# RRTMGP: A High-Performance Broadband Radiation Code for the Next Decade

Eli Mlawer
Atmospheric and Environmental Research
131 Hartwell Ave.
Lexington, MA 02421
phone: (781) 761-2226 fax: (781) 761-2299 email: mlawer@aer.com

Award Number: N000141310858
https://www.earthsystemcog.org/projects/espc-rrtmgp

## LONG-TERM GOALS

We developed a high-performance broadband radiation code for the current generation of computational architectures. This code, called RRTMGP, is a completely restructured (written in Fortran 2003) and modern version of the accurate RRTMG radiation code (*Mlawer et al., 1997*; *Iacono et al.,* 2008) that has been implemented in many General Circulation Models (GCMs) including the Navy Global Environmental Model (NAVGEM), the NCAR Community Earth System Model (CESM), and NOAA's Global Forecast System (GFS). Our development will significantly lessen a key bottleneck in these highly complex and coupled models, namely the large fraction of computational time currently required for the calculation of radiative fluxes and heating rates. We continue to work with a number of modeling centers to support the proper coupling of RRTMGP in these models.

## OBJECTIVES

The radiation calculations needed for climate simulations require many independent and complicated calculations, and are therefore an inviting target for new computing architectures such as Many-Integrated-Cores (MICs) and Graphical Processing Units (GPUs). We developed RRTMGP ('P' stands for 'parallel'), a modern version of the radiation code (RRTMG) used by many climate models, directed at the current generation of vector- and cache-based computational architectures. This code retains the high accuracy of RRTMG, but was developed from scratch to make it more flexible and amenable to optimization across a wide range of platforms. The objective has been to develop a single well-maintained, well-documented, and efficient radiation code that can be used by the modeling community for a diverse range of applications across a wide range of computing facilities. RRTMGP has been designed to exhibit profound improvements in speed for GPU and vector CPU machines and lesser, but still valuable, speed-ups on other CPU-based platforms relative to the current version of the code.

## APPROACH

*[technical approach, key individuals at your own or other organizations]*

The collaborating team consisted of scientists and programmers with detailed knowledge of RRTMG and its use within GCMs, as well as representatives of modeling centers that use RRTMG and plan to upgrade to RRTMGP. Eli Mlawer of AER, the lead developer of RRTMG, was the PI of this project and his team at AER included programmers with experience coding for modern computer architectures. In particular, programming expert Andre Wehe was instrumental to the project during its design and coding phases, and Mike Iacono, the leading expert on the implementation of RRTMG in GCMs, participated during the later stages of the project. Dr. Robert Pincus was the key figure in the design of the code and performed much of the coding. Dr. Frank Evans assisted with testing, especially of cloud cases.Brian Eaton of NCAR, which has employed RRTMG in its GCM (CESM/CAM) for nearly a decade, led a group at this modeling center participating in the project. Project collaborators Ming Liu and Tim Whitcomb of NRL represented the interests of the Navy GCM (NAVGEM).

Due to the expected wide impact of this development effort on climate and weather modeling, we have solicited feedback on this development effort from representatives of a number of different modeling centers (in addition to the collaborators on this project). We have provided access to the RRTMGP code repository at AER to John Michalakes and Yu-Tai Hou (NOAA), Robin Hogan (ECMWF), Leonidas Linarkakis and other colleagues at the Max-Planck Institute, Will Sawyer and Marcus Wetzstein (Swiss Supercomputer Center), V. Balaji (GFDL), Matthew Norman (ORNL), and Hans Johansen (LBNL).

## WORK COMPLETED

*[tasks completed, technical accomplishments]*

We completed a full implementation of RRTMGP code for both longwave and shortwave radiation. Longwave radiation can be computed assuming no scattering, using up to 4 Gaussian quadrature angles. Shortwave radiation is computed using the two-stream approximation for layer properties and adding to treat transport. The code is capable of including the impacts of cloud and/or aerosols and has been tested on idealized cloud cases, as well as for clear skies. The ability to include scattering in the longwave is nearly complete. The initial version of the code used older spectroscopic data in the mapping of atmospheric temperatures, pressures, and gas concentrations to spectrally-resolved optical properties, but we have upgraded the longwave gas optical properties to be based on state-of-the-science spectroscopic information, and the shortwave will be also soon be updated pending an upcoming release of an improved database.

