ISSN 2383-630X(Print) / ISSN 2383-6296(Online)
Journal of KIISE, Vol. 47, No. 1, pp. 1-10, 2020. 1
https://doi.org/10.5626/JOK.2020.47.1.1

KISTI 슈퍼컴퓨터 5호기 누리온 시스템을 이용한 지역해양모델의 병렬 최적화 연구

(A Study on Parallel Optimization of Regional Ocean Model Using KISTI 5th Supercomputer Nurion System)

김 동 훈 † 임 채 욱 †† 조 민 수 †††
(Dong-Hoon Kim) (Chaewook Lim) (Min-Su Joh)

안 준 언 †††† 문 일 주 ††††† 우 승 범 ††††††
(Jooneun An) (Il-Ju Moon) (Seung-Buhm Woo)

요 약 한국과학기술정보연구원 슈퍼컴퓨터 5호기 누리온 KNL(Knights Landing) 시스템의 장단점을 분석하여 ROMS (Regional Ocean Modeling System) 모델의 병렬 최적화를 시도하여 성능을 분석해 보았다. KNL 시스템이 Skylake 시스템 보다 노드 당 두 배 이상의 core를 가지고 있지만, 동일한 core를 사용 할 경우에는 KNL 시스템의 성능이 약 3배 정도 느린 것으로 알려져 있다. 그러나 누리온에서의 ROMS 모델의 KNL 성능은 Skylake 시스템에 비하여 약 2배 정도 차이에 그치는 것으로 나타났다. 이는 누리온의 KNL 시스템이 기존 KNL 시스템보다 약 1.3배 빠른 것을 의미하며, 누리온 시스템의 성능이 우수한 것을 말해 준다. 이러한 성능에 대하여 본 연구에서는 이상 실험 및 실황 실험의 결과를 토대로 성능 비교를 제시하였으며, KNL 시스템은 이상 실험과 실황 실험 모두에서 최대 코어 실험의 경우까지도 계속적으로 성능이 향상되는 결과를 보이고 있어서 초병렬 모델의 수행에 유리한 것으로 나타났다.

키워드: Nurion, ROMS, KNL, SKL

Abstract A parrallel optimiztion of the Regional Ocean Modeling system (ROMS) was conducted by analyzing the computational characteristics of the Knights Landing (KNL) system of Nurion, which is the fifth supercomputer at the Korea Institute of Science and Technology Information (KISTI). Although the KNL system comprises more than twice as many cores per node as the Skylake system,

· 이 논문은 2019년 해양수산부 재원으로 해양수산과학기술진흥원의 지원을 받아 시행된 연구임(해양수치모델링과 지능정보기술을 활용한 해양예측 정확도 향상 연구)
· 이 연구는 한국과학기술정보연구원 슈퍼컴퓨팅인프라센터가 수행하는 주요사업 "국가 슈퍼컴퓨팅 인프라 구축 및 서비스" 사업(K-18-L12-C01)의 지원으로 수행되었습니다.

† 비 회 원 : 인하대학교 해양과학과 교수
mail@dhkim.info
†† 비 회 원 : 인하대학교 해양과학과 학생
lichwo@gmail.com
††† 비 회 원 : 한국과학기술정보연구원 책임연구원
msjoh@kisti.re.kr
†††† 비 회 원 : 한국과학기술정보연구원 선임연구원
rocahn@kisti.re.kr
††††† 비 회 원 : 제주대학교 해양산업경찰학과 교수
ijmoon@jejunu.ac.kr
†††††† 정 회 원 : 인하대학교 해양과학과 교수(Inha Univ.)
sbwoo@inha.ac.kr
(Corresponding author임)

논문접수 : 2019년 6월 24일
(Received 24 June 2019)
논문수정 : 2019년 9월 20일
(Revised 20 September 2019)
심사완료 : 2019년 10월 16일
(Accepted 16 October 2019)

Copyright©2020 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.
정보과학회논문지 제47권 제1호(2020. 1)

the performance of the KNL system is known to be about three times slower than that of Skylake system, when using the same number of cores. However, the KNL performance optimized for the ROMS model in Nurion is only about twice that of the Skylake system suggesting that the KNL system of Nurion is approximately 1.3-fold faster than the normal KNL system. In this study, two types of numerical experiments (ideal & real cases) were conducted to compare the computational capabilities of the KNL and SKL systems; The KNL system has more computational efficiency. The KNL system shows continuous improvement in performance even in maximum core experiments under both ideal and real case simulation, which is an advantage for the simulation of super parallel numerical calculation.

Keywords: Nurion, ROMS, KNL, SKL

1. 서 론

한국과학기술정보연구원(Korea Institute of Science and Technology Information, 이하 KISTI)에서 이번에 도입한 슈퍼컴퓨터 5호기 누리온(Nurion)은 연산 속도가 이론성능과 실측성능 기준 25.7PF와 13.92PF에 이르고 계산 노드가 8305개(약 57만개 코어)로 구성되어 있어 기존의 4호기 성능보다 70배 이상 향상되었다. 이는 2018년 11월 기준 세계 상위 500대의 슈퍼컴퓨터 가운데 13위(실측성능 기준)에 해당하는 성능이다. 특히 CPU와 메모리, 디스크 간 데이터 이동을 고속으로 처리하기 위해 SSD 계층을 추가해 고속화 부분에 중점을 뒀다. 저장용량은 약 21PB이다.

누리온에 탑재된 Processor는 크게 두 가지로 구성되어 있는데, Intel Xeon Phi 7250 Knights Landing Processor을 탑재한 시스템(이후에는 “KNL”로 줄여 씀)과 Intel Xeon 6148 Skylake Processors를 탑재한 시스템(이후에는 “SKL”로 줄여 씀)으로 구축되었다. 본 연구에서는 결과 분석의 객관성을 높이기 위하여 일반 클러스터 시스템(제주대학교 대용량 클러스터, Hurricane)으로도 성능을 분석하였는데 이를 “CLS”로 줄여서 쓰도록 하겠다.

