신경망 컴퓨터의 실현은 물론 신경망 VLSI만으로 되는 것은 아니며 사람의 지적 능력 형성 과정에서도 그러하듯이, 신경망 컴퓨터 자체의 학습 방법이나 구조 설정의 면에 있어서 목적이나 용도에 따라서 해결하여야 될 문제점이 많다. 그러므로, 본 주제에서는 범용 신경망 VLSI에 대하여 기본 요소를 분석한다.

일반적으로 신경망 하드웨어의 기술적 목표를 요약하면

- 대규모 용량의 신경망
- 동작의 고속성 즉 고속 병렬 처리
- 가변의 프로그램 즉 학습에 따른 연결고리 강도의 변경
- 정밀도와 적절한 가격

을 들 수 있다. 이와 같은 목적의 소자는 LCD, GaAs 계열 반도체등의 광학적 구현도 생각할 수 있으나, Bipolar, CCD 특히 CMOS등의 기존의 VLSI기술인 실리콘 기술이 보편화되어 있다. 이는 실리콘 VLSI기술이 현재 가장 잘 발달된 기술이어서 실용화의 전제 조건인 상대적으로 저렴한 가격과 신뢰도를 가능하게 하기 때문이다.

실리콘에 집적화시키는 인공 신경망은 궁극적으로 방대한 규모의 연결고리(synapse)와 신경세포(neuron)가 연결된 망으로서, 연결고리는 곱셈과 덧셈 기능으로 모델링할 수 있으며 신경세포는 비선형 결정 함수로 모델링 할 수 있다. 이와 같은 기능을 전자회로로 신경망 VLSI를 구현하는 데는 그 회로 설계 개념에 따라 디지털과 아날로그로 구분할 수 있다. 디지털 방식과 아날로그방식은 그 시스템이나 VLSI 제조 공정에서 가지는 특성이 다르기 때문에 각각의 방식이 가지는 장,단점도 서로 상반된다.

디지털 방식으로 신경망 칩을 구현하면 우선 설계가 비교적 용이하다는 것을 제시할 수 있다. 기존의 각종 CPU나 MEMORY를 제작하는 규격화된 제작공정을 이용하므로 일반화될 수 있는 가격과 신뢰도가 가능하다는 장점도 있다. 그리고 디지털인 관계로 임의의 정밀도를 구현할 수 있고 프로그램이나 칩급이 용이하다는 점도 큰 장점 중의 하나이다. 그러나, 신

경망 VLSI 즉 하드웨어의 주 목적인 방대한 용량의 구현은 제한을 받는 데, 이는 연결고리의 기본 기능의 하나인 곱셈 연산에 상당한 칩 면적을 소비해야하기 때문이다. 그리고 전체적인 동작을 완전한 비동기 동작으로 구현하기가 어려워 구조적인 문제점이 되기도 한다.

아날로그 방식으로 신경망 칩을 구현하면 가장 큰 장점으로 집적도가 높다는 점을 들 수 있다. 회로 내용에 따라 다르겠으나 비동기성 동작을 자연스럽게 얻을 수 있고 넓은 동작 범위를 가질 수도 있다. 그러나 아날로그 방식의 가장 큰 단점은 특수 제작 공정을 필요로 하여 구현이 어렵다는 점을 들 수 있다. 그리고 기술적으로 높은 임의의 정밀도를 얻기가 어렵거나 프로그램에 제한을 받기 쉽다.

하이브리드 방식은 여러가지 형태를 생각 할 수 있으나, 근본 동기는 디지털 방식에서 얻을 수 있는 정밀도와 프로그램의 용이성을 얻으며 아날로그 방식의 장점인 소형 연결고리를 구현하고자 하는 것이다. 일반 디지털 제작공정을 사용하며, 연결고리는 아날로그로 동작하고 정밀도나 사용과정은 디지털에 준한다면 앞에서 언급한 디지털의 장점과 아날로그의 장점만을 가질 수 있게 된다.

