QuNova OFMO: Slot Error and GPU Compilation fixes

Aidan Pellow-Jarman, Rowan Pellow-Jarman, Shane McFarthing

May 2022

1   Introduction

This documentation describes the changes made inside the *OpenFMO_prv* directory to fix the slot error that occurs when increasing the number of processes to use and to compile the OFMO code to run on modern GPU's.

2   Using a Hostfile to fix the Slot Error

The slot error that was encountered is related to the argument **-np** which sets the number of processes. The mpi code checks that the number specified by this argument is less than or equal to the number of available slots on the system, which is set to the number of physical cores by default. As a result, if **-np** is set higher than the number of cores, a slot error occurs since over-subscription of the CPU is not allowed by default.

We fixed this by running OFMO using the **mpiexec** command and using a hostfile. The hostfile allows us to explicitly set the number of slots that we would like to use, and this can be set to a number larger than the number of cores if we wish to over-subscribe the CPU. The program will then proceed using aggressive over-subscription by default, which will provide better performance than degraded over-subscription in our case. This results from the VQE using the entire machine, which makes processes pausing to share CPU-time undesirable.

To create a hostfile for your local machine you must create a text file in the *OpenFMO_prv* directory, and write **localhost slots=x**, where **x** is the number of processes that you would like to use.

To run OFMO with this fix, you must use the following command:

mipexec -np 1 –hostfile *⟨insert path to your hostfile here⟩⟨insert the rest of the normal ofmo-master command here ⟩*

3 Makefile alterations to support compilation for modern GPU's

The original OFMO code included support for older GPU architecture's, however it did not support the newer GPU's that we are using. Originally, binaries were compiled for the specific architecture used and these binaries were not compatible with other architectures.

We added code in the OFMO Makefile to also compile to an intermediate representation that is compiled to machine-code at runtime to suit the architecture of the GPU being used.

These changes were made in lines **233, 237** and **242** of the Makefile in the *OpenFMO_prv* directory.