

분자동력학 모의실험과 병렬컴퓨팅

글 _ 오광진 · 슈퍼컴퓨팅응용실 · koh@kisti.re.kr

1. 서론

최초의 분자동력학 모의실험은 지금으로부터 약 50년 전 미국 로렌스 리버모어 연구소에서 근무하고 있던 Alder와 Wainwright에 의해 행해졌다. 그들이 행한 모의실험의 대상은 강체구(hard sphere)로 이루어진 단순한 시스템이었으며 UNIVAC과 IBM 704를 이용하였다.

그들의 분자동력학 모의실험 이후 모의실험의 대상은 좀 더 복잡하고 실제에 가까운 시스템으로 확대되어 갔다. 예를 들면, 아르곤 액체(1964년 Rahman), 이원자 분자로 이루어진 액체(1968년 Harp와 Berne), 물(1971년 Stillinger와 Rahman), 작은 강체 분자로 이루어진 액체(1971년 Barojas, Levesque, Quentrec), 탄화수소 액체(1975년 Ryckaert와 Bellemans), 단백질 분자(1977년 McCammon, Gelin, Karplus) 등이 있다.

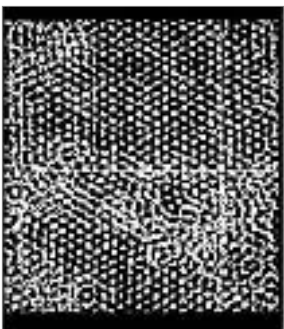
그런데 초창기 분자동력학 모의실험들은 대부분 수백 개

의 입자들로 구성된 비교적 작은 시스템에 대하여 수행된 것들이었다. 이는 컴퓨팅 파워의 부재에 기인한 것이다. 하지만 컴퓨터의 성능은 현재 분자동력학 모의실험이 태동하던 시기에는 상상하지 못할 만큼 비약적으로 발전하였다. 2004년 11월 발표된 TOP500 목록을 보면 세계에서 가장 빠른 IBM BlueGene/L은 무려 초당 7조번의 부동소수점 연산을 한다. 최초로 분자동력학 모의실험에 사용되었던 IBM 704와 비교한다면 무려 약 10,000,000,000 배 만큼 빠른 성능이다. 이와 같은 컴퓨팅 파워의 비약적인 발전은 분자동력학 모의실험 수행자들로 하여금 더욱 더 큰 규모의 분자동력학 모의실험을 수행하도록 고무시키기에 충분하였다. <그림 3, 4, 5>는 지금까지 행해진 대규모 분자동력학 모의실험의 몇몇 예들을 보여주고 있다.

본고에서는 앞서 언급한 분자동력학 모의실험에 대해 간단히 소개하고 대규모 분자동력학 모의실험을 가능케 하는 병렬처리에 대해 간단히 소개하고자 한다.



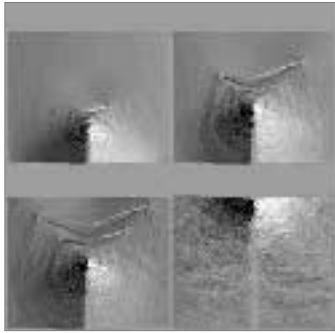
<그림 1>
미국 로렌스 리버모어 연구소에 있던 IBM 704 컴퓨터 - 부동소수점 연산 장치를 내장하고 있는 최초의 상업용 컴퓨터로 초당 약 5천 번의 부동소수점 연산 능력을 가짐



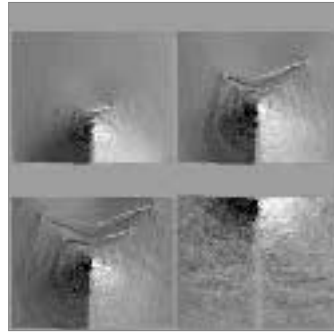
<그림 2>
Alder와 Wainwright가 분자동력학 모의실험으로부터 얻은 액체 - 고체 상전이 영역 근처에서의 강체구의 궤적



<그림 3>
일리노이 주립대학교의 술텐 그룹에서 수행한 ATP synthase의 분자동력학 모의실험(총 327,000개 원자로 구성됨). 술텐 그룹은 2002년 자체 개발한 NAMD라는 분자동력학 모의실험 코드를 사용하여 TFLOPS(초당 천억 번의 부동소수점 연산) 장벽을 넘었음



