

고성능 프로세서 기술 동향

Technology and Trends of High Performance Processors

김영우 (Y. W. Kim) 서버플랫폼연구팀 선임연구원
김성운 (S. W. Kim) 서버플랫폼연구팀 팀장

목 차

-
- I. 개요
 - II. 고성능 프로세서 동향
 - III. 관련기술 동향
 - IV. 결론

반도체 공정 기술의 발전으로 인하여 반도체 회로는 지속적으로 비약적인 성능의 발전을 가져오고 있다. 고성능 프로세서는 이와 같은 반도체 공정의 미세화에 따라 전력소모 및 발열 문제로 인하여 공정 및 속도 향상을 통한 성능 경쟁에서 탈피하여, 수십 개에서 수백 개의 코어를 내장하는 고도 병렬화/이기종화를 통한 성능 향상을 추구하는 시대로 접어들고 있다. 본 문서에서는 최근의 고성능 프로세서 동향을 중심으로 병렬/이기종화 기술 및 관련 기술의 최근 동향과 향후 발전 추세에 대하여 논의하고자 한다.

I. 개요

반도체 공정 기술은 2000년대 중반 100nm 이하의 고집적 기술의 시대로 접어들어 2010년 현재 35nm 기술을 사용하는 다양한 프로세서 제품군이 출시되고 있다. 반면, 실리콘을 중심으로 하는 반도체 공정에서 100nm 이하의 공정을 이용한 고집적 과정에서 누설전류(leakage current)의 증가로 인한 많은 전력 소모와 발열문제는 기존의 무어의 법칙을 변경하는 결과까지 야기하였다. 본 장에서는 프로세서 및 시스템의 발전을 저해하는 요소와 이를 극복하기 위한 업계 및 학계의 대응 전략과 최근의 동향에 대하여 간략히 기술하도록 한다.

1. 성능 향상의 장벽요소

일반적으로 고성능 컴퓨팅을 위한 기술에 있어 성능 향상을 저해하는 3가지 장벽요소를 이야기 할 수 있다. 이들 3가지 장벽요소(brick wall)는 명령수준 병렬화 장벽(ILP wall), 메모리 장벽(memory wall), 전력 장벽(power wall)이다.

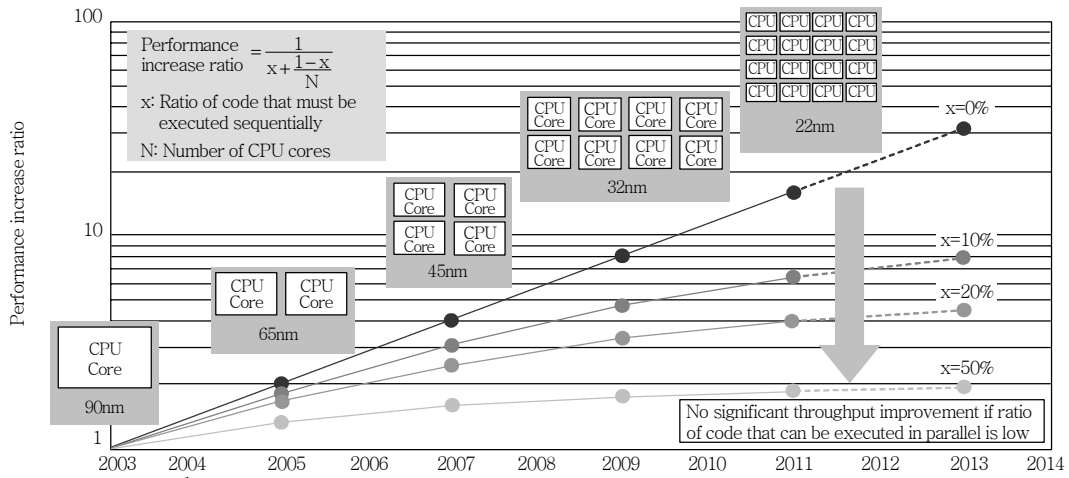
첫번째, 명령수준 병렬화 장벽은 명령어 수준의 병렬화(ILP)를 통하여 성능을 개선하는 전통적인 프로세서의 성능 개선 방법의 한계에 봉착할 수 있다는 것이다. 전통적인 프로세서의 성능 개선을 위하여 프로세서 내에서 많은 명령어가 동시에 실행 가능하도록 함으로써 성능을 개선하는 방법을 사용하여 왔다. 역사적으로 프로세서는 CISC 구조의 단일 명령어의 멀티사이클 실행, RISC 구조의 파이프라인을 통한 복수 명령어의 단일 사이클 실행(pipelining), 복수 실행 유닛을 채용한 슈퍼스칼라 처리, 명령실행 스레드의 병렬화(TLP)를 통한 성능 향상 등 고도의 병렬화를 통한 성능의 향상에 주력하여 왔다. 최근에는 반도체

기술의 발전으로 이전에 보드 및 칩 수준에서의 병렬 프로세서 구현이 단일 칩 수준에서 구현 가능한 수준으로 진행됨에 따라 6, 8개 이상의 코어를 내장한 프로세서를 쉽게 구하여 이용할 수 있는 수준으로 발전하였다.

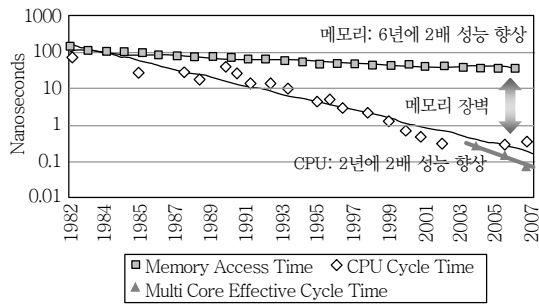
이와 같은 병렬화를 통한 성능의 향상이 최근 프로세서 설계의 주요 경향으로 자리매김하고 있으나, 병렬화를 통한 성능의 향상에도 한계가 존재한다. 암달의 법칙(Amdahl's Law)으로 대표되는 병렬화를 통한 성능향상은 물리적으로 프로세서를 고도로 병렬화 하더라도, 실행되는 소프트웨어의 병렬화가 지원되지 않을 경우 병렬화로 인한 성능향상의 이득을 얻을 수 없다는 것을 알 수 있다[1],[2]. 예로써(그림 1)과 같이 16개의 프로세서 코어를 사용하여 병렬처리 할 경우라도, 수행 프로그램의 20% 정도($x=20\%$)가 병렬처리가 되지 않고 순차처리 될 경우 전체 성능은 이론적인 성능의 약 40% 수준으로 떨어지게 됨을 알 수 있다.

두번째로, 메모리 장벽을 성능향상의 장벽요소로 들 수 있다. 메모리 장벽은 메모리의 성능향상 속도와 프로세서의 성능향상 속도의 차이에 기인한 장벽으로, 프로세서는 약 2년에 2배의 성능 향상을 보임에 반하여, 메모리는 6년에 2배 정도로 성능(속도)이 향상됨으로 인하여 2007년에 이미 프로세서와 메모리의 성능 차이가 100배 가까이 차이가 벌어져 있음을 알 수 있다(그림 2 참조)[3]. 이와 같은 메모리 장벽으로 인하여 프로세서의 성능이 획기적으로 향상된다고 하여도 전체 시스템의 성능이 그에 따라 향상되지 못하는 현상이 발생하고 있다.

