Marc Leef
Allen Malony
CIS 410
5/6/14

Homework 3

Data Reorganization:

1) The purpose of data reorganization is to minimize the data access time by decreasing the transfer of information across memory layers. Frequently, the main bottleneck of performance is not computation time, but rather, the time it takes to move data around the system. Parallelizing these reorganizations is a way to optimize the optimizations and guarantee the best data locality.
2) a) Memory latency issues will occur. Because the structs are stored in larger memory blocks, all three types of threads will be competing for access to essentially the same memory block if they are started at the same time and work similar speeds.
   b) Latency issues, while memory is being accessed by one thread, the other 2 will be blocked.
   c) The data could be partitioned so each thread works on one dimension of the cube at a time. That way, there will never be a simultaneous attempt to access the data stored in the cells.
3) Partitioning the data would greatly aid the stencil pattern. Dispatching different threads to independently work on separate subsets of the original data would eliminate concurrency issues and greatly improve performance.

Stencil Pattern:

1) a) The ghost cell regions would encompass the borders of each chunk. For the 500 x 500 x 500 chunk size, there would be 1500 ghost cells per chunk, for a total of 8 * 1500 = 12000 ghost cells that need to be updated every iteration.
   b) 64 threads would be required to make chunks of 250 x 250 x 250. 512 threads would be required to make chunks of 125 x 125 x 125.
   c) Our halo for the 250 x 250 x 250 chunks would be 750 * 64 = 48000. Our halo for the 125 x 125 x 125 chunks would be 375 * 512 = 192000. Our ratio of ghost to non-ghost cells decreases as we reduce the chunk size. This affects the computation by increasing the amount of memory required, but also reducing the number of communications required between threads. The overhead of additional memory versus additional communications would need to be further investigated.
2) We would want to increase the depth to reduce the number of communications between threads and increase overall thread independence. However, by doing so, we would also dramatically increase the amount of memory required and also incur many more redundant computations.
3) This operation could be parallelized by using a separating hyper plane, by identifying plane that that intersects the grid of intermediate results, parallel computation could still be performed without being hindered by any dependencies.