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 are developing a high-performance broadband radiation code for the current generation of computational architectures. This code, called RRTMGP, will be a completely restructured 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 proposed 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.

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 are developing 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 will retain the high accuracy of RRTMG, but is being developed from scratch to make it more flexible and amenable to optimization across a wide range of platforms. The objective is 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 will 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

The collaborating team is consists 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, is the PI of this project, and his team at AER includes programmers with experience coding for modern computer architectures, including the recent GPU-based implementation of RRTMG. Team member Robert Pincus of University of Colarado refactored RRTMG to create PSRad, which has served as a prototype for the current development. Drs. Mlawer and Pincus are leading the design and development of RRTMGP.

Brian Eaton of NCAR, which has employed RRTMG in its GCM (CESM/CAM) for nearly a decade, leads a group at this modeling center participating in the project. In the initial phase of the project, the NCAR focus has been on profiling the performance of RRTMG as used in their GCM. Project collaborators Ming Liu and Tim Whitcomb of NRL represent the interests of the Navy GCM (NAVGEM).

Due to the expected wide impact of this development effort on climate and weather modeling, representatives from a number of different modeling centers have contributed their perspectives on this development effort. This includes ECMWF, the Max-Planck Institute, and the Swiss Supercomputer Center. The project has also benefitted from the expertise of Dr. Jed Brown, an applied mathematician at Argonne National Lab, who has made (unfunded) contributions to the project with regards to interpolation schemes.

WORK COMPLETED

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). Finally, 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).

An overall modular structure was finalized for RRTMGP (see “Results” section).

Key aspects of the gas optics code were finalized, including the interpolation scheme and a straightforwardly parallizeable approach to handle bands with either 1 or 2 key species (see “Results” section). A draft version of the code was written.

A testing framework that can handle RRTMG, PSRad, and RRTMGP has been implemented.

RESULTS

Determination of new interpolation scheme

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 η). 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. Our investigation has determined that fitting the absorption coefficient data with Chebyshev polynomials will provide sufficiently accurate results and be straightforwardly parallelized.

Absorption coefficients were computed for an RRTMG band (700-820 cm^{-1}) for η values 0, 0.01, 0.02, ..., 0.99, 1.0 for all tropospheric pressures, temperature, and g-points (1040 sets of values), and the existing interpolation scheme was compared to Chebyshev polynomials of degree 3 through 9 that were fitted to these absorption coefficients. Figure KFIT shows the normalized percentage errors for a typical case (the set of absorption coefficients for which the overall error is at the 25th percentile for all the cases). This figure shows that, despite the 3-point interpolation employed near the upper boundary, the existing method still results in an error greater than 4% in this region. In contrast, 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 error analysis resulted in the decision to use a high-degree Chebyshev polynomial in the new code. This result is significant since the existing fitting scheme was not viable for modern computational environments, yet the accurate optical depths it produced were key to the accuracy of RRTMG. This new fitting approach will ensure both accuracy and parallelizability.