〈그림 4〉
IBM Almaden 연구소의 Abraham 그룹에서 수행한 이차원 Lennard - Jones 고체의 균열현상 분자동력학 모의실험 (총 2,027,776개의 원자로 구성됨)



〈그림 5〉
남가주대학교의 나카노와 그의 공동연구자들에 의해 수행된 실리콘/실리콘나이트라이드 나노픽셀의 분자동력학 모의실험 (총 10,000,000개의 원자로 구성됨)

2. 분자동력학 모의실험

분자동력학 모의실험이란 운동방정식의 해를 수치적으로 구하는 수치실험을 일컫는 말이다.

여기서 운동방정식은 고전역학을 지배하는 뉴우튼 방정식이나 양자역학을 지배하는 슈뢰딩거 방정식일 수도 있고 브라운 운동을 기술하는 랑주뱅 방정식과 같은 확률론적인 운동방정식일 수도 있다. 운동방정식은 분자동력학 모의실험을 통해 재현하고자 하는 시스템에 따라 또는 현상에 따라 선택하면 된다. 보통 분자동력학 모의실험하면 보통 뉴우튼 방정식을 사용하는 분자동력학 모의실험을 지칭하고 슈뢰딩거 방정식이 사용되면 양자 분자동력학 모의실험(ab initio molecular dynamics simulation)이라 불리고 랑주뱅 방정식이 사용되면 브라운 동력학 모의실험(brownian dynamics simulation)이라고 불린다.

여기에서는 뉴우튼 방정식을 사용한 분자동력학 모의실험에 대해서만 논의하도록 하고 양자 분자동력학 모의실험이나 브라운 동력학 모의실험은 생략하기로 하겠다. 뉴우튼 방정식은 주지하다시피 이차미분방정식이고 <가속도 = 힘 / 질량>의 형태로 주어진다. 여기서 가속도는 위치의 시간에 대한 2차 도함수이다. 따라서 뉴우튼 방정식의 해는 뉴우튼 방정식을 시간에 대해 수치적으로 적분하면 얻을 수 있다. 수치적분에 사용되는 알고리즘들 중 하나인 velocity verlet 알고리즘을 소개하면 다음과 같다.

1. 입자들의 초기 위치와 속도를 입력한다.
2. 각 입자들에 작용하는 초기 힘을 계산하고 가속도를 계산한다.

3. $t + \Delta t/2$ 에서 각 입자들의 속도를 계산하고 $t + \Delta t$ 에서 위치를 계산한다.

$$\text{속도}(t + \Delta t/2) = \text{속도}(t) + \Delta t/2 \times \text{가속도}(t)$$

$$\text{위치}(t + \Delta t) = \text{위치}(t) + \Delta t \times \text{속도}(t + \Delta t/2)$$

4. $t + \Delta t$ 에서 각 입자들에 작용하는 힘을 계산하고 가속도를 계산한다.

5. $t + \Delta t$ 에서 각 입자들의 속도를 계산한다.

$$\text{속도}(t + \Delta t) = \text{속도}(t + \Delta t/2) + \Delta t/2 \times \text{가속도}(t + \Delta t)$$

6. 3-5 과정을 반복한다.

여기서 Δt 를 작게 하면 할수록 우리는 정확한 해를 얻을 수 있다. 하지만 Δt 를 작게 하고 똑같은 시간만큼 모의실험을 진행시키려면 더욱더 많은 반복을 필요로 하게 된다. 따라서 Δt 는 정확도에 크게 영향을 미치지 않는 범위 내에서 가능하면 크게 잡는다. 생체분자에 대한 분자동력학 모의실험인 경우 대부분 1 또는 2 펨토초(100조분의 1초)를 사용한다.

가속도를 계산하기 위해서는 힘을 계산하여야 하는데 힘은 입자들 사이의 퍼텐셜에너지(상호작용 에너지)의 거리에 대한 1차 미분으로부터 계산한다. 그리고 퍼텐셜에너지는 상호작용에 대한 모델을 통해 계산된다. 예를 들면, 입자들 사이의 반 데어 발스 상호작용은 Lennard-Jones 퍼텐셜 함수를 통해 계산되는데 두 입자 사이의 거리의 함수로서 그 형태는 다음과 같이 주어진다.