마지막으로, 전력 장벽(그림 3 참조)은 반도체 공정 기술의 발전에 따른 공정의 미세화와 그에 따른 동작 전압의 저전압화에 따라 급격한 전력 소모 증가에 기인한 시스템 전력소모의 증가로 인한 장벽이다.

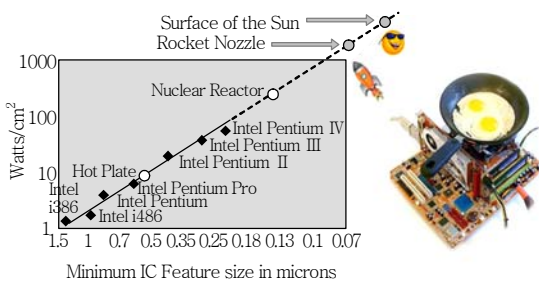


(그림 1) 암달의 법칙과 성능 향상[1]

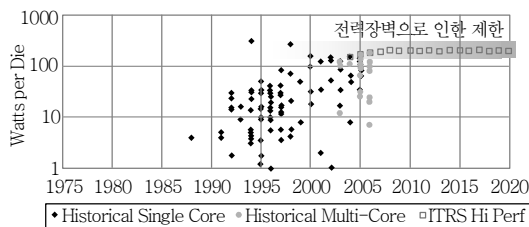


(그림 2) 메모리 성능과 프로세서 성능 비교[3]

지금까지의 반도체 공정은 전력소모를 증가시키지 않으면서 반도체 회로를 미소화하는 것에 중점을 두고 발전하여 왔다. 반도체 공정이 nm 수준으로 진입함에 따라 회로의 미소화로 인한 낮은 동작전압은 누설전류의 증가를 야기하였고, 공정의 미소화는 더 이상 동작전압을 낮추기 어려운 수준까지 발전하였다. 이로 인하여 단위 면적에 집적되는 회로 수 증가와 동작 주파수의 고속화, 낮은 동작전압으로 인하여 더 이상 반도체 공정의 개선만으로는 전력소모를 동일한 수준으로 유지하기가 어려워짐으로써 최근 저전력 구현에 대한 관심이 다시 대두되고 있다.



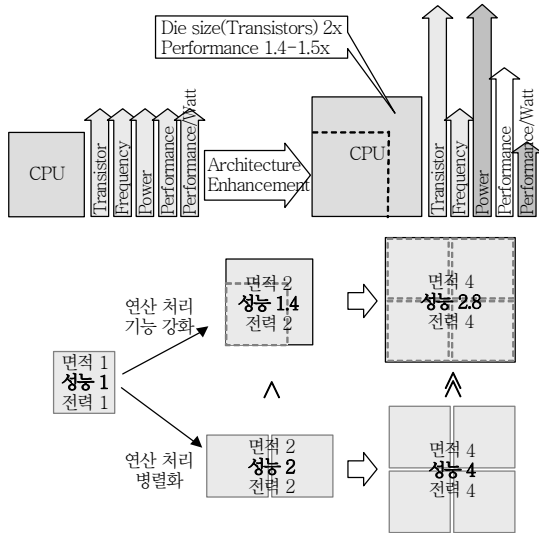
(a) CPU의 전력소모[4],[5]



(b) 프로세서의 전력소모[3]

2. 최근의 성능 향상 기술 동향

무어의 법칙으로 대표되는 반도체 장치의 성능 향상에 대한 법칙은 원래 18~24개월에 단위면적 당 소자의 수가 2배 증가한다는 개념이었으나, 미소 공정의 발전에 따라 24개월에 성능이 2배로 증가한다는 개념으로 변경되었다. 이와 같은 상황에서 2000년대 중반 AMD는 클럭 주파수로 대표되는 속도 중심의 성능 경쟁에서 탈피하는 것을 선언한 이후, 프로세서의 성능 발전은 단일 코어 프로세서에서 멀티코어, 매



(그림 4) 폴락의 법칙[6],[7]

니 코어 프로세서로 발전하고 있다. 이와 같은 이면은 인텔의 엔지니어인 폴락(Fred Pollack)에 의한 법칙(Pollack's Rule)으로 설명이 가능하다(그림 4) 참조[6].

폴락의 법칙은 반도체의 성능은 면적(복잡도)이 2배 증가할 때 1.4배 증가하며, 또한 전력소모는 면적에 비례한다는 법칙이다.

폴락의 법칙에 따르면, 성능 향상을 위하여 연산처리 기능을 강화함으로써 복잡도를 증가시키는 것 보다는 복잡도가 작으면서 적정수준의 연산능력을 가지는 프로세서를 복수 개 병렬로 사용하는 것이 더 효과적이라는 것을 알 수 있다.

폴락의 법칙에서 알 수 있듯이 최근 프로세서의 발전 방향은 개별 프로세서의 성능 향상(속도 향상, 기능 추가 및 복잡도 향상)보다는 복수 개의 프로세서 코어를 이용한 멀티코어 프로세서화, 기존의 고기능 다용도 프로세서 코어와 더불어 단순하면서도 높은 연산능력을 가지는 데이터 처리 전용 프로세서를 내장하는 이기종(heterogeneous) 프로세서화의 방향으로 진화하고 있다. 대표적으로 최근 NVIDIA의

GPGPU를 들 수 있다. GPGPU는 메인 프로세서는 아니나, 시스템의 성능 향상을 위하여 200~500여 개의 프로세서 코어를 이용하여 시스템의 연산능력을 극대화하고 있다. 이는 압달의 법칙에 따른 처리 프로그램에서의 데이터 처리 병렬도를 높임으로써 병렬처리 하드웨어의 성능을 최대화하는 것으로 ILP, TLP 보다 높은 수준의 병렬 처리 기법인 DLP를 구현한 것이다. 이와 같은 멀티코어/이기종 프로세서는 주요 프로세서 업체에서도 수 년 전부터 연구를 시작하여 2010년 현재 관련된 제품의 시제품을 출시하기에 이르러 향후의 고성능 프로세서에 대한 멀티코어/이기종 프로세서화가 더욱 가속될 것으로 예측된다.

II. 고성능 프로세서 동향

전술한 바와 같이 최근의 프로세서는 멀티코어화 이기종 코어화로 진행이 되고 있으며, 이와 같은 경향은 프로세서 제품군에 다양한 형태로 반영이 되고 있다. 이번 장에서는 프로세서 개발의 양대 산맥을 이루고 있는 인텔과 AMD의 최신 프로세서 개발 동향에 대하여 설명하도록 한다.

1. 인텔의 고성능 프로세서 개발 동향

인텔의 최근 프로세서 개발 동향의 설명에 앞서 간략히 인텔의 프로세서 개발 사이클에 대하여 알아보도록 한다.

