CPU 기술과 미래 반도체 산업 (III)

CPU Technology and Future Semiconductor Industry (III)

박상기 (Sahnggi Park, sahnggi@etri.re.kr) 소재부품원천연구본부 책임연구원

ABSTRACT

Knowledge of the technology, characteristics, and market trends of the latest CPUs used in smartphones, computers, and supercomputers and the research trends of leading US university experts gives an edge to policy-makers, business executives, large investors, etc. To this end, we describe three topics in detail at a level that can help educate the non-majors to the extent possible. Topic 1 comprises the design and manufacture of a CPU and the technology and trends of the smartphone SoC. Topic 2 comprises the technology and trends of the x86 CPU and supercomputer, and Topic 3 involves an optical network chip that has the potential to emerge as a major semiconductor chip. We also describe three techniques and experiments that can be used to implement the optical network chip.

KEYWORDS Photonic and optical interconnect, Optical network-on-chip, Optically interconnected CPU, Supercomputer architecture, CPU design and fabrication, Smartphone SoC, x86 CPU technology, Package on package, 2.5D package, 3D package

Ⅰ. 서론

컴퓨터 CPU는 인간이 도달할 수 있는 기술의 최고 수준을 나타내는 척도라고 할 수 있다. 인간이 상상하는 어떠한 종류의 지능형 모델과 프로그램도 CPU 성능에 구애 받지 않을 수 없다. 4차 산업뿐만 아니라 전체 IT 산업의 발전을 이끌어 가는 정점에 CPU가 있다. 2019년 출시된 스마트폰과 컴퓨터, 슈퍼컴퓨터에 사용된 최신 CPU의 기술과 특성 및 시장동향, 그리고 미국 주요기업과 대학교 전문가들의 연구방향, CPU 광인터커넥션 신기술을 3부에 걸쳐 기술하였다.

Ⅰ, Ⅱ부에 이어 Ⅲ부에는 다음과 같은 내용을 기술하였다. 미국의 주요기업과 대학교들은 광통신 기술이 CPU칩에 적용될 경우 채택할 수 있는 가장 유력한 네트워크 구조들을 제안하였고, 그

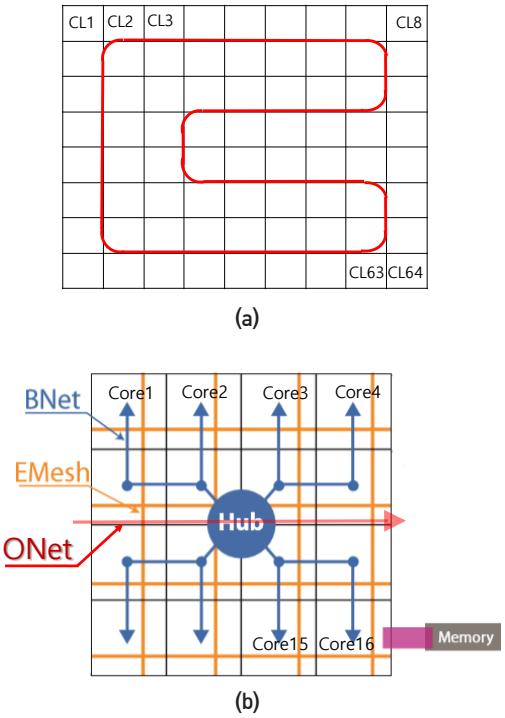
* DOI: https://doi.org/10.22648/ETRI.2020.J.350210

중 인용 빈도가 높은 MIT의 ATAC 구조와 HP의 Corona 구조에 대해 개략적으로 설명하였다. 그리고 이러한 구조가 칩에 실현되지 못한 주요 원인을 기술하였다[1–5]. 슈퍼컴퓨터에 활용되고 있는 VCSEL(Vertical Cavity Surface Emitting Laser, 표면방출레이저)과 PD(photodiode)를 도입하고 fat tree 네트워크를 그대로 적용하되 단지 광섬유만 SiON/SiN 물질의 광도파로로 교체하여 인텔 서버용 CPU에 적용하는 것이 가장 합리적이라 할 수 있다. 이를 위해 필요한 3가지 기술과 실험적으로 검증된 내용을 기술하였다[6–9].

Ⅱ. 온칩 광네트워크 기술

2000년대 초반부터 Intel, IBM, HP, Sun Microsystems, Samsung 등 반도체 선두 기업들은 광통신 기술을 반도체 칩상의 core–to–core 또는 CPU–to–memory 간 통신에 적용하기 위해 연구(silicon photonics 연구)를 시작하였다. 특히 CPU 설계와 제작을 주력으로 하는 기업들은 CPU 성능을 지속적으로 향상시키기 위해 CPU 내 광통신 기술의 도입이 불가피할 것으로 예측하고 경쟁적으로 대규모 연구 프로젝트를 진행하였다. MIT, Columbia University, HP, Sun Microsystems 등 학교와 기업들은 광통신 기술이 적용될 경우 가장 유력한 네트워크 구조(network–on–chip architecture)를 경쟁적으로 제시하였다. 그 중 인용 빈도가 높은 MIT의 ATAC 구조와 HP의 Corona 구조에 대해 개략적으로 기술한다[1–4].

ATAC 구조: MIT대학교 silicon photonics 연구팀에서 CPU 칩에 적용하기 위해 2010년 제안한 광통신 네트워크 구조이다. 1,024개 core로 구성된 단일 CPU 칩상에 광통신 기술을 적용하고 cache coherence와 CPU 성능을 전기통신의 경우와 비교 분석하였다. Clock 주파수 1GHz, in–order instruction, 16nm technology 적용 등 비교적 낮은 성능의 core를 대상으로 하였고, 이는 약 400mm² 면적의 칩에 1,024개 core를 집적하기 위해 단일 core의 면적(footprint)을 최대한 작게 가정해야 하기 때문이다.



출처 George Kurian et al., "ATAC: A 1000-Core Cache-Coherent Processor with On-Chip Optical Network," PACT'10, Sept.11-15, 2010.

그림 1 **ATAC 네트워크 구조. (a) 64 clusters, (b) cluster 내부구조**

칩 구성: 그림 1과 같이 총 64개의 clusters가 있고 각 cluster에는 16개의 cores가 있다. 각 cluster에는 memory controller가 있고 DRAM과 40Gb/s 통신속도로 연결된다.

네트워크: 각 cluster의 16개 core는 인텔의 mesh network와 유사한 전기신호 기반의 mesh network(EMesh)로 연결되어 있다. 64개의 cluster는 그림 1(a)와 같이 링형 광통신 network(ONet)로 연결되어 있고, 각 cluster에는 전기 신호를 광신호로,

표 1 ATAC의 통신라인 폭과 Latency

Core model	In-order, 1GHz, 1024 cores, 16nm technology
EMesh Hop Latency	2 cycles(router delay-1, link traversal-1)
ONet Hop Latency	3 cycles(E/O + O/E conversion-2, link traversal-1)
BNet Hop Latency	2 cycles(E/O conversion-1, link traversal-1)
EMesh	128-bit wide
ONet	128-bit wide
BNet	128-bit wide
Memory Bandwidth	64 memory controllers, 5GB/s per controller

광신호를 전기신호로 연결하기 위한 Hub가 있다. 예를 들어 core1이 cluster 내부의 core4로 data를 전송할 때에는 EMesh를 통해 신호가 전달된다. core1이 다른 cluster에 있는 core1024에 data를 보낼 경우 먼저 EMesh를 통해 cluster 1의 Hub에 신호가 전달되고 광신호로 변환되어 cluster 64의 Hub에 전달된다. 광신호를 전기신호로 변환하여 Hub에서 core1024로 신호를 전달할 때는 Bnet을 사용한다. ATAC 설계자들은 Hub에서 core로 가는 신호에 대해 routing 기능을 생략하고 전달과정을 단순화하기 위해 Bnet을 도입하였다. Bnet은 신호가 모든 core에 동시에 전달되는 broadcast 기능만 있다. 16개 cores를 두 그룹으로 나누어 신호의 특정 자리 수가 홀수일 때 1번, 짝수일 때 2번 그룹으로 전송한다.

통신라인 구성: EMesh와 Bnet의 data 신호 통신라인 수는 128개이고 clock cycle당 128bit 신호가 병렬적으로 전달된다. ONet은 data용 광도파로가 128개, control 신호용 광도파로 1개, 메타신호용 광도파로 다수, 총 129개 이상의 광도파로가 그림 1(a)와 같이 링형으로 각 cluster를 통과한다. 하나의 광도파로에는 64개 파장이 있고 각 cluster당 1개 파장이 할당된다. 따라서 각 cluster는 128개 광도파로에 1개 파장씩, 128bit data 신호를 동시에 병렬적으로 전송한다. 예를 들어, cluster 1에서 cluster 64로 data를 전송할 경우 128개 광도파로의 1번 파장에 신호를 실어 보낸다. Cluster 2~64는 각각 1번 파장의 신호를 1/63세기만큼 분기하여 수신하고 cluster 64를 제외한 나머지 cluster는 수신 후 무시하게 되며 cluster 64만 수신한 data를 채택한다. 서로 다른 파장을 사용하므로 다수의 cluster가 동시에 cluster 64에 신호를 보낼 수 있다.