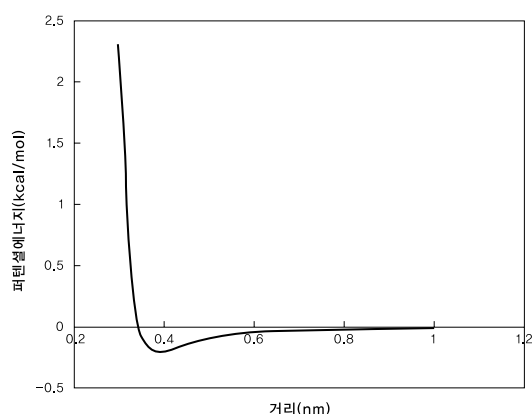
퍼텐셜에너지 (거리) =

$$4 \times \varepsilon \times [(\sigma/\text{거리})^{12} - (\sigma/\text{거리})^6]$$

여기서, ε 와 σ 는 경험적으로 얻어지는 상수이고 입자의 종류에 따라 다른 값을 갖는다. <그림 6>은 Lennard-Jones 퍼텐셜 함수로부터 계산된 아르곤($\varepsilon=0.238$ kcal/mol, $\sigma=0.3405$ nm)의 거리에 따른 퍼텐셜에너지 곡선을 보여준다.

분자동력학 모의실험에는 앞서 언급한 반 데어 발스 상호작용 외에도 정전기적 상호작용, 결합길이, 결합각, 이면

각의 변화와 같이 분자의 구조 변화에 따른 상호작용 등이 고려된다.



<그림 6> 아르곤의 거리에 따른 퍼텐셜에너지 곡선

3. 병렬처리

분자동력학 모의실험에서 대부분의 계산시간이 소요되는 부분은 퍼텐셜에너지 계산 부분이다. 계산시간을 줄이기 위한 특별한 알고리즘을 사용하지 않았을 경우, 퍼텐셜에너지를 계산하는 시간은 사용된 입자의 개수의 제곱에 비례하여 증가한다.

따라서 입자의 개수가 증가하면 계산시간은 기하급수적으로 늘어나게 된다. 예를 들면, 백 개의 원자로 이루어진 시스템과 백만 개의 입자로 이루어진 시스템을 생각해 보자. 후자의 경우 퍼텐셜에너지 계산시간은 전자에 비해 일억 배 만큼 더 느리게 된다. 따라서 이러한 장벽을 극복하기 위한 여러 가지 효과적인 알고리즘들이 제안되어 왔는데 대표적인 것으로는 potential truncation 방법이나 neighbor list 방법과 같은 비교적 간단한 방법에서부터 smooth particle mesh Ewald 방법이나 fast multipole 방법에 이르는 복잡한 방법까지 다양한 방법들이 있다. fast multipole 방법의 경우에는 계산시간이 단순 퍼텐셜에너지 계산처럼 입자의 개수의 제곱에 비례하는 것이 아니라 그냥 입자의 개수에 비례하는 것으로 알려져 있다.

그러나 앞에서 언급한 퍼텐셜에너지 계산시간을 줄이기 위한 여러 알고리즘들이 있다 하더라도 그 알고리즘들만으로는 하나의 프로세서에서 대규모 분자동력학 모의실

험을 수행한다는 것은 아직 어려움이 있다. 이에 연구자들은 병렬컴퓨터에서 수백 또는 수천 개의 프로세서를 효과적으로 이용할 수 방법들에 대해서도 많은 연구를 하고 있다.

병렬처리를 위해서는 작업을 프로세서마다 분배하게 되는데 대표적으로 원자분할기법과 영역분할기법이 있다. <그림 5>에서와 같이 원자분할 기법에서는 각 프로세서마다 일정한 수의 입자가 할당되고 분자동력학 모의실험 동안 프로세서 간에 입자의 전이 없이 그대로 유지된다. 원자분할기법은 프로그래밍 하기 수월하다는 장점이 있다.

