*DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.*

# Scalable Nonhydrostatic Unified Models of the Atmosphere with Moisture

Francis X. Giraldo
Department of Applied Mathematics
Naval Postgraduate School
833 Dyer Road
Monterey, CA 93943-5216
phone: (831) 656-2293 fax: (831) 656-2355 e-mail: fxgirald@nps.edu

Award Number: N0001408WR20136
http://faculty.nps.edu/fxgirald/projects/NUMA

## LONG-TERM GOALS

The long-term goal of this research is to construct a unified global and mesoscale nonhydrostatic numerical weather prediction (NWP) model for the U.S. Navy using new numerical methods specifically designed for modern computer architectures; this unified model is called the Non-hydrostatic Unified Model of the Atmosphere or NUMA. To take full advantage of distributed-memory computers, the global domains of NUMA are partitioned into local sub-domains, or elements, that can then be solved independently on multiple CPUs[1] and GPUs[2]. The numerical methods used on these sub-domains are local, high-order accurate, fully conservative, highly efficient, and geometrically flexible. Using these ideas we are developing global and mesoscale nonhydrostatic atmospheric models that will improve the operational models used by the U.S. Navy.

## OBJECTIVES

The objective of this project is to construct a new nonhydrostatic model, that uses high-order local methods, for the Navy's next-generation global and mesoscale nonhydrostatic NWP models. The high-order accuracy of these methods will improve the accuracy of the dynamics for the current global spherical harmonics model (NAVGEM) and the limited-area finite difference model (COAMPS). It is conjectured that improving the accuracy of the dynamics (including tracers) will increase the overall accuracy of the forecast; however, such improvements will also have to be made to both the physics and the data assimilation systems but these two topics are beyond the scope of this project. The objective of this project is to improve the accuracy of the dynamics, increasing the geometric flexibility in order to use any grid, and increase the scalability of this model on large core-count CPU and GPU computers. The scalability aspect of the new model is important because this means that it will be able to fully exploit next-generation large-scale computing architectures, which is important if we want to be well-positioned to exploit exascale computing in the near future. Computers will

[1] We define CPU as a central processing unit, or a typical commodity-type X86 processor.
[2] We define GPU as a graphical processing unit but, here, we mean any type of accelerator including Nvidia, AMD, or Intel (Xeon Phi) accelerators.

continue to offer massive amounts of compute-cores and if a model is not scalable today then it will have to be redesigned in a few years time when computers with billions of compute-cores will be available. The methods that we propose to use for NUMA are state-of-the-art and are not being used by either current or newly emerging NWP models.

## APPROACH

To meet our objectives we explore: 1) Unified high-order continuous Galerkin (CG) and discontinuous Galerkin (DG) spatial discretization methods; 2) fully unstructured and adaptive grids; and 3) scalable CPU and GPU implementations. The power of CG and DG methods is that they are high-order accurate yet are completely local in nature – meaning that the equations are solved independently within each individual element and, on parallel computers, may reside on separate compute-cores; here we use the term "compute-cores" loosely in order to include CPU and GPU based processors. The theoretical development of CG and DG methods is now well established and these methods are arguably the two most successful methods found in the literature for fluid flow problems, especially when considering large core-count CPU and GPU computers.

One final area that needs to be explored is the use of adaptive mesh refinement (AMR). We have now developed the necessary machinery (data structures) to allow NUMA to use nonconforming AMR both within the CG and DG formulations but have currently only achieved this in the two-dimensional version of NUMA. One advantage of AMR is that it may allow simulations not previously feasible without it. Another advantage of AMR is that, in principle, one can achieve the same level of accuracy as a uniform resolution simulation at a fraction of the computational cost. We explore such questions in this project.

## WORK COMPLETED

In this section, we describe the work completed this fiscal year. The NUMA team had a very good year both in terms of results achieved in the HIWPP studies and in model development. HIWPP is the High-Impact Weather Prediction Program funded by NOAA to scrutinize current weather prediction models for possible use in 2018 as the next operational National Weather Service model. Information on this program can be found at the website found in the footnote below[3]. Next, we shall describe the role of the NUMA team in the HIWPP and NGGPS[4] programs, in addition to the development work conducted to improve the capabilities of the NUMA model.

### 1. NUMA Team Contribution to HIWPP

*Introduction of a new Viscosity Operator*. The first phase of the HIWPP studies involved testing idealized test cases including baroclinic instabilities and flow over orography. This part of the HIWPP

[3] http://hiwpp.noaa.gov/

[4] NGGPS is the Next-Generation Global Prediction System and is the current name given to the next National Weather Service global NPW model. Information on this program can be found at http://www.nws.noaa.gov/ost/nggps/.

tests involved using the pure dynamics part of NEPTUNE, i.e. the NUMA dynamical core. The results published by NOAA can be found in the document cited at the footnote below[5].

Figure 1 shows the results for NUMA/NEPTUNE compared to all the other models in this study from the report in Footnote 5. Figure 1a shows that all the models look similar in the northern hemisphere but, at high resolution, Figure 1b shows that in the southern hemisphere, the models behave rather differently.