일반적으로 인텔의 프로세서 개발은 크게 2개의 연구팀이 약 4~5년의 개발 주기로 프로세서를 개발하는 것으로 알려져 있다. 고성능 서버용 프로세서의 개발은 미국 오레곤 힐스보로(Hillsboro)에 있는 연구팀에서 개발하며, 데스크톱 및 노트북용 프로세서

의 개발은 이스라엘 하이파(Haifa)에서 담당하고 있다[7]. 오래된 연구팀에서는 Nehalem 아키텍처의 서버용 프로세서 개발을 완료하고 현재 차세대 아키텍처인 Haswell에 대한 개발을 진행하고 있으며, 하이파에서는 2010년에 데스크톱용 프로세서인 Sandy Bridge에 대한 검증과 시제품을 출시하였다. 인텔의 프로세서 개발팀은 1년 아키텍처 개발, 2~3년 설계, 1년 검증 후 제품출시의 사이클을 바탕으로 2개의 팀이 시차를 달리하여 프로세서를 개발하고 있으며, 이를 통하여 인텔의 Tick-Tock 개발 주기를 달성하고 있다. 이 외에도 텍사스 오스틴에 소재한 연구팀에 의한 저전력 프로세서 개발팀이 있으나 현재 개발중에 있는 프로세서에 대하여는 알려진 바가 없다.

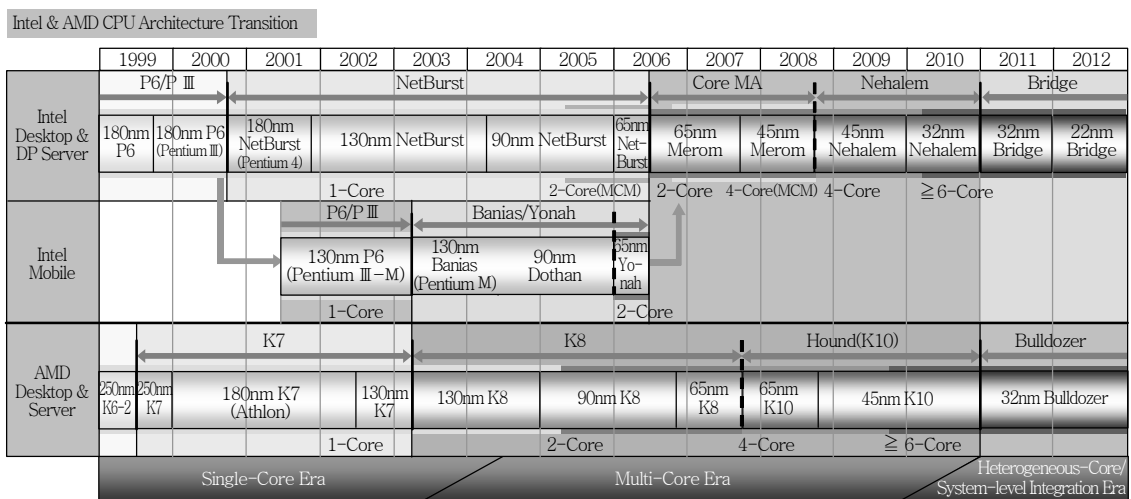
2010년은 Tick-Tock 주기 중 32nm 공정의 Tick 주기로서, 2008년에 발표된 Nehalem 프로세서의 32nm 버전과 2010년에 발표된 Sandy Bridge가 있는데, 이들 두 프로세서의 아키텍처 동향을 중심으로 인텔의 프로세서 개발 동향을 설명하도록 한다.

최근의 인텔 프로세서의 아키텍처 개발의 동향은 다음과 같은 몇 가지로 요약할 수 있다.

- 프로세서의 멀티 코어화
- 프로세서 고속화를 위한 주요 장치의 통합화
- 프로세서의 코어 아키텍처의 단일화
- 프로세서의 이기종 통합화

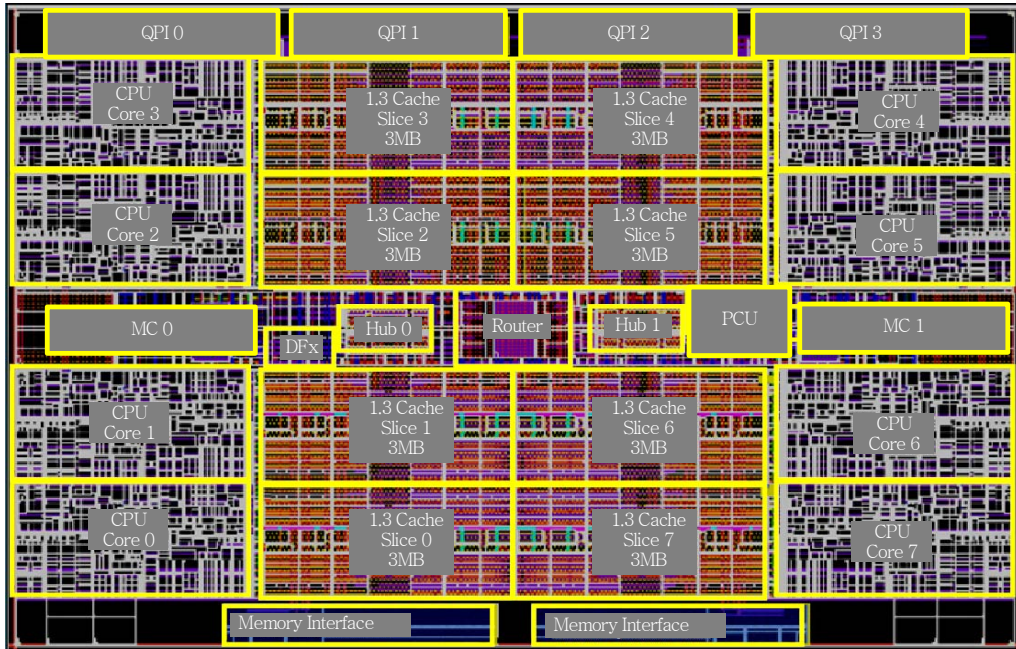
인텔은 2000년대 이전 Pentium 6 아키텍처 이후, NetBurst, Core 아키텍처를 거쳐 현재 Nehalem 아키텍처를 이용한 프로세서가 주요 제품으로 출시되고 있다(그림 5) 참조). 전술한 바와 같이 명령수준 병렬화 장벽, 메모리 장벽, 전력 장벽 등과 같은 장벽 요소로 인하여 인텔의 아키텍처 또한 멀티코어화 하고 있다. 인텔은 2000년대 초반 NetBurst 구조 프로세서의 후기 제품부터 듀얼 코어를 내장하기 시작하여 이후 Core Duo로 알려진 듀얼 코어 프로세서 제품을 출시하였고 2008년 이후 4개의 코어를 내장한 Nehalem 프로세서 제품이 출시되었다. 2010년에는 Nehalem-EX(Beckton)를 통해 총 8개의 코어를 내장한 프로세서를 출시하였으며, Nehalem의 32nm 버전인 Westmere의 경우 6개의 코어를 내장한 시제품을 소개하였다(그림 6) 참조).

