CS 575 -- Spring Quarter 2017

Paper Project

Xiao Tan

Instructor: Prof. Mike Bailey

Paper review: *Petascale WRF Simulation of Hurricane Sandy*

--- Deployment of NCSA's Cray XE6 Blue Waters

Hurricane has given us giant amount of damage; an accurate hurricane forecast is needed. I read a paper *Petascale WRF simulation of Hurricane Sandy*, is just a paper to validate a prediction model WRF on giant amount of data, WRF means Weather Research and Forecasting model. This title said Petascale because they applied a large storm prediction using over 4 billion points real date to simulate the landfall of Hurricane Sandy on the Cray XE6 "Blue water", which is one of the most powerful supercomputers in the world, and is the fastest supercomputer on a university campus at NCSA at the University of Illinois. A grid of size 9120x9216x48 (1.4Tbytes of input) was used, with horizontal resolution of 500 meters and a 2-second time step. 86 Gbytes of forecast data was written every 6 forecast hours at a rate of up to 2 Gbytes/second and collaboratively post-processed and displayed using the Vapor suite at NCAR.

This paper is written by Peter Johnsen, Mark Straka, Melvyn Shapiro, Alan Norton, Thomas Galarneau. Peter Johnsen is a Meteorologist in Performance Engineering Group in Cray, Inc., Mark Straka is a Sr. Research Programmer of NCSA, also Melvyn Shapiro, Alan Norton, and Thomas Galarneau are belong to NCAR which represents National Center for Atmospheric Research. Cray, Inc is a supercomputer manufacturer, for this paper, supercomputer Blue Water is used Cray XE6 as its cores, which contains an unprecedented 13680 nodes, it is 437,760 cores. And this supercomputer is located in NCSA, which represents National Center for Supercomputing Applications in University of Illinois at Urbana-Champaign. NCAR's missions is supporting, enhancing, and extending the capabilities of the university and the broader scientific community. They co-operated to finish this testify, so it has to be very professional and convincible.

In this paper, they introduce the Blue Waters supercomputer at first. Blue Waters, built from the latest technologies from Cray, Inc., uses hundreds of thousands of computational cores to achieve peak performance of more than 13 quadrillion calculations per second. Blue Waters is composed of 237 Cray XE6 cabinets plus 32 cabinets of Cray XK7 with NVIDIA® Kepler™ GPU computing capability. The Cray XE6 processor is a 16-core 64-bit AMD Opteron 6276 series (Interlagos). It features 8x64 KB of L1 instruction cache, 16x16 KB of L1 data cache, 8x2 MB of L2 cache per processor core, and 2x8 MB shared L3 cache. Up to 192 processors can populate a cabinet. All in all, this computer is powerful enough to handle petascale of data. WRF forecast model is The Weather Research and Forecasting (WRF) Model, which is a mature, multi-component application suite for mesoscale numerical weather prediction. WRF used the Blue Waters system do parallel as much as possible, the Hurricane Sandy's dataset have a horizontal grid of 9120 by 9216 points and 48 vertical levels, using WRF pre-processing utilities WPS, and additional experiments using 150 vertical levels instead of 48 progressing.

In this model, users can tune their performance in 3 realms, which are sourced code, runtime, and system. In source code layer, the WRF version 3.3.1 code was modified from the public distribution chiefly with concerns for I/O burden per MPI task. Moreover, WRF is a hybrid MPI/OpenMP code in runtime layer, it as such decomposes the global grid and distributes memory via rectangular subdomains called patches to the MPI ranks. In system layer, it contains the wide selection of application-independent MPICH environment variables, and Cray-specific topology aware take placement tools, which will be benefits to WRF and other applications on Blue Waters to varying degrees. The MPICH tunable parameters is the greatest impact for this project, it can span the gamut from communication protocol and message sizes to rank reordering. After these overview introductions, they showed preliminary experiments. First is load imbalance test, such as rain precipitation, and graupel precipitation. They also saw some periodic increases of up to 50% wall clock time in regular, which was not due to the code or input model, so they try some jitter by some methods, first is the use of core specialization, which is essentially allocating non-user resources by either deliberately idling a core module or explicitly using the –r1 option to the aprun command while still utilizing all cores on a node; then they were assuring that we were running on dedicated partitions of the torus; in the end, they were considering balanced injection: attempts were made to tune the injection bandwidth of the compute nodes with the network for certain communication patterns. However, by added these function, it will increase complexity. For now, they still not have a method to solve the Lustre ping effect, and they still need to balance large times with total integration steps. Then, in order to increase performance, they found the most effective way to run WRF on AMD Bulldozer core modules was to use MPI/OpenMP hybrid mode with 2 OpenMP threads per MPI rank. This puts 16 MPI ranks on each XE6 node. As table showed is performance that between default method and optimized method.

