Analysis of Workload and Load Balancing Issues in the NCAR Community Climate Model

John G. Michalakes

Abstract

Choice of an appropriate strategy for balancing load in climate models running on parallel processors depends on the nature and size of inherent imbalances. Physics routines of the NCAR Community Climate Model were instrumented to produce per-cell load data for each time step, revealing load imbalance resulting from surface type, polar night, weather patterns, and the earth's terminator. Hourly, daily, and annual cycles in processor performance over the model grid were also uncovered. Data from CCM1 suggested a number of static processor allocation strategies.

1 Introduction

The efficient application of hundreds or thousands of processors to large climate modeling problems will depend on the degree to which work can be most evenly distributed among the processors. Maximum speedup is achieved when all processors are engaged in useful work at all times during the execution of the climate model – a situation which, if it were possible, would be one of perfect balance. However, the inherent need to perform greater or lesser amounts of work in different areas of the model grid depending on features of the climate data requires that care be taken to allocate work to processors in a balanced way. Further adjustment may be necessary over time to maintain this balance. Understanding of the spatial and temporal distribution of processor load that occurs within a climate model will help determine appropriate strategies for balancing processor loads and keeping them balanced when such models are implemented on parallel processors.

This study involved instrumenting an existing model, the NCAR Community Climate Model (CCM1), to determine how many CPU seconds were spent in calculations within each individual latitude-longitude cell of the model grid. This information, carefully extracted from the sequential model, reveals patterns of load that may be expected when the model grid is decomposed into some set of groupings of grid cells and assigned to separate processors. The resulting portrait of load distribution within CCM1 can also be used to simulate loads for testing load balancing algorithms within prototype parallel climate models.

The load imbalances of interest for this study were those which resulted from variation in values representing physical phenomena over the CCM1 model grid. What, for example, was the effect of surface type – land, ice, or ocean – on the amount of work performed at a mesh point? Might load vary as a function of time of day? Time of year? Study was therefore limited to those sections of the model code which dealt directly with these physical quantities. Other sections of code, such as those dealing with transport computations where the amount of work would have little to do with the actual values being transported, were ignored.

Only a two-dimensional view of load over the three-dimensional CCM1 model grid was produced. The vertical dimension was ignored so that the load for each cell in the latitude-longitude mesh actually represented the load from all cells above that point in the vertical direction. For practical purposes, this approach reduced the amount of load data by a factor of 12, and simplified greatly the task of instrumenting DO loops within CCM1.

2 General Discussion

The quantity of interest for this study is L_{ijt}, the number of CPU seconds expended at each grid cell ij of the horizontal CCM1 model grid at each time step t in a given run of the model. Spatial load imbalances appear as variance in L over the indices i and j. Temporal imbalances appear over the t index.

The value of L_{ij} for a given time step is given by:

$$L_{ij} = I_{ij} + U_{ij}$$

where I_{ij}, the varying load, is the time spent in sections of the physics code where the amount of work performed is a function of the physical state of the grid point – that is, of the physical variables carried for the parcel of air represented by the grid point. U_{ij}, the unvarying load, represents the load in the remaining sections of the code – that is, code where the amount of work performed is constant with regard to the physical state of grid cell ij.

The physical calculations for a single time step in CCM1 can be represented schematically as a stream of statements that make up model physics. In Figure 1, the left side of the axis is the first statement executed; the right side is the last. Shaded sections of the statement stream are sections of code which contribute to I. The remaining sections of the stream represent the statements which contribute to U. The time spent to execute all instructions in the stream is L.

Inspection of the code showed that the number of statements which contributed to I was much smaller than the number of statements which contributed to U. In other words, in most of the code the computer was doing the same amount of work for each grid cell; only over a fraction of the code could the workload per cell vary. Therefore, I was easily determined for each cell ij by direct instrumentation. However, instrumenting the code to capture U directly would have been costly in terms of effort, would have increased the amount of load introduced by the instrumentation itself, and was unnecessary in any case, since the total time L (the sum of all L_{ij}s for a given time step) was easily measured.

The unvarying load U was determined by subtracting the sum of the varying loads for

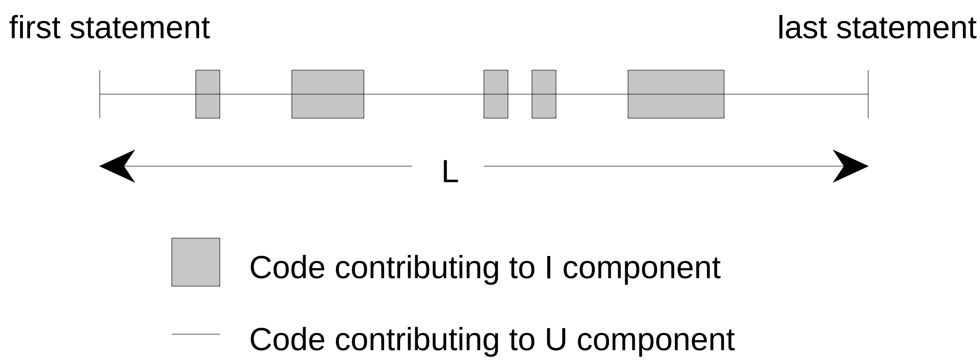

Figure 1: CCM physics statement stream

each grid cell from the total time L and then dividing the resulting time equally between all grid cells:

$$U_{ij} = \left(\frac{L - \sum_i \sum_j I_{ij}}{i * j} \right)$$

Thus, if L can be determined by starting and stopping a timer at the beginning and end of a complete physics calculation for a single time step and if I_{ij} can be determined by instrumenting the relatively small number of statements which exhibit varying load from cell to cell, the load for each cell L_{ij} can be determined as:

$$L_{ij} = I_{ij} + \left(\frac{L - \sum_i \sum_j I_{ij}}{i * j} \right)$$

In actually instrumenting the CCM1 code, it was also necessary to account for CPU load caused by the instrumentation itself. This is discussed in Section 3.3.

3 Method

The overall strategy for gathering load information from CCM1 was to modify the code to output timing data after each time step or number of time steps. The timing data consisted of one numeric value per mesh point representing the amount of time the CPU had spent over that point. Subsequent analysis of this output indicated the presence and magnitude of load imbalances over the model grid.

Thirty-nine routines make up the model physics of CCM1 and of these, ten routines required instrumentation; the call tree is shown in Figure 2. The routines were divided into nine "zones" to provide additional detail. The zones are listed in Table 1.

Instrumentation involved placing a timer at the start and end of DO loop bodies to record (1) the time spent in the loop at each iteration and (2) the location of the computation expressed in grid indices.