Latency: 신호 전달에 걸리는 시간을 각 통신라인별로 표 1에 나타내었다. EMesh의 경우 router에 1cycle이 소요된 후 hopping할 때마다 1cycle씩 추가된다. Core1에서 출발하여 core 16에 도착할 때까지 6번의 hopping이 필요하다. 이에 비해 Onet은 일괄적으로 3cycles이 요구된다. 즉 E/O 변환 1cycle, O/E 변환 1cycle, 광도파로 신호 이동은 거리에 관계없이 1cycle 이 요구된다. BNet은 O/E 변환에 1cycle 이 소요되고 hopping할 때마다 1cycle씩 추가된다. CPU clock이 1GHz이므로 전기선의 통신속도는 1Gb/s이고 광도파로의 경우 64개 파장이 있으므로 광도파로당 총 통신속도는 64Gb/s이다. 만약 CPU의 주파수가 3GHz인 경우 ATAC network은 전기선의 경우 3Gb/s, 광도파로는 192Gb/s의 통신속도를 갖는다.

광소자 구성: 광원은 CPU칩 외부에 있고 광섬유를 통해 CPU 칩 내부의 power 광도파로에 cw (continuous wave) 신호를 입사시킨다. 그 후 링공진기를 통해 power 광도파로에서 ONet 광도파로로 접속시킨다. 메타 신호를 제외할 경우 최소 129개 광도파로에 각각 64개 파장이 있으므로 외부 광원은 최소 64개, 최대 8,192개를 필요로 하게 된다. 각 cluster는 할당된 파장에 신호를 싣기 위해 도파로당 1개씩 변조기가 있고 Ge 기반의 Electro-Absorption(EA) 변조기를 사용한다. 이 경우 cluster당

최소 129개, 칩 전체에 8,256개 이상의 변조기가 집적된다. 검출기는 Ge 기반의 검출기를 사용하며 cluster당 최소 8,127개, 칩 전체에 52만 개 이상의 검출기가 집적되어야 한다. 링공진기는 변조된 신호를 광도파로에 접속하는 기능과 각 파장의 광신호를 1/63세기로 분기하는 기능을 한다. 따라서 링공진기는 변조기와 검출기의 합만큼 집적된다. 또한 링공진기는 power 광도파로에서 ONet 광도파로로 cw 신호를 접속하는 데 필요한 수만큼 더해야 한다. ATAC 설계자는 CPU칩과는 별도로 SOI 기판의 광네트워크 칩에 ONet과 광소자를 집적하고 CPU칩과 3D 패키지 기술로 접합하는 것을 가정하였다.

64개 cluster를 ONet 광통신으로 연결한 것과 256bit 폭의 Emesh 전기통신으로 연결한 경우에 대해 CPU의 성능 차이를 Splash2, Parsec, Synthetic 등 3개의 Benchmark 응용 programs으로 비교하였다. 이 계산에 cache coherence protocol은 ATAC 설계자들이 제안한 ACKwise protocol과 2개의 다른 기존 protocols을 적용하였다. 결론적으로 광통신 기반의 ACKwise protocol 이 전기통신 기반의 그 어떤 protocol 조합보다 60%에서 2.5배까지 우수한 성능을 보였다. 이와 같은 성능 차이는 core의 성능이 동일한 상태에서 광통신 기반의 cache coherence만으로 나타나는 차이이므로 의미 있는 차이라 할 수 있다. 또한 열 발생을 현격히 줄이고 Emesh로 거의 불가능한 범위의 연결을 광통신으로 가능하게 하는 데에 더 큰 의의가 있다고 할 수 있다[1].

Corona 구조: HP사 연구팀과 두 대학교의 참가자가 2008년 제안한 CPU 칩 내 광통신 네트워크 구조이다. CPU 칩의 구성과 네트워크 구조의 세부사항은 다음과 같다[2].

CPU 칩 구성: 각 cluster에는 4개의 cores가 있고 총 64개의 clusters가 광통신으로 연결되어 있다. Core의 clock 주파수는 5GHz, core당 2개의 L1 caches, 4개 cores당 하나의 L2 cache를 할당하였다. 각 cluster에는 memory controller가 배치되고 DRAM과 광통신으로 연결된다. Core당 4개 threads를 가정하여 CPU는 256개 cores, 1,024개 threads를 운영하는 것으로 하였다. L2 cache block을 공유하고 광통신으로 연결된 directory 기반의 MOESI cache coherence protocol를 사용하여 전기통신으로 연결된 경우와 성능을 비교하였다.

광통신 네트워크: Cluster 내부의 4개 cores 간 통신은 별도의 설명이 없으나 일반 상용 CPU와 동일한 bus형 전기통신으로 간주할 수 있고 64개의 clusters를 Corona network 광통신으로 연결하였다. Data 전송을 위한 광도파로 256개, broadcast 광도파로 1개, arbitration 광도파로 2개, clock 신호 전송 광도파로 1개, 총 260개의 광도파로가 있고, 하나의 광도파로에는 64개 파장의 독립된 신호가 전송된다. Data 전송을 위해 각 cluster는 4개 광도파로(256개 전송 채널)를 할당 받는다. 4개의 광도파로는 해당 cluster에서 출발하여 63개를 거친 후 원래 위치로 돌아온 후 이어지지 않고 끊어진 링형(broken ring) 구조를 갖는다. 모든 cluster는 이와 같은 형태의 광도파로 4개가 출발하고 끝맺는다.

예를 들어, 1번 cluster가 60번 cluster에 data를 전송할 경우 power 광도파로에서 1번 cluster에 할당된 광도파로로 cw 신호를 분기하여 cw 신호를 60번 cluster에 전송한다. 사전에 이미 1번이 60번에 data을 전송한다는 사실을 모든 cluster에 통고한 상태이고 60번을 제외한 나머지 cluster는 링공진기를 off로 둔다. 1번에서 출발한 cw 신호는 곧바로 60번에 도착하고 60번은 보내온 cw에 링공진기 변조기를 이용하여 data를 생성한 후 1번으로 보내게 된다.

Arbitration은 같은 cluster에 다수의 cluster가 동시

에 data를 전송하는 것을 피하기 위해 순서를 정하는 행위이다. 두 개의 arbitration 도파로 중 1번 도파로에 64개 파장, 2번 도파로에 1개 파장이 할당된다. 두 도파로는 끊어지지 않고 이어진 링형 구조이다. 64개 파장은 각 cluster당 1개가 할당된다. 1번 cluster가 60번에 data를 전송하기 위해 사전에 arbitration이 시행되고 이때 1번은 60번의 파장에 token 신호를 보낸다. 60번이 받으면 나머지 모든 cluster의 token 발행은 중지된다. 이때 필요한 통고 행위는 2번 도파로(1개 파장)를 통해 broadcast 형태로 진행된다. Arbitration은 기회를 공평하게 주기 위해 1번에게 기회를 주고 1번의 data 전송이 끝나거나 포기하면 2번에게 기회를 주고, 다시 3번, 4번, …, 64번까지 순차적(Round Robin 방식)으로 진행된다. 그 외 Corona 네트워크에는 clock 신호 전송을 위해 1개 도파로에 1개 파장, broadcast 신호를 위해 1개 도파로에 64개 파장이 할당되어 있다. 이는 Arbitration broadcast와 별도이고 일반적인 목적으로 broadcast 신호를 보낼 때 사용한다.

광소자 구성: Corona 설계자는 ATAC과 같이 외부광원을 사용하는 것으로 하였고, 이 경우 최소 64개에서 최대 16,514개의 레이저가 필요하게 된다. 신호 발생을 위해 링공진기 변조기를 사용하고 각 cluster당 data 전송에 256×63개, arbitration에 65개, broadcast에 64개 등 16,257개, 전체 cluster에 총 104만여 개의 링공진기 변조기가 필요하다. 검출기의 경우 cluster당 data 수신에 256개, arbitration에 65개, broadcast과 clock에 각 1개 등 323개, 전체 cluster에 총 2만여 개가 필요하다. 파장다중(WDM: Wavelength Division Multiplexing)을 위한 링공진기 필터는 변조기와 검출기를 합한 수와 파워 분기를 위한 386개 등 총 110만여 개가 필요하다.

Corona 성능: CPU의 target 동작속도는 10TFLOPS, cluster 간 data 통신라인 수는 256개(256bit-wide), 라인당 통신속도는 10Gb/s이다. CPU 내 광통신은 총 20TB/s, CPU와 DRAM 간에는 총 10TB/s의 통신폭(bandwidth)을 가진다. CPU의 전력소모는 82W, 면적은 423mm^{2}로 계산하였고, 이 중 광통신에 소요되는 전력은 39W로 계산하였다. 광통신과 전기통신에 대한 성능 비교를 위해 synthetic benchmark과 SPLASH2 benchmark를 사용하여 simulation을 수행하였고 결론적으로 2~6배의 성능차이를 보고하였다. 또 Corona 광네트워크 구조가 열발생(power wall)뿐만 아니라 core to core, CPU to memory 간 통신 장벽(bandwidth wall)을 풀 수 있는 유력한 해답이라고 논문에서 주장하였다.