누리온 KNL 시스템은 8,305 nodes, SKL 시스템은 132 nodes로 구축하였다. KNL은 매니코어 1세대 제품에서 독립적으로 운영체제를 실행할 수 있도록 인텔 xeon과 완전히 호환되도록 출시되었다[1]. 누리온 KNL 시스템은 CPU 당 코어의 수가 68개인 반면, 할당된 메모리가 96 GB이기 때문에 한 코어 당 약 1.4 GB 성능을 보인다. SKL 시스템은 CPU 당 코어의 수가 20개이고, 할당된 메모리가 192GB로 한 코어 당 약 9.6 GB 성능을 보인다. 이처럼 KNL 시스템은 상대적으로 메모리가 작고, CPU의 기본 주파수 또한 1.4 GHz로 SKL 시스템 및 일반 클러스터(CLS) CPU 속도(2.5~3 GHz) 보다 느려 일반 수치모델을 모의하게 되면 성능이 50% 이하로 떨어질 수 있다.

DELL EMC에서 2018년 6월에 제공한 백서1)인 “Application Performance of Intel Skylake and Intel Knights Landing Processors on Stampede2”에서는 여러 가지 수치모델을 이용하여 SKL 시스템과 KNL 시스템의 성능을 비교하였다. 이 백서에는 Weather Research and Forecasting model(WRF)의 성능 비교도 수행하였는데, 그 결과는 Fig. 1과 같다[2]. 이 실험에서는 WRF v3.6.1을 사용하여 CONUS(continental United States)를 2.5km 해상도의 benchmark 모형으로 구성하였으며 parallel-netcdf 1.8.1 라이브러리를 사

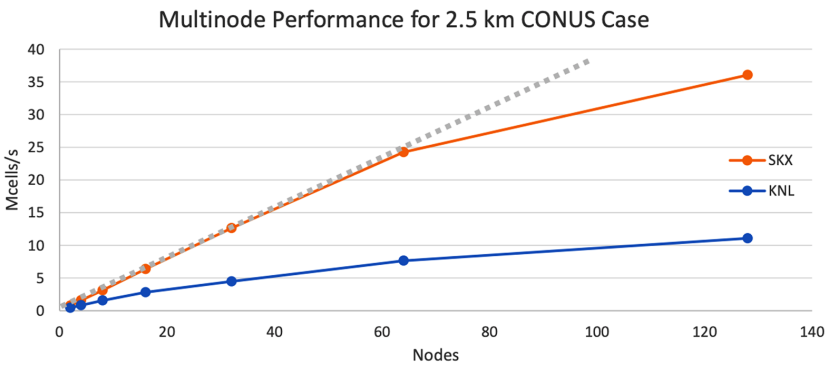


그림 1 KNL과 SKL(=SKX)에서의 WRF 성능 비교

Fig. 1 Comparison of WRF performance between KNL and SKL (=SKX)

용하여 입출력에도 병렬화를 적용하였다. 이것은 KNL 시스템과 같이 매니코어이면서 상대적으로 메모리의 용량이 적은 경우에 모델의 입출력에 사용하는 메모리를 분산시킬 수 있어서 유리하다. 비록 이 백서는 Skylake Processor의 우수성을 홍보하기 위한 것임을 염두에 두더라도, SKL 시스템에 비하여 기존의 KNL 시스템의 성능이 2~3 배 이상 느린 것을 알 수 있다. 이전 연구(노승은, 최지은, 남덕윤, 박근철, 박찬열, “매니코어 시스템에서의 병렬 프로그래밍 최적화를 위한 분석 도구 및 벤치마크 성능 실험,” 한국정보처리학회, 제25권 제1호 pp. 78-88. Jan. 2018)에서는 KNL 시스템을 사용하여 메모리와 적정 노드 수, MPI 프로세스, 스레드 수에 따른 성능 실험을 하였고 그에 따른 효율을 제시하였다. Cho et. al., (2017)에서는 KNL 시스템의 특징 중 하나인 온-패키지 메모리(Multi-Channel DRAM, MCDRAM)를 사용할 경우 동일 노드 상에서 병렬 MPI의 성능이 기존에 비하여 최대 272% 향상됨을 밝혔으며[3], Butcher et. al.,(2018)에서는 MCDRAM에 적합하지 않은 종류의 자료에서의 최적화 방법을 연구하였다[4]. 또한, Rho et. al., (2017)에서는 KNL 시스템에서 MPI 응용 프로그램들을 실행할 때 발생할 수 있는 성능 병목 현상을 해결하는 논의를 하였으며[5], Yoon and Song(2019)는 KNL과 같은 매니코어 아키텍처의 시스템 특징과 성능을 분석하였다[6]. 다만, 이러한 연구들은 실제로 연구자들이 사용하는 수치모델을 사용하여 실험한 것은 아니며 입출력이 미치는 영향에 대한 결과가 포함되어 있지 않다. 따라서, 본 연구에서는 SKL시스템 또는 CLS 시스템에 비하여 누리온의 주 시스템인 KNL 시스템의 성능이 실제로 어느 수준인지 분석하기 위하여 해양순환모형인 Regional Ocean Modeling System(ROMS)을 이용하였다. 이상(Benchmark) 실험과 실황 실험(태풍 차바)을 구성하고, 모의 결과를 이용하여 성능 비교를 진행하였다.

2. 실험 구성

누리온 시스템과 일반 클러스터 시스템의 사양을 요약해 보면 Table 1과 같다. 누리온 시스템에서 사용하는 기본 포트란 컴파일러인 Intel Fortran v17.x를 사용하였으며, MPI 라이브러리는 OpenMPI v3.x를 사용하고 모델 자료의 입출력에 사용되는 NetCDF 라이브러리는 v4.x를 사용하였다. 다만, 일반 클러스터 시스템인 CLS에는 동일 버전의 컴파일러가 설치되어 있지 않아서 최신 버전인 v18.x을 사용하였으나 버전 별 성능의 차이는 거의 없는 것으로 파악된다.