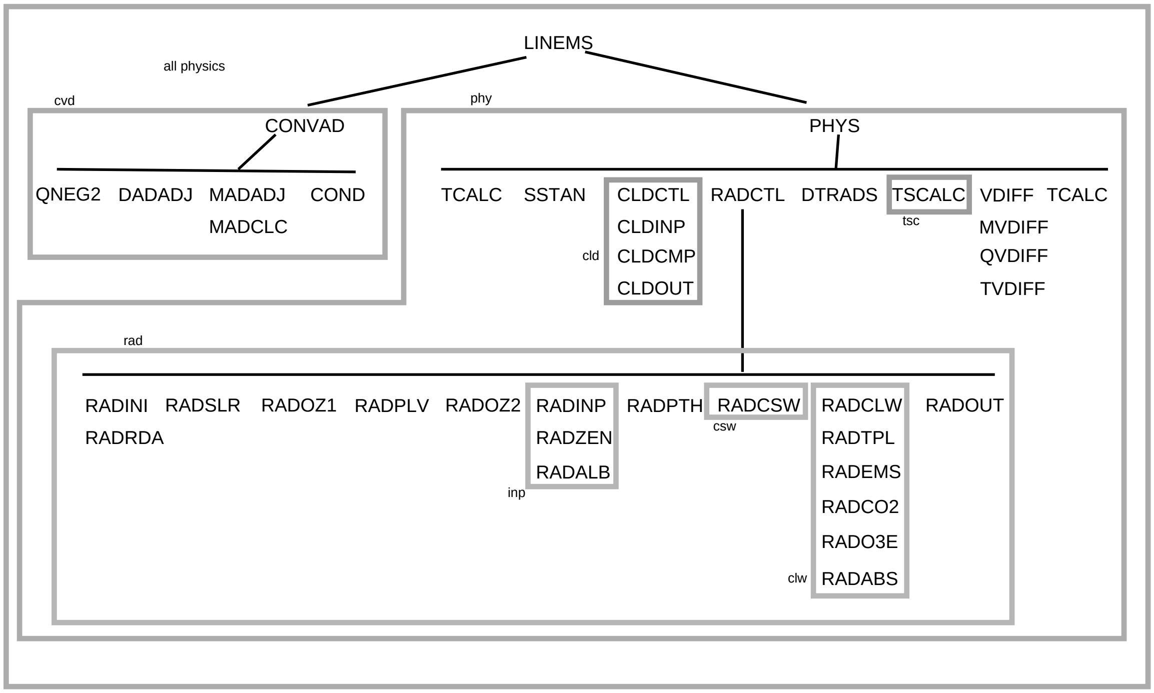

Figure 2: Call tree for physics routines of the NCAR Community Climate Model (CCM1), with zones boxed. Model physics consists of two main subtrees rooted at the subroutines CONVAD and PHYS. Zones were nested to provide additional detail.

It was deemed impractical and unnecessary to instrument all of the physics code since only certain sections were capable of contributing to a load imbalance. Each DO loop of the physics routines was inspected manually to determine whether it contained code that would contribute to a load imbalance over the cells of the model grid. The criteria for selecting DO loops to be instrumented were as follows:

- The loop must be over an index into the model grid,
- The body of the loop must contain conditional code which executes or does not execute based on the value(s) of some physical variable(s) within the model.

Without exception, loops within the physics code meeting the first criterion iterated over the longitudinal index, since the outer loop over latitudes was in a routine superior to the physics routines. In other words, physics is called for a single latitude at a time. The second criterion was met if the loop contained at least one IF statement or other conditional which would cause different computations to occur and if the conditional expression involved variables representing physical phenomena within the model.

Table 1: Division of CCM1 physics code into zones.

Root Routine	Zone Name	Description
LINEMS	all	all routines of model physics
CONVAD	cvd	convective adjustment
PHYS	phy	physical parameterizations
CLDCTL	cld	clouds
RADCTL	rad	radiation
TSCALC	tsc	surface temperature
RADINP	inp	zenith and albedo
RADCSW	csw	short wave radiation
RADCLW	clw	long wave radiation

3.1 Breaking Vectorization

Cray conditional vector merge (CVMG) operations qualified as conditionals and presented a special problem for this study. CVMGs are Cray extensions to FORTRAN which permit a DO loop containing conditionals to be vectorized, but which also have the unfortunate effect of hiding the very quantity being sought, potential load imbalance. For example, this statement appears in the body of the 328 DO loop of the subroutine TSCALC().

```
           DRIDTS(JL) = CVMGT(
     A        0.,
     A        -1.*GRAVIT*BUF(IHMIX+JL)/(BUF(ITNLEV+JL)*BUF(IVMAG+JL)**2),
     A        NLOGIC(JL)
     A        )
```

The first and second arguments are arithmetic expressions; the third argument is a Boolean expression. Depending on the value of the Boolean expression, either the first or the second expression is returned. The CVMG permits vectorization of the conditional by evaluating *both* arithmetic expressions for each iteration. The array on the left-hand side of the assignment is assigned through a vector mask generated by evaluation of the conditional. However, the second expression clearly represents more work: suppose relatively few of the 1,920 grid cells in a 48 by 40 grid require the second calculation. This is the type of load imbalance one would wish to discover, yet the CVMG operation would hide it completely. In a perverse way the load for model physics is perfectly "balanced" on a Cray; the processor does the largest possible amount of work whether it is needed at a particular cell or not!

To uncover load patterns hidden in this way, it was necessary to "break" CVMG operations by commenting them out and replacing them with conventional IF THEN ELSE constructs:

```
           if (NLOGIC(JL)) then
                DRIDTS(JL) = 0.0
```

```
          else
               DRIDTS(JL)=-1.*GRAVIT*BUF(IHMIX+JL)/
     A           (BUF(ITNLEV+JL)*BUF(IVMAG+JL)**2)
          endif
```

As an additional safeguard the compiler was explicitly directed not to vectorize the physics sections of CCM1.

Of the 39 subroutines in CCM1 physics, 10 contained 53 loops that were candidates for instrumentation.

3.2 Timer

Instrumentation consisted of a small library of routines written in C. These routines, when linked into the climate model, maintained an array of accumulators whose cells corresponded to the mesh points of the CCM1 model grid. The accumulator array was also dimensioned over the zones of the physics code so that timing data over the grid could be kept separate from zone to zone.

The library contained routines to initialize the instrumentation package, to dump the accumulators to output, and to reset the accumulators. The principal routine of the library was the timer itself, SYSTIME():

```
SYSTIME( zone, i, j, startstop )
    int *zone,          /* location in physics code */
        *i,             /* longitudinal index into model grid */
        *j,             /* latitudinal index into model grid */
        *startstop ;    /* 1=start, 0=stop */
```

Given the index of a longitude, i, and a latitude, j, SYSTIME() started or stopped the timer for the mesh point, depending on the value of startstop. Stopping the timer after starting it over mesh point ij in routine of zone *zone* caused the elapsed time to be added to the appropriate accumulator in the array.

Figure 3 shows the beginning and end of the 9999 DO loop of CLDCMP(), the routine which computes cloudiness in CCM1. The first call to SYSTIME() starts the timer at the beginning of each new iteration of the loop. The second call to SYSTIME() stops the loop and stores the elapsed time in the accumulator representing zone 1, longitude index i, latitude index $NROW$, and (always) level index 1. Since in CCM1 the cloudiness calculation is performed over oceans only every 24th time step, when full radiation calculations are performed, this loop is particularly interesting; both spatial and temporal load imbalances are strong possibilities.

3.3 Timer Correction

Since each invocation of the instrumentation cost a certain small amount of time, an empirically determined correction constant $CORR$ was subtracted from I and U quantities.

```
C
C     ****  CLOUDS NOT CALCULATED OVER OCEANS IF PARTIAL CALCULATION
C     ****  IN RADIATION
C
      DO 9999 I=1,NLON
         call systime(1,i,NROW,1,1)
C
      IF(FRADSW.OR.FRADLW.OR..NOT.OCEAN(I))THEN

   ... edited ...

C
      ENDIF
C
         call systime(1,i,NROW,1,0)
9999  CONTINUE
```

Figure 3: Instrumented loop 9999 of subroutine CLDCMP. The first argument to SYSTIME, 1, represents the zone of this routine, cloud physics.