실패원인: CPU에 적용하기 위한 광통신 기반의 network 구조가 ATAC과 Corona 외에도 다수 제안되었으나 실제 칩에 실험적으로 집적하고 시험할 수 있는 단계까지도 도달하지 못한 실정이다. 오히려 이와 같은 연구가 2015년경 이후 차츰 퇴색하는 경향으로 진행되어 왔다. 그 이유는 연구가 쌓여갈수록 실현 가능성을 낮게 하는 요인들이 점차 크게 부각되었기 때문이라 할 수 있다.

첫째, SOI 웨이퍼 사용의 불리함을 들 수 있다. 현재 CPU와 memory 칩 생산에 사용되는 웨이퍼는 일반 bulk 실리콘이다. 그리고 실리콘포토닉스 연구의 추진 동기가 되었던 전자소자와 광소자를 동일한 웨이퍼와 공정(CMOS process)으로 제작할 수 있다는 사실이 실제로는 별 이득을 주지 못 한다. 실리콘 광도파로를 사용할 경우 광도파로와 광소자는 CPU의 gate와 같은 실리콘 층에 배열되어야 한다. 이는 CPU의 gate 밀도를 높이는 데 불리할 뿐만 아니라 전자소자의 메탈층(13층의 metal layers) 설계를 거의 불가능하게 한다. 이러한 사실은 초창기부터 예견되었기 때문에 그 추진 동기와 달리 광소자와 CPU 포함 모든 전자소자를 서로

다른 웨이퍼에 제작하여 3D 패키지로 접합하는 구도를 취했다. 따라서 SOI 웨이퍼를 사용하는 데 대한 정당성이 인정되었으나 CPU의 집적도가 높아지고 면적당 발열량이 커짐에 따라 열전도율을 최대한 높이기 위해 SOI 웨이퍼의 oxide (SiO_2)층의 두께가 최대한 작아야 한다. 이는 광도파로 설계를 어렵게 하는 요인이 된다. 광도파로 손실을 줄이기 위해 oxide 층의 두께가 최소 1μm 이상이어야 하나 열전도를 높이기 위해서는 〈0.2μm 두께의 웨이퍼를 사용하는 것이 유리하다[5].

둘째, 실리콘 광도파로의 전송 손실이 지나치게 크다는 점을 들 수 있다. 실리콘은 굴절률(n=3.45)이 높아 광도파로의 크기를 최대한 작게 만들 수 있는 점은 유리하게 작용하지만 이는 오히려 전송 손실을 크게 만드는 단점이 된다. ATAC에서는 0.3dB/cm를 가정하였지만 논문에 보고된 최저값은 1.7dB/cm이고 일반적인 CMOS 공정으로 제작하면 2dB/cm보다 큰 값을 나타낸다[3]. ATAC에서 광도파로의 길이는 약 8cm이고 Corona는 16cm이다. 이 경우 전송 손실이 대략 16~32dB이고 이는 전송 불가를 나타낸다. 즉, 63개의 cluster가 1/63의 세기를 분기하여 data를 취하는 것이 불가능해진다.

셋째, 광변조기, 검출기, 링공진기의 수율과 성능이 크게 못 미친다는 점을 들 수 있다. ATAC은 Ge 기반의 광변조기와 검출기를 사용하고 Corona는 링공진기 기반의 광변조기와 Ge 기반의 검출기를 사용한다. Network이 작동하기 위해서는 각 소자의 실패율이 잉여소자 비율(통상 10%)보다 낮아야 한다. 통상 수율이 90% 이상이어야 한다. Ge 물질을 실리콘 웨이퍼에 결정 성장(epitaxial growth)을 하여 변조기와 검출기를 제작할 경우 소수의 선택된 소자가 성능에 도달하는 것은 가능하나 필요한 비율만큼 균일한 수율을 확보하는 것은 매우 어렵다. 특히 링공진기의 공진 파장과 같이 통계적 오차가 수반되는 경우 이론적으로도 불가능하게 된다. 이 경우 파장 튜닝을 위해 국소적으로 micro-heater를 부착하고 열을 가해야 한다. ATAC의 경우 약 50만 개, Corona의 경우 약 110만 개의 링공진기에 열을 가할 경우 광신호 도입의 효용가치가 없어진다. 또한 ATAC은 Ge 기반의 광변조기의 입력 손실을 1dB로 가정하였으나 현재 논문에 보고되는 해당 소자의 입력 손실은 약 3~6dB로 큰 차이를 나타낸다.

넷째, 광원 문제를 들 수 있다. 인텔사는 UC at Santa barbara 대학교 연구팀과 공동으로 InP 웨이퍼를 실리콘웨이퍼에 접합하는 방식으로 on-chip 레이저 광원을 개발하였으나 Ge 기반의 광소자처럼 수율 문제를 안고 있다. ATAC과 Corona 구조는 외부 광원을 사용하는 것을 가정하였으나 수백~수천 개의 레이저 광원을 광섬유 배열과 광도파로 배열의 정렬로 접속시키는 것은 거의 불가능할 뿐만 아니라 광섬유의 단면적을 고려할 때 CPU칩의 측면에 접합할 수 있는 공간 확보도 어렵다. 한 개의 광섬유를 실리콘 광도파로에 접속시킬 때 측정되는 접속손실은 최저 1~2dB까지 보고되고 있으나 이는 광도파로 입구에 특별한 구조를 덧붙이거나 변경하여야 가능하고 일반적인 경우 약 3~6dB 이상의 접속 손실이 발생한다. 2개 이상의 광섬유를 배열로 접속할 경우 모든 광섬유의 접속 손실을 균일하게 3dB 이하로 정렬하는 것은 단일모드 실리콘도파로에 대해 거의 불가능하다고 볼 수 있다.

Ⅲ. CPU 광인터커넥션 신기술

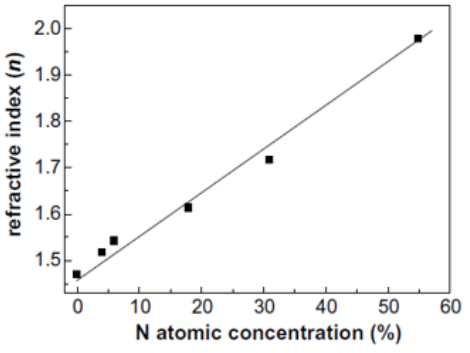
네 가지 실패 원인들은 실리콘을 광도파로로 사용하기 때문에 발생하는 문제들이라고 할 수 있으며, 만약 다른 물질을 사용하면 문제 해결이 가능

한지 살펴볼 필요가 있다. CMOS 공정에 사용되는 물질 중 광도파로에 적용 가능한 물질은 실리콘(Si), 실리콘나이트라이드(Si_3N_4), 실리콘옥시나이트라이드(SiO_xN_y), Oxide(SiO_2) 외 다른 물질은 존재하지 않는다. Oxide(SiO_2)는 도파로의 cladding으로 사용하므로 실리콘을 사용하지 않으면 후보 물질은 실리콘나이트라이드(Si_3N_4) 또는 실리콘옥시나이트라이드(SiO_xN_y)만 가능하다. 이 두 물질을 광도파로로 사용하여 광신호 전송에 관한 특성을 분석하고 광소자를 제작하여 소자특성을 측정한 논문은 많이 있다.

그러나 두 물질을 사용하여 CPU 칩 내 집적 가능한 network architecture를 제안하거나 실험한 논문은 보고된 바가 없다. 먼저 논문에 보고된 자료를 근거로 실리콘나이트라이드(Si_3N_4) 또는 실리콘옥시나이트라이드(SiO_xN_y)의 특성을 분석하고 위에 열거된 문제들을 해결하기 위해 가능한 네트워크 구조를 알아본다.

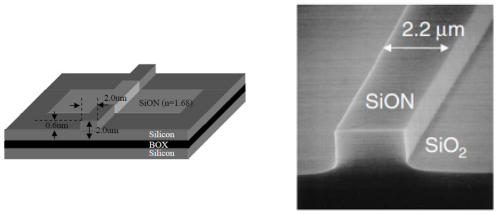
첫째, CMOS 공정에서 실리콘나이트라이드 또는 실리콘옥시나이트라이드 박막 도포(deposition)는 LPCVD(low pressure chemical vapor deposition) 또는 PECVD(plasma enhanced CVD) 공정으로 이루어진다. LPCVD는 화학적 성분 비율(Si_3N_4의 3과 4)이 정확하게 갖추어진(stoichiometric) 실리콘나이트라이드 박막을 도포할 수 있다. 도포 온도는 425~900℃ 범위에서 가능하나 CMOS 공정에 쓰이는 stoichiometric 박막의 경우 약 770℃ 근처에서 가장 양호한 특성을 나타내며 굴절률은 2.0이다. 광도파로로 제작할 때 전송 손실 역시 가장 양호하며 〈0.1dBcm의 측정값이 보고되어 있다. 단점은 tensile stress로 인해 200nm보다 두꺼울 경우 crack(박막에 갈라진 금)이 발생한다. CMOS공정의 metal layers에 사용되는 알루미늄과 구리의 녹는점이 각각 660℃, 1,084℃이고 구리의 경우 고온에서 다른 박막으로 침투되는 불순물 소스가 되고 알루미늄의 경우 융점보다 높으므로 LPCVD 박막은 첫 번째 metal layer 이전 단계까지만 사용된다. PECVD 공정은 200~400℃에서 도포가 가능하고 도포 속도가 780nm/min(LPCVD 4.8nm/min) 빠르므로 두꺼운 박막 형성에 유리하며 첫 번째 metal layer 이후부터 사용된다. PECVD 실리콘나이트라드 박막은 화학적 성분비가 맞지 않은 대신 stress가 상대적으로 작아 수 마이크로미터 두께를 crack 없이 도포할 수 있다. 광소자와 전자소자를 동일한 웨이퍼에 집적하기 위해 광도파로는 I 부 그림 5의 맨 위 보호막 층에 형성하는 것이 가장 유리하다. 따라서 CMOS 공정으로 광통신을 구현하기 위해서는 PECVD 박막을 사용하는 것이 가장 유리하고 앞서 제시된 첫 번째 실패 원인을 해결할 수 있게 된다.

