Asynchronous I/O Using the Earth System Modeling Framework

Matthew D. Turner
DoD HPCMP PETTT / Engility
matt.turner@engility.com

Kevin Viner
United States Naval Research Laboratory
kevin.viner@nrlmry.navy.mil

Tim Whitcomb
United States Naval Research Laboratory
tim.whitcomb@nrlmry.navy.mil

Earth system component models are computer models that simulate a specific portion of the Earth system. Each component model solves a set of equations representing the physical processes within the domain encompassed by the model, such as the atmosphere, ocean, sea ice, or land surface. There is a desire in the weather forecasting community for Earth system component models to achieve accurate sub-seasonal to seasonal forecasts. In order to obtain an accurate representation of the environment at this length of forecast, multiple component models need to be coupled together, and the component models need to run at increased resolutions. However, model resolution cannot be increased significantly without experiencing prohibitive performance degradation, a significant portion of which is a result of the massive increase in the amount of data being generated and written to disk.

Operational weather forecasting is the use of observational data of the environment, data assimilation, and numerical weather prediction (NWP) models to produce weather forecasts. To deliver weather forecasts to end users on a regular basis, each operational weather forecasting center specifies a runtime requirement for the NWP models.

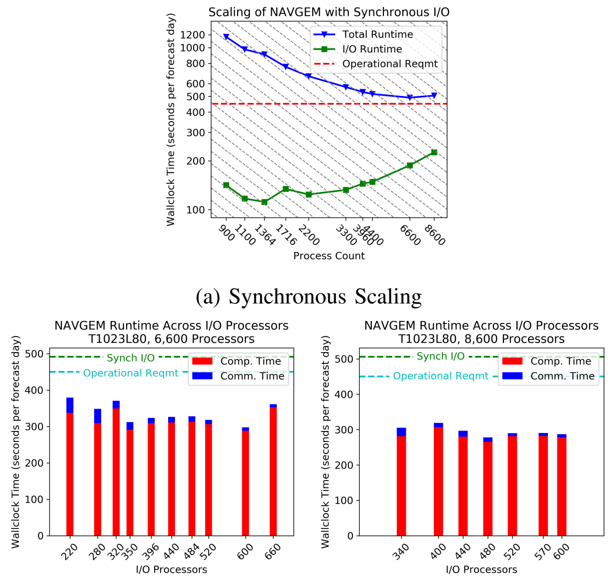
The Navy Global Environmental Model (NAVGEM) is the global NWP system designed and developed by the US Naval Research Laboratory, and run operationally by Fleet Numerical Meteorology and Oceanography Center. To obtain a more accurate forecast of the atmosphere, the NAVGEM model will be transitioning from a $31km$ horizontal resolution with 60 vertical levels to a $13km$ horizontal resolution with 100 vertical levels. The increase in model resolution results in simulations where the creation of restart files (via parallel HDF5) accounts for up to 45% of simulation runtime.

In an effort to obtain more accurate sub-seasonal forecasts, the US Navy is developing the US Navy Earth System Model (NESM), which couples NAVGEM to an ocean model, sea ice model, and a wave model. NESM utilizes the Earth System Modeling Framework (ESMF) as the coupling framework. ESFM is an open-source framework that includes a collection of software building blocks that can be added to Earth system component models, and allows for them to be combined into larger applications. In ESMF, major physical domains (e.g., atmosphere, ocean, land, sea ice, waves, etc.) are considered to be model “components”.

Traditionally, Earth system models perform the computing phase and the I/O phase sequentially. Due to the fact that the I/O phase outputs the current state of the simulation to disk, and the state (on disk) is not changed afterwards, it is possible for the computing phase to continue immediately without relying on the I/O phase to finish first. This overlap of computing and I/O is often referred to as asynchronous I/O.

Asynchronous I/O was implemented in the NAVGEM model by separating the I/O portion of the model from the compute section, and making it a separate ESMF component. In this sense, the portion of the model that writes restart files is considered a separate component model altogether. We use ESMF to handle the mapping between the domain decompositions of the computation component and the I/O component, and to perform the communication between the two.

This study explores the performance of NAVGEM with asynchronous I/O on a Cray XC50 system with Intel Broadwell cores on a Lustre file system. NAVGEM version 3 (beta) simulations were performed with a $13km$ horizontal resolution at the equator with 80 vertical levels on varying number of processors. For this resolution, each restart file has a size of 9.2 GB.



(a) Synchronous Scaling

(b) 6,600 Processors Asynch (c) 8,600 Processors Asynch

Fig. 1: Analysis of the performance of NAVGEM with (a) synchronous I/O, (b) asynchronous I/O on 6,600 total processors, and (c) asynchronous I/O on 8,600 total processors.

Prior to analyzing the performance of NAVGEM with asynchronous I/O, we performed a scaling analysis of NAVGEM with synchronous I/O. Fig. 1a shows that the model is incapable of meeting the operational runtime requirement with synchronous I/O for the process counts represented. Additionally, it is apparent that the amount of time spent in I/O increases as the process count increases, with I/O accounting for approximately 45% of the overall runtime for the 8,600 process count simulation.

Based on the synchronous I/O runtime analysis, it was chosen to analyze the performance enhancement of asynchronous I/O for simulations on 6,600 processors (Fig. 1b) and 8,600 processors (Fig. 1c). When running simulations with asynchronous I/O, a subset of the number of processors are withheld to perform I/O only. These processors are no longer available to the compute phase, resulting in the compute phase experiencing a slight performance degradation. However, the benefit of removing the I/O time from the simulation is shown to outweigh the slight loss of performance resulting from the reduction in the number of processors for compute.

We tested configurations with varying numbers of I/O processors to determine the optimal configuration for both the 6,600 processor case (Fig. 1b) and the 8,600 processor case (Fig. 1c). For the $6,600$ processor case (Fig. 1b), we observe an average runtime reduction of approximately 35% when compared to synchronous I/O. For the $8,600$ processor count case (Fig. 1c), we see an average runtime reduction when compared to synchronous I/O of approximately 43%. The most important conclusion from these results (from an operational mindset) is that every configuration tested is able to meet operational runtime requirements.