The SYSTIME() routine itself corrected the I quantity by subtracting the correction constant before adding the elapsed time to I_{ij}. Since U was computed during postprocessing, correction of the U quantity was done at that time as well. With correction, the equation giving U_{ij} became:

$$U_{ij} = \left(\frac{L - \sum_i \sum_j I_{ij} - S}{i * j} \right)$$

where S was defined as:

$$S = \sum_i \sum_j H_{ij} \times CORR$$

The quantity H_{ij} was the number of "hits" for a cell ij. This was the number of times a pair of SYSTIME() calls were invoked for that cell.

3.4 Experiments

Two runs of the instrumented CCM1 code were conducted on the Cray X-MP at Argonne. To gauge short-term effects, load data from 408 consecutive time steps were captured during an 8.5-day run of the model. A second experiment captured load data for the 0600 GMT time step every two weeks (672 time steps) during a 294-day run so that a total of 22 time steps were sampled during the test period. Because of cost,[1] the second experiment was

[1]Instrumented time steps were expensive to execute because vectorization was turned off. Whereas a normally executing version of CCM1 produced a day's worth of simulation in about 1 minute of CPU, a version with fully instrumented and unvectorized physics required quadruple that time.

"piggy-backed" onto an already running double carbon dioxide CCM1 simulation which was executing in the seventh year at the time the load data were captured.

In both runs, full radiation calculations and history tape dumps were set to occur every 24 time steps, beginning with time step 0. Resolution was set at R15, giving a 48 longitude by 40 latitude by 12 level grid. In both runs, annual cycles and hydrology were enabled. The initial data set had a base date of January 15, 1975, and was provided by NCAR. History data sets were retained to allow correlation of the load data with surface type data.

4 Analysis

Visualization and statistical analysis of the load data generated from CCM1 indicated the presence of both spatial and temporal imbalances during execution of physics routines. Spatial imbalances were patterns of greater or lesser load between cells or groups of cells independent of time. Temporal imbalances were patterns of greater or lesser load that depended on time and were independent of position within the model grid. There were four discernible patterns of spatial imbalance and three patterns of temporal imbalance. Spatial imbalances were correlated with

- surface type,
- polar night,
- weather patterns,
- day and night.

Observed patterns of temporal imbalance were

- hourly cycle (every 2 time steps),
- full radiation cycle (every 24 time steps),
- seasonal cycle.

The distinction between spatial and temporal imbalanced was difficult to maintain since several of the observed imbalances had both spatial and temporal dependencies. Load was observed to vary spatially depending on whether a cell was in a daytime or nighttime section of the globe; however, the boundaries of these regions moved from east to west with time. The seasonal cycle of load from equinox to solstice also had a spatial dependency, in that the imbalance occurred for cells in polar regions of the grid.

4.1 Full Radiation Cycle

Physics load jumps 20 fold every 24th time step when full radiation calculations are performed, causing a dramatic temporal imbalance and affecting the distribution of load as

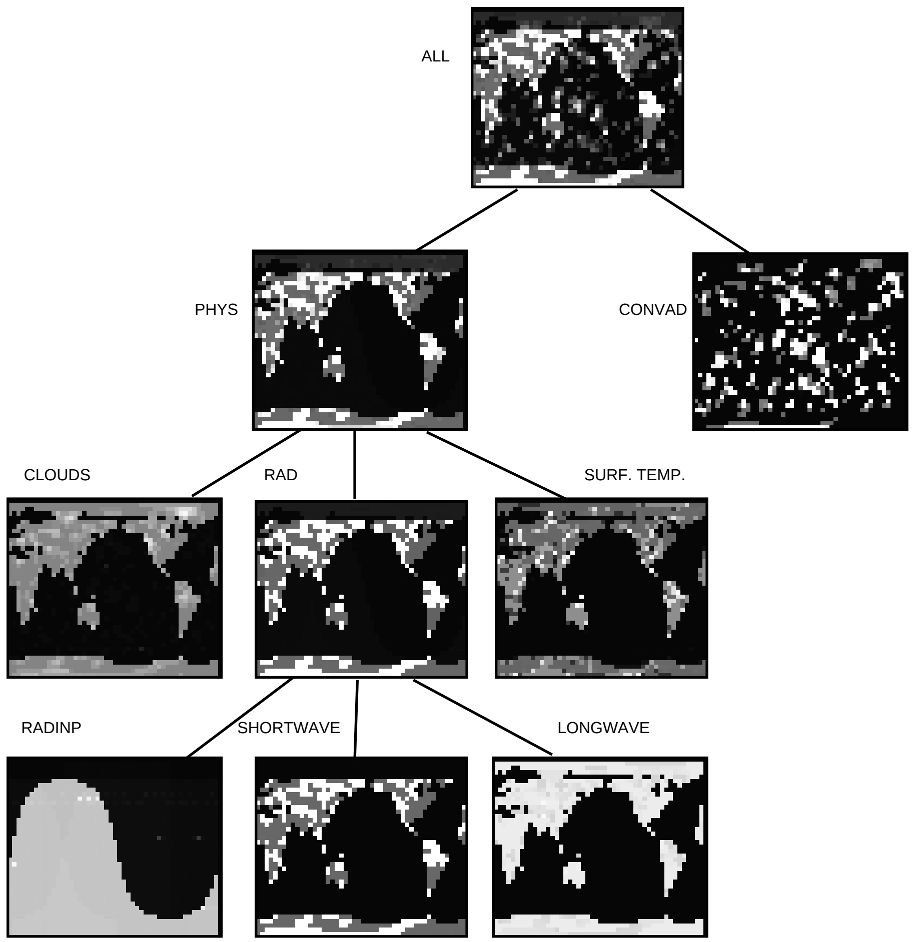

Figure 4: Physics load for a partial radiation calculation time step in January: $ts = 348$. Load for each zone of the physics code is shown as a density plot superimposed over the physics call tree, exposing the contribution of each zone to the overall load for all of model physics. Lighter shades indicate higher load. Load imbalance from surface type came from nearly all sections of the PHYS subtree. Weather patterns are contributed largely by the CONVAD subtree, particularly over the oceans. In the PHYS subtree, the oceans were virtually flat during partial radiation time steps. Reduced load from polar night is visible as a dark band of four latitudes across the tops of a number of the frames pictured. Each plot was scaled to its own mean.