둘째, PECVD 실리콘나이트라이드와 실리콘옥시나이트라이드 박막을 이용하여 광도파로를 형성하고 특성을 측정한 논문은 다수 있다. 실리콘옥시나이트라이드(SiO_xN_y)는 도포 시 주입 가스의 비율을 조절하여 박막의 굴절률이 특정값을 갖도록 할 수 있다. 그림 2에 나타나 있는 바와 같이 박막 내 질소 성분이 0%일 경우 PECVD 옥사이드(SiO_2)



출처 M. I. Alayo et al., "Deposition and charatrization of silicon oxynitride for integrated optical applications," Journal of Non-Crystalline Solids, 2004, pp. 76-80.

그림 2 질소 성분비와 굴절률 관계 그래프



출처 [a] Richard Jones et al., "Integration of SiON gratings with SOI," Conference: Group IV, Photonics, 2005.

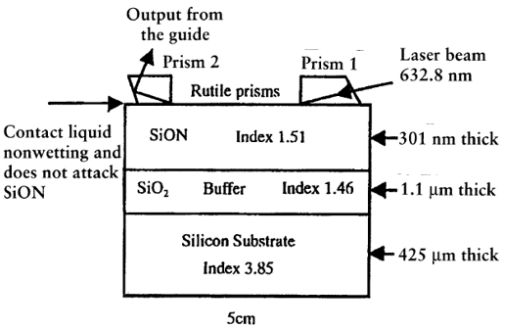
[b] Tai Tsuchtza, "Low-loss Silicon Oxynitride Waveguides and Branches for the 850nm Wavelength Region," Japanese Journal of Applied Physics, vol. 47, no.8, 2008, pp.6739- 6743.

그림 3 SiON 광도파로 (a) n=1.68, (b) n=1.515

박막과 같은 굴절률(1.45)을 나타내고 질소성분이 증가함에 따라 굴절률이 2.0까지 선형적으로 증가한다[6]. 따라서 PECVD 실리콘나이트라이드와 실리콘옥시나이트라이드, 둘 중 하나를 사용하거나 적절히 조합하여 광도파로를 제작하면 기존의 CMOS 공정규정을 준수하며 CPU와 동일한 면적에 입체적으로 집적할 수 있게 된다. 참고문헌 [7]에 따르면 PECVD SiON 박막의 굴절률과 전송손실은 그림 3(a)와 같은 광도파로 구조에 대해 각각, n=1.68, 0.3±0.15dB/cm이며 측정에 사용된 광파장은 1,310nm이다. 참고문헌 [8]에 따르면 그림 3(b)와 같은 광도파로 구조에 대해 굴절률과 전송손실이 각각, n=1.515, TE 모드 0.2dB/cm, TM 모드 0.3dB/cm이고, 측정에 사용된 광파장은 850nm이다. 따라서 PECVD 실리콘옥시나이트라이드 광도파로의 전파 손실이 약 0.3dB/cm 근처에 있다고 할 수 있고, 이는 ATAC 구조에서 설정한 광도파로의 전송손실과 동일한 값이다. 그리고 위에 제시된 두 번째 실패 원인이 해결됨을 의미한다.

셋째, 실리콘 물질의 광도파로보다 더 일찍 실리콘나이트라이드 광도파로에 대한 연구와 논문이 있었음에도 불구하고 실리콘포토닉스 연구에서 전자를 택한 이유는 전류 또는 전압을 가하여 작동하는 광 능동 소자(active optical devices; 예, 광변조기, 광스위치 등)의 성질을 후자는 보유하지 못하였기 때문이다. 그러나 실리콘의 경우 앞에서 지적한 바와 같이 광 능동 소자의 성질을 보유 하였어도 제작한 소자의 성능과 수율이 온칩 네트워크 구성을 위한 수준에 크게 미치지 못 한다. 이에 반해 표면방출레이저(VCSEL)의 기술은 획기적으로 발전하여 2000년 초·중반에는 직접 변조(direct modulation)가 5GHz 수준이었으나 2015년경부터 50GHz의 제품이 상용화되어 Ⅱ부 표 4에 나타낸 바와 같이 HDR 제품이 슈퍼컴퓨터 제작에 이미 사용되고 있는 수준에 이르렀다. 이 성능은 실리콘 광도파로로 구현된 광변조기보다 더 우수하다. 따라서 laser 광원과 광변조기를 서로 다른 소자로 구성하는 광네트워크보다 VCSEL을 이용하여 광원과 광변조기를 동일한 소자로 구성하는 것이 면적과 전력 측면에서 더 효율적이다. VCSEL를 사용하여 실리콘옥시나이트라이드 광도파로에 광신호를 직접 입사할 수 있는 기술과 파장다중(WDM)을 위해 링공진기 필터를 사용하지 않고 성능이 검증된 박막 필터(thin film filter)를 이용하는 기술이 확보될 경우 온칩 광네트워크는 기존의 검증된 소자와 기술로 구현 가능하게 된다. 또한 위에서 열거된 셋째, 넷째 실패 원인을 동시에 해결하게 된다.

ATAC과 Corona 구조를 비롯하여 기존 논문들에서 제안하고 있는 온칩 광네트워크 구조는 과도하게 높은 성능의 광소자와 광통신라인 폭을 요구하고 있다. 예를 들어 ATAC에서 최소 129개 광도파로와 각 도파로당 64개 파장의 파장다중 기술을 요구하고 있고, Corona는 이보다 더 높은 260개 광도파로에 파장다중 기술을 요구하고 있다. Cluster당 통신라인 폭에 해당하는 광소자의 수를 집적하는 것은 거의 불가능하다. 가장 합리적인 방법으로 단기간 내에 실현 가능한 광네트워크 구조를 찾아야



출처 B. S. Sahu et al., "Influence of hydrogen on losses in silicon oxynitride planar optical waveguides," Semicond. Sci. Technol. vol. 15, 2000.

그림 4 SiON 물질의 전파 손실 측정

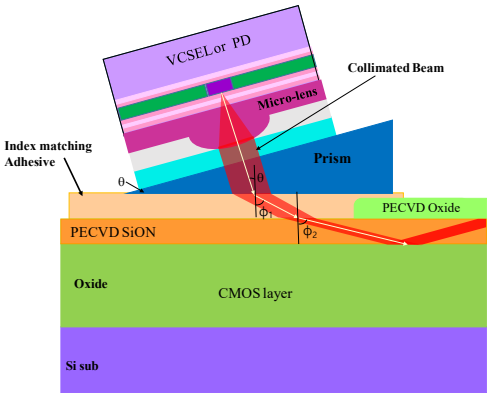


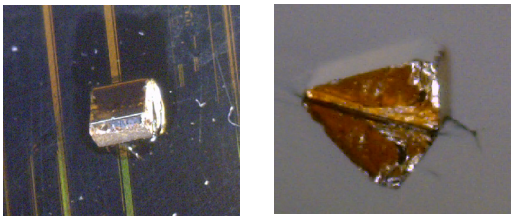
그림 5 SiON 광도파로에 대한 광신호 접속 개념도

한다. 슈퍼컴퓨터에 이미 활용되고 있는 VCSEL과 PD를 도입하고 fat tree 네트워크 구조를 그대로 적용하되 단지 광섬유만 실리콘옥시나이트라이드 광도파로로 교체하여 II부 그림 6부터 그림 9의 인텔 서버용 CPU에 적용하는 것이 가장 합리적이고 불확실성을 낮게 한다고 할 수 있다. 이를 위해 필요 충분한 3가지 기술을 한국전자통신연구원의 start-up 기업(필자가 설립함)이 개발하고 실험적으로 검증한 사실은 의미가 크다고 할 수 있다. 다음은 3가지 기술에 대한 실험과 인텔의 서버용 CPU에 적용할 수 있는 광네트워크를 기술한다.