Nehalem 칩 구조의 큰 특징으로는 이전의 인텔

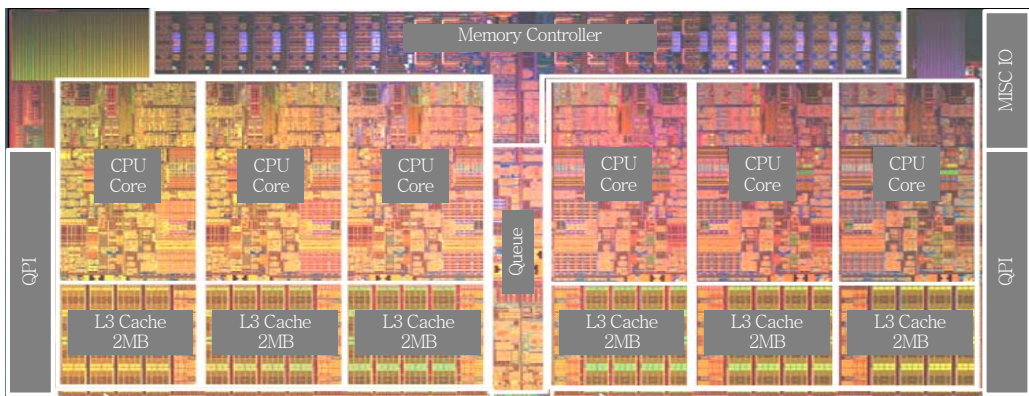


Copyright(c) 2009 Hiroshige Goto All rights reserved.

(그림 5) 인텔 및 AMD 프로세서 아키텍처의 변천[7]



(a) Nehalem-EX 다이



(b) Westmere 다이

(그림 6) Nehalem과 Westmere[7]

코어와 달리 메모리 제어를 프로세서 칩에 내장하였다는 점과 기존의 FSB 대신 QPI를 사용하였다는 점이다. 이와 같은 시도는 AMD에 의하여 먼저 시도된 방법(Opteron 프로세서에서의 메모리 제어기 내장 및 직렬 버스 - AMD의 경우 HT, Hyper Transport - 사용)으로써, 메모리 제어기의 내장에 따라 시스템의 성능(메모리 접근 성능)을 향상할 수 있으며,

고속의 직렬 버스(switched fabric 형태의)인 QPI를 이용함으로써 프로세서간의 통신 및 프로세서와 시스템 IO 통신을 보다 빠르게 수행하며, 프로세서의 확장(QPI를 통한 2, 4, 8-way 프로세서 구성 등)을 용이하도록 하였다.

최근의 인텔 프로세서 개발의 또 다른 특징 중 하나는 프로세서 아키텍처의 단일화 경향을 들 수 있다.

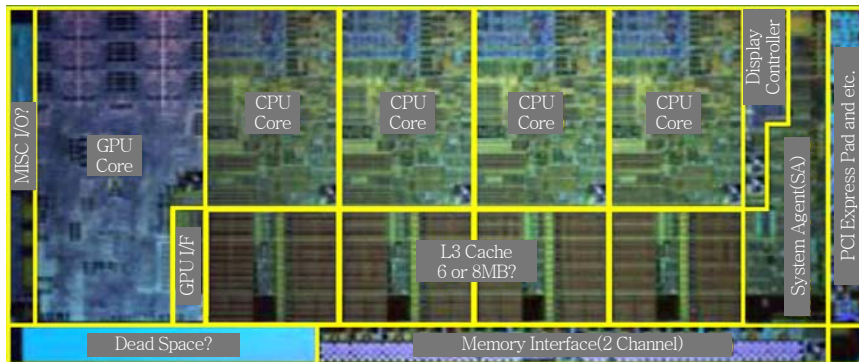
1990년대 및 2000년대 초반까지의 Pentium 및 Pentium Mobile 개발 시기와 달리 최근의 인텔 프로세서는 저전력 모바일, 데스크톱 및 고성능 프로세서의 코어에 동일한 아키텍처를 사용한 프로세서 코어가 적용되고 있다는 점이다. 이와 같은 프로세서 아키텍처의 단일화는 다양한 제품군의 개발에 따른 소프트웨어 제품의 최적화를 단기간에 가능케 함으로써 프로세서 아키텍처의 변경에 따른 소프트웨어의 최적화를 일관성 있게 수행할 수 있는 장점을 가진다는 점으로 향후의 인텔 프로세서 아키텍처에도 이와 같은 경향이 유지될 것으로 판단된다.

마지막으로 최근 IDF 2010 Beijing에서 발표한 인텔의 차세대 데스크톱 프로세서인 Sandy Bridge 코어를 통하여 나타난 인텔 프로세서의 개발 동향에 대하여 설명하도록 한다. Sandy Bridge 아키텍처의 가장 큰 특징은 무엇보다도 그래픽 처리 장치(GPU)를 프로세서 칩 안에 내장한 점을 들 수 있다(그림 7) 참조). 또 다른 특징으로는 새로운 벡터 연산 명령인 AVX 명령어를 실장하였다는 점이다. AVX는 최대 256비트 폭의 SIMD 명령을 통해 8개의 32비트 부동소수점 연산이 가능하다는 점이다.

인텔에서 최근 발표한 프로세서의 제품과 개발 추이를 볼 때 향후 인텔의 프로세서 개발 동향은 멀티코어를 내장하며, 경우에 따라서는 SIMD 구조 기반

의 벡터 연산이 가능한 코어를 복수 개 내장하는 이기종 멀티 코어로 발전할 것을 쉽게 예측할 수 있다. 다만 현 시점에서 인텔 Sandy Bridge의 경우는 단순 GPU를 통합한 것으로 NVIDIA나 ATI에서와 같은 GPGPU 기능은 제공하지 않는 것으로 추측되고 있다[8].

이와 같은 주류 고성능 프로세서의 개발동향과 더불어 몇 가지 추가적으로 논의하여 볼 만한 내용으로는 인텔의 멀티 코어 프로세서에 관한 연구를 예로써 들 수 있다. 예로써 상용 멀티 코어 제품으로 Larrabe에 대한 개발을 들 수 있다. Larrabe는 인텔이 Atom 프로세서의 개발과 거의 동시에 시작된 그래픽 칩 셋을 위한 멀티 코어 프로세서이다. 2009년 하반기까지의 Larrabe는 16개의 Larrabe IA 코어와 그래픽 하드웨어를 통합한 그래픽 칩으로 개발되어 왔다. 개별 Larrabe IA 코어는 IA 기반의 P54C (Pentium 4급) 코어와 512비트의 벡터 연산을 처리하는 SIMD 프로세서로 구성되며, 16개의 Larrabe 코어가 링버스를 통하여 연결되는 구조를 취하였다. 2009년 하반기에 인텔은 Larrabe를 이용한 그래픽 칩셋의 개발 로드맵을 변경하게 되었는데, 이와 같은 이면에는 그래픽의 성능과 전력 대비 성능이 나빠지는 업계의 후문을 통하여 추측하여 볼 수 있다[9]. 인텔은 Larrabe를 그래픽용으로 사용하는 대신, 범용



(그림 7) Sandy Bridge[8]

SIMD 프로세서로 활용하여 향후 인텔의 아키텍처인 Ivy Bridge(Haswell)에 적용할 것으로 예상된다.

주목할 만한 또 다른 인텔의 연구로서, 2010년 ISSCC에서 발표한 SCC를 들 수 있다[10]. SCC는 48개의 P54C급의 32비트 IA 코어를 사용한 연구용 프로세서로서, 2005년의 1TFlops, 80 코어 NoC 프로세서의 후속 연구 프로세서이다. SCC는 이전의 NoC와 달리 상용 IA를 사용하였다는 점과 이를 통하여 상용화의 가능성을 보였다는 점에서 큰 의미를 가진다고 볼 수 있다.