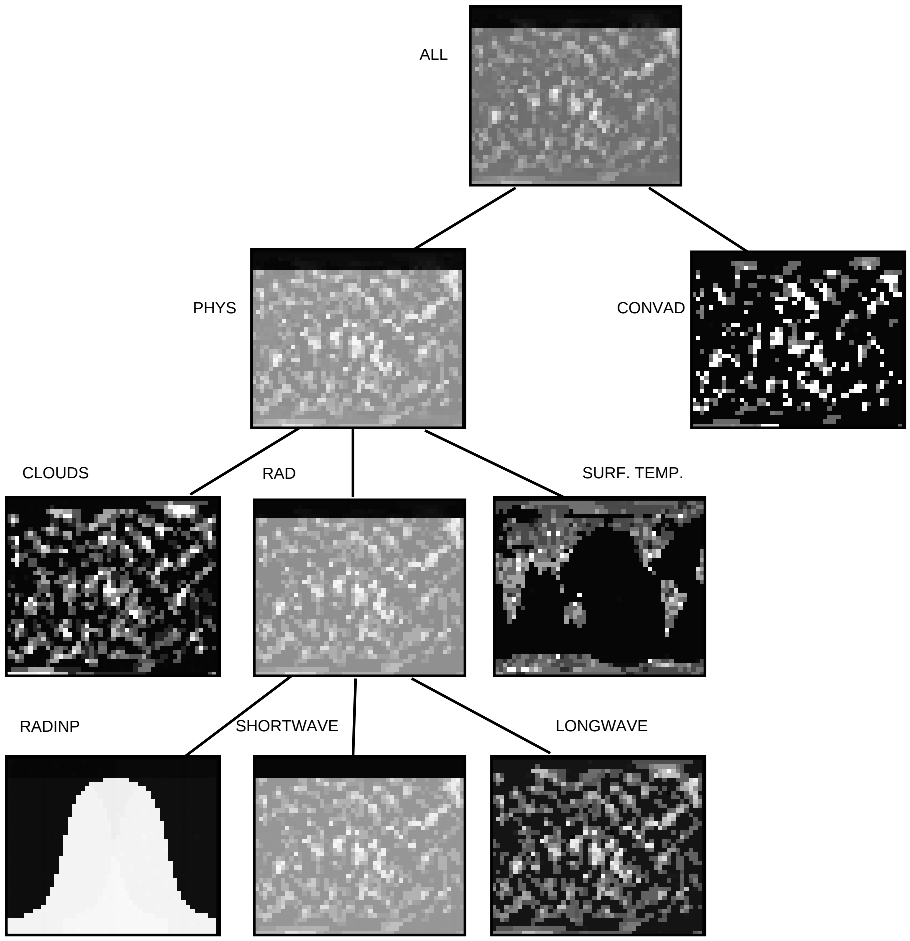

Figure 5: Physics load for a full radiation time step in January: ts = 336. The band of depressed load across the four northernmost latitudes visible during partial radiation time steps remained for full radiation time steps, indicating that load imbalance from polar night is a factor during both full and partial radiation time steps. However, surface related imbalance virtually disappears. Weather patterns were detectable but were not as pronounced. The CONVAD subtree was unaffected by the cycle of full and partial radiation calculations.

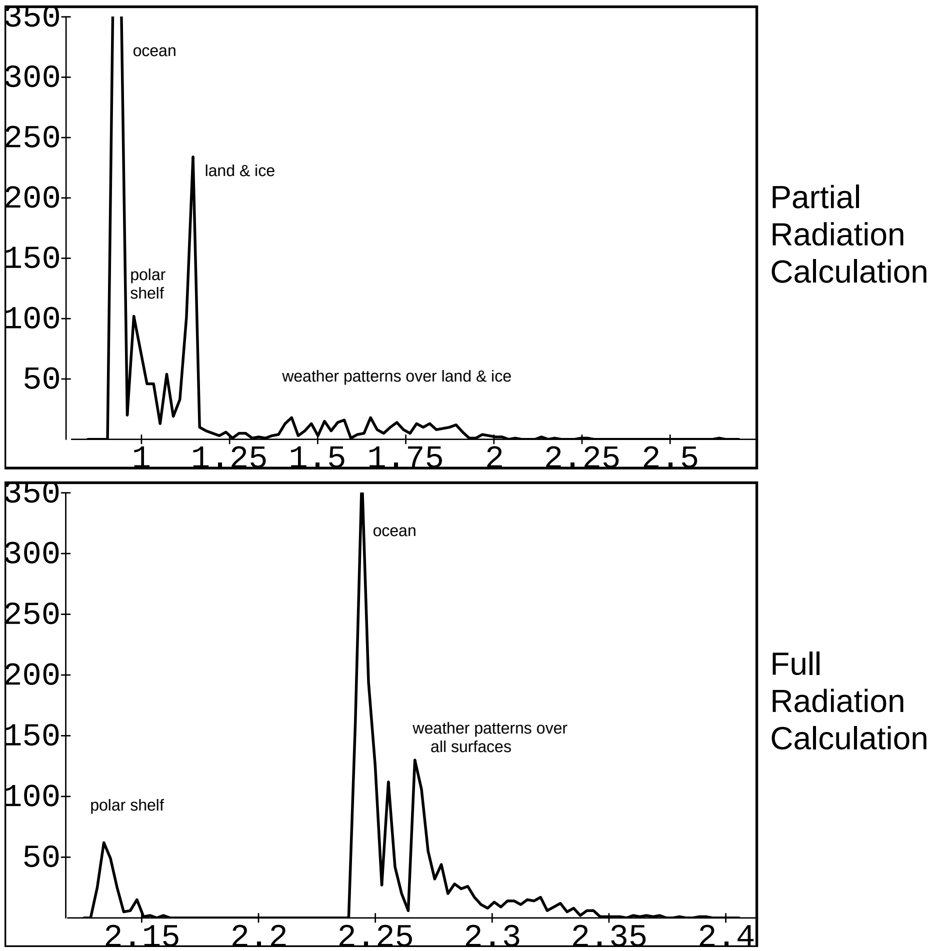

Figure 6: Frequency distribution of load across cells of model grid during physic calculations in time step 348 (partial) and time step 336 (full) of a 20-day run. During partial radiation time steps the imbalance in load caused by surface type is visible is spikes in the distribution over the major surface types: land, ice, ocean, and water. Load in cells involved in weather patterns was distributed more widely at the upper end of the distribution. The x axis is scaled in 10^{-3} seconds for the partial radiation time step and in 10^{-2} seconds for the full radiation time step. In each graph the y axis is the number of cells exhibiting a given CPU load.

a function of surface type. Average load during a partial radiation calculation time step was $1.1073 \cdot 10^{-3}$ seconds per cell (Table 2). Standard deviation was 24 percent. During the previous full radiation calculation, average load was $22.5051 \cdot 10^{-3}$ seconds per cell. Standard deviation dropped to 2 percent over the entire grid, a dramatic decrease in load imbalance.

Nearly all of the increase during full radiation calculations occurred in zones *csw* and *clw*, the short and long wave radiation routines, which accounted for substantial amounts of load at any time step. During partial radiation time steps, radiation routines accounted for 41 percent of total load. During full radiation calculations the amount of time spent in radiation routines jumped to 97 percent of physics load.

The dramatic increase in radiation calculations during full time steps obscured features of the load visible during partial radiation time steps (Figures 4 and 5). Spatial imbalance related to polar night remained visible, as did variations in load caused by weather patterns. Otherwise, however, the effect of full radiation time steps was to substantially flatten the load over the model grid during physics calculations.

4.2 Surface Type

Load was greater over land and sea ice than over open water, which provided the striking effect of continents standing out brightly against the darker oceans (Figure 4). Load levels over land and sea ice were, on average, 40 to 45 percent higher than over oceans except in regions of polar night. For example, load over oceans was $0.9687 \cdot 10^{-3}$ seconds per cell during partial radiation calculation time steps in non-arctic regions (Table 2). Load over land and sea ice was $1.3841 \cdot 10^{-3}$ and $1.4279 \cdot 10^{-3}$ seconds per cell, respectively.

