Nicholai Tukanov (ntukanov)
Chengyue Wang (wcy)

Major Changes: Have there been any major changes in the goals or implementation of your project since your project proposal? If so, please describe these changes and what brought them about.

- We don't have major changes in our goal. For the implementation part, when we wrote the proposal we thought we would use the infrastructure described in the paper Progressive Raising in MLIR, but it turned out that the infrastructure described in the paper can not be publicly accessed. So we changed our plan to create our own pass and pattern matching algorithm in the MLIR.

What You Have Accomplished So Far: Describe exactly what your group has accomplished so far. If you have any results so far, please post them to your web page, and point us to them. (You do not need to include your results in this writeup provided that they are available on your web page.)

- We are working on supporting conversion for 1 format. Specifically, the 1m format described by Van Zee [1]. We have also focused on using CPU kernels for the int16 data type.
- We have created the test models in python
- OperationPass with MLIR for replacing complex arithmetic and including a conversion routine
- Determine and write code for conversion routine specifications
- Create and test the complex GEMM kernels for GPU and FPGA (The kernel was developed based on our own research)
- Test the performance difference between the version developed and run by tensorflow and the version linked with the high performance complex kernels using C.

Meeting Your Milestone: Did you meet the milestone which you described in your original project proposal? If not, then please explain why.

- We are slower than expected. We needed to create a toy language in order for us to test since generating IR from tensorflow turned out to be more difficult than expected. Currently we have the results from a working prototype written in C, utilizing the complex GEMM kernels; so we have a good sense of what the results before and after replacing complex arithmetic in tensorflow. Furthermore, we had a DARPA project deadline last week, so we plan to devote more time over the next two weeks.

Surprises: Have there been any major surprises so far in your project? If they were bad surprises, how have you managed to work around them?

- No.

Revised Schedule: Describe what each member of your group is going to accomplish during your remaining time to successfully complete your project. If you are stuck on anything, please let us know. (In fact, if you are stuck, please send us email immediately rather than waiting for us to read this report.)

- For the last few weeks, we are working on getting an accurate representation of the complex arithmetic with our toy language. Using this toy language as a proof of concept. We will implement our conversion routine and model examples (Created by Nicholai). With these, we will use our OperationPass to optimize our toy IR (Created by Chengyue). Lastly, we will use our toy language to generate a binary to test our code and compare performance (Done together).

Resources Needed: Do you have all of the resources (e.g., software, benchmarks, simulators, etc.) that you need to complete your project? If not, then please explain how and when you will acquire these resources.

- Yes.