신경망 VLSI 성격으로 고려되는 다른 측면으로는 범용성 혹은 제한없는 확장성을 가지거나 혹은 특수한 용도 혹은 제한된 구조의 상태 여부에 따라 다른 장,단점을 가지게 된다. 이는 아직 특정 구조나 학습 방법 그리고 전기적인 신경 상태 변환 방식에서 유일하게 일반화될 해결안이 없어, 일반 기존의 컴퓨터에서와 같이 각각의 CPU 구조나 성능에 있어 서로 다른 기준의 장,단점을 가졌던 것처럼 신경망 VLSI도 서로 다른 기준의 장,단점을 가지므로 공통의 성능 기준으로 연결고리 연산속도나 정밀도및 용량을 단위로 하고 있다.

III. 신경망 VLSI의 고속화 그리고 고집적화

신경망은 인공 지능을 구현하고자하는 새로운 흐름으로 근본적으로 새로운 컴퓨터 구조에 개념을 두고 있는 관계로 초기 연구 단계로 부터 많은 연구가 진행되어 왔다. 그리고 범용성을 추구하면서도 기술적인 제약성과 다양한 학습이론 및 구조로 인하여 특수용도의 VLSI도 계속 추구되어 왔다. 기본적으로 신경망 구현에 VLSI를 사용하는것은 2가지 방향에서 시작되는 데, 하나는 아날로그든 디지털이든 어떤 학

습이론을 채택하였던 그 목적을 위한 새로운 신경망 VLSI를 준비하는 것이고, 다른 하나는 일반적 디지털 신호 처리 VLSI를 이용하여 신경망 VLSI로 사용하는 것이다. 최근에는 VLSI 기술의 발달에 따라 신경망 전용의 VLSI를 개발은 일반 범용 디지털 신호처리 VLSI와 비교하여 보다 고속화와 고집적화가 집중적으로 요구되고 있다.

초기의 신경망 하드웨어는 개별 소자나 기존의 디지털 로직, 연산증폭기를 이용하여 구성되었으며, 일부는 WIZARD와 같이 오래전에 상용화 되기도 하였다. 신경망 VLSI의 상용화 모델의 초기 단계는 Fujitsu사의 Analog Neuro Processor(ANP)나 Micro Device사의 Neuro Bit Slice를 들 수 있으며, 이들은 각각 1개의 뉴우론과 8개의 뉴우론을 가지고 있으며 연결고리의 등가적 구현은 한 순간에 하나씩 직렬 입력에 의한 시분할 방식으로 이루어졌다. 연결고리의 강도는 임의로 변경 가능하였고, ANP는 Bipolar 공정으로 제작된 특수한 곱셈형 D/A 변환기와 시분할 샘플링을 이용한 아날로그식이며 Micro Device사 제품은 MOS 공정으로 제작되었으며 AND 게이트를 이용한 1비트씩 직렬 곱셈형 연결고리로 시분할적 WIRED-OR의 개념으로 동작한다. 이에 대하여 비교적 범용성, 확장성에서 어느 정도 실용성을 찾을 수 있는 제품은 비교적 최근에 출현하였던 Neural Semiconductor사의 Neural Synapse Unit과 Nodal Unit를 들 수 있다. 이들은 펄스열 방식으로 연결고리 강도도 Stochastic Pulse Width로 구현하는 AND 게이트형 연결고리를 채택하여, Wired-OR를 통한 신경망을 형성하였다. 비교적 확장성을 가지는 이 제품은 칩 세트 개념으로 사용되며 32×32 규모의 단위 칩으로 구성된다. 이 VLSI의 제한점으로는 디지털 방식임에도 불구하고 비교적 정밀도가 낮고, 확장성을 가지나 그 규모가 Stochastic Pulse Operation에 따른 주기와 전체 크기에 의하여 제한을 가지는 점이다. 신경망 VLSI 그 자체로서 상품화 되어있는 것으로 인텔사의 ETANN 칩이 있으며, 상품화된 모델중 상위 성능과 가격을 가진다. 이는 아날로그 신경망 VLSI로서 특수한 CMOS 공정을 사용하여 Nonvolatile 아날로그 연결강도 메모리를 내장하고 있다. 208핀 VLSI로 칩에서 외부와의 연결은 시분할 방식으로 이루어지고 있다. 신경망 VLSI로 상품화는 되지 않고 있으나 상용화된 대단위 신경망 VLSI로 Adaptive Solution사의 CNAPS가 있다. 칩당 64개의 처리소자를 가진 무척 칩 크기가 큰 디지털 방식으로 워크스테이션의 가속기 형태로 상용

화 되어 있다.