During full radiation calculation time steps, the modulation of load by surface type virtually disappeared, as shown in Figure 5. Flattening of load occurred in both long and short-wave radiation and in cloud routines. The effect of surface type on load was also visible in frequency distributions of load values for full and partial radiation time steps (Figure 6).

4.3 Polar Night and Seasonal Cycle

If a section of the grid experienced no illumination for any fraction of the day, such as the arctic during northern winter, work performed in the radiation routines dropped significantly, producing a spatial imbalance between cells in the arctic and the rest of grid. During the equinoxes, when both poles were illuminated for some fraction of the day, the imbalance vanished. The appearance and disappearance of this spatial imbalance produced a temporal imbalance on a seasonal cycle.

The imbalance resulting from polar night was readily apparent as a "shelf" of decreased load over the four northernmost tiers of cells in the *all physics* and *rad* frames of Figures 4 and 5.[2] The source of this imbalance was the radiation routines of the PHYS subtree.

[2] At the solstices, the shelf consisted of five latitudes. Only four visible in Figures 4 and 5 because the time step shown was approximately 30 days after the winter solstice and the northern shelf had begun to disappear. Five full latitudes of reduced load are seen in the June and December frames of Figure 7.

As shown in Table 2, average load dropped about 10% from $1.1193 \cdot 10^{-3}$ seconds per cell for most of the globe down to $1.0000 \cdot 10^{-3}$ seconds per cell in the arctic load shelf. The table also shows that the drop-off occurred primarily over land and sea ice; ocean was not affected. During full radiation time steps, load dropped about 5 percent as a result of the arctic load shelf.

Plots of load data from the half-year run of CCM1 confirm that polar night was strongly correlated with the polar load shelf, since the shelf followed precisely the movement of polar night from pole to pole over time (Figure 7). As simulation time approached the vernal equinox, the northern load shelf waned gradually until disappearing completely at the equinox. After the equinox, a new southern load shelf was observed to wax northward from the south pole as the northern-summer solstice approached. At the solstice, a southern load shelf identical in size with the January northern load shelf encompassed the grid cells in the five southernmost Gaussian latitudes. This seasonal migration of the polar night was a strong source of temporal load imbalance in the model as shown in Figure 8.

4.4 Day and Night

A very slight but definite variation in load was found to exist between cells in daylight sections of the grid and cells in nighttime sections. The imbalance was detectable over oceans in the PHYS module, where load was otherwise fairly flat. In animations of load data for sequential time steps, the boundary was visible as a sine wave over the flat projection of the earth which moved steadily from east to west, completing a cycle every 48 time steps (every 24 hour period of the simulation). Figures 4 and 5 show clearly that the variation is coming from the RADINP subtree of the radiation module. Manual inspection of the code suggested that the most probable source of the imbalance was the subroutine RADZEN, which calculated zenith angles.

The variation between day and night was consistently $0.008 \cdot 10^{-3}$ seconds regardless of whether the time step was a full or partial radiation calculation. This represented an about 0.7 percent variation during partial radiation time steps. The variation during full radiation time steps was insignificant, only several hundredths of a percent.

4.5 Hourly Cycle

Average load for odd-numbered time steps was consistently greater than average load for even-numbered time steps by approximately 2.5 percent (Figure 10). The effect appeared in load data for the zone rooted at PHYS, but was not apparent in any of its subzones, indicating that the temporal imbalance must have been coming from a routine which was not assigned to any zone when instrumentation of the physics code was performed. Manual inspection of the physics code revealed that the imbalance most likely arose from a section of code in the subroutine DTRADS which was overlooked during instrumentation. DTRADS performed physics budget calculations, but a portion of the calculation was set to execute only during odd-numbered time steps.

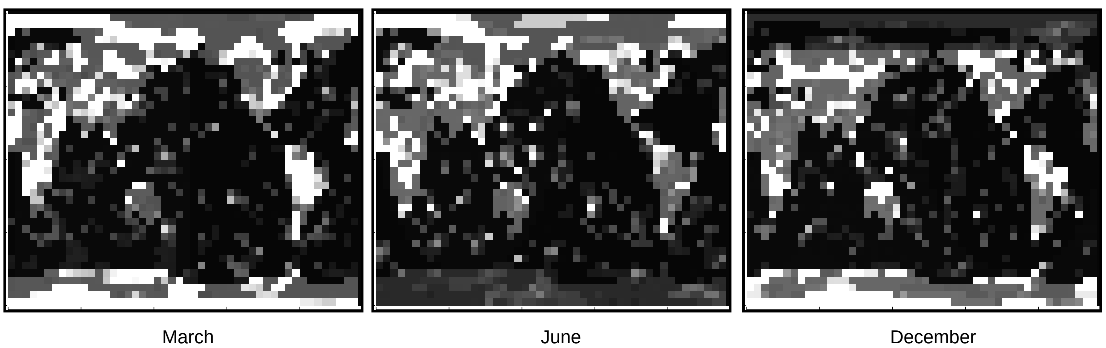

Figure 7: Physics load during for northern vernal equinox and the summer and winter solstices. Notice complete absence of a “load shelf” at either pole during spring when both poles were illuminated for part of the day. During northern summer, the load shelf appeared at the south pole and during northern winter at the north pole. Five latitudes represented the Arctic Circle and five represented the Antarctic Circle.