PECVD 실리콘옥시나이트라이드 광도파로에 빛을 입사시킬 수 있는 방법 중 silicon photonics 연구에 많이 도입되었던 광섬유 끝과 광도파로 끝을 맞대는 방법(butt coupling)과 광도파로 표면에 회절격자(grating)를 새겨 빛의 회절(grating coupling)을 이용하는 방법은 여기에 적용하기 어렵다. Butt coupling은 칩의 변두리에서만 신호 생성이 가능하므로 변두리에 위치하지 않는 core의 신호는 생성이 불가능하다. Gating coupling의 경우 격자가 새겨지는 물질, 즉 실리콘옥시나이트라이드(SiON)와 옥사이드(SiO_2)의 굴절률차가 작아 (Δn=0.1~0.5) 접속율이 낮을 뿐만 아니라 링공진기와 같이 통계적 오차가 발생하는 방법이므로 수율 문제를 극복하기 어렵다. 따라서 위 두 방법을 제외하면 유일하게 가능한 방법은 프리즘을 이용하는 방법이다. 이는 광학 도서와 다수의 논문에서도 용도는 다르지만 소개되어 있다.

그림 4는 실리콘옥시나이트라이드 물질을 통과하는 빛의 전파손실을 측정하기 위해 프리즘을 사용하고 있다[9]. 이 경우 단순히 굴절률 정합(index matching oil)에 필요한 오일을 프리즘과 실리콘옥시나이트라이드 물질 표면에 바르고 빛을 굴절시켜 측정할 수 있다. 그러나 광도파로에 빛을 입사시키는 소자로 사용하기 위해서는 다음과 같은 3가지 기술을 해결하여야 한다.

첫째, 굴절률 정합 조건을 만족하고 투명하며 200~400℃에서 접착 강도를 유지하고 CMOS 세척공정에 사용되는 BOE, isopropanol, aceton 등 용제에 영향을 받지 않는 접착 기술이 개발되어야 한다. 상용제품 중 굴절률이 1.6보다 크고 투명한 접착제는 판매되고 있지 않다. 세척 용제와 400℃까지 접착 강도를 유지하는 접착제는 논문에서도 보고되어 있지 않다. 따라서 상용 접착제를 사용하지 않고 프리즘을 광도파로 표면에 접착할 수 있는 기술을 개발하여야 한다.



(a) (b)

그림 6 SiON 광도파로에 프리즘을 접착한 실물사진

둘째, 그림 5와 같은 조건에서 평행광을 만들 수 있는 micro-lens가 개발되어야 한다. 상용제품은 Quartz와 실리콘 재질의 micro-lens가 판매되고 있으며 굴절률이 현재의 용도에 부합하지 않는다.

셋째, VCSEL과 PD에 대해 초소형 비방습형(non-hermetic) 패키지 기술이 개발되어야 한다. 현재 상용 제품의 패키지는 laser 빛의 출구와 렌즈 간 일정거리를 유지한 후 질소 또는 아르곤 가스를 채우고 방습(hermetic) 밀폐한 구조이다. 이와 같은 패키지 공정은 부피를 과도하게 차지하기 때문에 수십~수백 개의 VCSEL과 PD를 CPU칩 면적 위에 집적할 수 없다. 패키지 후 단일 광소자의 면적과 높이가 최대 〈0.7mm²과 〈1.0mm 정도이어야 한다. ETRI의 start-up 기업은 폴리이미드의 특수한 성질을 이용하여 위 특성을 만족하는 접착 기술과 마이크로렌즈, 그리고 초소형 비방습형 패키지 기술을 개발하였다.

표 2 프리즘 접착 특성

특성	
굴절율	1.69~1.72 at 850nm
투과율	＞99% for 2μm thickness at 850nm
고온저항	400℃(compatible with CMOS metal process)
솔벤트 저항	BOE, Aceton, Isopropanol 등(unaffected by alcoholic cleaning)
접착강도	＞1N/mm² (not detachable before breaking GaP prism at 25℃)

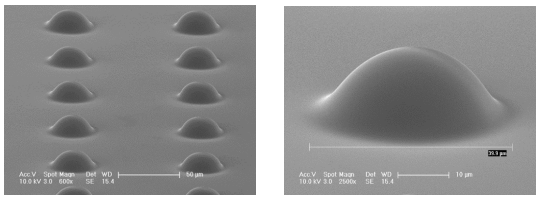


그림 7 마이크로 렌즈 SEM 사진

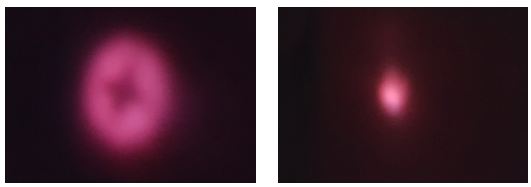


그림 8 마이크렌즈 부착 전 후 VCSEL 출력광 사진

그림 6(b)는 GaP 물질의 프리즘을 SiON 광도파로 표면에 접착한 현미경사진이다. (b)는 측면의 모습을 확대한 사진이다. 프리즘과 웨이퍼 간 접착 특성을 표 2에 나타내었다. 접착에 필요한 두께, 2μm에서 99% 이상의 투과율을 나타내었고 굴절률은 고온처리 조건에 따라 1.72까지 측정되었다. 접착 강도는 측면에서 힘을 가할 경우 프리즘이 부서지는 경우가 떼어지는 경우보다 더 많이 일어난다. 폴리이미드 물질의 특성에 따라 고온저항과 솔벤트 저항은 표 2와 같이 정해져 있고 실재 측정에서도 동일한 결과를 보여준다.

VCSEL의 공기 중 방사각은 통상 20~40° 범위에 있다. 중간 값인 30°에 대해 마이크로렌즈는 평행광을 만들 수 있는 굴절률 조건을 만족해야 한다. 그림 7은 제작한 마크로 렌즈 배열의 SEM(scanning electron microscope) 사진과 하나를 확대한 사진이다. 마이크로렌즈의 지름은 약 30μm이다.

그림 8과 같이 25GHz VCSEL 칩에 마이크로렌즈를 부착한 후 출력광의 모양을 관찰하였다. VCSEL에서 수직거리 7cm 위치에 적외선 카드(적외선 감지용 infrared card)를 두고 출력광을 찍은 사

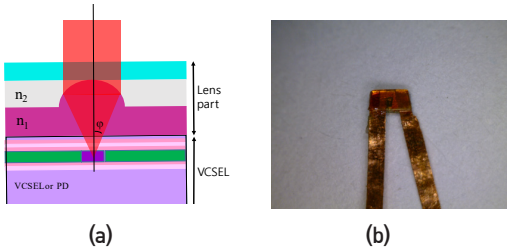


그림 9 VCSEL의 비방습 패키지 모듈, (a) 패키지 구조도, (b) 실물 사진

진이다. VCSEL에서 광도파로까지 프리즘을 통과하는 거리는 〈1mm이므로 사진에 나타난 평행광의 성능은 충분히 발휘되는 수준에 있다.

마이크로렌즈의 시험결과를 이용하여 비방습 패키지를 완성할 수 있다. 그림 9(a)는 패키지 구조도를 나타낸다. 렌즈 부분을 별도로 제작한 후 최종 단계에서 VCSEL에 접착하였다. 그림 9(b)는 비방습 패키지를 수행한 VCSEL의 현미경사진이다. 밑면이 넓이가 1.2×1.0mm^2, 높이가 640mm 이다. 광접속 검증용으로 제작하였기 때문에 최소 크기를 나타내지 않는다. VCSEL의 크기가 0.25×0.25mm^2 이므로 〈0.7mm^{2}는 어려운 수준이 아니다. 엄격한

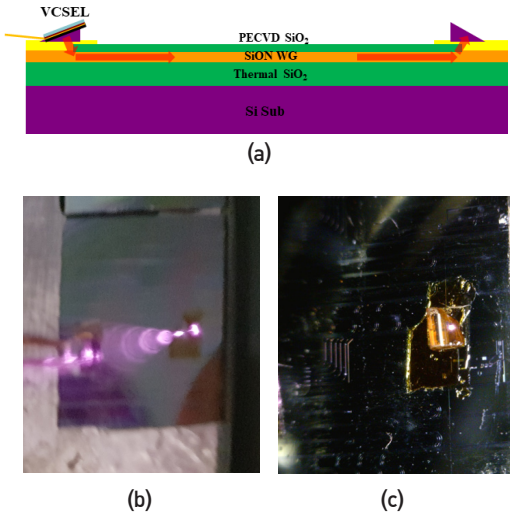


그림 10 광신호 접속 실험 (a) 실험 구도, (b) 실험 사진, (c) 반대편 프리즘에 빛이 출력되는 사진

수준의 신뢰성 검사를 거치지 않은 상태이나 hot plate에서 온도 범위 25~150℃, 고온 지속시간 〉60min, 10회 이상 반복 후 뚜렷한 출력 세기 변화가 나타나지 않았다.