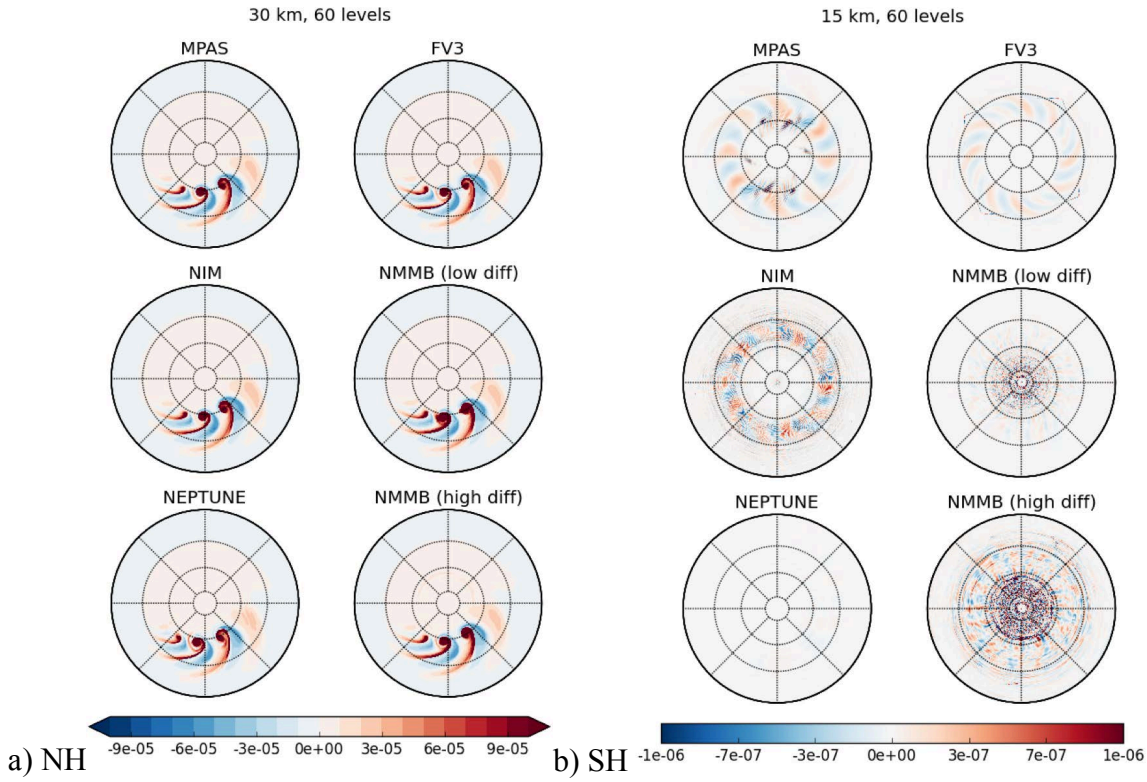



***Figure 1: The a) Northern Hemisphere and b) Southern Hemisphere results for all the models in the HIWPP study for the baroclinic instability problem.***

Note that all the models, except for NUMA/NEPTUNE, exhibit numerical (grid) noise. Most of the models exhibit unacceptably large values of grid noise (for some models, the artificial signal in the southern hemisphere is almost as large, within 10%, of the true signal in the northern hemisphere). Although the simulations shown in Figure 1 were run by the NRL-Monterey team, the NUMA team provided the dynamical core (this test only uses the dynamical core) and the NUMA team had to develop a new viscosity operator for these tests to run successfully. The diffusion operator in NEPTUNE used the full 3D viscous operator and, since NUMA/NEPTUNE is the only fully 3D spectral element code in the world, the viscous operator was acting across all three dimensions and thereby dampening the perturbation that eventually leads to the breaking wave. Without diffusion, the model did not give comparable results to the other models (since all the models used diffusion) and with diffusion it was annihilating the perturbation. The NUMA team had to re-derive the viscous

[5] http://hiwpp.noaa.gov/docs/HIWPP_Report-Year1-final-withAppendices-v2.pdf.

operators in order to apply diffusion to the horizontal operators only while leaving the vertical operators untouched. This required new mathematical techniques because NUMA uses unstructured grids whereby the grid is not aligned with the spherical geometry (NUMA uses Cartesian coordinates). The solution to this problem by the NPS team allowed the NRL team to perform the idealized cases for the HIWPP experiments. Unfortunately, the contribution of the NUMA team was not acknowledged in the HIWPP report.

*Discussion of NOAA Scaling Results for NUMA/NEPTUNE*. Figure 2 shows the scalability study performed by NOAA. The NUMA/NEPTUNE results are shown by the dark blue line; let us review these results carefully. All the NUMA/NEPTUNE results are obtained with double precision arithmetic while the MPAS results (yellow line) are at single precision only.