Below is a brief year-by-year description of the work completed:

Year 1:

- NCAR hosted a kick-off meeting for the project on February 12-13, 2014. Attendees were Mlawer, Pincus, Berthiaume, Eaton, John Dennis (NCAR), Jim Edwards (NCAR), Tim Whitcomb (NRL), Jed Brown (ANL), and Tom Henderson (NOAA). Mlawer presented the motivation for the project, key information about radiation calculations in GCMs, and details about RRTMG and its stored tables and interpolations algorithms. Pincus presented his refactoring of RRTMG into PSRad, a much better structured code, and the work done at AER to port RRTMG to a GPU (RRTMGPU). Last, NCAR host John Dennis spoke about issues related to the NCAR codes that may impact the direction of the RRTMGP development. The in-depth discussion that followed centered on a draft modular structure for the new code, key issues in the current code that inhibit vectorization (most notably the interpolation scheme in the gas optics code), the need for a modular code structure that facilitates unit testing, and the potential for a framework that that could run efficiently on MICs, GPUs, and regular CPU processors. One outcome of the meeting was that NCAR would investigate the performance of the current RRTMG code making use of the PORT offline driver (see "Results" section).

We spent a fair amount of time considering a new interpolation scheme for the code to eliminate the branching caused by the conditional scheme used in RRTMG. In RRTMG linear interpolation in log-pressure and temperature is performed to calculate the needed absorption coefficients, but in bands with more than one major absorbing species there is an additional interpolation in the 'binary species parameter' (referred to as $\eta$). The type of interpolation is conditioned on the value of this parameter, with 3-point interpolation used near the extremes (0 and 1) of this space and linear interpolation elsewhere. This conditional test is not optimal with respect to code pararellization. We investigated fitting the absorption coefficient data with Chebyshev polynomials to provide sufficiently accurate results and be straightforwardly parallelized. Our evaluation demonstrated that the existing method resulted in an error greater than 4% near the endpoints of the range of the variable $\eta$, while the errors from the Chebyshev fits of all degrees are smaller than 3%, although the typical error throughout the domain is higher for the lower degree Chebyshev fits than for the interpolation scheme due to the oscillatory nature of the polynomial fit. This mixed result led to our decision to use purely linear $\eta$-interpolation in RRTMGP, using a strategic definition of $\eta$ to minimize the prevalence of cases with $\eta$ values near the endpoints in this space.

- The structure of two main routines in RRTMGP_LW, the driver and the gas optics code, were determined. A decision was made to utilize some object-oriented features to simplify the calling structure. An optical properties class will implement the optical properties needed for a simple absorption-only calculation (optical depths), and will be able to be extended to handle 2-stream scattering calculations (optical depths, single-scattering albedo, and asymmetry parameter), and multi-stream scattering calculations (optical depths, single-scattering albedo, and moments of phase function). The spectral configuration type was designed to include the number of g-points, number of bands and their spectral boundaries, and the mapping between g-points and bands. The gas optics module contains the gases considered and whether they are considered major, minor, or inactive in each band, and allows the user to specify the mass mixing ratio as a scalar, z-dependent field, or x-z dependent field. The LW solvers will include a solution without scattering and a 2-stream solution including scattering. As in RRTMG, RRTMGP was designed to contain modules or examples for the: random number generator (state/seed); cloud state; determination of the spectrally-resolved cloud optical properties given the cloud state, spectral configuration, and random number generator; aerosol state; determination of the spectrally-resolved aerosol optical properties given the aerosol state and spectral configuration; surface state; spectrally-resolved surface reflectivity/emissivity properties. The modular nature of the code will support users implementing alternate specifications of these quantities. The gas optics code was designed to incrementally include the contribution from the minor species, foreign water vapor continuum, water vapor self continuum, and major species. As is done in PSRad, a decision was made to store the absorption coefficient data in netcdf files to be read in during initialization. For the gas optics code, a method was devised to identically structure the code for all bands through the use of a dummy second major species for bands with only one major species. For RRTMG, the different structure of the gas optics code for bands with one vs. two major species was a big impediment to parallelization. These advances allowed coding of the individual modules to begin in the second year of the project, although draft versions of some modules were coded in the first year.

- An initial test framework that can handle RRTMG, PSRad, and RRTMGP was implemented.

- A preliminary code profiling investigation of RRTMG was carried out at NCAR making use of the performance measurement and analysis tools Extrae and Paraver from the Barcelona

Supercomputer Center. Hardware counters were used to measure several performance metrics, including the number of double-precision (DP) floating-point operations (FLOPs), and the number of vector double-precision floating-point operations (VEC-DP). An early finding is that on average the RRTMG code is executing about 0.2 DP FLOPs per CPU cycle. Experience with production science code is that it is possible to achieve execution rates in the range of 0.5 to 1.0 DP FLOPs per cycle. Looking at the ratio of vectorized DP FLOPs to total DP FLOPs we saw that for most of the execution time the percentage of DP FLOPs that are vectorized is less than 50%. This reinforced our belief that here were significant opportunities to increase the execution rate of RRTMG by making more efficient use of the vector instructions.