그림 10(a)에 나타낸 바와 같이 SiON 광도파로의 양 끝에 프리즘을 접착하고 한쪽 프리즘에 그림 9(b)의 VCSEL 모듈을 부착하였다. VCSEL의 빛이 광도파로에 입사하고 반대편 프리즘으로 나오는 빛의 세기를 측정하는 실험을 하였다. 그림 10(b)는 실험에 사용된 레이저 빛이 광도파로에 입사한 뒤 반대편 프리즘으로 나오는 모습을 보여주는 현미경사진이다. 왼편에 두 선의 레이저 전극을 볼 수 있다. 그림 10(c)는 주변의 밝기를 더 세게 하여 산란된 레이저 빛이 보이지 않는 상태에서 반대편 프리즘으로 나오는 레이저 빛을 보여주는 현미경사진이다. 프리즘의 일부가 파손된 상태에서도 레이저 빛이 프리즘 빗면 중앙에서 강하게 나오는 모습을 볼 수 있다. 다수의 모듈과 광도파로에 대해 측정한 결과 접속 손실 평균값은 약 3.54dB이고, 이는 프리즘에 반사 방지 박막(AR: Anti-Reflection coating)을 도포하지 않은 상태의 값이다. 반사 방지 박막(AR coating)을 프리즘 양면에 도포할 경우 이론적으로 2.6dB의 이득을 얻게 되므로 접속 손실은 약 0.94dB가 된다. ATAC에서 설정한 접속 손실이 3dB이므로 훨씬 더 양호한 값을 얻을 수 있다. 입사광이 평행광이고 굴절률 정합 조건을 정확히 맞출 경우 반사 손실을 제외한 이론적 최저값은 0dB 근처에 있다.

그리고 위 실험에서 1dB 손실이 발생하는 정렬유격(alignment tolerance)은 광도파로와 평행한 방향으로 18.9mm, 수직한 방향으로 7.2mm의 측정값을 나타내었다. 수직한 방향의 정렬유격은 측정에 사용한 광도파로(폭 30mm)보다 넓은 광도파로를 사용하여 자유롭게 조정이 가능하므로 광도파

로와 평행한 방향의 정렬 유격이 중요하고 측정한 값(18.9mm)은 대량생산 시 자동화할 수 있는 범위에 있다고 볼 수 있다.

실리콘 포토닉스 연구에서 하나의 광도파로에 다수의 광파장을 전송하기 위해 링공진기 파장다중 필터(WDM filter)를 가정하였으나 그 성능이 열악하여 on-chip뿐만 아니라 off-chip에서도 상용제품으로 사용되지 않고 있다. 현재 off-chip 광통신에 사용되는 상용제품 중 성능이 가장 검증된 파장다중 필터는 굴절률이 서로 다른 두 박막을 번갈아 적층한 박막 필터이다. 그림 11(a)는 프리즘 밑면에 파장다중 박막 필터를 코팅하여 4개의 광파장을 하나의 광도파로로 전송하는 구도를 나타낸다. 예를 들어, λ_3의 경우 해당 파장의 필터가 도포된 프리즘은 통과하여 도파로 내부로 입사되고 다른 파장의 필터가 도포된 프리즘에서는 프리즘 밑면에서 반사되어 도파로를 따라 그대로 진행한 후 해당 파장의 프리즘에서 방출된다.

그림 11(b)는 파장다중 필터의 구조와 빛의 진행경로를 도식적으로 나타낸다. 파장다중 박막 필터를 설계하기 위해 자주 사용되는 TiO_2와 Ta_2O_5 박막을 이용하여 컴퓨터 프로그램으로 도출한 반사

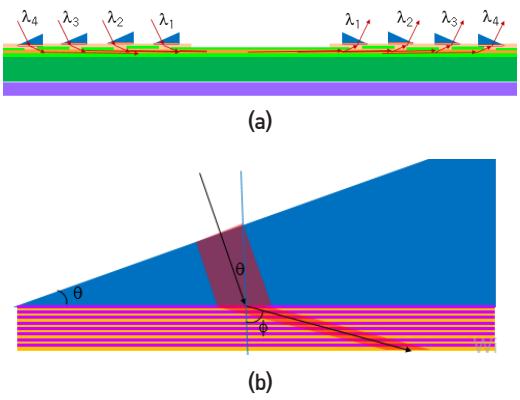
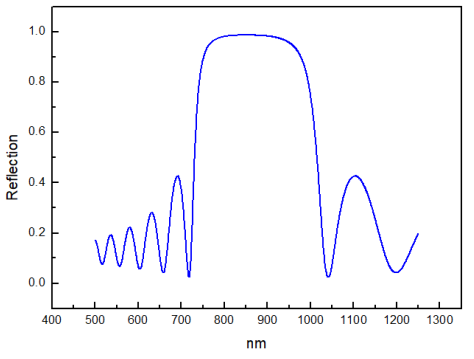
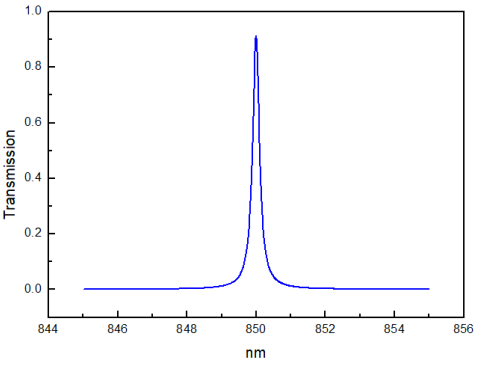


그림 11 박막 필터를 이용한 4채널 파장 다중
(a) 빛의 입출력 구도, (b) 박막 필터와 빛의 경로



(a)



(b)

그림 12 컴퓨터로 계산한 박막 필터 스펙트럼
(a) 반사 스펙트럼, (b) 투과 스펙트럼

스펙트럼과 투과 스펙트럼 그래프를 그림 12에 나타내었다.

표면방출레이저의 파장 850nm에 대해 프리즘의 굴절률은 $n_{prism}=3.16$, TiO_2 박막의 굴절률은 $n_H=2.5086$, Ta_2O_5 박막의 굴절률은 $n_L=2.0908$이다. 프리즘의 각도가 θ일 때 빗변에 수직으로 입사한 빛은 밑변에 θ로 입사하고 굴절률 조건에 의해 박막에서 ϕ로 진행한다. 각 박막은 해당 각도로 진행하는 경로가 λ/4가 되도록 두께가 정해지며 TiO_2 박막은 T_H, Ta_2O_5 박막은 T_L로 표기한다. 그림 12(a)는 T_H 박막과 T_L 박막을 번갈아 가며 8회 반복$[(T_HT_L)^8]$할 경우 측정되는 반사 스펙트럼이다. 그림 12(b)의 투과 스펙트럼은 T_H 박막

과 T_L 박막을 8회 반복하여 형성한 두 반사 거울 사이에 4λ 두께의 스페이서를 삽입[$(T_HT_L)^8(T_HT_H)^8(T_LT_H)^8$]할 경우 측정되는 투과 스펙트럼이다. 이와 같은 컴퓨터 계산은 기술적으로 매우 성숙되어 있고 필터 제작 후 측정 그래프와 정확히 일치하며 상용제품을 생산하는 데 사용되고 있다. ATAC과 Corona 구조에서 요구하는 파장다중 성능을 상용 제품의 박막 필터는 만족하는 수준에 있다.

ATAC과 Corona 구조는 과도하게 높은 성능과 광통신라인 폭을 요구하고 있기 때문에 실현성이 낮다고 볼 수 있다. 그러나 실험적으로 검증한 광접속 기술과 컴퓨터 계산을 통해 증명한 파장다중 필터 기술을 활용하면 슈퍼컴퓨터의 fat tree 네트워크 구조를 응용하여 ATAC과 유사한 성능의 CPU, 즉 1,024개 core로 구성된 광네크워크를 웨이퍼에 구현할 수 있다. 그리고 현재 슈퍼컴퓨터에 파장다중 기술을 사용하지 않고 광능동케이블(AOC)을 이용한 일대일(point-to-point) 통신만을 사용하므로 광접속 기술과 실리콘옥시나이트라이드 광도파로는 슈퍼컴퓨터의 fat tree 네트워크를 CPU 칩으로 가지고 오기 위한 충분한 구성 요소가 된다.

Ⅱ부 그림 6에 보인 바와 같이 Intel 서버용 CPU die는 694mm²에 최대 28개 cores가 mesh network의 전기통신으로 연결되어 있다. Desktop 용 CPU die

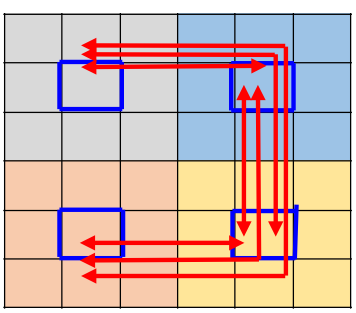


그림 13 36개 core의 CPU와 광통신의 적용 (청색선: 전기통신 ring bus, 적색선: 광통신)