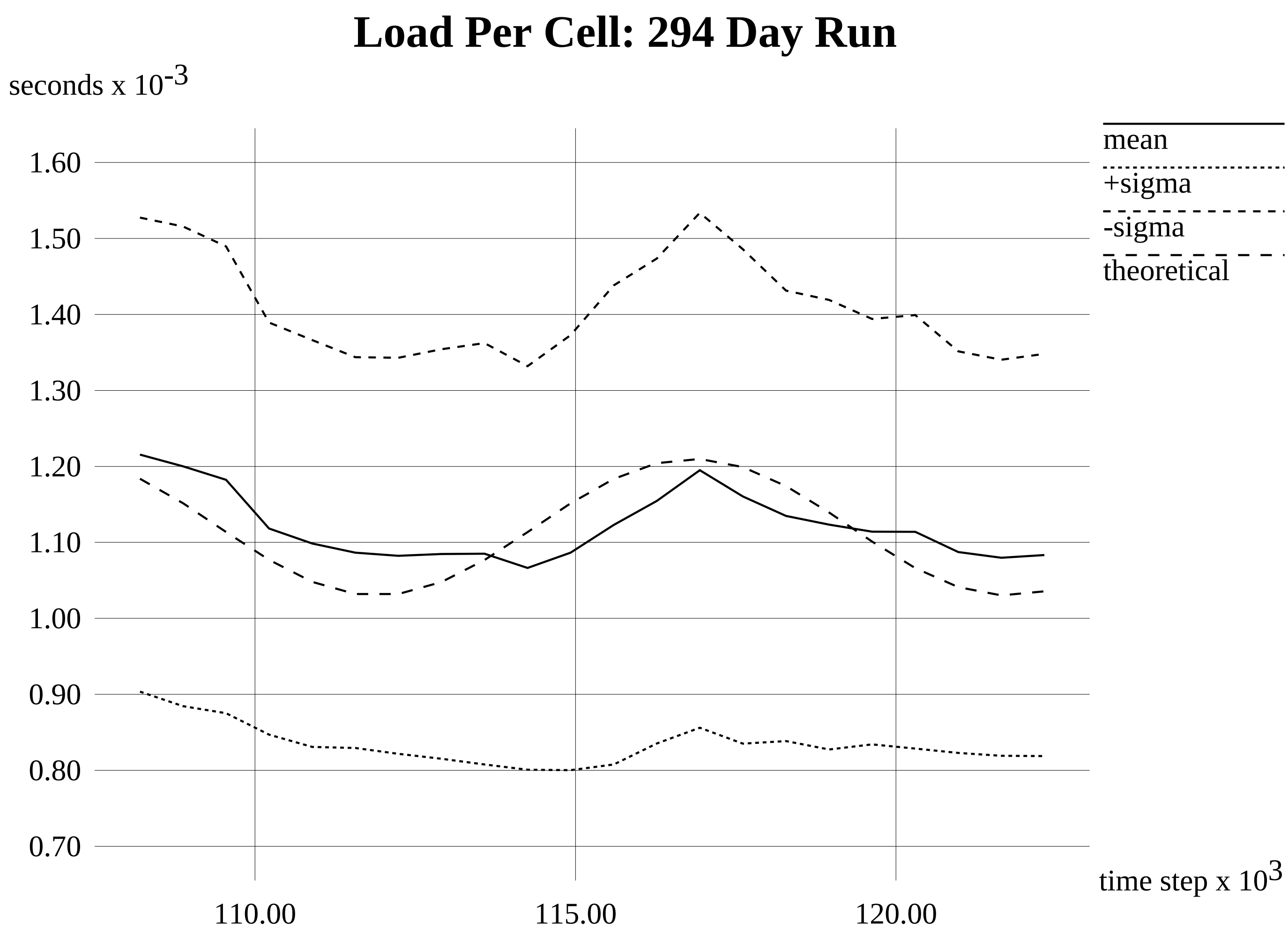

Figure 8: Physics load (PHYS and CONVAD) per grid cell for 294 days (14,112 time steps) of CCM1. Average load and load one standard deviation above and below the mean were plotted. Data were captured every 2 weeks (672 time steps) during the seventh year of a doubled carbon dioxide run of CCM1; sampled time steps were partial radiation calculations. The first sample, time step 108,204, was within a day or two of the vernal equinox for that

Table 2: Mean CPU load per cell in 10^{-3} seconds over ocean, land, and sea ice cells. Data in parentheses are standard deviations as a percentage of mean. Data were collected for both a partial and full radiation calculation during the same day in January. Surface-type data was recovered from CCM1 history tapes for the time period. For arctic latitudes, there were 42 ocean cells, 33 land cells, and 117 sea ice cells; for non-arctic latitudes there were 1102 ocean cells, 623 land cells, and 3 sea ice cells. For all latitudes there were 1144 ocean cells, 656 land cells, and 120 sea ice cells. Note that loads recorded for each zone do not account for all physics load because not all routines were assigned a zone.

January Load as a Function of Surface Type and Polar Night									
	Zone	Ocean		Land		Sea Ice		All	
non-arctic	Partial Radiation Time Step								
	All Physics	0.9687	(9.8)	1.3841	(21.7)	1.4279	(14.5)	1.1193	(25.0)
	PHYS	0.6942	(0.6)	1.1280	(24.1)	1.1932	(17.4)	0.8515	(31.1)
	CONVAD	0.2739	(34.6)	0.2556	(21.8)	0.2341	(0.0)	0.2672	(31.1)
	CLDCTL	0.0393	(0.5)	0.0516	(3.1)	0.0518	(1.8)	0.0437	(13.8)
	RADCTL	0.3242	(1.2)	0.7047	(37.7)	0.7971	(27.7)	0.4622	(52.6)
	TSCALC	0.1426	(0.1)	0.1836	(11.9)	0.1561	(9.0)	0.1574	(15.0)
	RADINP	0.0618	(6.5)	0.0633	(6.0)	0.0661	(0.1)	0.0624	(6.4)
	RADCSW	0.1725	(0.0)	0.5149	(51.8)	0.6049	(36.6)	0.2967	(77.5)
	RADCLW	0.0750	(0.0)	0.1117	(1.2)	0.1113	(0.8)	0.0883	(19.9)
	Full Radiation Time Step								
	All Physics	22.6407	(1.2)	22.6147	(0.9)	22.5526	(0.4)	22.6312	(1.1)
	PHYS	22.3699	(1.0)	22.3638	(0.8)	22.3179	(0.4)	22.3676	(0.9)
	CONVAD	0.2702	(30.3)	0.2504	(18.2)	0.2341	(0.0)	0.2630	(27.2)
	CLDCTL	0.0519	(3.5)	0.0516	(3.5)	0.0507	(0.7)	0.0518	(3.5)
	RADCTL	21.9703	(1.0)	21.9138	(0.8)	21.8851	(0.5)	21.9498	(0.9)
	TSCALC	0.1585	(0.1)	0.2090	(13.3)	0.1928	(11.3)	0.1767	(16.7)
	RADINP	0.0586	(6.7)	0.0574	(7.0)	0.0618	(0.1)	0.0582	(6.9)
	RADCSW	2.1882	(8.8)	2.1402	(6.7)	2.1133	(4.7)	2.1708	(8.2)
	RADCLW	19.6934	(0.1)	19.6861	(0.1)	19.6798	(0.1)	19.6908	(0.1)
arctic	Partial Radiation Time Step								
	All Physics	0.9326	(2.3)	1.0445	(4.4)	1.0117	(2.0)	1.0000	(4.6)
	PHYS	0.6901	(0.1)	0.7753	(1.1)	0.7731	(1.5)	0.7553	(4.8)
	CONVAD	0.2419	(8.7)	0.2687	(16.6)	0.2380	(6.3)	0.2441	(10.8)
	CLDCTL	0.0392	(0.2)	0.0539	(5.5)	0.0513	(2.8)	0.0491	(11.3)
	RADCTL	0.3201	(0.0)	0.3562	(0.3)	0.3568	(0.3)	0.3487	(4.4)
	TSCALC	0.1428	(0.2)	0.1771	(5.3)	0.1769	(6.6)	0.1694	(10.2)
	RADINP	0.0577	(0.1)	0.0577	(0.1)	0.0577	(0.1)	0.0577	(0.1)
	RADCSW	0.1725	(0.0)	0.1725	(0.0)	0.1725	(0.1)	0.1725	(0.1)
	RADCLW	0.0751	(0.0)	0.1112	(1.0)	0.1118	(1.1)	0.1036	(14.6)
	Full Radiation Time Step								
	All Physics	21.3319	(0.2)	21.4281	(0.3)	21.3684	(0.2)	21.3706	(0.3)
	PHYS	21.0920	(0.1)	21.1609	(0.2)	21.1282	(0.1)	21.1259	(0.2)
	CONVAD	0.2394	(6.0)	0.2666	(13.8)	0.2396	(8.3)	0.2442	(10.2)
	CLDCTL	0.0515	(2.7)	0.0541	(5.9)	0.0514	(3.2)	0.0519	(4.2)
	RADCTL	20.6927	(0.1)	20.7200	(0.1)	20.6879	(0.1)	20.6945	(0.1)
	TSCALC	0.1584	(0.0)	0.1975	(10.3)	0.1995	(5.2)	0.1902	(10.8)
	RADINP	0.0535	(0.1)	0.0535	(0.1)	0.0535	(0.1)	0.0535	(0.1)
	RADCSW	0.9238	(0.0)	0.9238	(0.0)	0.9238	(0.0)	0.9238	(0.0)
	RADCLW	19.6853	(0.1)	19.7126	(0.1)	19.6805	(0.1)	19.6871	(0.1)
all lats	Partial Radiation Time Step								
	All Physics	0.9673	(9.7)	1.3671	(22.1)	1.0221	(7.4)	1.1073	(24.2)
	PHYS	0.6940	(0.6)	1.1102	(24.9)	0.7836	(9.5)	0.8418	(30.1)
	CONVAD	0.2727	(34.2)	0.2562	(21.6)	0.2379	(6.2)	0.2649	(30.1)
	CLDCTL	0.0393	(0.4)	0.0517	(3.4)	0.0513	(2.8)	0.0443	(14.0)
	RADCTL	0.3241	(1.2)	0.6872	(39.3)	0.3678	(21.0)	0.4509	(51.7)
	TSCALC	0.1426	(0.1)	0.1832	(11.7)	0.1763	(6.9)	0.1586	(14.7)
	RADINP	0.0617	(6.5)	0.0630	(6.2)	0.0579	(2.3)	0.0619	(6.6)
	RADCSW	0.1725	(0.0)	0.4976	(54.3)	0.1833	(41.5)	0.2843	(77.9)
	RADCLW	0.0750	(0.0)	0.1116	(1.2)	0.1118	(1.1)	0.0898	(20.0)
	Full Radiation Time Step								
	All Physics	22.5926	(1.6)	22.5550	(1.4)	21.3980	(0.9)	22.5051	(2.0)
	PHYS	22.3230	(1.4)	22.3033	(1.4)	21.1579	(0.9)	22.2435	(1.9)
	CONVAD	0.2691	(29.9)	0.2512	(18.0)	0.2395	(8.2)	0.2611	(26.2)
	CLDCTL	0.0519	(3.5)	0.0518	(3.8)	0.0514	(3.2)	0.0518	(3.6)
	RADCTL	21.9234	(1.5)	21.8538	(1.4)	20.7178	(0.9)	21.8242	(1.9)
	TSCALC	0.1585	(0.1)	0.2084	(13.3)	0.1994	(5.4)	0.1781	(16.3)
	RADINP	0.0584	(6.8)	0.0572	(7.0)	0.0537	(2.4)	0.0577	(7.0)
	RADCSW	2.1418	(14.2)	2.0790	(14.5)	0.9536	(19.5)	2.0461	(20.1)
	RADCLW	19.6931	(0.1)	19.6875	(0.1)	19.6805	(0.1)	19.6904	(0.1)