Table 1의 “Compiler Options”에서 보는 바와 같이 KNL 시스템은 Many Integrated core(MIC) 최적화 옵션인 “-xMIC -AVX512”를 사용하였으며, SKL은 “-xcore -AVX512”를 사용하였다. 모델 성능 분석을 위하여 수백 개의 코어를 사용하여 병렬화 할 경우에 1 코어 당 계산해야 하는 격자의 수가 너무 적으면 CPU의 성능보다는 CPU 간 또는 Node 간의 네트워크 성능에 크게 좌우되므로 경험상 1 코어 당 최소한 100(위도방향)x100(경도방향)=10000개 이상의 격자가 할당되도록 모델을 구성하여야 한다. 예를 들어, 100개의 CPU 코어를 사용하는 경우에 ROMS 모델의 구성을 최소 1000×1000 격자로 구성되는 초고해상도 모델로 구성해야 하는 것이다. 초고해상도 모델의 실험은 입출력이 없이 계산만 수행하는 이상 실험의 경우에 비하여 실제 지형과 실제 외력을 반영한 실험의 경우는 입출력 자료의 방대함과 모델의 overflow를 방지하기 위한 많은 노력을 기울여야 하므로 신중하게 구성하여야 한다.

표 1 슈퍼컴퓨터 5호기 누리온과 일반 클러스터의 시스템 사양 요약

Table 1 Summary of system specifications for 5th supercomputer (Nurion) and general cluster

Category	KNL	SKL	CLS
Manufacture and model	Intel Xeon Phi 7250 KnightsLanding	Intel Xeon 6148 Skylake	Intel Xeon Gold 6140
Number of nodes	8305	132	16
CPU × cores per node	1 × 68 = 68	2 × 20 = 40	2 × 18 = 36
Clock speed	1.4GHz	2.4GHz	2.3GHz
Main Memory	16GB (MCDRAM), 96GB	192GB	252GB
File System	Lustre	Lustre	Lustre
Compiler	Intel v17.0.5	Intel v17.0.5	Intel 18.0.4
MPI Library	openmpi v3.1.0	openmpi v3.1.0	openmpi v3.1.1
NetCDF Library	netcdf v4.6.1	netcdf v4.6.1	netcdf v4.4.4
Compiler Options	- fp-model precise - ip - O3 - fPIC - xMIC-AVX512	- fp-model precise - ip - O3 - fPIC - xcore-AVX512	- fp-model precise - ip - O3 - fPIC

2.1 Configuration of Ideal(Benchmark) Case

ROMS는 해양의 순환을 모의하기 위한 3차원 수치 순환 모델이며, 지역 규모의 해양 모의에 보편적으로 사용되는 수치 모델이다. ROMS는 free-surface와 terrain-following, primitive equation을 사용하며 계산 효율성을 위하여 정수역학, 경·수압 계산 시간 간격 분할 기법 등을 사용한다[7]. 본 연구에서는 ROMS version 3.6을 사용하였다.

이상 실험은 Fig. 2와 같이 이상적인 상황에서의 전 지구 동경 0°~동경 360°, 남위 70°~북위 70°의 영역을 대상으로 8192 × 1024 격자로 구성하였다. 경도 방향으로는 약 2.6' 간격, 위도 방향으로 약 8.2' 간격으로 세로로 넓은 직사각형 모양으로 구성하여 동일한 사각형 지역을 분할할 경우 경도 방향으로 약 4~8배 정도 코어를 더 할당하여 계산 효율성을 높이도록 하였다. 참고로, 다음 절에 설명하는 실제 지형의 경우는 계산 성능보다는 계산 결과를 더 중시해야 하므로 정사각형에 가깝게 구성할 것이다.

모델의 전체 수심은 4000m로 동일하게 설정하였으며, 전지구 형태로 동서경계는 서로 연결하고 남극 경계의 수심은 500m로 낮게 구성하여 남극을 간단하게 표현하고 연직 층수는 30개로 나누었다. 초기장은 해수면 높이(η)와 해류(U, V, W)를 모두 0으로 초기화하고, 수온(T)은 표층으로부터 저층까지 3.5℃~0℃로 순차적으로 줄여 나감으로써 밀도 성층화를 표현하였으며, 염분(S)은 모두 35psu의 값으로 동일하게 구성하였다. 시간 간격은 15초로 하여 성능 분석에 필요한 적절한 적분 시간인 3시간 동안을 적분하였다. 이상 실험은 계산 결과에 큰 의미를 부여하지 않기 때문에 결과의 분석은 생략하고 계산 성능만을 비교하여 제시하도록 하겠다.

2.2 Configuration of Real(Typhoon Chaba) Case

북서태평양(North Western Pacific ocean, NWP) 실험은 실제 지형과 실제 외력을 고려하여 구성하였다. 우리나라를 중심으로 동경 117°~150°, 북위 20°~52°의 영역을 대상으로 1400 × 1600 격자로 구성하였다. 격자

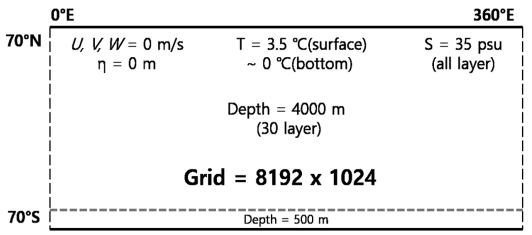


그림 2 ROMS를 사용한 초고해상도 이상 실험 구성도
Fig. 2 The configuration of high-resolution experiment using ROMS

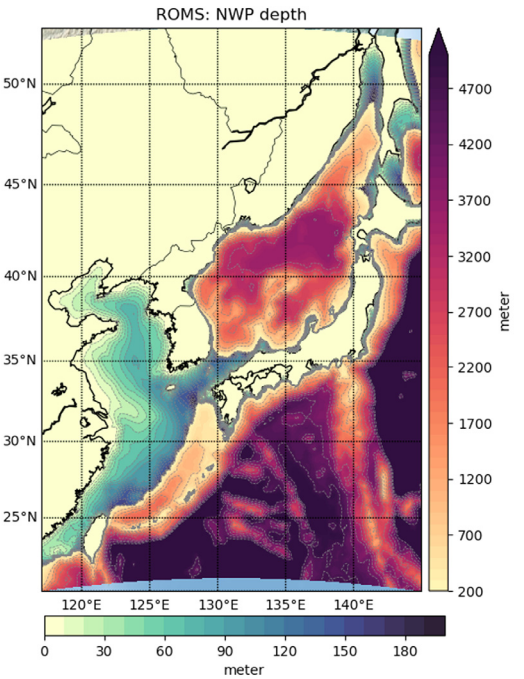


그림 3 ROMS의 NWP 영역 및 수심 구조
Fig. 3 NWP area and depth of ROMS

당 간격은 약 0.5km로 직사각형에 가깝게 구성하였다. 모델의 수심은 20" 간격의 해상도를 가지는 ETOPO20의 자료를 사용하여 격자에 맞게 내삽(Fig. 3)하고 연직으로 30개의 층으로 나누었다.