이와 같이 여러 형태의 신경망 VLSI가 상용화 되어 있으나, 이 외에도 실험실에서 여러가지 방식으로 대규모 집적화된 신경망 VLSI가 연구되고 있다. 특수용도로서는 Hitachi인 6인치 웨이퍼 스케일 집적화와 Masthusida의 1 비트 연결고리 강도를 가지는 39,000 연결고리의 단일 칩 집적화를 들 수 있다. 범용 대규모 신경망 VLSI로는 NTT, BT(British Telecom)의 디지털 방식과 한국통신의 하이브리드 방식이 있다.

III-1. 디지털 신경망 칩

디지털 신경망 칩은 디지털 신호처리 VLSI의 특별한 신경망 기능 추가와 병렬화로 구현하는 방향과 Stochastic Pulse를 사용하거나 Pulse 신경 상태의 Wire-OR와 같은 새로운 개념의 디지털 처리 소자를 도입하는 방향으로 크게 나눌 수 있다. 초기의 Micro Device사 칩은 멀티플렉서와 직렬식 신경상태 및 연결강도의 AND기능으로 구성되고 칩 외부에 연결강도 저장용 MEMORY를 가지고 있으며, 신경망 크기의 확장은 멀티플렉서를 이용한 칩간의 정보 전달로 구현한다. 이에 대하여 Neural Semiconductor사의 칩은 4비트 연결고리 강도 1,024개를 칩 내부에 저장하는 완전 디지털 개념의 Pulse Train 방식이다. 여기서 Pulse Train 방식은 신경 연결고리에서 필수적인 곱셈기능의 디지털 회로를 고집적화하기 위한 노력으로 직렬식 곱셈기가 초기부터 출현함에 따라 생겨났다. 신경 상태나 연결고리 강도를 펄스열로 하는 것으로, 디지털 방식에서는 기본적으로 곱셈기능은 AND 게이트에서 이루어지며 펄스의 형태는 펄스 수요, 유효한 단위 주기내의 펄스 폭 총 합으로 나타낸다. 이에 대하여 Adaptive Solution사의 디지털 칩 N64000은 VLSI라기 보다 ULSI로 주장하듯이 칩 크기가 0.8 마이크로 CMOS 공정에서 2.54cm×2.54cm 규모로 Single Instruction data Multiple Data stream 구조인 1차

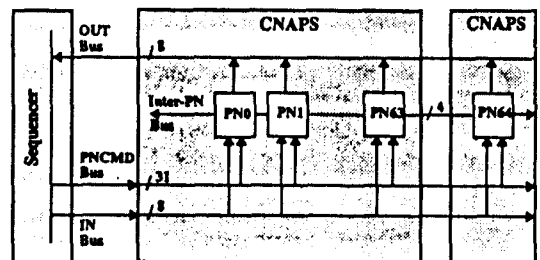


그림 1. 멀티 PN 구성

원적 처리소자 array의 집합체로 칩당 64개의 처리소자(Processor Node: PN)를 가동할 수 있다. 이 칩역시 외부의 전용 컨트롤러와 세트를 이루어 처리소자의 다중화는 그림 1과 같이 구성하며, 칩은 그림 2의 구조를 가지고 있다. 칩당 속도는 1.6 Giga cps (connections per second)이며 모두 4개의 칩이 모여 CNAPS를 구성한다.