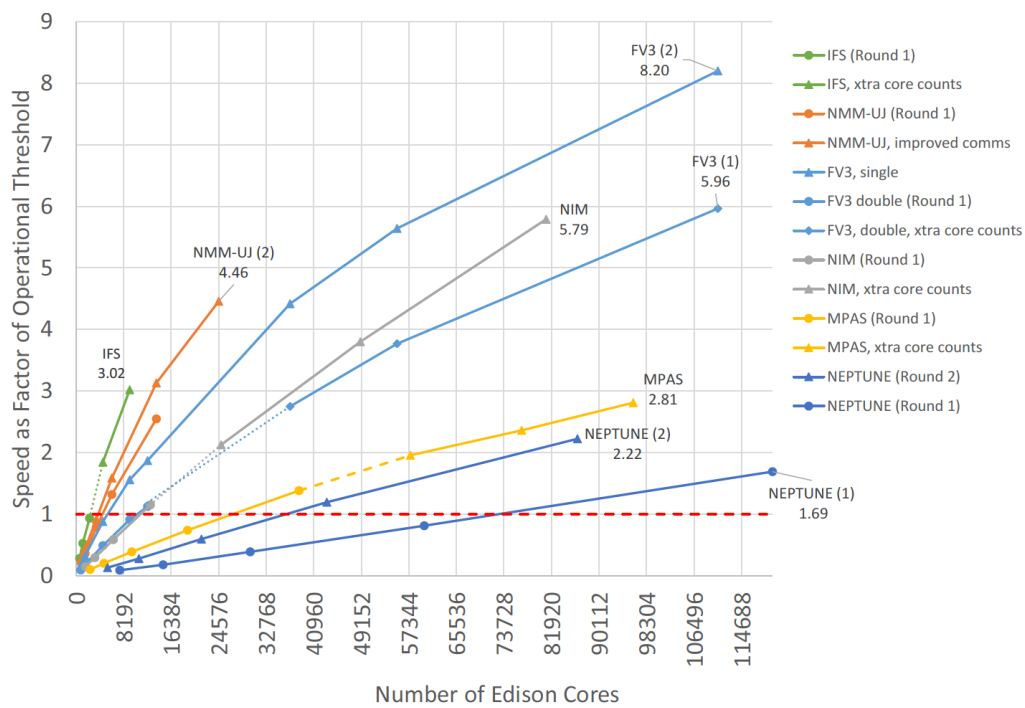



**Figure 3:** Speed as a function of number of cores, normalized to benchmark operational forecast integration rate of 8.5 min/day, shown as dotted red horizontal line at y=1.0. (13 km workload; higher is better metric.)

***Figure 2: The scalability study performed by NOAA on all the models in the HIWPP tests[6].***

Our in-house computation shows that the use of double versus single precision accounts for as much as a factor of 1.5 speedup – with these numbers, NUMA would be faster than MPAS at the same precision. The FV3 results (light blue lines) also support the 1.5 factor speedup when switching from double to single precision. Another point that needs review is that NUMA/NEPTUNE was only run

[6] This figure comes from the NOAA report found in: http://www.nws.noaa.gov/ost/nggps/DycoreTestingFiles/AVEC Level 1 Benchmarking Report 08 20150602.pdf

with less than 90K cores even though NUMA (at this resolution) is able to fully utilized 300K compute-cores. In our own scalability study (a paper is in preparation), we will show that NUMA is much faster than these NOAA results show. We will also show this in the Results section below. Finally, another point that was never discussed in these reports is that NUMA is the only model in the NOAA study that uses the deep-planet equations. This is important because it means that NUMA is the only model that carries a full three-dimensional grid while all the other models do not. It is further important to understand this point because it means that NUMA/NEPTUNE has the potential to be used not just for NWP but also for important research studies in high-altitude physics.

*Development of a Parallel Grid Generation Strategy*. The NUMA/NEPTUNE scalability simulations were performed by the NRL team and by NOAA, however, it was found during these scaling studies that NEPTUNE was not able to use more than 60K cores because the initialization (grid generation step) was done serially and so the root processor ran out of memory before it could farm out the grid to the rest of the processors. The NUMA team had to work for 2 weeks non-stop (long days and over weekends in late December 2014) to develop a new approach based on parallel grid generation. The NUMA team worked in close collaboration with the P4est[7] team (based at NPS and the University of Texas-Austin) to extend their parallel grid generation library to be able to build grids that would be usable for the HIWPP experiments. This version of the P4est library was built exclusively for NUMA and then this new version of NUMA was given to the NRL team in order to be able to perform these studies. Unfortunately, neither the NUMA team nor the P4est team was acknowledged in the NOAA report. Another missing data point in the NOAA results is that they only measure the time of the solvers without including the initialization step. The initialization step is quite substantial in any simulation and NUMA is able to do this fully in parallel that is a major advantage that, unfortunately, was not mentioned in the NOAA report.

## 2. NUMA Development Accomplishments

We now turn to the discussion on the basic research and development accomplishments concerning the NUMA model. The topics discussed below include: 1) the unification of the CG/DG models, 2) adaptive mesh refinement, 3) positivity-preservation mechanisms, 4) scalability study of the NUMA CG model, and 5) the porting of NUMA to the GPU.

*Unification of CG/DG.* One of the most important development milestones achieved in the NUMA project, is the unification of the continuous Galerkin (CG) and discontinuous Galerkin (DG) models. The idea of unifying these two methods was conceived in the course of this ONR project and is an important theoretical breakthrough because it means that these two methods can now be viewed to be quite similar and can be constructed in a unified way. The algorithmic breakthrough to unite these two methods within the same computer code was also conceived during the course of this ONR project. The first step in this direction occurred in the previous years where we developed 2D versions of the unified CG/DG NUMA (see Refs. [5], [7], and [8]). The barrier that needed to be overcome this year stemmed from the parallel versions of the 3D codes. In the past, we treated CG and DG differently and, therefore, required different parallel communication stencils that led to entirely different computer codes.

[7] Information on the P4est library can be found at: http://www.p4est.org/

Figure 3 shows the DG (panel a) and CG (panel b) storage formats. In DG storage (panel a) we use duplicate storage values at a given corner point while in CG storage (panel b) we use unique storage points. The breakthrough in being able to incorporate both the CG and DG methods within NUMA came from the fact that we could use DG storage for both CG and DG. However, CG storage is also included in NUMA by a clever use of data structures. This concept is explained in Ref. [11]. Although too technical to describe in this report, let us highlight the importance of this new approach. Although the unified CG/DG NUMA model is radically different from its CG-only predecessor, NUMA remains backward compatible in that the original CG-only NUMA used for the HIWPP experiments (and currently being used by NRL-Monterey in its NEPTUNE model) can be run exactly as before. The new NUMA model allows the use of various forms of the Euler equations. For example, the CG-only NUMA uses density-momentum-potential temperature in non-conservation form (which we call set 2NC) while the unified NUMA carries set 2NC as well as its conservation form (set 2C) as well as a conservation form that uses total energy (set 3C). To our knowledge, no other model is able to carry so many different forms of the governing equations in such a straightforward manner as is done in NUMA.