4.6 Weather Patterns

Small clusters of cells exhibiting heightened load appeared uniformly distributed over the model grid. These clusters arose from many areas of the physics code, exhibited significantly heightened load, and were the major source of dynamic load imbalance observed. The exact cause or meaning of these clusters has not been confirmed; but the way in which these clusters formed, moved, and dispersed over time in animations suggested nothing so strongly as cloud formations or other weather patterns.

Load arising from "weather patterns" is visible in Figure 4. The patterns appeared as splotches of load across the map for all physics. In the PHYS subtree of physics, the load patterns disappeared entirely over ocean but remained over land. In the CONVAD subtree, weather patterns were present over the entire map at all times and were, in fact, the only noticeable source of imbalance.

The range of variation in load from these patterns was wide, as indicated in frequency distributions for physics load data (Figure 6). During partial radiation time steps, base loads over land and ice (for those cells not part of a weather pattern) were centered at about $1.15 \cdot 10^{-3}$ CPU seconds per cell. Load attributed to weather patterns began at $1.25 \cdot 10^{-3}$ CPU seconds per cell and continued up to a maximum load of $2.64 \cdot 10^{-3}$ CPU seconds per cell. Given that the mean load per cell for this time step was $1.11 \cdot 10^{-3}$ CPU seconds per cell, the variation from weather patterns was 125 percent of the mean! Relatively few – about 16 percent – of the 1920 grid cells were involved in a weather pattern.

During full radiation calculations, load variation from weather patterns was present but much less severe. The base load over all surfaces was centered around $22.4 \cdot 10^{-3}$ CPU seconds per cell. The variation due to weather patterns affected some 400 cells (about 20 percent) but was only about 5 percent of the mean $22.5 \cdot 10^{-3}$ CPU seconds per cell.

The shapes and sizes of the load clusters were irregular, but the clusters tended to be globular and range in size from 2 to 6 cells in diameter. The clusters were also subject to a spreading out in an east-west direction at extreme northern and southern latitudes as a result of the mapping of grid points on the sphere.

Load imbalance from weather patterns was large; and unlike other observed imbalances, it was unpredictable. On the other hand, the effect had a relatively small grain size and the features were distributed evenly over the model grid at a given point in time. Static allocations of processors may be sufficient to balance the load provided the size and shape of the grid cell to processor allocations are carefully chosen. Figure 9 shows the result of an experiment to test several allocations of processors to the model grid in which the size and shape of the cell to processor mappings were varied. Each processor received the aggregate load of the cells it was assigned and communication cost between processors was not considered. The quality of an allocation was measured as the mean processor load divided by the maximum processor load. The results were that larger numbers of cells per processor provided better quality by putting a better mix of loads into the workspace of each processor. Long, thin allocations provided better quality than block-shaped allocations because they sliced up globular clusters of heightened load; block-shaped allocations tended to engulf clusters, creating wider variations in load between processors. Allocations were more effective when oriented in the north-south direction because they were not affected by east-west broadening of load clusters at extreme latitudes.

5 Conclusions

The largest source of imbalance in the physics routines of CCM1 was temporal, the 24 time step cycle of radiation calculations. At each such time step, load jumped more than 20 fold and – discounting polar effects – became nearly flat over the model grid. Given that one full radiation time step took about the same amount of time to compute as all the partial radiation time steps in the cycle combined, load is flat half the time the model is running and requires no special balancing (but see the next paragraph). During the other half, balancing strategies can improve load balance but the benefits must be weighed against the costs of formulating and moving to a different cell to processor mapping as the climate model executes.

The effect of the seasonal cycle of polar night did present a slowly changing and extremely predicatable cycle of reduced load at the poles during solstices and full load at both poles during equinoxes. The effect was very slow, only 1 cycle per year if summer and winter solstices are considered separately. Several static allocations of processors may be sufficient to handle different parts of the year. Load imbalance from polar night was present during both full and partial radiation time steps, so some adjustment of a straight processor allocation during full radiation time steps may be useful.

Load imbalance related to surface type was significant, providing a 30 percent variance between cells over land and ocean. However, surface related load imbalance was static except for seasonal variation in sea ice at extreme latitudes. Surface related load imbalance was only a factor half the time, during partial radiation calculations.