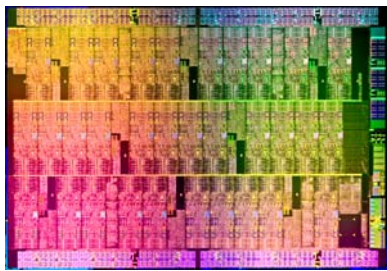
이와 같은 Larrabe 및 SCC에 관한 연구는 최근 국제 컴퓨팅 컨퍼런스에서 MIC 구조의 상용화 매니 코어 칩인 Knights Corner의 발표로 이어졌다[11]. (그림 8)과 같이 Knights Corner의 시제품인 Knight

Ferry는 22nm 공정을 사용하여 32개의 IA 코어를 내장하고 있으며, 2011년 상용 제품이 출시될 것으로 예상되고 있다. MIC 아키텍처의 제품은 현재 50 개 코어 이상으로 확장 가능한 것으로 알려지고 있는데, 현재 개발 키트 형태로 일부 사용자에게 제공되고 있는 Knights Ferry의 경우 32개 코어에서 총 128 개의 스레드를 동시 수행할 수 있는 것으로 알려져 있다(약 153GFlops 정도의 성능으로 추측됨). 인텔의 MIC 아키텍처는 현재 각광받기 시작한 GPGPU 시장을 목표로 하고 있는 것으로 보이며, 기존의 NVIDIA, ATI와 달리 IA 코어를 사용함으로써 기존 인텔 계열의 프로세서에서 사용되었던 코드를 활용할 수 있다는 것이 큰 장점으로 작용할 것으로 보인다.

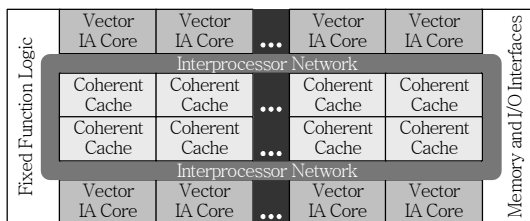
2. AMD의 고성능 프로세서 개발 동향

다음으로는 후발 주자이면서도 다양한 기술적인 시도를 통하여 인텔을 추격하고 있는 AMD의 프로세서 기술 동향에 대하여 알아보도록 한다.

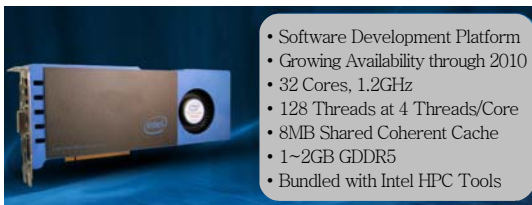
AMD는 1975년 인텔의 8080프로세서를 역엔지니어링을 통하여 개발함으로써 프로세서 시장에 참여하였다. 이후 인텔의 프로세서를 라이선싱하여 프로세서 사업을 수행하였으나, 저작권 문제로 인한 인텔과의 많은 분쟁으로 인하여 1996년 K5 프로세서 이후로 독자적인 구조를 가지는 x86 호환 프로세서 개발업체로 자리매김하면서 현재 x86 프로세서 업계의 2인자의 자리를 차지하게 되었다[13]. AMD는 인텔을 뒤따라가는 후발주자임에도 불구하고 기술 발전의 측면에서 볼 때, 때때로 인텔을 능가하는 선견 지명과 전략을 구사하였다. 몇 가지 예를 들면, 인텔보다 먼저 듀얼 코어 프로세서 및 4, 6코어 프로세서를 출시하였으며((그림 5) 참조), 2000년대 초반에는 당시 프로세서의 성능으로 대표되던 클럭 주파수 속



(a) Aubrey Isle의 다이(Knights Ferry)



(b) MIC 아키텍처



(c) Knights Ferry

(그림 8) Intel의 MIC[12]

도에 의한 성능 표시를 버리고 독자적 성능 메트릭을 채용하였던 점, 메모리 제어를 프로세서와 통합하고 시스템 버스로 HT를 사용하여 시스템 성능을 향상하였던 점을 들 수 있다. 그러나 무엇보다도 놀라운 점은 2006년 그래픽 칩셋 업체인 ATI를 인수한 이후에 Fusion이라고 하는 프로세서와 GPU를 통합하는 전략을 발표하여 기술 및 전략적인 측면에서 인텔보다 앞서나가는 전략을 발표하였다는 것이다.

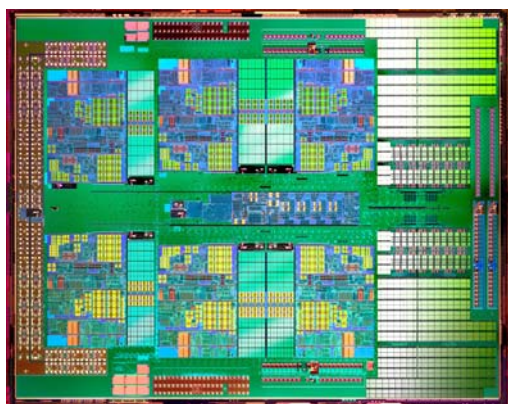
이와 같은 AMD의 프로세서 아키텍처 개발 동향은 다음과 같은 몇 가지로 요약할 수 있다.

- 프로세서의 멀티 코어화
- 프로세서의 코어 구조의 최적화
- 프로세서의 이기종 통합화

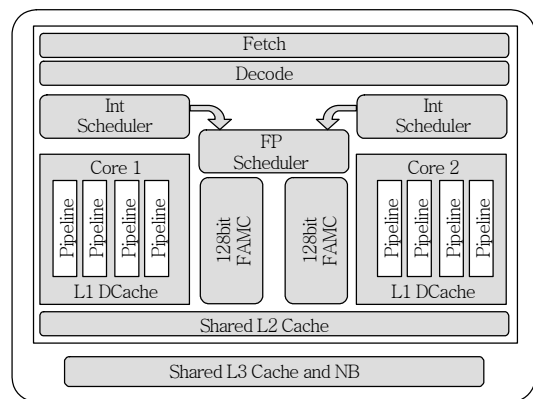
AMD는 인텔에 앞서 2, 4, 6코어 제품을 소개하여 왔는데, 2010년 현재 AMD는 45nm 및 32nm 공정에 기반한 K10(Hound) 프로세서 제품 군을 출시하고 있다. K10 코어는 개별 코어에서 최대 6개의 uOP을 1사이클에 실행할 수 있는 코어로 AMD의 Many Cours 제품 군으로 최근 6개 코어를 내장한 Phenom II X6 프로세서를 발표하였으며(그림 9 참조), 2개의 Phenom II X6를 MCP화 하여 총 12개의 프로세서 코어를 내장한 제품을 발표하였다[14]. Phe-

nom II X6는 총 6개의 프로세서 코어, 4개의 HT 연결 및 메모리 제어를 내장하고 있으며, 내장 코어 수의 증가에 따라 칩 소켓의 규격에도 변화가 발생하였다.