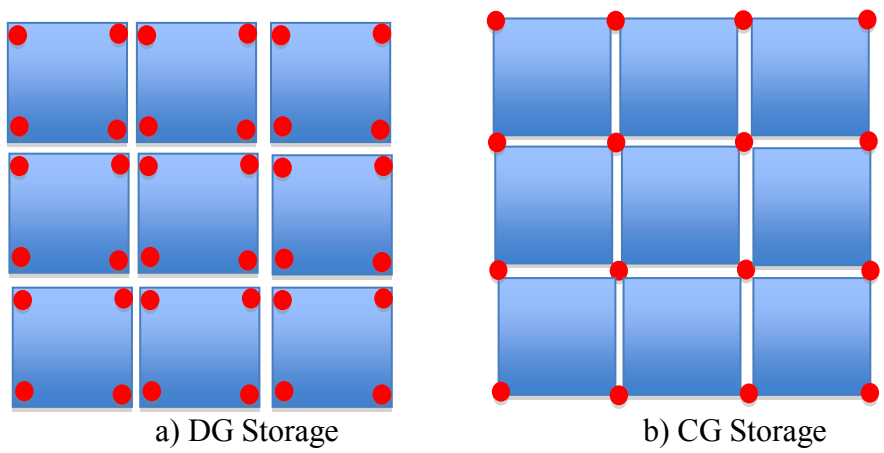



***Figure 3: The a) DG Storage and b) CG Storage formats.***

If the reasons for constructing a unified CG/DG version of NUMA remain unclear, let us briefly state that this unification is both an important theoretical breakthrough as well as an algorithmic one. For example, without such a code-unification, it would be quite difficult to include AMR in both methods. Because AMR for DG is more straightforward, we were able to do this task first and then use the working algorithm to extend non-conforming AMR to CG as well. Another advantage of this unification includes the addition of positivity-preserving mechanisms. It is well known that positivity-preservation is much easier for DG, yet DG methods tend to be more expensive than CG methods. By unifying both methods, we have been able to improve each method using the strengths of the other. An example is the recent development of a new quasi-positivity-preserving method for DG that we published in Ref. [2]. This idea has now been extended to CG so that we can get the same type of positivity-preservation even with CG. Note that the converse is also true; the stabilization method that we originally designed for CG in Ref. [3] has now been extended to DG.

*Adaptive Mesh Refinement*. We have stated in the course of this project that the best next-generation models will be those based on element-based Galerkin (EBG) methods such as the spectral element (SE/CG) and discontinuous Galerkin (DG) methods. However, we have only partially showed the

benefits of this approach such as: high parallel efficiency and high-order accuracy. Two years ago, we began a study of adaptive non-conforming quadrilateral grids (we published our results for non-conforming AMR for DG in Ref. [7] and AMR with CG and DG in Ref. [5]). One of the attractions of NUMA is that it uses either CG or DG methods. In other words, the CG and DG spatial discretizations have been written in a unified way which allows the model to use either method and both methods can now be run with AMR. In 2D, our AMR algorithm only takes 1% of the total runtime. Although non-conforming AMR for both CG and DG has been presented before, the contribution of our work is in showing how to extend this approach to nonhydrostatic atmospheric models. Moreover, to our knowledge, the results published in Refs. [5] and [7] are the only known results showing the conservation of mass up to machine precision for non-conforming AMR. In addition, in Ref. [5] we prove rigorously the conservation of mass for CG with non-conforming AMR – this is the first proof ever shown on this topic. Finally, we have been extending our 2D AMR machinery to the 3D code. While this may not seem like a huge undertaking, it is actually quite daunting because constructing efficient and parallel data structures to handle AMR is non-trivial. We now have preliminary results for 3D static AMR that we show in the Results section.

*Scalability of NUMA.* One of the biggest advantages of NUMA over any other nonhydrostatic NWP model is that it has been designed, from the beginning, to be highly scalable on ANY computer architecture. To prove this point, in the Results section we show scaling results for NUMA on both CPU and GPU based computers. In fact, the results shown in these sections were run on the two fastest (non classified) computer systems in the U.S.; i.e., we show results on Mira (the 5$^{th}$ fastest computer in the world) and Titan (the 2$^{nd}$ fastest computer in the world). The difference between NUMA and other models is that we use the same code to run on both systems. In fact, NUMA is capable of running on any type of hardware without too many modifications[8]. Out of all the models tested by NOAA in the HIWPP/NGGPS experiments NUMA and NIM are the only two models with a clear path towards using GPUs. In fact, NIM was designed from scratch to run on such hardware and, here, we show that NUMA has also been designed from the very beginning with this in mind even though the HIWPP reports never mention this feature about NUMA.

**RESULTS**

The main results that we wish to discuss for this year are: 1) adaptive mesh refinement for both CG and DG; 2) positivity-preservation for both CG and DG; 3) scalability on CPU and GPU based computers.

*Adaptive Mesh Refinement.* One of the distinguishing features of NUMA is that it is the only model from all the ESPC-considered models that has the capacity for adaptive mesh refinement (AMR). In [5] and [7] we examined the 2D AMR scheme with both the CG and DG method and showed that it is efficient, accurate, and conserves mass to machine precision. Currently, we are working on the 3D version of this algorithm and a preliminary result is shown in Fig. 4. This figure shows the rising thermal bubble in a 3D domain with a refined region in the middle of the domain. The bubble rises from the coarse region at the bottom, through the refined area in the middle to the coarse region at the top of the domain without causing reflections or mass conservation errors. This is ongoing work that will result in a fully dynamically adaptive mesh refinement scheme for both the CG and DG methods

[8] The capability of NUMA to run on GPU hardware was accomplished thanks through another ONR funded project whose sole purpose was to extend existing weather and climate models to GPUs.

in 3D NUMA. Although it is not clear if AMR can be used for all types of NWP simulations, it is certainly a good candidate for high-resolution hurricane simulations. Although the fully dynamic AMR (when the grid changes automatically) is not yet ready, we have all the pieces in place in order to make dynamic AMR a reality. Note that this is non-trivial as one needs to ensure that the number of elements are distributed equally among all the processors. To do this effectively, we will rely on the P4est library mentioned previously. One of the main challenges will be how best to couple NUMA with the P4est library so that the scalability of NUMA (shown below) will not be hampered by changing the grid on the fly. This topic will be left for a future project (for FY16).