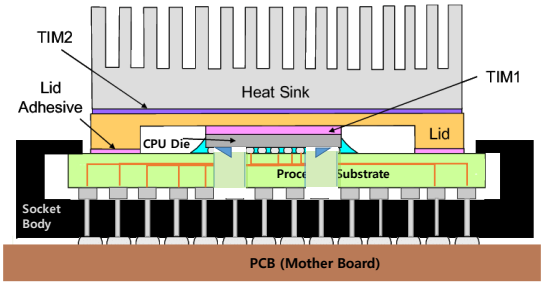


그림 14 프로세서 기판 가공 후 flip-chip bond 패키지

는 ~174mm²에 최대 8개 core가 링형 bus로 연결되어 있다. 서버용 CPU에 광통신을 도입할 경우 그림 13과 같이 9개 core를 링형 bus로 연결하고 36개 core를 4개의 cluster로 나누어 광통신으로 연결할 수 있다. 링형 bus는 인텔 desktop 상용 CPU의 전기통신 구조를 그대로 사용한다. 광통신은 crossbar 구조를 취하도록 한다. 즉, 각 cluster는 나머지 3개의 cluster와 일대일 통신라인을 갖는다. 따라서 동시에 2개 이상의 cluster가 같은 cluster에 data를 전송할 수 있고 arbitration이 요구되지 않는다. 또한 모든 광통신라인은 동일한 latency(3cycles)를 갖는다. 이 두 성질은 ATAC과 동일하다. CPU가 3GHz, 64bit을 지원할 경우 통신라인당 최소 통신속도는 192Gb/s 이다. 즉 그림 13의 각 통신라인은 단 방향 256Gb/s, 양방향 512Gb/s의 통신속도를 필요로 한다. 이는 ATAC에 비해 2배 큰 통신속도이다. 각 cluster는 50Gb/s VCSEL과 PD 각 15개씩, 총 30개의 면적이 필요하다. 실험에서 VCSEL과 PD의 개별 패키지 면적이 〈1mm²이므로 최대 〈30mm²이다. 인텔의 상용 CPU 면적을 고려하면 각 cluster당 면적은 대략 170mm²이므로 충분히 여유가 있다. 동일한 면적에 대해 3GHz, 256bit CPU까지 지원 가능하다. 그리고 모든 광도파로와 광소자는 metal 층 상부의 보호막 층에 형성되므로 CPU core 및 전기 통신선과 같은 면적 위에 올 수 있다.

그림 14와 같은 패키지 구조는 CPU를 프로세서 기판에 뒤집어 접합(flip-chip bond)하므로 프로세서 기판을 특별히 가공할 필요가 있다. 즉, 광소자가 있는 부분을 잘라내고 도선과 전극을 적절히 배치한다. 그림 13의 각 통신라인은 10개의 광도파로로 구성되고, 5개의 VCSEL, 5개의 PD가 각 화살표에 배치되는 것을 나타낸다. 파장다중 기술을 적용할 경우 1개 도파로에 5개 파장의 신호를 동시에 전송할 수 있고, 이는 현재 VCSEL 성능으로 가능한 기술이다. 다만 면적과 비용을 고려하여 파장다중 적용 여부를 결정할 필요가 있다.

그림 13에는 총 6개의 광통신라인이 있고 통신속도 전체 합은 3.072Tb/s 이다. 이를 근거로 동일한 통신속도에 대한 전기통신과 광통신의 전력소모를 비교하면 다음과 같다. 전기통신의 경우 1Gb/s당 약 2mW를 소모하고 평균 5번 hopping하므로 총 30W를 소모한다. 광통신의 경우 1Gb/s당 약 0.08mW를 소모하고 1회 hopping하므로 총 0.25W를 소모한다. 28개 core CPU에 대한 총 소비전력이 약 200W 임을 고려하면 30W는 TDP 기준 한계 성능을 산정하는 데 상당한 비중을 차지한다. 광통신은 50Gb/s당 3cycles의 latency가 발생하므로 총 184cycles의 latency가 발생한다. 전기 통신의 경우 3Gb/s(3GHz clock 주파수)당 평균 5cycles의 latency가 발생하므로 총 5,120cycles의 latency가 발생한다. 즉, 대략 5,000cycles의 시간 이득을 얻게 되고 이는 cache coherence에 반영된다.

그림 15는 1,024개 core를 광통신으로 연결하는 fat tree 네트워크 구조의 한 예를 나타낸다. 각 Part(P1, P2 등)는 8개의 코어를 포함하는 cluster를 나타내지만 광통신에 연결되는 memory controller, GPU 등을 포함할 수 있다. CU(Controller Unit)는 슈퍼컴퓨터의 Infiniband와 같은 기능의 32ports (하향 16ports, 상향 16ports) 스위치이고 내장되는

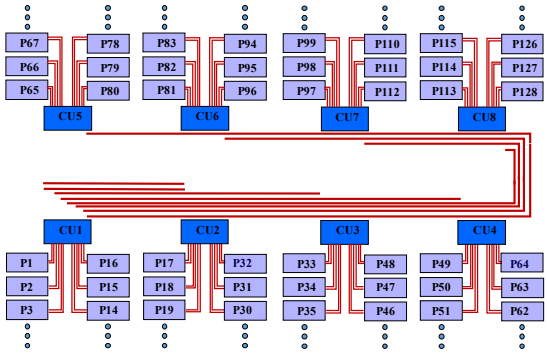


그림 15 1,024개 core의 fat-tree 네트워크

processor 칩과 스위치 칩, 각종 아날로그 칩은 CPU core와 동일하게 집적된다. 8개의 CU가 있고 각 CU에는 16개의 parts가 연결되어 있으며 총 core 수는 1,024개이다. 각 part의 8개 core는 인텔의 desktop CPU와 동일한 링 bus 전기통신으로 연결되어 있다. CPU가 3GHz, 64bit을 지원할 경우 앞에서와 같이 각 part와 CU 간 통신라인은 양방향 512Gb/s의 통신속도를 나타낸다. 모든 CU와 CU 간 통신은 2개 ports씩 할당되고 인접 CU 간은 4개 ports가 할당된다. 예를 들어 CU1와 CU2 사이는 양방향 2048Gb/s의 통신속도가 할당되고 CU1과 CU3~CU8에는 양방향 1,024Gb/s의 통신속도가 할당된다. 그림에는 CU1에서 출발하는 통신라인만 그려져 있으나 동일한 통신라인이 CU2~CU8에도 반복된다. 편의상 평면에 펼쳐 그렸지만 CPU core와 광도파로, 그리고 광도파로와 광패키지 모듈은 서로 다른 층에 형성되므로 동일한 면적에 겹쳐 질 수 있다. 따라서 한 개 part(8개 cores)의 면적이 약 50mm^{2}가 되도록 CPU를 설계할 경우 각 part에 필요한 광도파로와 광소자는 그 면적 위에 올 수 있고, 1,024개 cores를 포함하는 웨이퍼 전체 크기는 8cm×8cm가 된다. 현재 최고 성능의 슈퍼컴퓨터에 사용되는 광통신은 fat tree 구조에 양방향 200Gb/s을 사용하므로 최소 2배 이상 빠른 통신속도를 가

지고 있을 뿐만 아니라 모든 core 간 통신라인 폭 64bit/cycle을 유지하게 된다. 다만 광통신을 통과할 때마다 serialize/deserialize(SerDes) 과정을 거쳐야 하고 이는 슈퍼컴퓨터뿐만 아니라 AMD의 Infinity fabric(전기통신)에도 사용되고 있다.

모든 광통신라인을 다 사용할 경우 발생하는 통신속도 합은 98Tb/s이다. ATAC은 8.192Tb/s, Corona는 160Tb/s의 광통신속도 합을 가지고 있다. 개별 통신라인에 대한 통신속도가 2배 빠르므로 CU를 1회(3cycles) 또는 2회(6cycles) 거치며 발생하는 latency를 제외한 나머지 특성은 대부분 ATAC과 비슷하거나 우위에 있다. 다만 전기적 스위치를 거치며 발생하는 latency는 기술에 따른 불가피한 특성이다. 앞에서와 같이 동일한 속도 합을 전기통신으로 전송할 경우 발생하는 전력소모와 latency를 비교하면 다음과 같다. 전기통신의 경우 1,024개 cores를 32×32metric mesh network으로 연결하면 대략 평균 32회 hopping(32cycles latency)한다고 가정할 수 있다. Cluster 내부 이동에 따른 평균 4회 hopping을 제외하면 약 28회 hopping이 광통신 이동거리에 해당한다. 광통신의 경우 모든 광통신라인은 1회 hopping이므로 위 네트워크의 경우 모든 광신호는 2회 또는 3회 hopping 후 목적지에 도달한다. Latency는 최소 7cycles, 최대 12cycles를 가지므로 평균 9cycles를 가정한다. 98Tb/s에 대해 전기통신의 전력소모량은 대략 5,488W, 광통신의 전력소모량은 약 23.5W이다. Latency 합은 전기통신의 경우 약 915,000cycles, 광통신의 경우 약 18,000cycles이다. 소모 전력과 latency 모두 큰 차이를 나타냄을 알 수 있다. 이와 같은 수치 차이 보다 광통신이 적용될 경우 구현이 가능하지만 전기통신의 경우 통신라인 폭 64bit/cycle을 유지하며 1,000개 cores를 웨이퍼에 구현할 수 없다는 것이 더 큰 차이라 할 수 있다. CMOS 포토공정에서 한

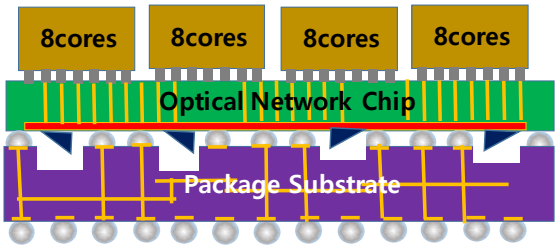


그림 16 2.5D 패키지와 광 네트워크 칩

번에 노광할 수 있는 크기(노광면적)는 대략 2cm×2cm이다. 그림 15에서 광도파로는 선폭(30~50μm의 멀티모드)이 충분히 넓어 한 번에 노광할 수 있는 크기 내에 있지 않아도 패턴 연결이 가능하지만 전기통신 선은 거의 불가능하기 때문이다.

