

새로운 벤치마크 SPEC95

최 근 표준성능평가협회(SPEC)는 Unix 워크스테이션의 속도를 비교하는데 가장 널리 사용되어 온 벤치마크인 SPEC92의 새 버전 SPEC95를 발표하였다.

SPEC92의 문제점은 최근의 강력한 워크스테이션이 몇몇 테스트를 수행하는데 2.5초도 채 안 걸리기 때문에 컴퓨터 성능의 큰 변화는 측정하기 어려운 반면 작은 변화만 과장되기 쉽다는 것이다. SPEC95는 실행 시간이 더 길어졌다. SPEC의 새로운 테스트에 의하면 가장 강력한 단일 프로세서 부동 소수점 기계는 IBM 3CT 워크스테이션인 것으로 나타났는데, 이 기계에서 몇 개의 부동 소수점 테스트를 제외하고는 4분 이상 소요되었다(사실 SPEC측에서는 아무리 뛰어난 기계라 해도 각각의 테스트를 마치는 데 최소한 5분은 걸릴 것이라고 예상했었다). 컴퓨터의 성능이 12 내지 18개월마다 두 배로 증가하는 상황에서 평가 대상이 되는 32개의 SPEC 회원사가 모두 동의하는 테스트 프로그램을 계속해서 만들어 내는 것은 쉬운 일이 아니다. 비록 목표였던 1994년 말까지 완료하지는 못했지만, SPEC95는 헌신적인 노력의 산물이다. SPEC의 대표인 IBM의 Kaivalya Dixit는 "그래도 Windows 95보다는 며칠 앞서 출시되지 않았느냐"라며 웃었다.

1. 컴파일러 레지스턴트(Compiler-Resistant)

요즘의 컴퓨터 시스템에서는 파이프 라인의 깊이 라던가 캐시의 구조, 명령 순서, 레지스터 종류 등에 따라 실행 속도가 달라진다. 현대의 컴파일러들은 프로그래머가 컴파일시 이러한 요인들을 조절할 수 있게 되어 있다. Dixit은 새로운 테스트가 컴파일러의 최적화를 어렵게 함으로써 '공평한 게임의 룰'을 제공하고 있다고 말한다. "우리는 컴파일러 레지스턴트한 프로그램을 만들었습니다"라고 그는 말한다. 새로운 테스트는 "상수 전달"을 방해한다. CPU 레지스터에 데이터를 로드할 때 프로그램이 캐시나 주

메모리 사이를 몇 번이나 들락거리게 되는데, 상수 전달은 이 횟수를 줄여줌으로써 실행 속도를 향상시키는 컴파일러 기법이다. 예컨대, SPEC92 프로그램 코드는 $Pi=3.141592$ 와 같은 상수를 정의하고 있는데 컴파일러는 코드를 훑어본 후에 Pi 를 CPU 레지스터에 한번 로드하고, 루틴에서 Pi 를 필요로 할 때마다 속도가 늦은 캐시나 주메모리로 접근하지 않고 CPU 레지스터에 로드된 Pi 를 이용한다. SPEC95의 코드에서는 상수를 쓰는 않고 실행 시에(캐시나 주메모리에 있는)입력 파일로부터 필요한 데이터를 읽도록 하고 있다. 따라서 컴파일러는 이러한 종류의 최적화를 하지 못한다.

SPEC은 벤치마크 최적화 문제를 다루려고 노력해 왔으며 특히 작년에는 SPECbase 규정을 발표했다.

«CAD Report September 1994»

SPECbase 규정에 의하면 벤더들은 각각의 테스트에 대해 별도의 최적화 플래그를 사용하는 대신, 모든 테스트에 대해 하나의 플래그를 가지고 컴파일하게 되어 있다. 플래그를 단 한번만 설정하도록 하는 것은 응용 프로그램 개발자들이 프로그램을 컴파일하고 최적화하는 방식을 잘 반영하도록 하기 위한 것이다.