2011년 투입 예정인 AMD의 차세대 프로세서 코어 아키텍처는 코드명 Bulldozer로서, 형상 측면에서 이전 프로세서 코어와 달리 2개의 정수코어와 공유 부동소수점 유닛을 하나로 통합한 Bulldozer 모듈로 구성되는 점이 큰 차이를 보이고 있다(그림 10 참조)[15]. Bulldozer의 아키텍처상 가장 큰 특징은 지금까지의 정수코어의 성능향상을 목표(ILP의 향상)로 추구하던 설계 방향에서, 스레드 및 데이터 병렬화를 통한 성능향상을 목표로 프로세서의 아키텍처 설계 방향이 변경되었다는 점이다. 구체적으로는 이전의 Hound 아키텍처에서 정수 유닛은 총 6개의 uOP와 부동소수점 유닛에서 3개의 uOP를 처리할 수 있었던 점에 반하여, Bulldozer 아키텍처에서는 하나의 정수 코어에서 총 4개의 uOP를 처리함으로써 개별 정수코어에서 처리되는 uOP의 수가 2/3로 줄어들게 되었다. 반면, Bulldozer 모듈 전체로는 2개의 정수 코어를 사용하여 동시에 2개의 독립적인 스레드의 실행이 가능하게 되었다. 부동 소수점 유닛 또한 이전의 Hound 아키텍처에서 크게 변경되어 Bulldozer 아



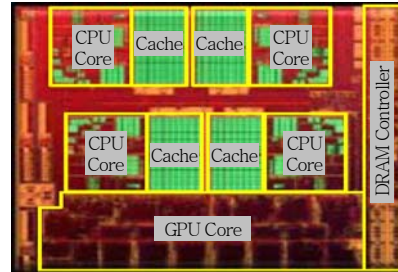
(그림 9) Phenom II X6[13]



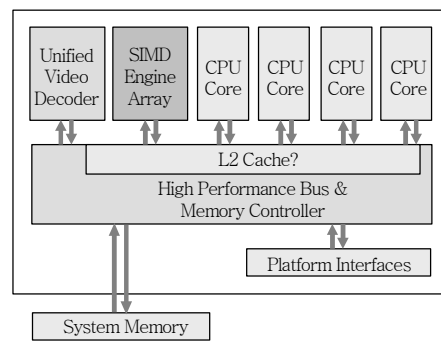
(그림 10) Bulldozer 아키텍처[14]

키텍처에서는 2개의 정수코어가 1개의 부동 소수점 유닛(2개의 128bit 파이프를 사용, 총 4개의 부동 소수점 uOP의 실행이 가능한 것으로 추정되고 있음)을 공유하는 구조로 변경되었다. 이전 아키텍처에 비하여 부동 소수점 유닛의 수가 줄어들게 된 이유로는 일반적인 응용프로그램(HPC 응용을 제외한)의 경우 부동소수점 유닛의 활용도가 크지 않으므로 인함에 기인한 것으로 알려져 있다. 즉, Bulldozer 아키텍처는 이전 아키텍처에 비하여 단순화된 2개의 정수 코어를 사용하여 스레드 수준의 병렬화(TLP)를 강화하고, 부동소수점 유닛의 비트폭을 확장함으로써 데이터 수준의 병렬화(DLP)를 동시에 추구하는 방향으로 아키텍처가 변화하였다[15]. 이와 같은 구조 변경의 배경에는 아마도 전술한 폴락의 법칙에서와 같이 정수코어의 기능강화(면적 증가)를 통한 성능향상 보다는 동일 면적에 더 많은 코어를 집적함으로써 얻어지는 성능 및 전력 향상을 추구한 것으로 생각된다.

ADM 프로세서의 개발 동향에 있어서 가장 많은 관심을 끌고 있는 것 중 하나는 APU(구 Fusion) 전략이다. APU 전략의 기준은 프로세서와 GPU를 통합하는 것으로서, APU 전략 이전의 메모리 제어를 프로세서에 내장한 것 또한 넓은 의미의 APU 전략으로 생각할 수 있다. 최초의 Fusion 전략은 2006년에 발표되어 최근 APU라는 이름으로 변경되었는데, 2010년 6월 Computex에서 APU 전략을 채용한 최초의 프로세서인 Liano 프로세서를 발표하였다(그림 11) 참조[16]. AMD의 Liano는 인텔의 Sandy Bridge와 유사한 구조를 가지고 있다. 즉, 복수의 프로세서 코어와 GPU를 통합한 구조이다. 반면, 아직까지 구조상으로 명확히 밝혀진 것은 아니나 인텔의 Sandy Bridge의 경우는 GPU 코어의 범용 응용(GPGPU로서)을 고려하지 않고 있다는 데 반하여, AMD의 Liano APU는 GPU 코어의 범용 프로그램



(a) AMP Liano 다이



(b) Liano 아키텍처

(그림 11) AMD의 APU Liano[16]

에서의 응용이 가능하다는 점을 강조하였다는 것이다. 즉 AMD의 Fusion APU 전략은 단순한 GPU의 통합을 넘어서 GPU의 SIMD 코어를 활용한 DLP를 추구하고 있다는 점을 명확하게 하였다. 이와 같은 차이점은 당분간 인텔과 AMD의 프로세서 아키텍처 전략에 영향을 미칠 것으로 기대된다.

3. 향후의 프로세서 개발 동향

앞에서 알아본 바와 같이 향후의 프로세서 아키텍처에서의 개발 동향은, 복수 개의 코어를 내장하는 멀티 코어화와 이기종 코어의 채용을 통하여 DLP를 통한 성능향상을 추구하는 이기종 멀티코어화의 방향으로 진행될 것으로 예측할 수 있다. 향후 1~2년 안에 메인 프로세서에서는 최대 16코어 이상을 내장한 프로세서가 출시될 것으로 예측되고 있으며, 프로세서 코어와 GPU의 통합이 더욱 가속화될 전망이다.

인텔과 AMD 모두 멀티코어 및 DLP를 강화하기 위한 SIMD 코어를 채용한 이기종 멀티코어화로 진행할 것으로 예측되고 있으나 이들 두 업체의 전략에 있어서는 약간의 차이점을 보이고 있다.

인텔의 멀티코어 전략은 현시점에서는 Sandy Bridge 아키텍처에 GPU를 통합하고 있으나 전술한 바와 같이 NVIDIA, ATI의 경우와 달리, GPU(SIMD)의 프로세서 측으로의 명령어 지원을 통한(AVX와 같은) 통합을 중심으로 진행할 것으로 예측되며, 현시점에서 명확한 언급은 없으나 차세대 아키텍처인 Ivy Bridge(Haswell)에서 Larrabe를 통한 연구를 바탕으로 보다 많은 SIMD 코어를 내장할 것으로 추측되고 있다.

AMD는 인텔보다 먼저 이기종 멀티코어화를 선언하고(Fusion 전략) 이를 추진중에 있으며, GPU의 통합 및 GPGPU 기능의 지원과, 차세대 Bulldozer 아키텍처의 확장을 통한 이기종 멀티코어화를 지속적으로 진행할 것으로 예측되고 있다. 다만, 현시점에서 인텔의 AVX와 같이 GPU의 명령어 수준 통합에 관한 언급은 없으나, 인텔과 유사하게 명령어 수준에서의 SIMD 코어 지원을 추구하고 있는 것으로 예상하고 있다.

III. 관련기술 동향

이번 장에서는 GPGPU와 같은 성능 가속 기술과 시스템 연결망 기술 등 고성능 프로세서 동향과 관련된 기술들의 최근 동향을 간략히 정리하도록 한다.