모델의 초기장과 외력은 자료의 입수 및 편집 여건을 고려하여 2016년에 발생한 태풍 차바의 발생기간을 선정하였다. 2016년 9월말에 발생한 제18호 태풍인 차바는 5등급의 슈퍼 태풍으로 활동기간은 2016년 9월 28~10월 6일로 최저 기압은 905hPa, 최대 풍속은 50 m/s 이상을 기록하였다. 우리나라에 피해(사망자 10명)를 준 차바가 제주도에 최 근접한 시기인 2016년 10월 5일 01시~04시를 선택하였다(Fig. 4). 모델의 초기장은 3차원 해양의 U, V, SSH, T, S 등이 필요하며, 이와 함께 모의 기간 동안의 동서남북 연직 경계층의 경계치가 시계열로 필요하다. 이를 위하여 Simple Ocean Data Assimilation ocean/sea ice reanalysis(SODA.[8])의 최신 버전인 SODA3.12.2 자료 중 5일 평균된 2016년 10월 5일 자료를 NWP 모델 영역에 맞게 내삽하여 사용하였다(Fig. 5). 또한, ROMS의 표층 대기 외력은 동 시간대의 1시간 간격의 Modern-Era Retrospective analysis for Research and Applications, Version 2(MERRA2, [9])의 자료를 사용하였다(Fig. 5). ROMS에서 필요로 하는 표층 대기 외력은 Surface Wind, Heat Flux(short

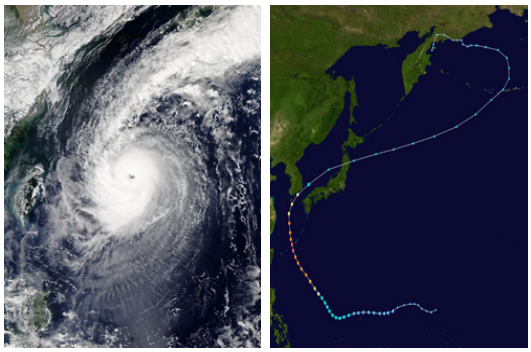


그림 4 2016년 제 18호 태풍 차바(CHABA)

Fig. 4 Satellite image and route of CHABA (2016 18th Typhoon)

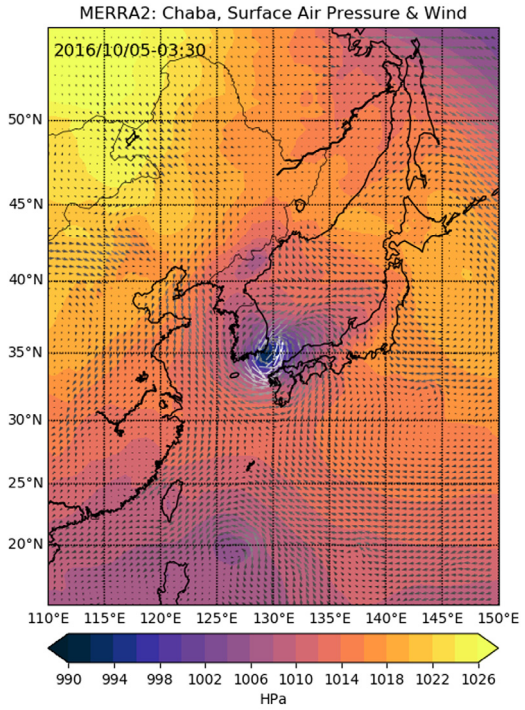


그림 5 태풍 차바 기간의 표층 외력(Pair, U10wind from MERRA2)

Fig. 5 The surface layer external force (Pair, U10wind from MERRA2)

wave, long wave), Fresh Water Flux(precipitation, evaporation), Surface Air Pressure, Cloud, Albedo 등이며 ROMS의 입력 형식에 맞게 변환하여 입력하였다. 적분 시간 간격은 10초로 하여 성능 분석에 필요한 적절한 적분 시간인 3시간 동안만 적분하였다. 본 실험은 모의 결과에 큰 의미를 부여하지 않기 때문에 성능

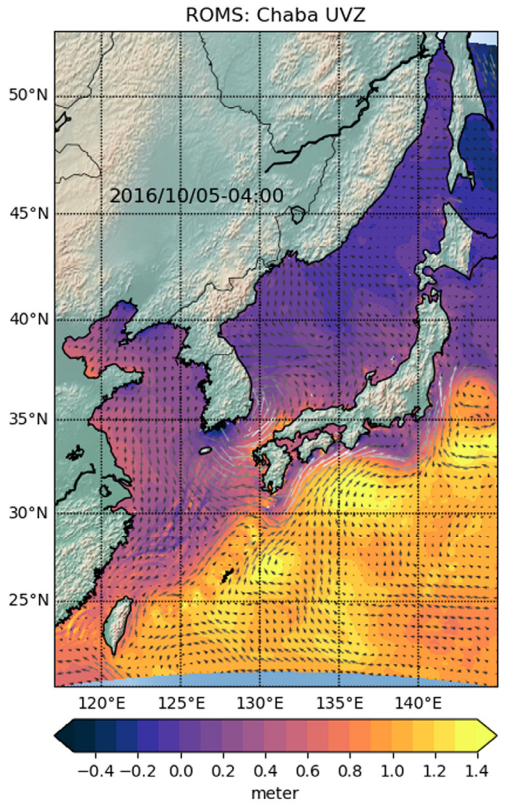


그림 6 태풍 차바 기간의 ROMS 모의 결과(SSH, Surface U & V)

Fig. 6 ROMS simulation results during CHABA period (SSH, Surface U & V)

분석에 필요한 짧은 기간만을 모의하였다. 따라서 본 연구에서는 모의 결과 분석은 하지 않고 계산 성능만을 비교하여 제시하도록 하겠다. 또한, 실제 지형과 실제외력을 사용한 초고해상도 해양모델 구성에 대한 상세한 방법은 본 연구의 목적에서 벗어나므로 설명을 생략하였다. 참고로, 모의 결과 중에서 KNL시스템(448 코어 사용)에서 모의한 결과인 표층 해류와 해수면 높이를 Fig. 6에 제시한다. 결과에서 확인할 수 있듯이 태풍 차바가 제주도를 지나가는 시기의 표층 해류와 해수면 높이가 잘 모의되고 있다.