하지만 퍼텐셜에너지를 계산하기 위해서는 매번 모든 프로세서 간에 정보를 서로 교환해야 하기 때문에 프로세서의 수가 증가하면 통신에 따른 시간도 증가하게 된다. 따라서 프로세서의 수가 증가하게 되면 병렬처리 효율성이 떨어진다. 반면에 <그림 6>에서와 같이 영역분할 기법에서는 모의실험 상자를 프로세서의 숫자만큼 작은 영역으로 분할하고 각 프로세서는 할당된 자신의 영역을 담당한다. 이 경우에는 입자들의 운동에 기인하여 프로세서 간에 입자들의 전이가 빈번하게 이루어지나 인접한 프로세서와의 통신만 하면 된다. 따라서 원자분할기법에서처럼 통신으로 인한 병렬처리 효율성의 저하는 없다. 이것이 가능한 것은 분자동력학 모의실험에서는 보통 일

정 거리 이상에서는 입자 간에 상호작용을 하지 않는다고 가정하기 때문이다. 즉, 영역의 크기를 그 거리 이상으로 잡으면 퍼텐셜에너지를 계산할 때 인접한 영역에 있는 입자들과의 상호작용만 고려하면 되기 때문이다.

여기서 한 가지 더 언급하고 싶은 것은 원자분할기법이나 영역분할기법을 사용하여 작업을 분배하여 병렬처리를 해도 프로세서 간에 작업의 분배가 균등하게 이루어지지 않으면 병렬처리 효율성이 저하된다는 점이다. 따라서 병렬처리를 할 경우에는 부하의 균형도 고려할 중요한 요소이다.

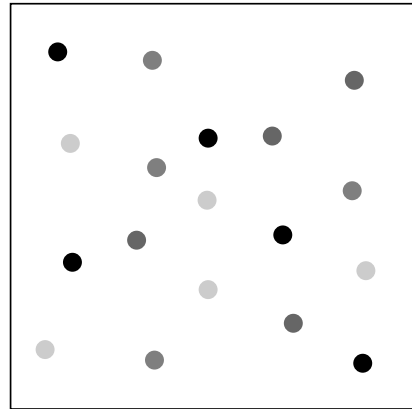
코드의 병렬처리 효율성을 측정하는 하나의 척도로서 속도향상률이 사용된다. 속도향상률이란

$$\text{속도향상률 (N)} = \frac{\text{한개의 프로세서를 사용했을 때 걸린 시간}}{\text{N개의 프로세서를 사용했을 때 걸린 시간}}$$

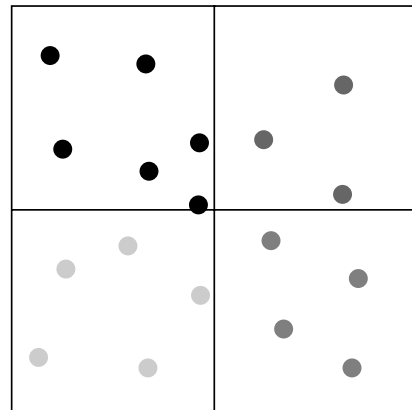
로 정의된다. 위의 식으로부터 알 수 있듯이 이상적인 코드는 속도 향상률의 기울기가 1인 코드이다. 즉, N개의 프로세서를 사용하였을 경우 N배만큼의 속도가 향상되는 코드이다. 하지만 프로세서간의 통신 때문에 실제로는 이런 일은 일어나지 않는다.

<표 1>은 현재 많이 사용되고 있는 분자동역학 모의실험 코드들 중 CHARMM, AMBER, NAMD에 대해서 스크립스 연구소의 브룩스 그룹에서 속도 향상률을 비교해 놓은 표이다. 사용된 컴퓨터는 샌디에고 슈퍼컴퓨팅 센터 Blue Horizon (IBM Power3 프로세서, 375 Mhz)이고 사용된 시스템은 물과 단백질로 총 23,558개의 원자로 구성돼 있다.

<그림 7>과 <그림 8>은 KISTI 슈퍼컴퓨팅 센터에서 자체 개발하고 있는 분자동역학 모의실험 코드의 속도 향상률을 보여준다. <그림 7>은 모델 세포막에 대하여 IBM p690에서 검사한 결과를 보여주고 <그림 8>은 이차원 Lennard-Jones 고체에 대하여 리눅스 클러스터에서 검사한 결과를 보여준다.