1. 시스템 연결망 기술

시스템 연결망 기술은 시스템 내부에서 프로세서와 프로세서, 프로세서와 IO 장치의 연결 등 다양한

장치간의 연결에 사용되는 기술로서 종전에는 버스 방식의 연결망이 널리 사용되어 왔다. 2000년대 초반부터 시스템과 IO 장치를 연결하는 연결망으로써 HT, PCI Express 등이 널리 사용되기 시작하여 종전의 버스 시스템을 대체하였으며, HT를 시작으로 프로세서와 프로세서 간의 연결에도 직렬 버스 방식의 연결망이 사용되기 시작하였다(AMD의 Opteron). 인텔은 Nehalem 아키텍처에서 종전의 FSB를 탈피하여 QPI라고 하는 직렬 버스를 채택하였다. QPI는 인텔의 주도 하에 개발되었던 PCI Express의 프로세서 연결망 버전이라고 생각하면 쉽게 이해될 수 있으며, 주요한 차이점으로는 프로세서 간의 캐시 일관성 유지를 위한 프로토콜을 포함하고 있다는 점이 큰 차이점이다. 현재 QPI의 상세 스펙은 공개되고 있지 않으나, 최대 20개의 lane(연결)을 통하여 6.4Gtps(약 25.6GB/s 상당)의 데이터 전송 성능을 가지고 있다(참고로, PCI Express x16 Gen2의 경우는 약 8GB/s, HT3.1 x16(uni-dir)의 경우는 약 12.8GB/s의 성능을 가짐). QPI는 AMD의 HT와 마찬가지로 다양한 연결 형상을 통하여 2, 4, 8개 등의 프로세서를 연결하여 프로세서의 성능을 확장하도록 하고 있다[17].

QPI는 기본적으로 프로세서 칩과 칩간의 연결망으로 사용되고 있는데, 최근에 MCP를 사용하여 원 칩화 하고 있는 인텔 프로세서에 QPI를 적용하여 최대 10Gtps, 약 40GB/s의 성능을 내고 있는 것으로 추정되고 있다.

이와 같은 시스템 연결망 기술들은 프로세서간의 통신뿐 아니라, 내장된 메모리 제어를 통한 메모리의 접근을 보다 빠르게 수행할 수 있도록 지원함으로써 고성능 프로세서의 성능을 최대화 할 수 있도록 지원한다는 점에서 프로세서 기술과 더불어 중요한 기술적인 위치를 점유하고 있다고 할 수 있다.

2. 성능 가속 장치의 기술 동향

최근에 HPC 분야에서 가장 각광을 받고 있는 기술로서는 GPGPU 기술을 들 수 있다. GPGPU는 그래픽 처리기인 GPU의 프로그래밍 가능한 벡터 코어 부분을 범용 소프트웨어의 처리에 응용하고자 한 기술로서, 2000년대 초반부터 꾸준히 연구되어져 온 분야 중 하나이다.

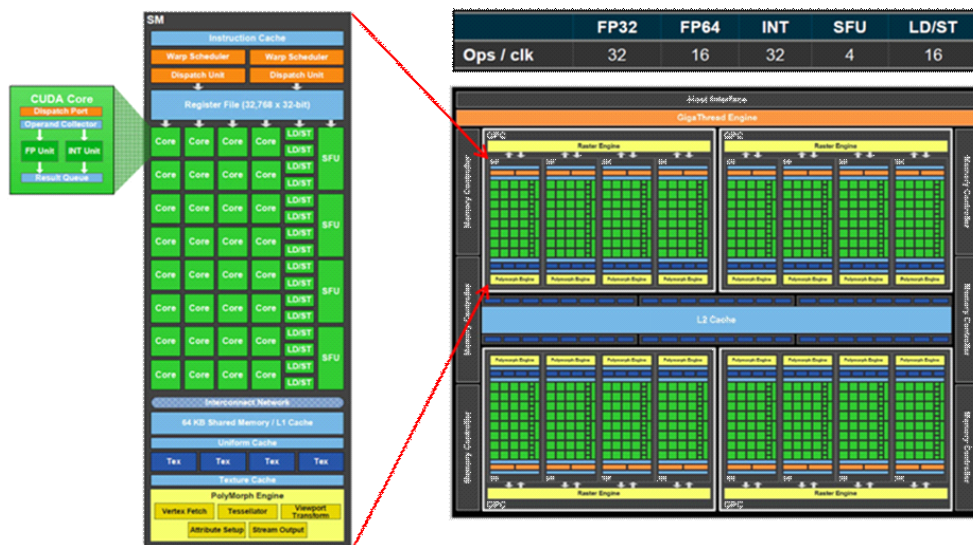
이 분야가 최근 들어 각광을 받게 된 것은 실시간 게임 등과 같은 고성능 그래픽 처리요구에 따른 GPU의 성능 발전에 기인하며, NVIDIA의 시의 적절한 공격적인 GPGPU 전략(GPGPU라는 용어도 NVIDIA에서 처음 사용함) 또한 시장에 주효하였던 것으로 보인다.

NVIDIA는 2006년 Tesla 아키텍처를 채용한 Tesla 8 시리즈를 발표하였고, 2009년 Fermi 아키텍처를 채용한 Tesla 10 시리즈를 통하여 단일 GPU에서 1TFlops의 단정도 부동소수점 연산 성능을 구현함으로써 두각을 나타내었다.

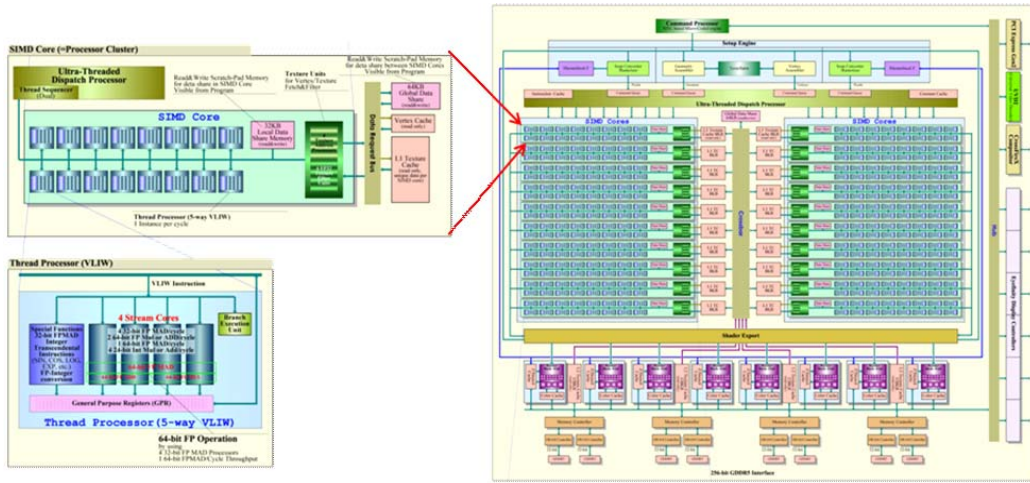
NVIDIA의 Fermi 아키텍처는 하나의 FP 유닛과

하나의 정수 유닛으로 구성되는 CUDA 코어를 최대 512개 사용하여 단정도 연산은 최대 1.2TFlops, 배정도 연산은 최대 630GFlops를 처리하는 SIMD 구조의 벡터 연산 프로세서이다(그림 12) 참조). NVIDIA는 Fermi 아키텍처를 바탕으로 2017년 ExaScale 성능을 목표로 하는 HPC 환경에서 1개의 GPU 노드(CPU+ GPU+ memory)에 약 7,200개의 FP 유닛을 사용하여 단정도 연산에서는 40TFlops, 배정도 연산에서는 13TFlops를 구현할 수 있을 것으로 예측하고 있다[18].