*Positivity-Preservation*. In order to successfully couple NUMA with moist physics requires including tracers into NUMA. This capability already exists in NUMA and NUMA is able to use an infinite number of tracers (via an input parameter).

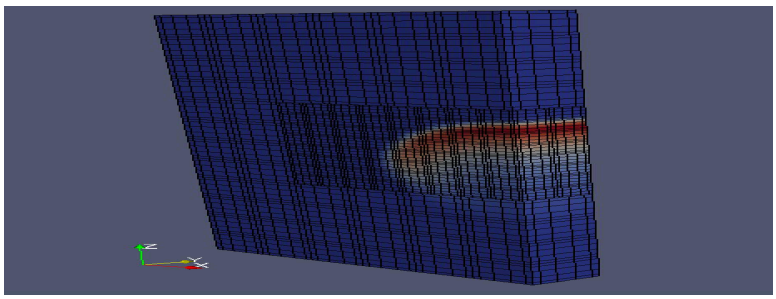
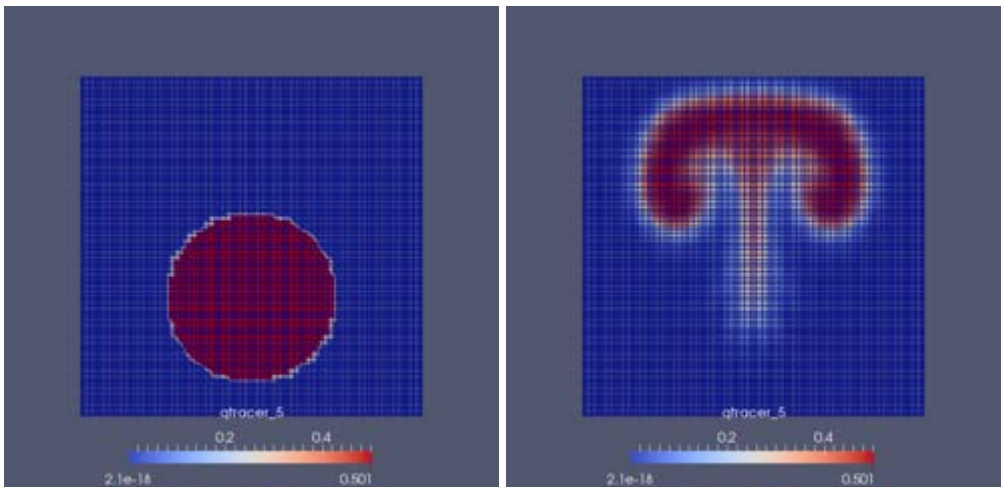


***Figure 4: A slice through the rising thermal bubble simulation in a 3D domain with non-conforming static mesh refinement.***

In order to successfully run an atmospheric model with moist processes (and other tracers) requires the addition of positivity-preserving methods since having negative tracers is unphysical. Of course, one could simply remove these negative values but this will have consequences on the conservation of the tracer fields. Since NUMA is fully conserving, we must construct positivity-preserving machinery that also is fully conserving. This year, we were able to devise two new approaches that are quasi-positivity-preserving; these two new methods are: 1) Local Adaptive Viscosity (LAV) and 2) Dynamic Sub-Grid scales (DynSGS). We derive all of our stabilization and quasi-positivity-preserving schemes as diffusion operators that use a non-constant and adaptive viscosity coefficient. This viscosity coefficient is based on the flow features of a specific field (e.g., for LAV) or on the residual error of the full equations (e.g., for DynSGS). The attraction of this idea is that these and all other stabilization methods rely on the exact same machinery in NUMA – the only difference is in how the viscosity coefficient is computed. Finally, we can make the tracer fields completely positive-definite by the inclusion of a very simple limiter (based on the work by Zhang and Shu).




a) Tracer at the initial time. b) Tracer after 700 seconds.

***Figure 5: The evolution of a tracer field within the rising thermal bubble experiment. A non-smooth tracer field is shown at a) initial time and b) after 700 seconds. The grid resolution is 10 meters for the DG method with the LAV + Limiter. Similar results are achieved with the CG solver.***

Figure 5 shows an example of our new stabilization methods with positivity-preservation. The details of these methods can be found in Refs. [2] and [3]. We have augmented these results with a true positivity-preserving limiter from the literature (Zhang and Shu, *Journal of Computational Physics*, v229 (2010), pp.8918-8934). The left panel (Fig. 5a) shows the initial tracer field (which is between 0 and 0.5) and the right panel (Fig. 5b) shows the tracer field after 700 seconds. Note that the initial tracer field is completely discontinuous – meaning that it consists of a circle (a cylinder in 3D) of fluid with no tracer values outside the circle. This is a notoriously difficult problem for any numerical method, especially a high-order one. Between our stabilization and the positivity-preserving limiter, we are able to achieve excellent results since there are no undershoots (negative values) while the tracer field remains well behaved. There is an overshoot but this is far better than one would achieve without stabilization or positivity-preservation. With these results in place, we can now handle moisture in the unified CG/DG NUMA without having to clip negative values as is typically done by much of the physical parameterizations.