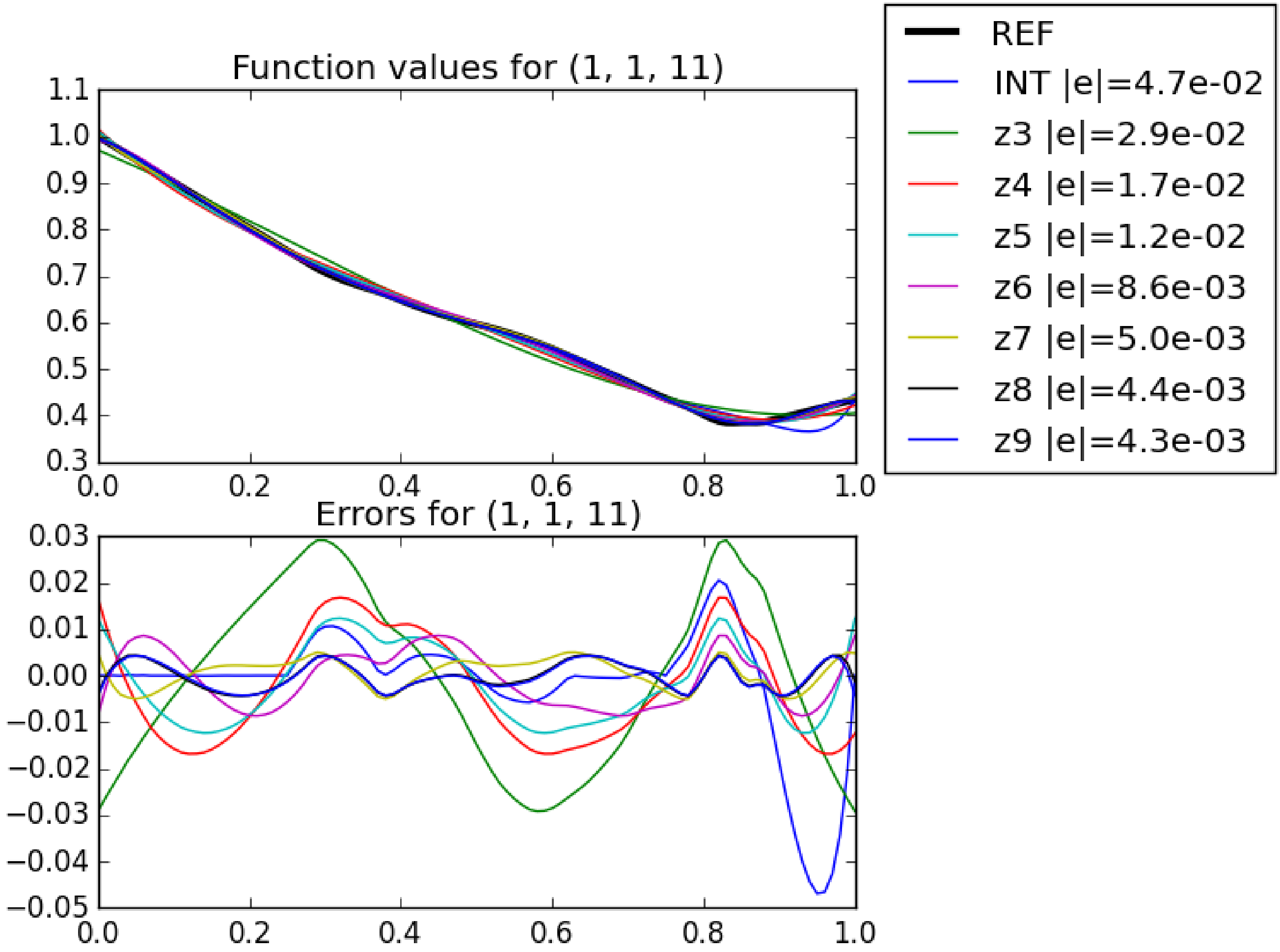

Figure KFIT. (top) Normalized reference absorption coefficients and values obtained using various fitting techniques as a function of □ for band 700-820 cm^{-1} and P,T,g = 1,1,11 (bottom) Normalized percentage errors for each fitting technique – "INT" is the current interpolation scheme and "zn" refers to the Chebyshev polynomial fit of degree n.

Outline/pseudocode for main and gas optics routines
The structure of two main routines in RRTMGP_LW, the driver (see Figure RRTMGP) and the gas optics code (not shown), has been determined. All code will be written in Fortran 2003. Some object-oriented features will be used to simplify the calling structure, although this will be limited by the desire to have the code be readable by scientists in the field, including graduate students.