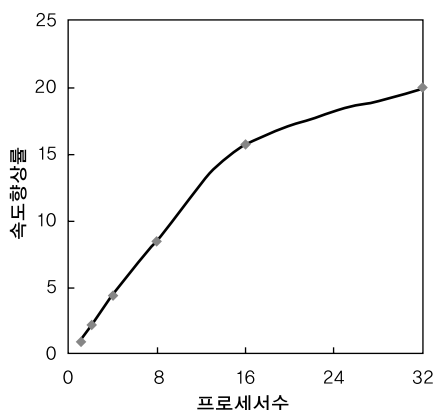
<그림 7> 원자분할기법 (다른 색깔은 다른 프로세서를 의미함)



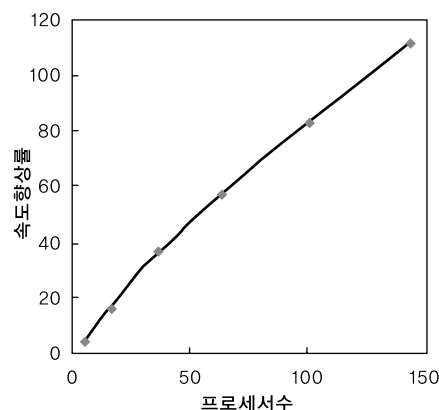
<그림 8> 영역분할기법 (다른 색깔은 다른 프로세서를 의미함)

프로세서수	CHARMM c29b1	AMBER 7.0	NAMD 2.4
1	2643	2538	2986
2	1371 (1.9)	1229 (2.1)	1809 (1.7)
4	724 (3.7)	658 (3.9)	930 (3.2)
8	414 (6.4)	375 (6.8)	516 (5.8)
16	255 (10.4)	228 (11.1)	308 (9.7)
32	195 (13.6)	162 (15.7)	202 (14.8)
64	182 (14.5)	124 (20.5)	124 (24.1)
128			126 (23.7)

<표 1> CHARMM, AMBER, NAMD의 계산시간 비교. 괄호안은 병렬처리에 따른 속도향상률



〈그림 9〉 모델 세포막(총 12,487개의 원자로 구성)에 대한 분자동력학 모의실험의 속도 향상률



〈그림 10〉 14차원 Lennard-Jones 고체(총 2,000,000개의 원자로 구성)에 대한 분자동력학 모의실험의 속도 향상률

4. 결론

분자동력학 모의실험은 나노과학이나 생명과학 등 여러 분야에 걸쳐 매우 유용하게 사용되고 있다. 아주 좋은 현미경으로도 관측할 수 없는 미시세계에서의 현상에 대한 자세한 정보를 제공해 주기도 하고 실험을 수행하기 힘든 극한 상황에서의 현상을 재현해 주기도 한다. 그리고 연구자들이 제안한 이론을 검증해 주기도 한다.

하지만 분자동력학 모의실험이 모든 문제를 해결해 주는 것은 아니다. 분자동력학 모의실험은 크게 두 가지 한계를 갖고 있다. 하나는 길이의 한계이고 다른 하나는 시간의 한계이다. 우선, 길이의 한계부터 논하고자 한다. 우리가 눈으로 보는 물질들은 대략 아보가드로 수(6×10^{23}) 만큼의 입자를 포함하고 있고 그 크기는 보통 밀리미터 이상이다. 하지만 분자동력학 모의실험에서는 지금까지 시도된 가장 큰 길이이라 할지라도 마이크로미터(십만분의 1미터) 정도이다. 따라서 거시세계의 길이 단위에서 일어나는

모든 현상을 분자동력학 모의실험이 재현해주지는 못한다. 다음, 시간의 한계에 대해서 언급하고자 한다. 현재까지 수행된 분자동력학 모의실험 중에서 가장 오랜 시간동안 진행된 것이 마이크로초(십만분의 1초) 정도이다. 하지만 거시세계에서 일어나는 현상은 보통 초 단위 이상에서 일어난다. 뿐만 아니라 미시세계라 할지라도 단백질과 같은 생체분자의 경우에는 마이크로초 단위 이상에서 일어나는 현상도 많다. 이러한 현상들은 분자동력학 모의실험으로 재현할 수 없다. 이와 같은 분자동력학 모의실험의 한계를 극복하기 위해서는 앞으로 더욱더 많은 연구가 이루어져야 할 것으로 생각된다.

지금까지 분자동력학 모의실험과 그 병렬처리에 대해서 간단히 소개하였다. 지면 관계 상 자세한 내용에 대해서는 생략하였지만 분자동력학 모의실험에 생소한 분들에게 조금이나마 분자동력학 모의실험을 이해하는데 도움이 되었으면 하는 바람이다. 