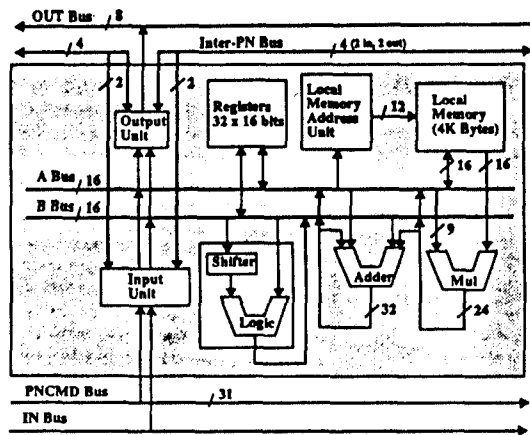


그림 2. PN의 구조

BT의 신경망 칩인 HANNIBAL은 칩당 8개의 처리소자와 최대 8,192개의 연결고리가 있으며 20MHz 클럭에 의하여 동작된다. 0.7 마이크로 공정으로 9mm×11mm 크기의 칩으로 제작되며 칩당 160 Mega cps의 성능을 가지고 칩의 단위 모듈은 그림 3의 구성을 가진다.

III-2. 아날로그 신경망 칩

일반적으로 아날로그 신경망 칩의 문제는 아날로그 연결고리 강도의 저장 문제, 그리고 정밀하고 극소형화된 아날로그 곱셈기 구현을 들 수 있다. 이와 함께 부하에 따른 특성 변화 대비가 필요하며, 기존의 아날로그 VLSI 제작 공정은 디지털 VLSI 제작공정보다 까다롭고 집적도도 낮아진다. 이와 같은 문제점중 연결고리 강도의 저장 문제는 DRAM에서 하듯이 연결고리 강도를 커패시터에 저장하고 주기적으로 갱신함으로써 해결한다. 아날로그 곱셈기의 극소형화는 MOS 트랜지스터의 비포화 영역에서의 비선형성을 무시하거나, 포화 영역에서 부분 선형화된 모델을 적용함으로써 집적화 하고 있다. 아날로그 신경망칩의 대표적인 예로는 인텔사의 ETANN이 있으며 연결고리 강도를 2개의 차동증폭단 전류원 기준 전압

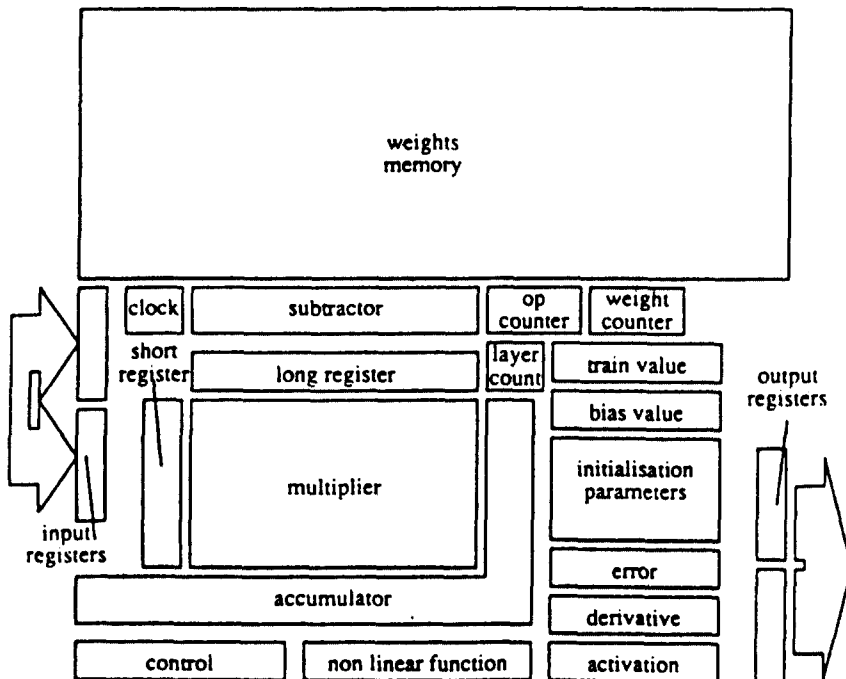


그림 3. HANNIBAL의 주 기능 블록도

차이로 구현하고, EEPROM 제작시와 유사한 공정을 사용하여 제작되어 전류원 트랜지스터 게이트에 인가되는 연결고리 강도 기준 전압이 Nonvolatile로 유지된다. 아날로그 입,출력을 가지는 완전한 아날로그 상태로 동작하며 1.5W의 전력소모로 2 Giga cps의 성

능을 가지고 있으며, 연결강도를 변경시키는 과정은 기존 기준 전압 차이의 상태를 감지하여 추가분의 전하 펄스로 이루어지므로 다소 시간이 걸리고 역시 전용 컨트롤러가 필요하다. 개념적인 기능도는 그림 4와 같다.