GPU의 성능 측면에서 볼 경우, 최초로 1TFlops의 성능을 구현한 것은 ATI였다. ATI 또한 GPGPU 기능을 제공하는 GPU를 제작하고 있으며, 최신 ATI Cypress GPU의 경우, 1,600개의 Thread Processor라고 하는 VLIW 코어를 사용하여 2.7TFlops의 단정도 연산 성능과 544GFlops의 배정도 연산 성능을 가지고 있는 것으로 알려져 있다(그림 13) 참조)[19]. ATI의 경우, NVIDIA와는 달리 AMD의 Fusion 전략에 맞추어 독립된 GPGPU 제품 보다는



(그림 12) NVIDIA의 Fermi 아키텍처[18]



(그림 13) AMD Cypress GPU의 구조[19]

AMD 프로세서의 SIMD 코어로 통합되는 형태의 GPGPU 전략을 사용하고 있지 않은가 생각된다.

성능 가속 장치는 대표적인 GPGPU 이외에도 현재 100개의 프로세서 코어를 온 칩 네트워크로 연결하여 구성된 Tiler 프로세서와 같은 다양한 형태의 가속 프로세서가 시장에 속속 등장하고 있다.

IV. 결론

최근의 프로세서 개발의 동향은 이기종 멀티코어화라는 단어로 정리할 수 있다. 현시점에서 기존의 프로세서 코어는 인텔의 Nehalem, AMD의 Phenom II X6와 같이 전통적인 정수 코어의 기능 확장 및 복수 코어의 채용을 통한 멀티코어화가 진행되고 있으며, 인텔의 Sandy/Ivy Bridge, AMD의 Bulldozer의 예와 같이 GPU와 같은 벡터 연산에 강한 SIMD 기반의 고속 연산 코어의 프로세서 통합에 대한 시도가 진행 중이다. 향후 1~2년 안에는 메인 프로세서에 최대 16코어 이상을 내장한 프로세서가 출시될 것으로 예상되며 프로세서 코어와 GPU의 통합이 더욱 가속화될 전망이다.

전통적인 프로세서 코어의 발전과 더불어 GPGPU로 대표되는 성능 가속 장치의 개발도 더욱 가속화할 것으로 예측된다. GPGPU는 이미 단일 칩에서 1TFlops 이상의 성능을 보임으로써 전통적인 프로세서 코어의 성능을 수 배~수십 배 뛰어넘는 성능을 보이고 있다. GPGPU는 프로세서의 이기종 병렬화와 더불어 상당기간 시장에서 큰 영향력을 발휘할 것으로 예상된다.

● 용어 해설 ●

장벽요소(Brick Wall): 컴퓨터의 성능을 제한하는 컴퓨터 구조상의 문제점을 통칭하는 컴퓨팅 분야의 용어. 명령수준 병렬화 장벽, 메모리 장벽, 전력 장벽의 3가지로 통칭하며, 최근 병렬화에 따른 병렬 프로그래밍의 어려움으로 인한 프로그래밍 장벽(Programming Wall)이라는 용어도 등장하였음

약어 정리

APU	Advanced Processing Unit
AVX	Advanced Vector Extensions
CUDA	Compute Unified Device Architecture
DLP	Data Level Parallelism
Flops	Floating point operations per sec

FSB	Front-Side Bus
GFlops	Giga Flops
GPGPU	General Purpose GPU
GPU	Graphic Processing Unit
HPC	High Performance Computing
HT	Hyper Transport
ILP	Instruction Level Parallelism
MC	Memory Controller
MCP	Multi-Chip Package
MIC	Many Integrated Core
NoC	Network on Chip
PCU	Power Control Unit
QPI	QuickPath Interconnect
SCC	Single-chip Cloud Computer
SIMD	Single Instruction Multiple Data
TFlops	Tera Flops
TLP	Thread Level Parallelism
tps	transactions per sec

참고 문헌

- [1] "Multiplying Multicores to Boost CPU Power," Nikkei Electronics Asia, 2007. 11.
- [2] Mark D. Hill and Michael R. Marty, "Amdahl's Law in the Multicore Era," IEEE Computer, July 2008.
- [3] P. Kogge et al., ExaScale Computing Study: Technology Challenges in Achieving ExaScale Systems, DARPA Information Processing Techniques Office (IPTO) sponsored study, 2008.
- [4] Intel IDF 2004.
- [5] Core 2 Duo E6400 Overclocked to 3.33GHz, Tom's hardware, <http://www.tomshardware.com/reviews/cheap-thrills,1335.html>
- [6] Pollack's Rule, Wikipedia, http://en.wikipedia.org/wiki/Pollack's_Rule
- [7] "Intel의 2013년의 CPU 「Haswell」へと続く Nehalem 개발의 舞台裏," 後藤弘茂의 Weekly 海外 뉴스, Impress Watch, 2010. 5., http://pc.watch.impress.co.jp/docs/column/kaigai/20100521_368266.html
- [8] "4種類に増える Sandy Bridge 世代の CPU ソケット," 後藤弘茂의 Weekly 海外 뉴스, Impress Watch, 2010. 4., http://pc.watch.impress.co.jp/docs/column/kaigai/20100426_363357.html
- [9] "仕切り直しとなった Larrabee の何か問題だったのか," 後藤弘茂의 Weekly 海外 뉴스, Impress Watch, 2010. 6., http://pc.watch.impress.co.jp/docs/column/kaigai/20091209_334552.html
- [10] Howard et al., "A 48-Core IA-32 Message Passing Processor with DVFS in 45nm CMOS," *In Proc. of ISSCC*, 2010.
- [11] "Intel Unveils New Product Plans for High-Performance Computing," Intel Press Release, <http://www.intel.com/pressroom/archive/releases/2010/20100531comp.htm?cid=rss-90004-c1-254992>
- [12] Kirk Skaugen, "Petascale to Exascale," ISC 2010 Keynote Presentation, Intel, 2010.
- [13] Advanced Micro Devices, Wikipedia.
- [14] AMD Financial Analyst Day 2009, AMD.
- [15] "整数演算パフォーマンスを犠牲にして効率を取る AMD の 「Bulldozer」," 後藤弘茂의 Weekly 海外 뉴스, Impress Watch, 2010. 2., http://pc.watch.impress.co.jp/docs/column/kaigai/20100205_346902.html
- [16] "COMPUTEX で FUSION プロセッサを初公開した AMD," 後藤弘茂의 Weekly 海外 뉴스, Impress Watch, 2010. 6., http://pc.watch.impress.co.jp/docs/column/kaigai/20100604_371991.html
- [17] Intel QuickPath Interconnect, Wikipedia.
- [18] Toru Baji, "Future Direction in GPU Computing," SIGGRAPH Asia, Dec. 2009.
- [19] "ワンチップで 2.7TFLOPS をたたき出す AMD の Cypress の秘密," 後藤弘茂의 Weekly 海外 뉴스, Impress Watch, 2009. 9., http://pc.watch.impress.co.jp/docs/column/kaigai/20090924_317309.html