Table 1. Halo exchange messaging statistics for a single WRF time step on 13,680 XE6 nodes (218,880 MPI ranks). *Total Bytes Exchanged* assumes each message contains two packed 3D single precision variables and a halo region of 5 slices

Placement Method	*Total Messages*	*Total Bytes Exchanged*	*On-node Messages*	*Off-node Messages*	*Off-node Bytes Exchanged*
Default Placement	3.6E07	1.5E12	1.8E07	1.8E07	1.1E12
Optimized MPI rank Ordering	3.6E07	1.5E12	2.4E07	1.2E07	2.8E11

In results, they got good performances in strong scaling simulation, overhead is still relatively low, even at extreme scale. As the following graph shows.

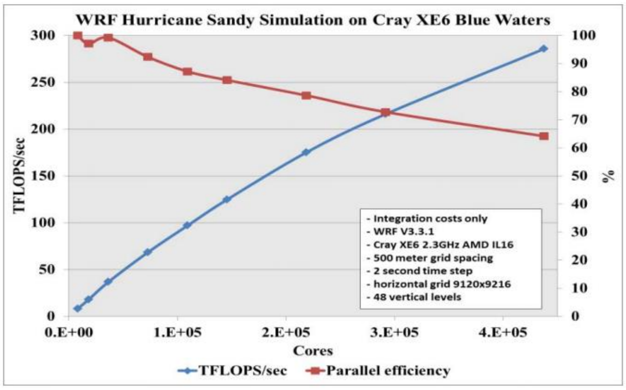
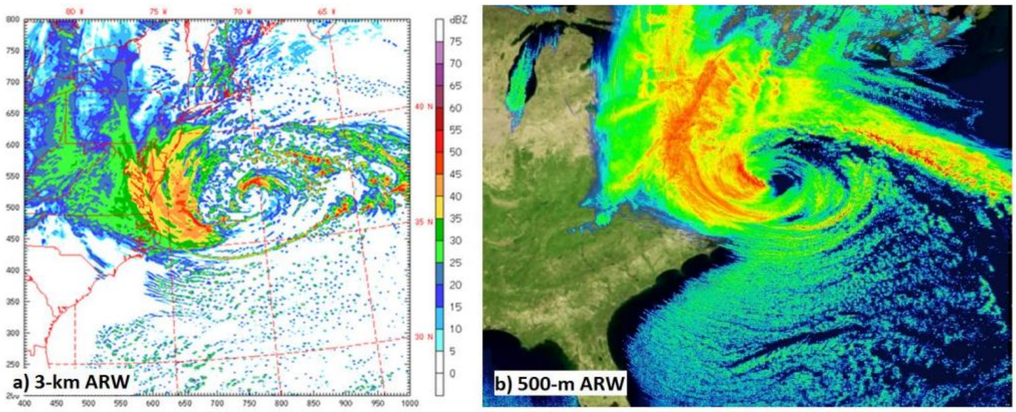


Figure 4. Strong scaling of Hurricane Sandy run. Sustained performance in Tflops/second (y-axis, left) and parallel efficiency over base run on 8,192 cores (y-axis, right) are shown.

In addition, for weak scaling, it still achieves efficiency.

Their system still compares with the real-time data, they gave the sample of the landfall of Hurricane Sandy along the New Jersey shoreline at 2330 UTC 30 October 2012, this Hurricane cased a catastrophe, they use the simulation system to generate radar graph at 3-km and 500-m horizontal resolution compare with the maximum radar reflectivity verifying at 1500 UTC 29 October 2012. In both simulations, a broad region of heavy precipitation is located on the west and southwest side of Sandy, and is organized in a region where warm moist northeasterly flow intersects a northwesterly surge of cold continental air, as the following picture showed.



In the end, the results show this model accuracy was extremely validated against actual atmospheric measurements. Moreover, there still have some new scientific discoveries by the simulations of hurricane Sandy, it can support the need for increased resolution in these models, along with architectures such as the Cray Blue Waters system, and the codes map exceptionally well.

In this paper, I knew it is so important to use parallelism, even in supercomputer where have such great number of cores, and with it high calculate speed and parallelism, it can achieve peta level data size calculations. Nowadays, with analyzing the data of hurricanes, we can get a mostly certain forecast of hurricanes, it will help millions of people to seek shelter before hurricane attack their home, it will save millions of lives, which is great gifts for people.

However, as the authors mentioned, the performances were optimized, but still have room for improvement, they may could come up with a better algorithm to improve their system.

If I were the researchers, I would like to try this system to analysis earthquake wave. Earthquake also a catastrophe for people, we would like to have a more precise forecast to survive.