*Optimization of NUMA for the Supercomputer Mira.* In the last four months of this project, we were able to make the main part of NUMA 13 times faster (Fig. 6) on Mira; Mira is an IBM Blue Gene Q computer with 780K compute-cores, each with 4 threads – it is currently the 5$^{th}$ fastest computer in the world. Vectorizing the code and improving memory access allowed us to achieve this 13X speedup. The fastest runtime was obtained by using machine specific vector instructions. This does not limit the potential for optimizations on other machines since similar vector instructions are available on all current computer architectures.

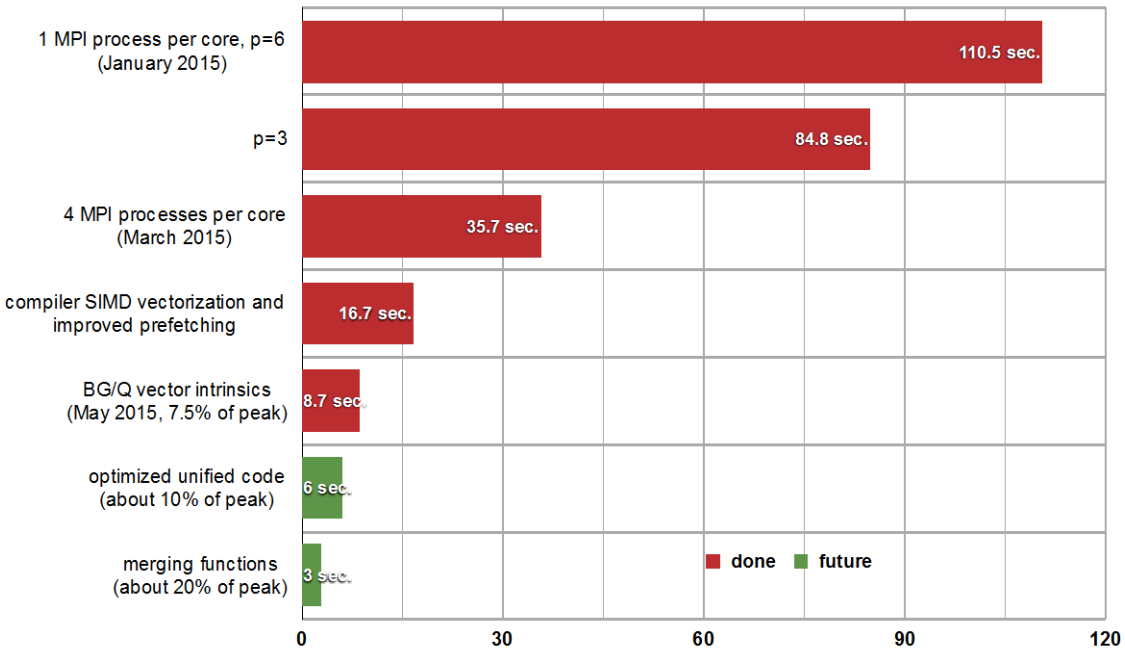



***Fig. 6: Optimization of the main part of the code. Shown is the runtime of the main part of the code for a baroclinic instability test case using 30720 CPUs for the different stages of our optimization.***

The main bottleneck of our code is now that the data is not always read in the order in which it is stored in memory. Since the processor always has to read and write entire blocks of data (cache lines) this slows our simulations down. The recently finished unified CG/DG version of NUMA allows us to store the data in the order in which we need it. This will reduce the memory bandwidth needed and should give us another significant speedup. The unified code will also allow us to merge different functions of our code. This will further reduce the memory bandwidth that we use since we can do more operations without loading new data from memory. NUMA is already 13 times faster than the version used for HIWPP and we expect to make NUMA 40 times faster than the HIWPP version in the near future.

Figure 6 shows the history of the optimizations performed. The 2$^{nd}$ bar from the top (84.8 secs) represents the version of NUMA used for the HIWPP experiments reported by NOAA in Fig. 2 that were run in December 2014. By March 2015, we were able to speedup the code by more than a factor of 2 (third bar from the top with 35.7 secs). Between SIMD compiler vectorization and using BG/Q vector intrinsics we have been able to increase the speedup by another factor of 4. This means that, as of today, NUMA is almost a factor of 10 faster than the version used in the HIWPP scalability studies. This is an important point because it means that NUMA today would be faster than all the models in the HIWPP studies by more than a factor of 2. The details of this performance improvement will be described in an upcoming publication (see Ref. [13]).

*Scalability of NUMA on Mira.* NUMA achieves an unprecedented strong scaling efficiency of 99% on the entire 3.14 million hardware threads of Mira (Fig. 7). This means that these 3 million threads can run the simulation almost 3 million times faster than a single processor with a single thread. Together

with our code optimizations this excellent scalability allows us to run a one-day forecast of a baroclinic instability test case at 3.0 km global horizontal resolution and 31 vertical layers within 4.16 minutes runtime. This means that NUMA can run at 3.0 km within the timeframe required for operational weather prediction as dictated by NOAA operational requirements.