3. 실험 결과

누리온의 KNL 시스템과 SKL 시스템의 노드 당 CPU는 각각 1개와 2개이며, 1 CPU 당 코어 수는 68개와 20개이다. 또한, 일반 클러스터인 CLS 시스템은 노드당 2개의 CPU와 1 CPU 당 18개의 코어를 가지고 있다. 따라서, 객관적인 성능 평가를 위하여 각 CPU의

모든 코어를 사용하여 가능한 적은 노드를 사용하는 경우(Full Load, FL)와 모든 시스템이 동일한 코어와 노드를 사용하는 경우(Optimized Load, OL)로 나누어서 실험을 수행하였다. 특히, 각 시스템의 특성 상, 일반 클러스터 시스템인 CLS는 단독으로 사용이 가능하지만 슈퍼컴퓨터의 경우는 다수의 사용자와 함께 사용하는 경우가 많아서 객관성을 부여하기 위하여 KNL과 SNL 시스템의 사용률이 각각 90% 이상일 경우와 60% 이하일 경우로 나누어 중복 실험을 하여 그 평균값을 취하였다. KNL 시스템은 설치 환경이 안정적인 것으로 파악되며 실험의 수행 시의 환경에 따른 성능의 차이가 크게 나지 않지만, SKL은 실행 시의 환경에 따라서 그 성능의 차이가 있는 것으로 파악된다. 따라서, KNL의 실험은 각각 2번만을 수행하고, SKL의 실험은 경우에 따라 2~5번을 수행하여 그 평균값을 취하였다.

3.1 Result of Ideal(Benchmark) Case

이상 실험은 격자의 구성(8192×1024×30)을 고려하여 Table 2와 같은 경우로 나누어 실험하였다. 이 실험은 이상적인 형태의 모델 구성을 계산상에서 준비하기 때문에 자료의 입출력이 없어서 계산자원만의 성능을 비교하기에 적합하다. FL실험과 OL실험은 각각 64개, 128개, 256개, 512개를 사용한 실험을 수행하여 병렬 시스템의 성능을 분석하였다.

노드의 모든 코어를 사용(FL실험)한 이상 실험(Fig. 7)에서 보이는 바와 같이, KNL 시스템은 SKL 또는 CLS에 비하여 전반적으로 1.82배 정도 느린 성능을 보이고 있다. 이는 1장에서 설명한 Dell EMC Solutions 사의 백서에서 보여준 WRF의 성능 차이(약 3배)에 비해 상당히 빠른 성능을 보이고 있는 것이다. 또한, 모델의 격자 구성을 반영한 최적의 코어를 사용(OL실험)한 실험(Fig. 8)에서는 KNL 시스템이 약 1.78배 느린 것으로 나타나서 FL 실험에 비해 약 2.2% 정도 빠른 것으로 나타났다. 이 차이는 KNL 시스템에서 64 코어 실험의 경우에 FL의 경우에 2시간 30분 소요되는 것이 OL의 경우에는 2시간 4분으로 약 26분의 시간 절약을 의미한다.

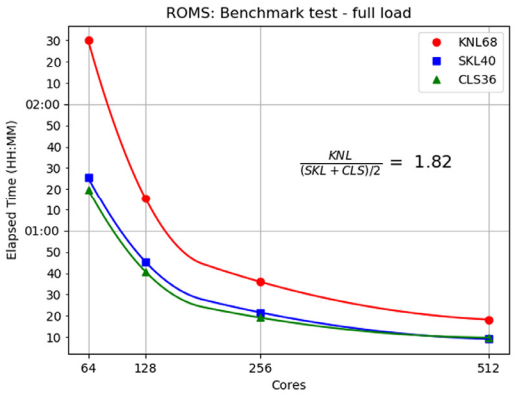


그림 7 Node의 모든 코어를 사용한(FL) 이상 실험
Fig. 7 Ideal case with full cores of node (FL)

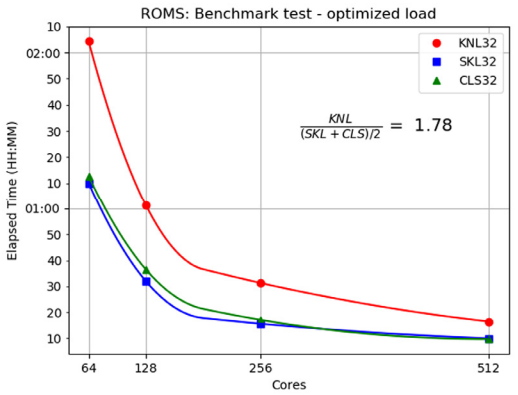


그림 8 Node의 최적 코어를 사용한(OL) 이상 실험
Fig. 8 Ideal case with optimization core of node (OL)

CPU 배분에 따른 성능의 차이를 분석하기 위하여 FL 실험(Fig. 7)과 OL실험(Fig. 8)을 합성하여 Fig. 9에 나타냈다. 점선은 FL실험을 나타내며, 실선은 OL실험을 나타낸다. CPU 배분에 따른 최적화 성능은 SKL이나 CLS에 비하여 KNL이 약 2%(OL/FL=0.98) 정도 더 효율적인 것으로 나타났으나 큰 차이는 없은 것으로 판

표 2 Benchmark 실험의 계산 자원 배분(FL: Full Load, OL: Optimal. Load, n: node, c: 코어)
Table 2 Computational resource allocation of ideal case (FL: Full Load, OL: Optimal. Load, n: node, c: core)