Year 2:

- A working version of the new longwave code was completed. This version is capable of computing fluxes and heating rates and has the anticipated final organization and structure of RRTMGP. This version was provided to our NASA GSFC collaborators and NCAR colleagues for evaluation and profiling, as well as colleagues in the Swiss climate modeling center to support their initiative to adapt this code to run on a GPU for implementation into their GCM (based on the Max Planck Institutes's ICON model).

- A completely new driver was developed for this version, interfacing with the calling GCM with required (for each atmospheric column) inputs: layer average temperatures, layer average pressures, layer edge pressures, surface skin temperature, and gas concentrations (derived type) for all modeled gases. Optional inputs are the layer column amounts of dry air and layer edge temperatures. A fully functional gas optics module was developed; in addition to its code, this module contains an object class that specifies the spectral bands, their g-points, all stored coefficients, etc. This module has these same inputs, and its outputs are (per g-point and atmospheric column) gas optical depths, the internal (i.e. Planck) layer source irradiances, the internal source irradiances for increasing (relative to the layer ordering provide by the user) layer edges, the internal source irradiances for decreasing layer edges, and the source irradiances from the surface. The gas optical depth calculation includes contributions from major species, minor species (though not all have stored coefficients yet), and the water vapor foreign and self continua. The interpolation scheme needed to compute absorption coefficients from gas concentrations, temperature, and pressure was completely redesigned compared to RRTMG. We increased (compared to RRTMG) the high range of surface pressures to 1100 mb so that we encompass the highest sea level pressure recorded below 750 m (1083.8 mbar). Also we expanded the range of Earth-based temperatures to 160 – 355 K so our absorption coefficients are valid for more extreme cases. Making these changes in RRTMGP required changes to the systems and scripts that generate the stored k-distributions.

- A new longwave absorption-emission solver was developed, with exact treatment of the linear-in tau approximation. Since the solver operates on each g-point independently while the gas optics code operates most efficiently operates with g-point as its inner loop, at the gas optics routine the index order of its input is reversed before performing its calculations and then reverses the order of its output after its calculations are performed. All needed components for longwave scattering calculations were developed for the future use by the longwave scattering solver.

- All output from the gas optics module, including the components of the optical depth (e.g. major, continuum), was validated with respect to the corresponding output from the development code, which has the same coding structure as in RRTMG. This validation exercise was successfully performed for layering that goes from surface to TOA, as well as TOA to surface. A script was developed that calls the driver using input in netcdf format, allowing validation of the calculated flux and cooling rates.

- A new build system was developed for easier integration into existing projects. It supports various Fortran compilers, automatically determines and displays dependencies, and allows for a parallel build.

- Co-PIs Pincus and Wehe attended the ESPC AOLI meeting in Boulder, CO, in November 2014, which was hosted by Pincus. At this meeting, Pincus presented an overview of our project's goals and status. (PI Mlawer could not attend due to a scheduling conflict with another meeting.)

Year 3:

- We implemented a systematic testing suite using a standardized netCDF file for convenience. This includes three classes of tests: 1) an end-to-end test that compute fluxes given atmospheric and boundary conditions; 2) a series of "unit tests" in which significant computational task is executed in series; and 3) a set of validation tests against known values and/or independent implementations. The end-to-end and unit tests are automated, such that new code can be run against previous versions and any differences reported by executing a single script that makes use of a standardized Python environment. The verification tests include comparisons of two-stream results against a code being developed for ECMWF and comparisons of RRTMGP transport algorithms against the plane-parallel version of the high angular-resolution Spherical Harmonics/Discrete Ordinate Method (SHDOMPP) code developed by Frank Evans. We also developed validation tests that consist of comparisons between RRTMGP flux and heating rate calculations against those calculated by the benchmark radiation model LBLRTM. These tests consist of a set of scripts we developed to compare the calculations of the two models for numerous atmospheric columns representative of the present atmosphere. Also included are profiles corresponding to pre-industrial and future conditions and gas concentrations, allowing the code's radiative forcing performance to also be validated. This rigorous testing provided us with solid confidence in the correctness of our code.