SPEC은 벤더가 SPECbase95 벤치마크를 출판하는 것을 장려하고 있다. SPEC95에서는 벤더로 하여금 SPEC95의 결과를 공표하기 전에 미리 SPEC에 기초 결과를 보고하도록 하여 하드웨어 구입자나 언론 매체 종사자들이 SPEC으로부터도 그 결과를 얻을 수 있도록 규정하고 있다. 이것은 최적화 및 기초 벤치마크 결과를 동시에 SPEC에 보고하여 협회 소식지에 실도록 규정한 SPEC92의 경우에 비하면 한층 나아진 것이다. 과거에는 몇몇 업체로부터 제때에 SPECbase92의 결과를 받는데 문제가 있었다. 우리는 벤더들이 벤치마크 결과를 소식지나 광고지에 내는 경우, 기초 결과를 최적화된 결과와 함께 신도

록 계속 독려할 것이다. 그러면 구매자들이 SPEC으로부터 기초 결과를 얻기 위해 별도의 노력을 쏟을 필요가 없을 것이다.

Dixit는 SPEC95가 더욱 컴파일러에 저항적이기 때문에 최적화된 벤치마크와 기초적인 벤치마크와의 차이가 SPEC92에 비해 작다고 말했다. SPEC은 SPEC95를 발표할 때 14개의 컴퓨터 벤더들로부터 받은 기초 결과를 첨부했다. 최적화된 결과는 제외되었으며, 최적화된 실행으로부터 얻어진 성능 향상 폭은 모두 10%미만이었다. 하지만 앞으로 벤더들이 새로운 테스트의 상세한 내용을 파악하게 되면 최적화로 인한 성능 향상 폭은 점점 증가될 것이다.

2. SPEC95의 세부 내용

SPECint95는 8가지 정수 테스트 점수의 기하 평균 값을, SPECfp95는 10가지 부동 소수점 테스트의 기하 평균값을 나타낸다. SPEC92에서는 DEC의 VAX 11/780을 기준으로 한 반면 SPEC95는 SPARCstation 10, 40 MHz를 사용한다. 이 워크스테이션에서는 첫 번째 정수 테스트로서 인공지능 프로그램인 바둑 게임 Go를 수행했는데 4,600초가 소요되었다. 반면에 가장 빠른 정수 연산 기계인 DEC의 Alpha AXP 3000 모델 900에서는 768초가 소요되었다. 이 테스트에서 Alpha 컴퓨터가 받은 점수는 5.99(=4,600/768)이다. 세 번째 실험 항목인 SPECrate95는 주어진 시간 내에 동일한 테스트를 몇

번이나 실행할 수 있는가를 측정한다. 이는 다중 프로세서 컴퓨터를 평가하는데 주로 사용된다.

SPEC95 부동 소수점 패키지에 큰 변화가 있었는데 배정도 연산에 대한 비중이 높아진 것이다. SPECfp92는 14개 테스트중 단지 하나만이 배정도에 관한 것이었는데, SPEC95에서는 10개 중에서 5개가 배정도에 관한 것이다. 배정도는 단정도에 비해 2배의 메모리가 필요하다. 따라서 배정도 연산에서는 CPU와 메모리 사이의 데이터 흐름에 부하가 커진다. 일반적인 컴퓨터가 64 bit 메모리 버스를 가지고 있는 반면, IBM 3CT는 128 bit 메모리 버스를 가지고 있기 때문에 벤치마크 부동 소수점 테스트에서 뛰어난 점수를 나타내고 있다.

과거의 SPEC92는 6개의 정수 테스트와 14개의 부동 소수점 테스트를 포함하고 있었다. SPEC95 정수 테스트에는 Go 게임, 데이터베이스 프로그램, 비디오 또는 JPEG 파일 작업이 추가되었고, 스프레드시트 프로그램과 로직 디자인 테스트인, "espresso"와 "eqntott"는 사라져 버렸다. 부동 소수점 테스트에서는 회로 설계 테스트인 "spice"를 없애고, 모토로라 마이크로 프로세서의 모의 실험을 정수 테스트에 추가하였다.