Grid (8192×1024×30)	KNL	SKL	CLS
64 Core (32×2)	- FL: 1n×68(64)c - OL: 2n×32c	- FL: 1n×40c+24c - OL: 2n×32c	- FL: 1n×36c+28c - OL: 2n×32c
128 Core (32×4)	- FL: 1n×68c+60c - OL: 4n×32c	- FL: 3n×40c+8c - OL: 4n×32c	- FL: 3n×36c+20c - OL: 4n×32c
256 Core (32×8)	- FL: 3n×68c+52c - OL: 8n×32c	- FL: 6n×40c+16c - OL: 8n×32c	- FL: 6n×36c+40c - OL: 8n×32c
512 Core (64×8)	- FL: 7n×68c+36c - OL: 16n×32c	- FL: 12n×40c+32c - OL: 16n×32c	- FL: 14n×36c+8c - OL: 16n×32c

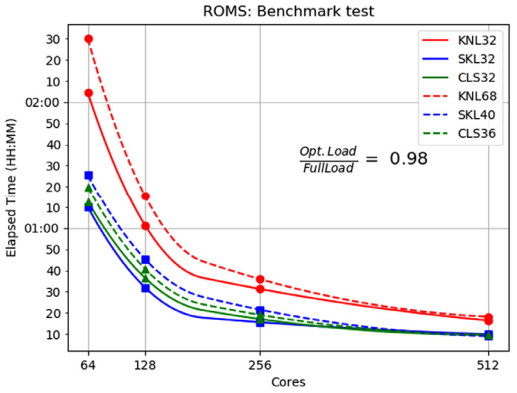


그림 9 FL, OL 이상 실험 비교
Fig. 9 Comparison of ideal case (FL, OL)

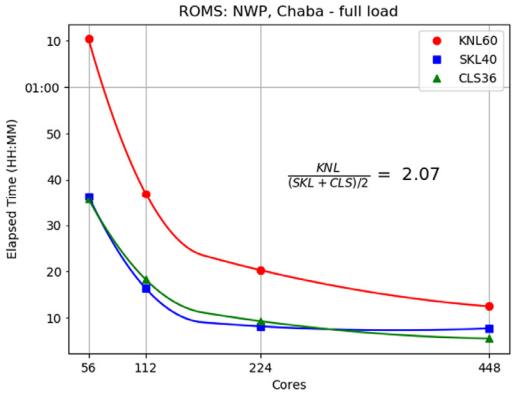


그림 10 Node의 모든 코어를 사용한(FL) 실제 실험
Fig. 10 Real case with full cores of node (FL)

단된다. 또한, 코어 수가 적은 경우에 KNL과 SKL 모두 CPU 배분에 따른 성능차이가 큰 것으로 나타났으나 코어 수가 늘어나면 네트워크 성능이 중요시되면서 그 차이는 무의미하게 줄어드는 것으로 나타났다. 이는 초병렬 모델의 수행에서는 노드의 모든 코어를 사용하여 사용 노드 수를 줄이는 것이 더 경제적인 것을 알 수 있다. 다만, 다음 절의 Fig. 12에서 알 수 있듯이, 실제 지형을 사용한 NWP 모델에서는 각 코어 간의 계산량이 다르고 입출력의 영향이 있어서 Benchmark 실험과는 결과가 다른 것을 유의해야 한다. 일반 클러스터인 CLS에서는 CPU 배분에 따른 성능차이가 크지 않은 것으로 나타나는데, 이는 SKL 시스템의 132 노드에 비하여 16 노드의 간단한 네트워크(InfiniBand)로 구성되었으며 모델을 독점적으로 수행함으로써 얻는 이득인 것으로 판단된다.

3.2 Result of Real(Typhoon Chaba) Case

북서태평양(NWP) 모델을 구축하여 2016년 제18호로 발생한 태풍 차바가 제주도를 지나가는 시기인 10월 5일 1시~4시 사이의 초기장과 해양-대기 외력을 사용하여 모의해 보았다. NWP실험은 격자의 구성(1400×1600×30)을 고려하여 Table 3과 같은 경우로 나누어 실험하였다. 이 실험은 실제지형과 실제상황을 모의하기 때문에 계산자원 뿐만 아니라 시스템의 IO 성능에도 영향을 미쳐서 시스템의 전반적인 성능 평가에 적합하다. FL실험과 OL실험은 각각 56개, 112개, 224개, 448개를 사용한 실험을 수행하여 병렬 시스템의 성능을 분석하였다. NWP의 고해상도 실험은 입출력 병렬화가 가능하지 않은 상황에서 KNL 시스템에서는 60코어를 넘기게 되면 메모리 부족으로 모의가 가능하지 못하게 된다. 따라서 Full Load 실험에서는 부득이하게 CPU 당 60코어만을 사용하였다.

Node의 모든 코어를 사용(FL실험)한 NWP 실험(Fig. 10)에서 보이는 바와 같이, KNL 시스템은 SKL 또는 CLS에 비하여 전반적으로 2.07배 정도 느린 성능을 보이고 있다. KNL 시스템이 전반적으로 느린 성능을 보이고 있다. 이는 Benchmark 실험보다는 다소 성능이 떨어지는 것으로써 자료의 입출력에 기인하는 것으로 보인다. 슈퍼컴퓨터에서 다수의 사용자가 동시 작업을 할 경우의 자료 입출력이 성능에 미치는 영향은 Fig. 10과 11의 SKL과 CLS의 결과를 보면 알 수 있다.