There are a number of aspects of this outline that are of interest:

- The 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 will 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 will list 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 the current code, RRTMGP will 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 will incrementally include the contribution from the minor species, foreign water vapor continuum, self water vapor continuum, and major species. As is done in PSRad, the stored absorption coefficient data will be in netcdf files that will 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.

The advances will allow coding of the individual modules, planned for the second year of the project, to begin.

```
SUBROUTINE RRTMGP_LW(ncols, ndim, nlev, pressure, temperature, gas_concs, cld_state,
rng_state, aer_state, sfc_state, flux_up, flux_dn)
  USE mo_kinds,      ONLY: wp       ! Real KIND parameter

! Optical properties class
  USE mo_optics,   ONLY: opt_prps, opt_prps_2str, opt_prps_nstr, +,=
  ! Spectral configuration
  USE mo_spec_cfg, ONLY: spec_desc

  ! Model-specific user-supplied descriptions of clouds, aerosols, surface
  USE mo_cld_desc,   ONLY: cld_desc
  USE mo_aer_desc,   ONLY: aer_desc
  USE mo_sfc_desc,   ONLY: sfc_desc
  ! Model-specific  user-supplied optics given descriptions of clouds, aerosols,
surface and current spectral configuration
  USE mo_cld_optics, ONLY: cld_optics
  USE mo_aer_optics, ONLY: aer_optics
  USE mo_sfc_optics, ONLY: sfc_props

  USE mo_lw_gas_optics, ONLY: gas_optics, planck, spec_desc
  USE mo_lw_solvers,    ONLY: solve_lw ! Can wrap multiple solvers

  IMPLICIT NONE
  ! ----------------------------------------------------
  ! Input and output variables
  !
  INTEGER,                          INTENT(IN   ) :: ncols, ndim, nlev
  REAL(WP), DIMENSION(ndim, nlev),  INTENT(IN   ) :: pressure, temperature
  CLASS(gas_desc),                  INTENT(IN   ) :: gas_concs
  CLASS(cld_desc),                  INTENT(IN   ) :: cld_state
  CLASS(rng_desc), DIMENSION(ndim), INTENT(INOUT) :: rng_state
  CLASS(aer_desc),                  INTENT(IN   ) :: aer_state
  CLASS(sfc_desc),                  INTENT(IN   ) :: sfc_state
  REAL(WP), DIMENSION(ndim, nlev),  INTENT(  OUT) :: flux_up, flux_dn

  ! ----------------------------------------------------
  ! Local variables
  !
  INTEGER         :: ngpts
  CLASS(opt_prps) :: gases, aerosols, clouds, clear_sky, all_sky
  REAL(wp), DIMENSION(ndim,nlev+1)        :: temp_levs
  REAL(wp), DIMENSION(:,:  ), ALLOCATABLE :: sfc_emis, sfc_temp ! Dimension ndim,
ngpts
  REAL(wp), DIMENSION(:,:,:), ALLOCATABLE :: src_lay, src_lev   ! Dimension ndim,
nlev or nlev+1, ngpts
  REAL(wp), DIMENSION(:,:  ), ALLOCATABLE :: src_sfc            ! Dimension ndim,
ngpts

  ! ----------------------------------------------------
  ! 1 -- Initialization
  ! Determine the spectral configuration including number of g-points
  ngpts = spec_desc.ngpts() ! Is this the right syntax?
```

```
! 2 -- Optical properties of atmospheric components
  !
  CALL gas_optics(ncols, nlev, ngpts, pressure, temperature, gas_concs, spec_cfg,
gases)
  CALL aer_optics(ncols, nlev, ngpts, temperature, aer_state, spec_cfg, aerosols)
  CALL cld_optics(ncols, nlev, ngpts, temperature, cld_state, rng_state,  spec_cfg,
clouds)
all_sky = gases + aerosols + clouds

  CALL sfc_props(ncols, ngpts, sfc_state,  spec_cfg, sfc_emis, sfc_temp)
  ! ----------------------------------------------------
! 3 -- Solution
  ! Interpolate temperature to levels
  temp_levs = ...

  ! Planck source function at layers, levels, surface
  CALL planck(ncols, ndim, nlev,   ngpts, spec_cfg, temperature, src_lay)
  CALL planck(ncols, ndim, nlev+1, ngpts, spec_cgf, temp_levs,   src_lev)
  CALL planck(ncols, ndim,         ngpts, spec_sfc, sfc_temp,    src_lay)

  CALL solve_lw(ncols, ndim, nlev, ngpts, all_sky, src_lay, src_lev, sfc_emis,
sfc_temp, flux_up, flux_dn)

END SUBROUTINE RRTMGP_LW
```