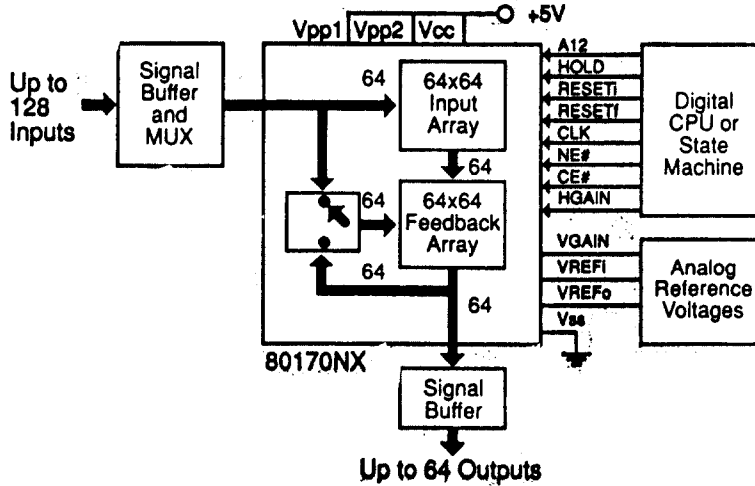


그림 4. ETANN의 기능도

III-3. 아날로그-디지털 하이브리드 신경망칩

아날로그나 디지털로 신경망 칩을 구현하면 집적화 과정에서 여러 가지의 차이가 있으며 각각의 서로 상반된 장점과 단점, 즉 어느 한 쪽의 장점은 다른 편에서는 단점이 된다. 이와 같은 문제점을 극복하는 하나의 방법으로 기본 단위 회로의 아날로그 설계 및 동작과 디지털의 신경상태 및 외부 접속을 허용하는 한국통신의 하이브리드 방식이 제시되었으며, 이는 디지털 표준 공정으로 구현 가능하다. 여러 가지의 하이

브리드 방식중 그림 4의 한국 통신 방식은 아날로그 단위 블록을 사용하지 않고, 정적인 전류원과 스위치만을 사용하므로 가장 빠른 속도를 보이고 있다. 그리고 연결고리가 가지는 아날로그 동작에서의 상호 독립성은 칩과 칩 사이의 확장에 제한이 거의 없으며, 따라서 모듈별로 임의의 구조를 구성하거나 정밀도를 증가시킬 수 있다. 한국통신의 하이브리드 신경망 칩을 사용하는 측면에서는 전원과 함께 몇몇 기준 전압을 인가하는 이외에는 디지털 신경망칩과 동일하게 사용할 수 있으며 타이밍 상의 제한이 거의 없이 완벽한 비동기성을 가진다. 칩의 구현은 신경 상태의 정의에 있어 단순한 디지털 이외에 여러가지 펄스식 동작도 가능하므로 집적화의 특수성을 고려하여 연결고리와 처리소자의 칩 세트 개념으로 추진하고 있다. 처리소자의 크기는 연결고리 2개 정도의 크기이고 이는 submicron의 채널 길이를 가지는 트랜지스터 제작에 사용되는 1.0 마이크로 16 MDRAM 공정으로 집적화할 때 70미크론 직경이내로 구현가능한 고집적화 특성을 가진다. 1차, 2차에 걸쳐 제작된 수백-수천 연결고리 규모인 시험 칩 제작 및 시험 결과를 통하