표 3 NWP 실험의 계산 자원 배분(FL: Full Load, OL: Optimal. Load, n: node, c: 코어)
Table 3 Computational resource allocation of real case (FL: Full Load, OL: Optimal. Load, n: node, c: core)

Grid (1400 × 1600 × 30)	KNL	SKL	CLS
56 Core (7 × 8)	- FL: 1n×60(56)c - OL: 2n×28c	- FL: 1n×40c+16c - OL: 2n×28c	- FL: 1n×36c+20c - OL: 2n×28c
112 Core (7 × 16)	- FL: 1n×60c+52c - OL: 4n×28c	- FL: 2n×40c+32c - OL: 4n×28c	- FL: 3n×36c+4c - OL: 4n×28c
224 Core (14 × 16)	- FL: 3n×60c+44c - OL: 8n×28c	- FL: 5n×40c+24c - OL: 8n×28c	- FL: 6n×36c+8c - OL: 8n×28c
448 Core (14 × 32)	- FL: 7n×60c+28c - OL: 16n×28c	- FL: 11n×40c+8c - OL: 16n×28c	- FL: 12n×36c+16c - OL: 16n×28c

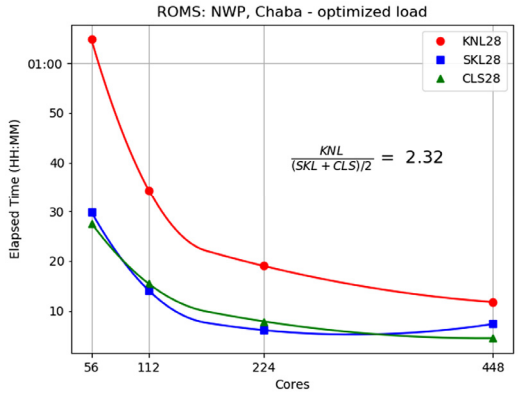


그림 11 Node의 최적 코어를 사용한(OL) 실제 실험
Fig. 11 Real case with optimization core of node (OL)

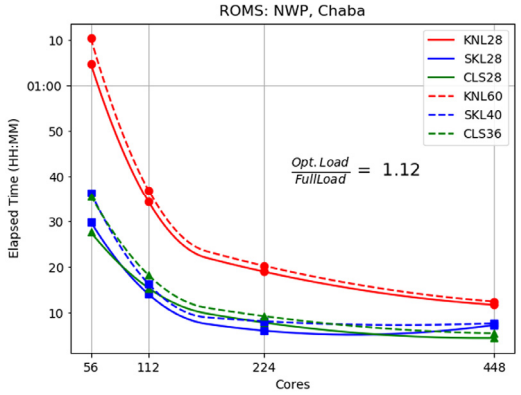


그림 12 FL, OL 실제 실험 비교
Fig. 12 Comparison of real case (FL, OL)

SKL 시스템의 결과는 Benchmark 실험의 결과에 비하여 코어 수가 증가함에 따라 도리어 성능이 저하되는 것을 알 수 있다. 실제로 SKL 실험은 모의 시의 상황에 따라 10% 이상 성능 차이가 나타나서 3~5번의 실험을 하여 평균한 값을 사용한 것이다. 이에 비하여 KNL 시스템은 모의 성능이 안정적인 것으로 나타나 2번의 실험을 하여 평균하였다.

모델의 격자 구성을 반영한 최적의 코어를 사용(OL실험)한 실험(Fig. 11)에서는 KNL 시스템이 약 2.32배 느린 것으로 나타나서 오히려 SKL이나 CLS 와의 격차가 더 벌어진 것으로 나타났다. 이것은 Fig. 12의 OL/FL = 1.12인 것에서도 알 수 있는데, 이는 KNL시스템이 68개의 코어를 사용할 수 있음에도 모델 구성상의 최적화를 염두에 두어 28개만을 사용함으로써 시스템의 효율을 떨어뜨린 것에 기인한 것으로 판단된다. 다만, 수치 자체는 크게 유의미한 차이를 보이는 것은 아니며 Benchmark 실험에 비하여 코어 수가 증가함에도 계속적으로 최적화의 성능이 모든 시스템에서 나은 것으로 나타났다. 특히, KNL 시스템의 경우는 최다 코어 실험인 448 코어 실험까지도 그 성능이 계속적으로 높아지고 있어서 초병렬 시스템에 더 적합한 시스템인 것으로 판단된다.

4. 결 론

누리온의 KNL 시스템은 일반 클러스터 시스템에 비하여 CPU의 처리 속도가 느리고, 노드당 사용할 수 있는 메모리가 적다는 단점을 가지고 있는 반면, GPGPU의 특징을 가지고 있으며 1 CPU당 68개의 매니코어 시스템으로 개발되었다. 따라서 매트릭스 기반의 모델이 아닌 일반 격자 모델의 경우에 KNL시스템이 다소 느린 성능을 보인다. 그러나 Dell EMC Solutions사의 백서에서 제시한 WRF의 성능비교 (KNL이 SKL에 비하여 약 3배 느림) 보다는 누리온에서의 ROMS의 KNL 성능은 SKL에 비하여 약 2배정도 차이에 그친다. 이것은 일반 클러스터시스템과의 비교에서도 동일한 것으로 나타나기 때문에 누리온의 KNL시스템 성능이 우수한 것으로 판단된다. 또한, 누리온의 KNL 시스템은 설치 환경이 안정적인 것으로 파악되며 실험의 수행 시에 따른 성능의 차이가 크게 나지 않지만, SKL은 실행 시의 환경에 따라서 그 성능의 차이가 있는 것으로 파악되어 여러 번의 실험을 거쳐야 하였다.

이상 실험에서는 KNL시스템이 약 1.8배 정도 느린 것으로 나타났으며, 코어를 모두 사용하는 것에 비해 최적화하는 경우가 조금 더 효율이 높은 것으로 나타났다. 그러나 코어 수가 증가할수록 네트워크 성능이 중요 시되면서 이 차이는 무의미하게 줄어드는 것으로 나타났다. 따라서, 초병렬 모델의 수행에서는 노드의 모든 코어를 사용하여 사용 노드 수를 줄이는 것이 더 경제적인 것으로 나타났으며, 동일 노드에서 KNL 시스템이 SKL 시스템에 비해서 2배 이상의 core를 장착하고 있다는 것을 감안하면 초병렬 모델에서의 KNL 시스템의 효율이 좋을 것으로 판단된다.