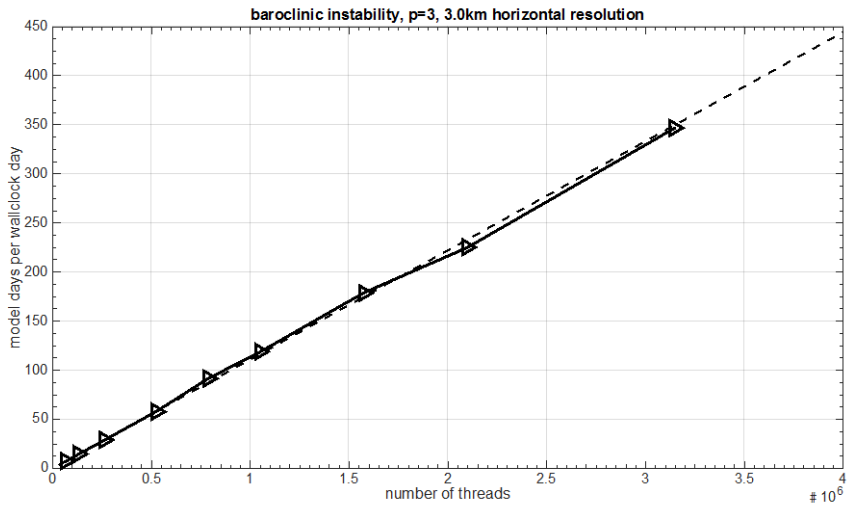



***Fig. 7: Strong scalability of NUMA for a baroclinic instability test case at 3.0 km global horizontal resolution. We show the number of model days that can be simulated within one wallclock day as a function of the number of threads used on the supercomputer Mira. The last data point uses the entire 3.14 million hardware threads of Mira.***

<u>*Porting NUMA to the GPU*</u>. NUMA has been ported to the GPU using algorithms suitable for the single instruction multiple thread (SIMT) architecture of GPUs[9]. The time to solution has been reduced by two orders of magnitude on a GTX Titan Black GPU card as shown in Table 1. On average the speed up is about 150X on single precision calculations and about 110X on double precision calculations. For portability to heterogeneous clusters, we used a new programming language called OCCA that can be cross-compiled at runtime to either OpenMP, OpenCL, CUDA, etc. The accuracy and performance of our GPU implementations have been verified using several benchmark problems representative of different scales of atmospheric dynamics. In Figure 8, we show a GPU accelerated simulation of an acoustic wave propagation test on the sphere using explicit time integration.

[9] Although the porting of NUMA to the GPU is officially part of a different ONR-NOPP project, we report the results here because we are leveraging the work from that project to improve the NUMA code and wish to report the current state of affairs with the NUMA model development.

***Table 1: Speed-up comparison between CPU and GPU for single precision and double precision calculations. The GPU card is GPU Titan Black and the CPU is Intel Xeon E5645 @2.4GHZ.***

| Test case | Double precision | | | Single precision | | |
|---|---|---|---|---|---|---|
| | CPU | GPU | Speedup | CPU | GPU | Speedup |
| 2D rising thermal bubble | 930.1 | 8.87 | 104.9 | 612.3 | 13.4 | 131.0 |
| 3D rising thermal bubble | 4408.9 | 41.47 | 106.3 | 3097 | 54.5 | 165.8 |
| Acoustic wave propagation | 3438.8 | 26.72 | 128.7 | 2379.9 | 44.4 | 152.9 |

*Scalability of multi-GPU NUMA on Titan supercomputer* . The second phase of the GPU project involved implementing and testing a multi-GPU version of our code. NUMA has already the necessary MPI-infrastructure required for communicating between CPUs. Recently, NUMA's scalability on a cluster of CPUs has been tested on the MIRA supercomputer with 3.1 million MPI ranks (see Figure 7). The multi-GPU implementation reuses this tested and proven method for indirect communication between GPUs by copying data to the CPUs. The scalability of the multi-GPU implementation has been tested on a GPU cluster, namely, the Titan supercomputer that has 18688 Nvidia Tesla K20 GPU accelerators. The scalability result in Figure 8 shows that the GPU version of NUMA is able to achieve 90% scaling efficiency on tens of thousands of GPUs.

## IMPACT/APPLICATIONS

This work targets the next-generation NWP systems for massively parallel computer architectures. NUMA has been designed specifically for these types of computer architectures while offering more flexibility, robustness, and accuracy than the current operational systems. Additionally, the new models are expected to conserve many important quantities such as mass and use state-of-the-art time-integration methods that will greatly improve the capabilities of the Navy's forecast systems.

## TRANSITIONS

Improved algorithms for model processes will be transitioned to 6.4 as they are ready, and will ultimately be transitioned to FNMOC.

## RELATED PROJECTS

Some of the technology developed for this project could be used to improve NAVGEM in other NRL projects. The work on the mesoscale models will help improve COAMPS. An example is the time-integration methods that we are exploring for the new models may well be incorporated into the current operational version of COAMPS. In a separate Department of Energy (DoE) project, the Mathematics and Computational Science group at Argonne National Laboratory has interfaced NUMA with their highly scalable software PETSc (Portable Extensive Toolkit for Scientific Computing). This union will vastly increase the capability of NUMA through the access to the large suite of time-integrators, preconditioners, and solvers contained in PETSc. Furthermore, having NUMA scrutinized by software and parallel computing scientists will allow us to improve it further.