Heightened load associated with weather patterns was the only dynamic source of imbalance observed. It was a major source of imbalance during partial radiation calculations. It was noticeable but not severe during full radiation calculations. Size and shape of cell to processor mappings affected the quality of static allocations to achieve load balance.

Study of load patterns within CCM1 suggests that a carefully chosen static allocation of processors will be sufficient to achieve acceptable load balance in physics routines of a parallel climate model. Ideally, assuming zero cost to move work to new processors, several interchangeable allocations would be desirable to account for changes in load as a result of the seasonal cycle and the full radiation calculation cycle. In reality, however, cost to move work between processors will be a factor, forcing limits on the frequency with which reallocations can be made.

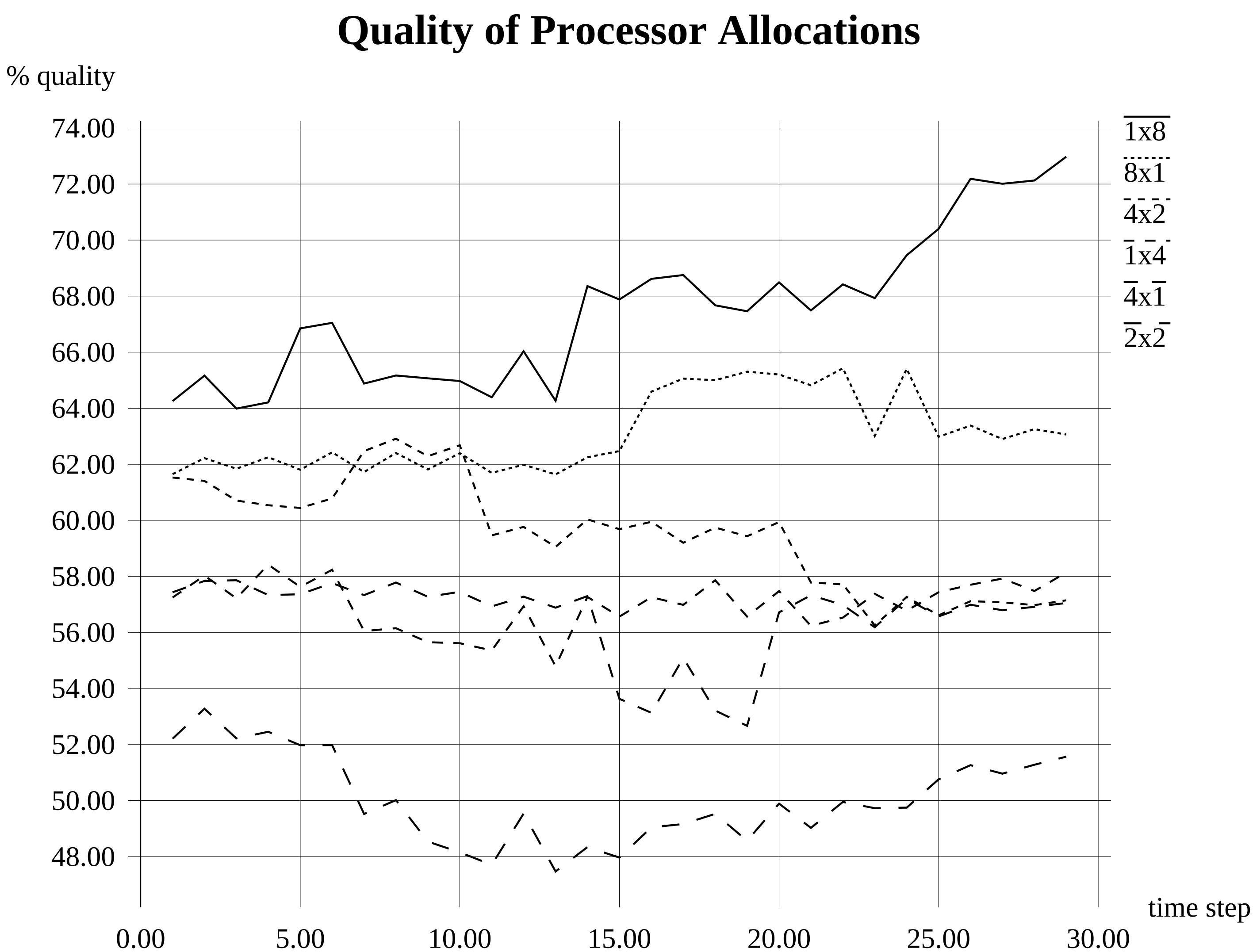

Figure 9: Shape and size of simulated cell-to-processor mappings affected load balance quality during partial radiation time steps. Grid cells were grouped into rectangular or square tiles over the model grid, then assigned to "processors." The sum of the loads in a tile was the load for the assigned processor. Quality of the mapping was measured by dividing the average load per processor by the maximum load per processor. Two sizes

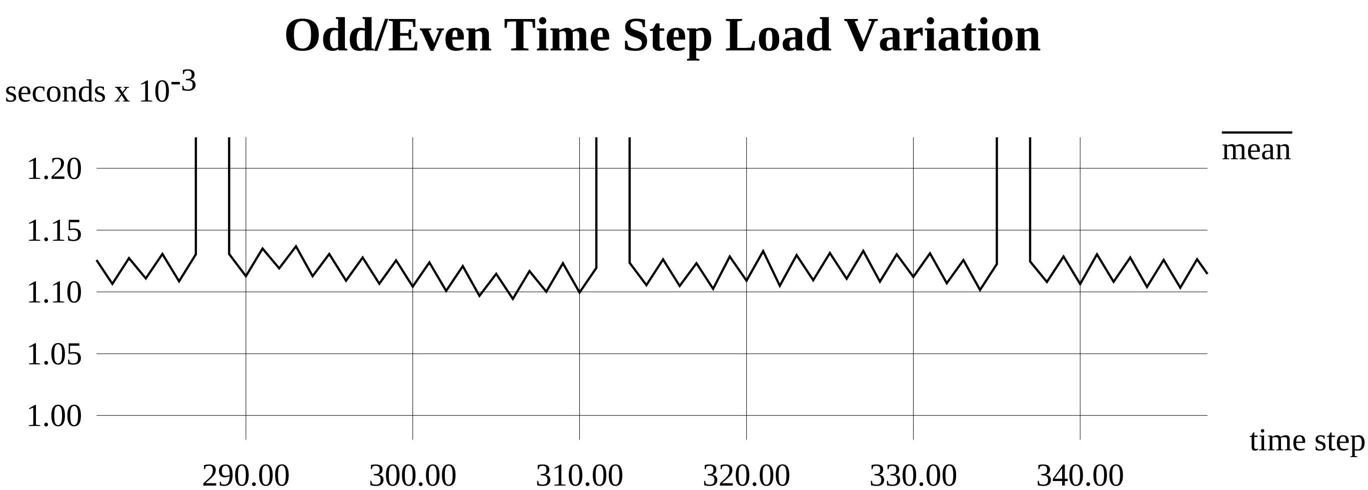

Figure 10: Average load over several dozen consecutive time steps from 8.5 day run. Load during odd-numbered time steps was consistently 2.5 percent above load for even-numbered time steps, generating a saw-tooth pattern except during full radiation calculations, shown as spikes. All of the effect is traceable to the routine DTRADS in the PHYS subtree of model physics.