태풍 차바에 대한 북서태평양(NWP) 실험에서는 실제 지형과 실제 초기장 및 외력의 자료를 읽어야 하며 계산 결과를 출력해야 하기 때문에 메모리가 적은 KNL 시스템에서는 초고해상도로 개발한 모델을 수행할 수 없었다. 이에 대한 해결책은 다른 연구에서 논하는 것으로 미루고, 본 연구에서는 차선책으로 이 보다 한 단계 낮은 고해상도 모델을 개발하여 수행하였다. NWP 실험에서는, KNL은 SKL이나 CLS에 비하여 약 2.2배 정도 느린 것으로 나타났으며 Benchmark 실험(=1.8배)에 비해 성능이 떨어지는 것은 각 CPU 당의 계산 할당량이 다르게 주어지고, 입출력에 관한 시스템 부하가 영향을 미치는 것으로 판단된다. 또한, CPU 당 코어를 모두 사

용하는 것에 비해 최적화하는 경우가 오히려 효율이 떨어지는 것으로 나타났는데 이는 모델 구성상의 최적화를 염두에 두어 노드 당 68개 코어 중 28개만을 사용하여 시스템 특성을 잘 이용하지 못한 것에 기인하는 것으로 판단된다. 다만, 수치 자체는 크게 유의미한 차이를 보이는 것은 아니며 Benchmark 실험에 비해 코어 수를 크게 늘려도 최적화 효율이 유지된다는 것은 의미를 부여할 수 있겠다. 또한, 슈퍼컴퓨터에서는 다수의 사용자가 동시에 작업을 하는 상황이기 때문에 자료의 입출력은 모의 수행 상황에 따라서 변동이 생긴다. 이러한 영향은 SKL 시스템에서 더욱 크게 나타나는데, SKL 시스템에서는 코어 수가 크게 증가함에 따라 오히려 성능이 떨어지는 현상이 보이기도 하였다.

특히, KNL 시스템은 Benchmark 실험과 NWP 실험 모두에서 최다 코어 실험의 경우까지도 계속적으로 성능이 향상되는 결과를 보이고 있어서 초병렬 모델의 수행에 유리한 것으로 나타났다.

References

[1] A. Sodani, *Knights landing (KNL): 2nd Generation Intel Xeon Phiprocessor*, Hot Chips 27Symposium (HCS) IEEE2015, 2015.

[2] Dell EMC Solutions, *Application Performance of Intel Skylake and Intel Knights Landing Processors on Stampede2*, 2018.

[3] J-Y. Cho, H-W. Jin and D. Nam, "Using the On-Package Memory of Manycore Processor for Improving Performance of MPI Intra-Node Communication," *Journal of KIISE*, Vol. 44, No. 2, pp. 124-131, Feb. 2017.

[4] N. Butcher, S. L. Oliver, J. Berry, S. D. Hammond and P. M. Kogge, "Optimizing for KNL Usage Modes When Data Doesn't Fit in MCDRAM," *ICPP 2018 Proceedings of the 47th International Conference on Parallel Processing Article*, No. 37, Aug. 2018.

[5] S. Rho, S. Kim, D. Nam, D. Park and J-S. Kim, "Enhancing the Performance of Multiple Parallel Applications using Heterogeneous Memory on the Intel's Next-Generation Many-core Processor," *Journal of KIISE*, Vol. 44, No. 9, pp. 878-886, Sep. 2017.

[6] J. W. Yoon and U-S. Song, "System Characteristics and Performance Analysis in Multi and Many-core Architectures," *Journal of Digital Contents Society*, Vol. 22, No. 3, pp. 597-603, Mar. 2019.

[7] A. F. Shchepetkin and J. C. McWilliams, "The regional oceanic modeling system(ROMS): a split explicit, free surface, topography following coordinate oceanic model," *Ocean Modelling*, Vol. 9, Issue 4, pp. 347-404, Dec. 2005.

[8] J. A. Carton and B. S. Giese, "A reanalysis of ocean climate using simple ocean data assimilation (SODA)," *Monthly Weather Review*, Vol. 136, pp. 2999-3017, Aug. 2008.

[9] R. Gelaro et al., "The Modern-Era Retrospective Analysis for Research and Applications, Version-2 (MERRA-2)," *Journal of Climate*, Vol. 30, pp. 5419-5454. Jul. 2017.



김 동 훈

1994년 성균관대학교 토목환경공학과 학사 1996년 성균관대학교 해안·해양공학 석사 1999년 성균관대학교 해안·해양공학 박사 2000년 연세대학교 대기과학과 박사후 연구원. 2002년 서울대학교 대기과학과 박사후 연구원. 2005년 (일본)전력중앙연구소 박사후 연구원. 2012년 기상청/기상연구소 기상연구관. 2014년 연세대학교 대기과학과 연구교수. 2017년 차세대 도시·농림 융합스마트 기상서비스 개발 사업단 본부장. 현재 인하대학교 해양과학과 연구교수. 관심분야는 해양-대기 수치 순환, 인공지능 예측



임 채 욱

2011년 인하대학교 해양과학과 학사. 2013년 인하대학교 해양과학과 석사. 현재 인하대학교 해양과학과 박사과정. 관심분야는 해양 수치 모델링, 인공지능 예측



조 민 수

1987년 연세대학교 대기과학과 학사. 1991년 연세대학교 대기과학과 석사. 1996년 연세대학교 대기과학과 박사. 2018년 한국과학기술정보연구원 슈퍼컴퓨팅서비스센터장, 재난대응 HPC 연구센터장. 2018년 국가과학기술심의회 전문위원. 현재 한국과학기술정보연구원 책임연구원, 한국기상학회 부회장. 관심분야는 기상기후 수치모델링, 슈퍼컴퓨팅 응용, 인공지능 활용



안 준 언

1999년 경북대학교 컴퓨터공학과 석사 현재 한국과학기술정보연구원 선임연구원 관심분야는 슈퍼컴퓨팅, HPC 기반 클라우드 컴퓨팅



문 일 주

1991년 부산대학교 해양학과 학사. 1994년 서울대학교 해양학과 석사. 2000년 서울대학교 해양학과 박사. 2005년 미국 로드아일랜드 대학 박사후 연구원 및 연구교수 현재 제주대학교 해양산업경찰학과 교수, 태풍연구센터 센터장. 관심분야는 태풍-파랑-해양에 대한수치모델링, 상호작용, 기후변화 영향 연구



우 승 범

1993년 성균관대학교 토목환경공학과 학사. 1995년 성균관대학교 해안·해양공학 석사. 2001년 Cornell University 해안·해양공학 박사. 2003년 Cornell University 박사후 연구원. 2004년 한국건설기술연구원 선임연구원. 현재 인하대학교 해양과학과 정교수. 현재 경기씨그랜트사업단 단장. 관심분야는 연안 공학, 해양-대기 수치 순환 연구