Figure RRTMGP. Draft outline of main routine

Code profiling at NCAR.

A preliminary investigation has been carried out making use of the performance measurment 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 see (Figure PROF) that for most of the execution time the percentage of DP FLOPs that are vectorized is less than 50%. Hence we believe there are significant opportunities to increase the execution rate of RRTMG by making more efficient use of the vector instructions.

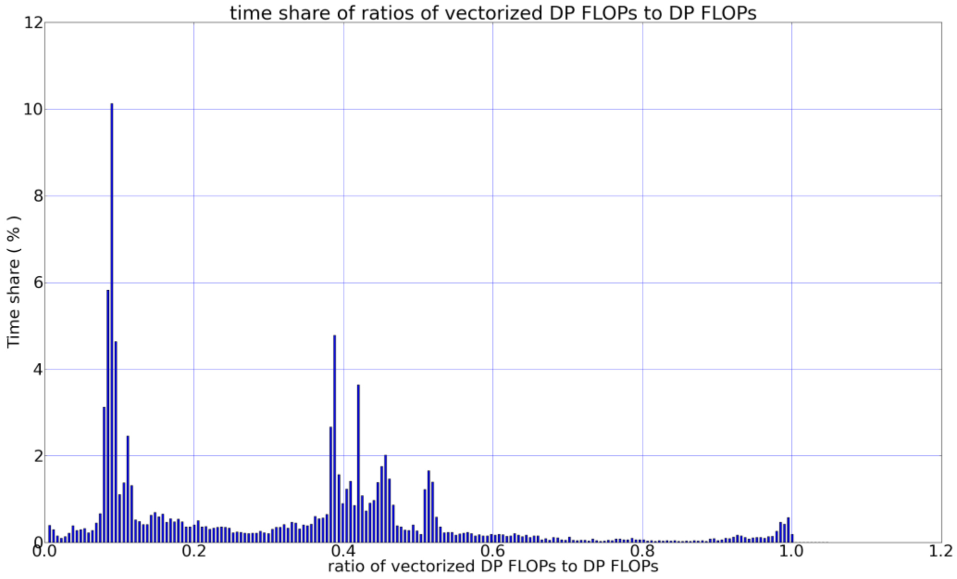

Figure PROF. Frequency of occurrence of the ratio of vectorized floating point operations to total floating point operations for RRTMG in the NCAR model.

Detailed performance analysis work is ongoing to identify the specific loops in the RRTMG code which will provide the most benefit from additional vectorization.

RELATED PROJECTS

We have continued our discussions with the Swiss Supercomputer Center (CSCS) in Lugano about their potential utilization of a GPU-compatible version of RRTMG in the next version of their prediction model. They would like to build a GPU version of the ICON LES model in a shorter time frame than this ONR-funded project. Previously we had provided them with our GPU codes RRTMGPU_LW and SW, but the fact that only the shortwave code was developed in OpenACC prevented CSCS from using these existing codes. Among the options discussed with CSCS were developing an OpenACC version of the longwave code based on RRTMGPU_LW, developing a OpenACC version of the longwave code based on PSRad, developing a version of the RRTMGP based on the spectroscopy in RRTMG, and extending the time line. These discussions are continuing.

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.