Launch kernel 1 with 512 threads and ceil(nglob / 512) blocks

cudaThreadSynchronise()

loop on outer and then inner elements

END

CONTINUE

If done with outer elements, use a kernel to get MPI buffers from the device and start the non blocking MPI calls

loop on sets of elements of same color

END

CONTINUE

Launch kernel 2 with 128 threads and number of blocks = size of color set

MPI_Test()

TRUE

FALSE

Process next step of MPI assembly

cudaThreadSynchronise()

MPI_WAIT()

If not finished, process MPI buffers received

Use a kernel to copy the assembled MPI buffers to the device

Launch kernel 3 with 512 threads and ceil(nglob/512) blocks

cudaThreadSynchronise()