전자 CAD(ECAD) 응용을 다루는 엔지니어는 전자 회로 설계 프로그램들이 SPEC95에 잘 반영되지 않았다는 것을 어떻게 받아들여야 할까? SPEC의 CPU 소위원회 의장인 인텔의 Jeff Reilly는 워크스테이션들의 계산 능력을 평가하는 테스트의 전반적인 신

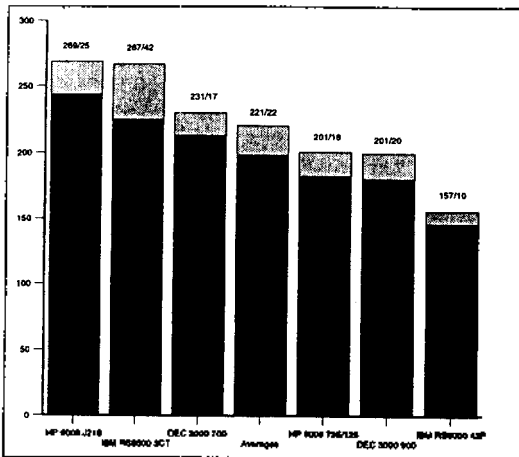


그림 1. SPECfp92의 기초 결과(검은색)와 최적화에 의한 증가치(회색).

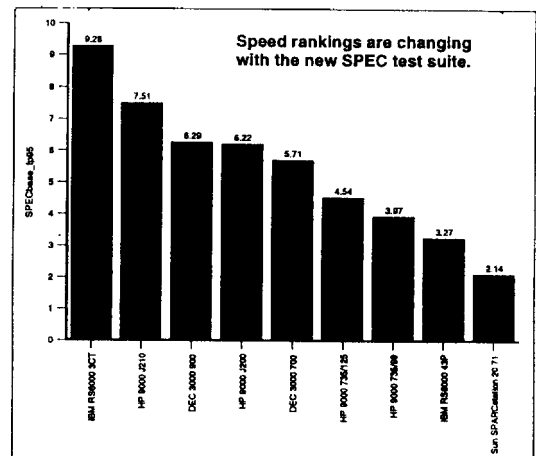


그림 2. SPECrate95의 기초 결과.

회도 항상에는 동조하면서도 회로 설계 엔지니어에게는 새로운 테스트가 "좀 실망스러운" 점이 있을 수도 있다는 것을 인정하고 있다. 그러나 Dixit은 그렇지 않다고 주장한다. 예컨대, 인공지능 테스트인 099.go는 논리 회로를 최적화 하는 것만큼 많은 "what if" 시나리오를 컴퓨터에 설정해야 한다. 147.vortex 데이터베이스 프로그램은 객체 지향 데이터베이스에서 데이터의 조작이 필요한데, 이것은 ECAD에서 설계 라이브러리를 골라내고 모의 실험 시 데이터를 앞뒤로 옮기는 과정과 매우 유사하다. 또한 그는 가우시안 소거나, 3차원 그리드에서의 상수 계수 방정식의 풀이 등의 많은 부동 소수점 테스트가 ECAD를 위한 것이라고 말한다.

3. 윈도우즈 NT

Dixit는 새로운 벤치마크 프로그램이 Unix OS 시스템이 아닌 Windows NT나 디지털(Digital)의 Open VMS에 대한 이식성이 더 크다고 말한다. NT 용은 곧 출시될 예정이다. 그러면, SPEC과 NT기반 플랫폼 벤더로 하여금 NT에 대한 벤치마크 결과를 내도록 하여 Unix 기계의 결과와 비교할 수 있을 것

이다. 예컨대, Alpha-based 워크스테이션이 32 bit OS인 NT를 탑재할 때와 64 bit OS인 Digital Unix를 탑재할 때와의 성능의 차이를 알아 볼 수 있다.

SPEC95 벤치마크의 첫 번째 테스트에서는 몇 가지 주목할 만한 사실이 발견되었다. SPEC92의 부동 소수점 테스트에서는 Hewlett-Packard의 J210 워크스테이션이 IBM의 가장 강력한 Power2 RS/6000 워크스테이션인 3CT를 근소한 차이로 따돌리고 있었다. 그러나 새로운 SPECbase_fp95 테스트에서는 3CT가 거의 25%나 빨랐다. 뿐만 아니라 64MB RAM과 1MB 보조 캐시를 장착한 133MHz 펜티엄 시스템이 똑같은 RAM과 L2 캐시를 장착한 SPARCstation 20-71 보다 정수와 부동 소수점 테스트에서 모두 앞섰다는 사실이다.

 <<CAD Report September, 1995. pp. 11-14>>

~~~~~  
 본 기사는 경상대학교의 전차수 편집위원이 "CAD Report"에서 발췌하였으며 출판사인 CAD/CAM Publishing Inc.의 연락처는 다음과 같다.

- Fax: 1-619-488-6052
- e-mail: Cadcirc@aol.com