인텔 패키지 구조를 이용하여 앞에 언급한 바와 같이 프로세서 기판을 특별히 가공하여 동일한 구조로 패키지할 수 있다. 그러나 그림 16과 같이 2.5D 패키지 구조를 도입할 경우 생산 능률을 획기적으로 높일 수 있다. 그림 16의 2.5D Interposer는 실리콘 웨이퍼이고 CMOS 공정을 사용하여 제작한다. Interposer의 아랫면에 광도파로와 광소자를 형성하고 윗면에는 flip-chip bond 전극만을 형성하여 TSV로 윗면과 아랫면을 연결한다. 윗면에는 CPU와 DRAM, GPU, SoC 등 필요한 칩을 flip-chip bond로 접합한다. 아랫면은 적절한 높이와 배치를 맞추어 패키지 기판과 BGA로 접합한다. 이렇게 할 경우 Interposer는 광 네트워크 칩(ONC: optical network chip)으로써 제3의 기업이 독자적으로 제작하고 CPU와 memory 기업으로부터 해당 칩을 공급 받아 패키지 단계에서 결합할 수 있다. x86 서버용 CPU뿐만 아니라 기술과 수요가 늘어남에 따라 mobile PC용 CPU와 스마트폰 SoC까지 다양한 유형으로 광통신을 이용할 수 있게 된다. 또한 CPU 제작에서 개별 core에 대한 수율이 90%이라 하여도 28개가 모두 양호할 확률은 5%로

떨어진다. 따라서 수십~수백 개 cores를 가진 CPU 칩을 생산하는 것보다 4~8개 cores를 가진 양호한 CPU 칩을 생산하여 결합하는 것이 훨씬 경제적이며 제작 원가를 낮출 수 있다.

표 3(a)와 같이 인텔사가 현재 판매하고 있는 서버용 28개 코어 CPU 가격은 $10,000~$18,000 범위에 있다. 그리고 desktop용 8개 코어 CPU 가격은 $440~$600 범위에 있다. 패키지 전 8개 코어 CPU die 원가를 약 $300로 추정하고 4개 die를 패키지 단계에서 flip-chip bond로 조립하는 경우, 그리고 그림 16과 같이 광 네트워크 칩(ONC)으로 4개 die를 연결할 경우 현재 판매되는 광소자 가격을 근거로 생산가격을 산출하면 다음과 같다. 2.2GHz, 64bit CPU에 대해 64bit 신호 전송을 위한 최소 통신속도는 140.8Gb/s이다. Lane당 25Gb/s, 6개 lanes, 150Gb/s가 필요하다. 표 3(b)에 나타낸 바와 같이 현재 판매되고 있는 가격(4개 lane $308)을 기준으로 6개 lanes의 광소자 가격은 $462이다. 그림 13의 광통신 네트워크는 6개의 통신라인으로 구성되고 각 통신라인은 양방향 150Gb/s(6개 lanes)를 필요로 한다. 6개 통신라인에 대한 광소자 비용 총액은 $2,772이다. AOC의 패키지 비용을 빼면 실재 광소자 비용은 최대 $2,000 미만이다. VCSEL과 PD 모듈 및 프리즘을 제외한 ONC 웨이퍼 가격은 최대 $1000 미만으로 추정한다. ONC 웨이퍼에는 CMOS 공정으로 광도파로, LD driver, TIA, TSV 등이 배치된다. 이상을 종합하면 광네트워크 연결 32개 코어 CPU 가격은 최대 $4,200이고 이는 현재 판매되는 28개 코어 서버용 CPU의 약 1/3이다. 전력소모와 latency를 고려하면 월등한 성능과 가격 경쟁력을 갖추게 된다. 스마트폰의 경우에도 SoC, 5G 모뎀, SDRAM, External GPU와 NPU 등 mother board에 있는 다수의 전자소자를 ONC에 배치할 수 있고, 이 경우 전체 하드웨어 크기뿐만 아니라 가격 경쟁에서도 점차 우위를 점할 것으로 예측된다.

표 3 (a) 인텔 CPU, (b) 100Gb/s QSFP

모델	i9-9980HK	Xeon Platinum 8276M
코어수	8	28
주파수	2.4GHz	2.2GHz
Photo	14nm	14nm
현재 가격	$583	$11,722

(a)

Form factor	QSFP28 AOC Duplex
Wavelength	850nm
Data rate/lane	25Gb/s
Length	25m
2019. 12. 20. 현재가격	$308

(b)

ATAC, Corona, 그리고 위 계산에서 알 수 있는 것처럼 코어 수는 4~16개가 적당하다. 다수의 cluster를 광통신으로 연결하는 것이 CPU의 성능을 높이고 비용을 낮추는 데 유리하며 슈퍼컴퓨터는 이미 그렇게 설계되고 있다. Silicon photonics 연구의 광 네트워크 설계자들이 기대한 방향으로 기술이 발전하면 컴퓨터 및 IT 기기의 중요 구성 소자는 CPU, Memory, 광 반도체 칩(ONC)이 되어야 한다. 이 중 광 반도체 칩의 시장 비중은 기술이 진보할수록 높아지게 되고 궁극적으로 CPU 시장보다 더 커질 수 있다. 슈퍼컴퓨터에 적용된 광소자와 광 네트워크를 CPU 칩에 도입하기 위해 필요한 3가지 기술이 실험적으로 검증된 것은 의의가 크다고 할 수 있다.

Ⅳ. 결론

2000년대 초반부터 Intel, IBM, HP, Sun Mi-

crosystems, Samsung 등 반도체 선두 기업들은 광통신 기술을 반도체 칩의 core-to-core 또는 CPU-to-memory 간 통신에 적용하기 위해 대규모 연구비와 인력을 투입하였다. MIT, Columbia U., HP, Sun Microsystems 등 학교와 기업들은 광통신 기술이 적용될 경우 가장 유력한 네트워크 구조를 제안하였지만 실제 칩에 실험적으로 집적하고 시험할 수 있는 단계까지 도달하지 못하였다. 실패의 주 원인은 실리콘 물질을 광도파로로 사용하기 때문이라 할 수 있고 PECVD SiON 물질의 광도파로를 사용할 경우 해결이 가능하다.

슈퍼컴퓨터에 이미 활용되고 있는 VCSEL과 PD를 도입하고 fat tree 네트워크를 그대로 적용하되 단지 광섬유만 PECVD SiON 광도파로로 교체하여 인텔 서버용 CPU에 적용하는 것이 가장 합리적이라 할 수 있다. 이렇게 할 경우 현재 인텔 서버용 28개 코어 CPU 가격보다 약 1/3만큼 싼 값으로 공급할 수 있다. 이를 위해 필요 충분한 3가지 기술을 한국에서 개발하고 실험적으로 검증한 사실은 의미가 크다고 할 수 있다. 광 네트워크 설계자들이 기대한 방향으로 기술이 발전하면 컴퓨터 및 IT 기기의 중요 구성 소자는 CPU, memory, 그리고 광 반도체 칩이 되어야 한다. 이 중 광 반도체 칩의 시장 비중은 기술이 진보할수록 높아지게 되고 궁극적으로 CPU 시장보다 더 커질 수 있다.

참고문헌

[1] George Kurian et al., "ATAC: A 1000-Core Cache-Coherent Processor with On-Chip Optical Network," PACT'10, September 11-15, 2010.

[2] Dana Vantreas et al., "Corona: System implications of Emerging Nanophotonic Technology," ISCA, 2008. 35, 2008.

[3] Aleksandr Biberman et al., "Photonic Network-on-Chip Architectures Using Multilayer Deposited Silicon Materials for High-Performance Chip Multiprocessors," ACM Journal on Emerging Technologies in Computing Systems, vol. 7, no. 2, June 2011.

[4] Ashok V. Krishnamoorthy et al., "Computer Systems Based on Silicon Photonic Interconnects," Proceedings of the IEEE, vol. 97, no. 7, July 2009.

[5] Chen Sun et al., "Single-chip microprocessor that communicates directly using light," Nature, vol. 528, no. 24, December 2015.

[6] M. I. Alayo et al., "Deposition and charatrization of silicon oxynitride for integrated optical applications," Journal of Non-Crystalline Solids, 2004, pp. 76-80.

[7] Richard Jones et al., "Integration of SiON gratings with SOI," Conference: Group IV, Photonics, 2005.

[8] Tai Tsuchtza, "Low-loss Silicon Oxynitride Waveguides and Branches for the 850nm Wavelength Region," Japanese Journal of Applied Physics, vol. 47, no.8, 2008, pp.6739-6743.

[9] B. S. Sahu et al., "Influence of hydrogen on losses in silicon oxynitride planar optical waveguides," Semicond. Sci. Technol. vol. 15, 2000.