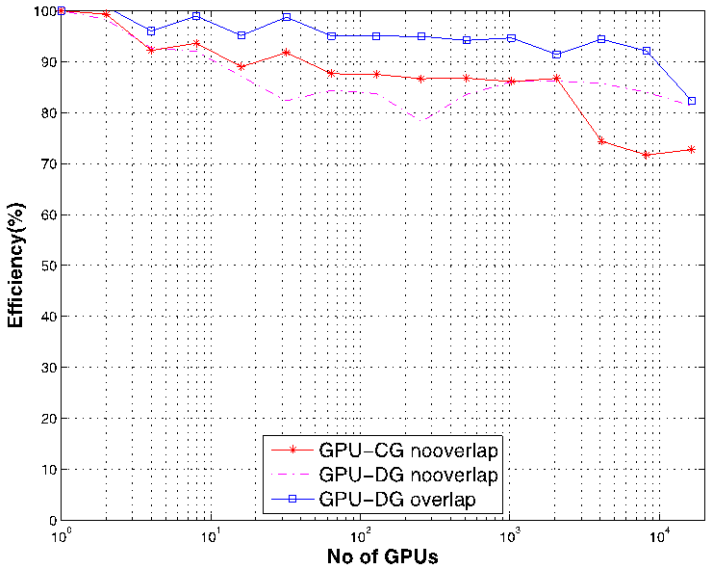



***Figure 8: The scalability of Unified CG/DG NUMA using CUDA and MPI on Titan.***

## PUBLICATIONS

### Journals

1. Co-author of the article, THE EARTH SYSTEM PREDICTION SUITE: Toward a Coordinated U.S. Modeling Capability, Bulletin of the American Meteorological Society (in press, 2015).
2. M. Yu, F.X. Giraldo, M. Peng, Z.J. Wang, Localized Artificial Viscosity Stabilization of Discontinuous Galerkin Methods for Nonhydrostatic Mesoscale Atmospheric Modeling, Monthly Weather Review (in press, 2015).
3. L.E., Carr, C.F. Borges, and F.X. Giraldo, A Polynomial-based Nonlinear Least-Squares Optimized Preconditioner for a Discontinuous Galerkin discretization of the Euler Equations, Journal of Scientific Computing (in press, 2015).
4. S. Marras, M. Nazarov, and F.X. Giraldo, A Conservative, Consistent, and Grid Adaptive Stabilized Spectral Element Method based on a Dynamic LES Model for the Compressible Euler Equations: the Case of Low Mach Stratified Flows, Journal of Computational Physics, Vol. 301, 77-101 (2015).
5. M.A. Kopera and F.X. Giraldo, Mass Conservation of Unified Continuous and Discontinuous Element-based Galerkin Methods on Dynamically Adaptive Grids with Applications to Atmospheric Modeling, Journal of Computational Physics, Vol. 297, 90-103 (2015).
6. S. Marras and F.X. Giraldo, A Parameter-free Dynamic Alternative to Hyper-Viscosity for Coupled Transport Equations: Applications to the Simulation of 3D Squall Lines using Spectral Elements, Journal of Computational Physics, Vol. 283, 360–373 (2015).
7. M.A. Kopera and F.X. Giraldo, Analysis of Adaptive Mesh Refinement for IMEX Discontinuous Galerkin Solutions of the Compressible Euler Equations with Application to Atmospheric Simulations, Journal of Computational Physics, Vol. 275, 92-117 (2015).

8. S. Marras, M.A. Kopera and F.X. Giraldo, Simulations of Shallow Water Jets with a Unified Element-based Continuous/Discontinuous Galerkin Model with Grid Flexibility on the Sphere, Quarterly Journal of the Royal Meteorological Society, Vol. 141 Issue 690 1727–1739 (2015).
9. S.J. Choi, F.X. Giraldo, J. Kim, and S. Shin, Verification of a Non-hydrostatic dynamical core using Horizontally Spectral Element and Vertically Finite Difference Methods: 2D aspects, Geoscientific Model Development, Vol. 7, 2717-2731 (November 2014).
10. E. Hendricks, M.A. Kopera and F.X. Giraldo, Evaluation of the utility of static and adaptive mesh refinement for idealized tropical cyclone problems in a spectral element shallow water model, Monthly Weather Review (in review, 2015).
11. D.S. Abdi and F.X. Giraldo, Efficient Construction of Unified Continuous and Discontinuous Galerkin Formulations for the 3D Euler Equations, (in preparation).
12. D.S. Abdi, L.C. Wilcox, T. Warburton, F.X. Giraldo, GPU NUMA on Titan, (in preparation).
13. A. Mueller, F.X. Giraldo, L.C. Wilcox, T. Isaac, CPU NUMA on Mira, (in preparation).

**Plenary Talks**

1. Giraldo, F.X.: A Unified CG/DG Model for Non-hydrostatic Atmospheric Modeling, International Center for Mathematical Sciences, Edinburgh Scotland (March 2015).

**Invited Talks**

1. Giraldo, F.X.: An Element-based Galerkin Atmospheric Model with an eye towards Exascale computers, Department of Mathematics, University of California at Berkeley (Sept 9, 2015).
2. Mueller, A., M. Kopera, S. Marras, F. X. Giraldo: Parallel performance of the atmospheric model NUMA using element-based Galerkin methods, Institute for Numerical Simulation, University of Bonn, Germany (June 11, 2015)
3. Mueller, A., M. Kopera, S. Marras, D. S. Abdi, F. X. Giraldo: Efficiency of high-order continuous and discontinuous Galerkin methods, German Weather Service DWD, Offenbach, Germany (December 15, 2014)
4. Mueller, A.: Challenges and Opportunities of upcoming Supercomputers for NWP, Institute for Atmospheric Physics, University of Mainz, Germany (December 12, 2014)

**Conference Abstracts**

1. Mueller, A., D. S. Abdi, M. Kopera, L. Wilcox, F. X. Giraldo: Towards operational weather prediction at 3.0 km global resolution with the dynamical core NUMA, Workshop on the Solution of Partial Differential Equations on the Sphere, KIAPS, Seoul, South Korea (October 2015)
2. Mueller, A., M. Kopera, F. X. Giraldo: Strong scaling on more than 700,000 cores with the atmospheric model NUMA, Workshop Galerkin methods with applications in weather and climate forecasting, ICMS, Edinburgh, Scotland (March 2015)
3. Mueller, A., M. Kopera, S. Marras, F. X. Giraldo: Does high order and dynamic adaptive mesh refinement improve the efficiency of atmospheric simulations? Coupling physics and dynamics workshop, CICESE, Ensenada, Mexico (December 2014)