- The code was systematically organized into three layers. One layer interacts with users and the host model, managing initialization, user choices, error handling, and especially the translation of the host model's physical description of the atmosphere to the optical description required by the radiative transfer calculation. A lower layer acts as a broker, managing flow control to implement user choices, for example by choosing which low-level routines to call. The foundation of RRTMGP is a set of low-level computational kernels. These are highly-optimized, fully-vectorized routines that do the heavy computational work with sanitized (know-good) inputs. The kernels all have C bindings. This means they may be called by other languages including C, C++, and Python, but also so they may be replaced by other implementations without affecting the interface to the host model. Our experience suggests that this division allows for both flexibility and efficiency, the project's two main goals.

- We modestly enhanced the flexibility of the RRTMGP computational ecosystem by modifying two Fortran 2003 "classes" (objects) representing gas concentrations and optical properties (optical thickness, single-scattering albedo, and phase function – the variables used in the radiative transfer equation). Much of RRTMGP's efficiency comes because it is vectorized across the user-settable "column" dimension, so that problems can be optimally sized for a given platform. The column dimensions of the gas concentration and optical properties must be consistent with this size; we added a way to extract subsets of these objects along the column dimension, similar to the Fortran ability to extract array sections, so that the objects may conveniently be used in loops over a larger problem.

## RESULTS

*[Describe meaningful technical results achieved in the report fiscal year. Make the significance clear. Emphasize what was learned, not what was done. This should be a summary of significant results and conclusions, and, especially, any "new capabilities" generated.]*

The completion of the current version of RRTMGP constituted the major deliverable of the project and, therefore, was a significant milestone for the project. The developed code strikes a balance between readability and comprehensibility (for its anticipated scientific user base) and advanced computational features and flexibility (critical for parallelization). RRTMG makes extensive use of Fortran 2003 classes and abstract interfaces, and code efficiency is gained from exposure of fine-grained parallelism and algorithmic simplification. The code and the stored data it utilizes are independent of each other; the data is provided at run time.

### *Development activities at NCAR*

The longwave code, which has been mature for longer, has been fully implemented in the NCAR CESM by Brian Eaton. Brian provided very valuable feedback to the development team that resulted in several major modifications to the code, including a move from using subroutines and explicit error handling to implementing tasks as functions that return error strings, and a move to having initialization accept arrays rather than read directly from files. These choices should make it easier to use RRTMGP in existing model infrastructures.

## IMPACT/APPLICATIONS

The development of RRTMGP should have a significant impact on the ability of GCMs, including NAVGEM and CESM, to perform efficient and accurate simulations of climate and weather.

## TRANSITIONS

*[An S&T product has sufficiently matured and some organization (acquisition, industry, customer) outside of ONR is doing something with it. "Product" includes equipment, prototypes, original ideas/theories, and equations. Include 'who' that 'organization' is, how they are using it, and when it is expected to be used. It is of special interest if it is already being used or has had acquisition funds committed. Examples are 'products' entering acquisition, being used by industry, or being used by other S&T organizations such as DARPA)]*

Access to RRTMGP_LW was provided to colleagues at a number of modeling centers, including NRL, NCAR, NOAA (including GFDL), CSCS, and DOE (ORNL and LBNL). Our colleagues continue to analyze the new code's structure for a number of different compilers and provide us with useful feedback. They will also be profiling the code's computational performance.

## RELATED PROJECTS

*[Identify closely related projects and briefly describe the nature of each relationship (include web links as appropriate/available)]*

Our colleagues at the Swiss Supercomputer Center (CSCS) in Lugano are continuing to work on developing a GPU version (OpenACC) of this code for use in the ICON LES model. In 2015 AER programmer Andre Wehe went to Lugano to work on a GPU version of RRTMGP with OpenACC.

This effort did not result in a functioning GPU version since the the relevant Fortran compilers didn't support a number of the features employed in RRTMGP and, at the time, the computational kernels in the code hadn't been as cleanly isolated as they now are. Our Swiss and German collaborators are working on an implementation using GridTools, a software framework that allows for optimal computational layouts transparently across a range of platforms.

## REFERENCES

Iacono, M.J., J.S. Delamere, E.J. Mlawer, M.W. Shephard, S.A. Clough, and W.D. Collins, Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models, *J. Geophys. Res.*, **113**, D13103, doi:10.1029/2008JD009944, 2008.

Mlawer, E.J., S.J. Taubman, P.D. Brown, M.J. Iacono, and S.A. Clough, Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.,* **102**, 16,663-16,682, 1997.

## PUBLICATIONS

Pincus, R., E. J. Mlawer, et al., Radiative flux and forcing parameterization error in aerosol-free clear skies, *Geophys. Res. Lett.*, http://onlinelibrary.wiley.com/doi/10.1002/2015GL064291/pdf