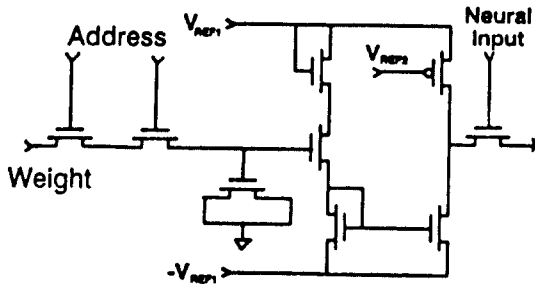


그림 4. URAN의 단위 연결고리

여 2단계인 고집적 신경망 칩 제작이 진행 중이며, 그 규모는 135,424 연결고리 규모로 그림 5의 내부 블록을 가지고 있다. 처리 속도는 칩당 340 Giga cps를 가지며 48개의 칩으로 PCB를 구성하면 16 Tera cps의 성능이 가능하다.

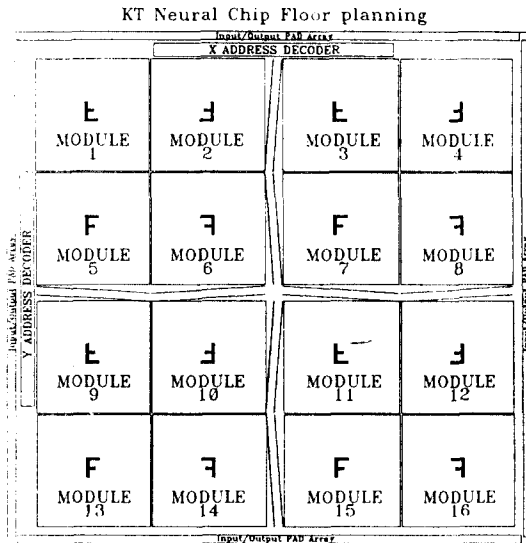


그림 5. 135,424 연결고리 신경망 칩의 블록도

IV. 결 언

신경망 VLSI는 현재 여러 방향으로 발전 중이며, 어느 경우나 소규모 응용에 있어서 기술적 한계점은 거의 해결되어가고 있다. 현재의 추세로 보면 조만간 방대한 규모의 신경망 칩이나 이를 이용한 하드웨어가 실현될 것이고, 이때는 속도 혹은 정밀도 외에 새로이 제시되는 소프트웨어를 함께 집적화 하는 것이 문제가 된다. 즉 기존 컴퓨터의 범용성을 지니면서도 신경망의 고속성을 추구하는 양면성이 요구되며, 점점 고속화되는 디지털 신호처리 VLSI와의 성능 대비가 주요하여지고 있다. 이런 점에서 특수한 아날로그 회로 구현으로 일반 디지털 신호처리 VLSI를 능가하는 고집적성, 고속화를 추구하는 동시에 디지털 인터페이스를 제공하는 하이브리드 신경망 칩은 기존의 컴퓨터와 접속하거나 디지털 신호처리 VLSI와 혼용성 등의 많은 새로운 측면을 가지고 있다.

참 고 문 헌

1. Dann Hammerston and Nguyen, "An Implementation of Kohonen's Self-Organizing Map on the Adaptive Solutions Neurocomputer," Artificial Neural Network, 1991, pp.715-720.
2. D. J. Myers et al, "HANNIBAL: A VLSI Building Block for Neural Networks with On-chip Back-propagation Learning," BT, pp.171-181.
3. 한일송, "하이브리드 신경망 칩의 확장성과 특성," JCE ANF, pp.526-521.

韓 一 松

- 工學博士, 한국통신 연구개발단 책임연구원
- 1975년 3월 ~ 1979년 2월 : 서울대학교 공과대학 전자공학과
- 1979년 3월 ~ 1981년 2월 : 한국과학기술원 전기 및 전자공학과 석사학위
- 1981년 3월 ~ 1984년 2월 : 한국과학기술원 전기 및 전자공학과 박사학위
- 1984년 5월 ~ 1985년 2월 : 한국과학기술원 전자공학부 연수연구원
- 1985년 2월 ~ 현재 : 한국통신 연구개발단 단부
- 1989년 4월 ~ 1990년 3월 : 영국 EDINBURGH 대학교 VISITING ACADEMIC